Models Scientific Practice in Context An Educational Module

Prepared by

Elizabeth Fisher Professor of Environmental Law Corpus Christi College, Oxford University

Pasky Pascual Environmental Protection Agency

Wendy Wagner Joe A. Worsham Centennial Professor University of Texas at Austin School of Law

For

### Committee on Preparing the Next Generation of Policy Makers for Science-Based Decisions Committee on Science, Technology, and Law

# The National Academies of SCIENCES • ENGINEERING • MEDICINE

June 2016

The copyright in this module is owned by the authors of the module, and may be used subject to the terms of the <u>Creative Commons Attribution-NonCommercial 4.0 International Public License</u>. By using or further adapting the educational module, you agree to comply with the terms of the license. The educational module is solely the product of the authors and others that have added modifications and is not necessarily endorsed or adopted by the National Academy of Sciences, Engineering, and Medicine or the sponsors of this activity.

# Contents

| Instructors' Guide for Models: Scientific Practice in Context                                | 1 |
|--|---|
| How Is the Module Organized and How Many Class Sessions Does It Cover?                       | 1 |
| What Concepts Will Students Learn in this Module?  | 2 |
| How Might I Structure My Classes and Pre-Class Assignments?                                  | 5 |
| Readings   | 5 |
| Class Format   | 5 |
| What Additional Resources or Issues Fit with this Material that I Could Integrate on My Own? | , |
|  | 8 |
| How Might I Evaluate the Students' Performance?  | 9 |
| What Are the Direct Links to Online Materials Used in the Course Packet?                     | 0 |
| Models: Scientific Practice in Context1  | 1 |
| 1. Introduction1   | 1 |
| 1.1 Road Map to This Module1   | 2 |
| 2. What Models Are And Why We Use Them in Regulatory Decision Making                         | 3 |
| 2.1. Models – Inputs, Outputs, and Assumptions1  | 3 |
| Check Point A: Models as Simple Descriptions of Causal Relationships                         | 5 |
| 2.2. Models: Useful Simplifications of Reality1  | 5 |

| Check       | Point B (with Side Trip): The Relationship between Models and Their Purpose 1 | 16 |
|-------------|---|----|
| 3. Models   | as a Scientific Practice  | 18 |
| 3.1. Scie   | ntific Utility of Models 1  | 18 |
| 3.1.1.      | "Big Data" and Modeling Capacity1   | 18 |
| 3.1.2.      | Modeling as Compared to Other Scientific Techniques 1                         | 19 |
| 3.2. Mo     | dels, Causation, and Probabilities2   | 20 |
| Check       | <pre>c Point C (with a Side Trip) – Thinking about Correlations in Data</pre> | 22 |
| 3.3. Epis   | stemic Frames: The Fundamental Importance of Model Assumptions                | 24 |
| 3.4. All N  | Models Are Wrong, Some Are Wronger: The Importance of Model Evaluation 2      | 25 |
| Check       | Point D: Modeling as a Scientific Practice in the Round                       | 27 |
| 4. Models   | in Policy Making  | 28 |
| 4.1. Moo    | dels Are Everywhere2  | 28 |
| 4.2. Extr   | racting the Framing and Choices Embedded in Models                            | 31 |
| 4.2.1.      | Framing   | 32 |
| 4.2.2.      | Assumptions and Data  | 32 |
| 4.2.3.      | Evaluation Now and in the Future  | 33 |
| Check       | <pre>c Point E: Putting the Inside and the Outside Together</pre>             | 33 |
| 5. Oversigh | ht and Review: Models and Accountability Processes                            | 36 |
| 5.1. Ove    | ersight Processes within EPA  | 37 |
| Check       | x Point F: Ad Hoc vs. Systematic Oversight Processes?                         | 11 |
| 5.2. Moo    | dels and Judicial Review  | 12 |
| 5.3. Acc    | ountability Processes and Their Impact on Modeling Practice                   | 17 |

| Check Point G: Attack a Model Not to Improve It, but to Win the Debate         | 52 |
|--|----|
| 5.4. An Exercise for Recapping and Reflection                                  | 52 |
| Moment of Truth Check Point H: Can You Really Understand a Model if You Didn't |    |
| Create It and Aren't a Modeler Yourself?                                       | 52 |
| 6. Ambient Air Quality Standards Case Study                                    | 53 |
| 6.1. Air Quality Regulation – A Brief History and Overview                     | 53 |
| Check Point I: Asthmatics Stay Indoors?  | 58 |
| 6.2. The Process Required for Setting Ambient Air Quality Standards            | 59 |
| Check Point J: Extending EPA's NAAQS Process to other Settings                 | 63 |
| 6.3. NAAQS under the Microscope: Criticism and Judicial Review                 | 63 |
| 6.4. The Development of Particulate Matter NAAQS Standards                     | 69 |
| Check Point L: Reading a Scientific Paper                                      | 76 |
| 7. Conclusions   | 76 |

# Instructors' Guide for Models: Scientific Practice in Context

This instructors' guide offers suggestions on ways the Models Module could be taught to graduate students or advanced undergraduate students who have little to no advanced scientific or mathematical training. Given the diversity of pedagogical approaches across graduate schools (e.g., journalism vs. law vs. public policy), we do not offer a step-by-step lesson plan, but instead suggest various options for teaching the material that we believe will fit the diverse settings in which (we hope) the module will be offered.

# How Is the Module Organized and How Many Class Sessions Does It Cover?

The first part (Sections 1–4) of materials introduces students to the basic concepts of models. The second part (Sections 5–6) then settles this understanding within applied legal and policy settings and includes a more detailed case study of the Environmental Protection Agency's (EPA) use of models to set ambient air quality standards (Section 6). If a course is not particularly policy oriented, the final section—Section 6 case study on air quality standards—could be omitted, leaving more time for student engagement with the basic concepts and material.

We believe that the entire module could be taught over two or three class periods (note that four periods may be needed if the instructor uses break-out groups during the class period). The materials could even be taught in a single 2-hour class for students who already have some basic familiarity with either computational models or the legal system or both. An instructor attempting this fast track will not have time for class discussion of most of the checkpoints, however.

If the instructor wishes to dedicate even less class time to the Models Module, the materials for Sections 1–4 should provide students with a solid foundation on models that are self-contained. The second part of the materials assumes that students have already mastered the preceding sections. Thus, while instructors can teach Part 1 without Part 2, we have not prepared the materials with the reverse option—it would be very difficult to teach the last two sections of the materials without first reading Part 1.

A conceptual map of the module material is provided in the figure on the next page and discussed in more detail in first few pages of the student materials. Although this material will provide a useful orientation for students, we expect it to be even more valuable to instructors in explaining the scope and relationship of the various sections in the materials.

| Co<br>us<br>po   | ore lessons on<br>sing models for<br>blicy-making  |  | Key ic<br>policy | deas on using models for<br>y-making   |
|------------------|--|--|------------------|--|
| 2.<br> <br> <br> | Models describe, in a<br>mathematical way,<br>how components of<br>a process or system<br>influence each other.  | Key questions and<br>ideas on the art and<br>science of modeling   | 4.1              | Agencies use models to<br>issue regulations,<br>permits, enforce rules<br>and evaluate the<br>impacts of their<br>decisions.     |
| 3.1              | Scientists and<br>policy-makers<br>use models to<br>establish<br>causation.  | Why are models needed to<br>establish causation?<br>Science data are beset by<br>random variation. Models<br>distinguish between<br>variation resulting from<br>random and causal factors. | 4.2<br>Case      | Policy-makers should<br>ask modelers to explain<br>the basis for their<br>modeling choices.<br>e study:                          |
| 3.2              | Models rely on   | How do models establish<br>causation?<br>Using an EF, modelers<br>mathematically approximate a<br>process or system.   | 6                | Using models to establish air<br>quality standards.<br>1 History of air quality<br>protection.                                   |
| 3.3              | Models are evaluated,<br>based on<br>appropriateness of and<br>internal consistency with<br>their EF, as well as<br>consistency with<br>observations.                                    | All models are<br>wrong. But they can<br>be evaluated for<br>their usefulness in<br>generating credible<br>inferences.   |                  | <ul> <li>6.2 The law, judicial</li> <li>6.3 review and modeling</li> <li>6.4 of air standards for particulate matter.</li> </ul> |
| 5.1<br>5.2       | When agencies use<br>models to develop<br>regulations, they are held —<br>accountable by both<br>internal (within the<br>agency) and external<br>(within the court system)<br>processes. | <b>Example:</b><br>Using Bayesian analysis to<br>model baseball games.   |                  |  |
| 5.2              | Courts will evaluate the<br>rational basis of a policy<br>based on an agency's<br>appropriate use of<br>models.  |  |                  |  |

SOURCE: Courtesy of the authors.

# WHAT CONCEPTS WILL STUDENTS LEARN IN THIS MODULE?

There are several essential topics relevant to models that the students should master. Some of the topics are self-explanatory (e.g., the use of models in policy; the different processes agencies have developed to scrutinize models). Other topics are more esoteric and will benefit from class discussion or instructor guidance to underscore their content as well as their centrality to understanding models. We offer a few sample topics that instructors could explore in class to ensure that the concepts are adequately understood. Many of these topics could form the basis for learning goals, depending on the objectives of the larger class and instructor.

A. "All Models are wrong and some are useful." This is a well-known statement by George Box, a modeling expert (see Section 3.4). We think this adage is an excellent way to engage students in the basic definition and value of models. Students should be able to understand why "all models are wrong" and yet why some are useful and necessary for both science and policy. In the course of probing this paradox, students should also be able to provide a working definition of what a computational model is.

Instructors might also explore the ways that models are useful by providing specific examples. The materials in Section 2 highlight several features of models that make them particularly valuable to science and policy. To explore their value, instructors could begin by discussing the decline effect excerpt—both what it is and its implications for science and policy. Then students could be prodded to consider the value of models as a way to study both the trajectory of experiments as well as extract from individual studies some larger value, even while the findings may become less and less significant as the studies accumulate.

The materials in Section 2 also alert the student to what the world might look like without computational models. Students should think about this connection and be pressed to consider how we might organize and make sense of data without models. The students' awareness of big data also helps identify the challenges faced by modelers as well as underscoring their importance.

B. Basic components of models (aka model development). After appreciating the significance of models, students should be able to identify the main features of models and understand their working parts. The materials discuss the anatomy of models later in the first part (see Section 4.2) after immersing students in working models using the simple climate model. Indeed, without this type of experiential piece, it will be too easy for students to abstract from models—a tendency that may make it more difficult for them to anticipate and troubleshoot the many practical issues that arise when models are ultimately in use—like understanding how the model was designed and evaluated and understanding the choices that modelers made. Forcing students to experience firsthand a model—both with the simple climate model and later with water quality—forces students to place their abstract understanding into the real world where models can be much messier in terms of how they are constructed and how they should be evaluated.

The materials can also be easily supplemented or even switched out with regard to the experiential exercises. If the focus of the course is on a more specific topic—for example, children's health and toxins or more general economic concepts in public health—the instructor can easily insert an alternate model that fits the course topic and ask many of the same questions. The point as a pedagogical matter is to force the students to work with models and to see them in practice.

C. Model evaluation. In both science and policy, the "rubber hits the road" with respect to how one evaluates a model (or compares alternative models). The materials at Section 5 discuss some common processes used for the evaluation of models—such as subjecting them to diverse scrutiny and various forms of peer and public review. Students should be pressed to consider these processes for the evaluation of models. Some general questions include (1) Does the evaluation process for a model—and the processes for that evaluation—depend on its use? Are there different paths for models? (2) What are some of the glitches in practice in ensuring rigorous model evaluation? (3) What if models are not evaluated? (4) How do the "results" of model evaluation get fed back into science, policy, and the public sphere? Can they be communicated at all?

Of course, developing a rigorous evaluation process is helpful, but underneath that process must be a framework or way to consider how to evaluate the working parts of models. Again, particularly if the course is policy focused, students should be pressed to grapple with how one might evaluate the assumptions embedded in a model. Should modelers be required to explain each significant assumption/algorithm and the alternative plausible assumptions? If so, should alternative models be run to test each assumption's significance? When does one stop in this process of infinite regress? And what should evaluators do with this information?

Similar questions arise with respect to framing the model itself. Once the question is framed for the model, it often drops into the background without further critical thought. Should the framing question itself be up for grabs in a model evaluation? Of course if too much is put on the table, then the modeling process may never reach an interim conclusion but instead would be run in a continual loop with adjustments that never end.

Finally, the inputs or data can be critical to model outputs. Often the data is difficult to judge with regard to its reliability or collection methods. Instructors could even show sample data sets that illustrate some of these problems (e.g., they are seasonal; partial; fall outside the model's specific domain, etc.). And if there are multiple datasets collected in different ways, how should they be combined? What about outliers? And so on.

D. Incentives. Running throughout the model development and model evaluation topics is the question of incentives. What incentives do model developers have to communicate their choices, uncertainties, use of data, and so on? The same for the evaluators? Many of the challenges posed by models arises from their tendency— both for social and scientific reasons—to be black boxed in ways that resist more rigorous, holistic examination by outsiders who nevertheless should ideally be part of the model development or at least the evaluation stages. Thus, to the extent that students feel intimidated or alienated by the experiential pieces of the models, particularly the experience with the water quality models, this offers an excellent opportunity to discuss the black-boxing problem—both why it occurs, how it occurs,

and whether and how it can be counteracted.

E. *Models and process*. Putting these key themes together, students should be pressed (hard) to consider whether and how an agency or any person for that matter can develop processes that ensure that the model is used responsibly. Students will repeatedly encounter how models are built on assumptions which in turn must often be based on raw value choices that involve difficult trade-offs. If that is the case, how does a modeler (or those responsible for the model, like an agency administrator) explain these features honestly? And when criticisms begin pouring in, how or does the agency (or other) defend the model against attacks that the model is "soft, malleable, and subjective"? (In other words, it is not only wrong, but made up and not reliable.) These types of challenges are almost inevitable if a model is explained properly in a political process. Is the alternative—shrouding the model in mystique once the modeler feels comfortable with the algorithms and data sources—a better solution? These are important and very real questions that arise regularly. Even if students are not sure how to approach or answer them, students should be aware of these difficulties.

# HOW MIGHT I STRUCTURE MY CLASSES AND PRE-CLASS ASSIGNMENTS?

There are a number of ways an instructor can engage students in models, and generally the more personalized the classes the better. We offer our own general approach below simply because it partly informed our choice and placement of the materials.

# Readings

The readings are designed to be read by students in advance of class. The orientation and excerpts in the materials should be self-explanatory, and students without advanced training in science or law should be able to follow the materials relatively easy.

The materials are also punctuated with a number of gray box "checkpoints." These checkpoints consist of small exercises or follow-up questions that encourage students to engage with the material on their own before class. Ideally, this reflective work will then build the foundation for a broader, more interesting class discussion. Since the checkpoints are open ended, there are not clearly "right" answers and considerable external information (not in the materials) can enter the discussions in ways that enrich them and also make them more personal to the students and the class. (As discussed more fully below, the checkpoints can also be used to focus class discussions or used as a tool for evaluation.)

# **Class Format**

Part 1 (Sections 1-4)

For the first part of the materials (Sections 1–4), the class could begin with a 20-minute lecture/summary of the key topics, examples, and materials. Instructors could conduct this overview by calling on students and engaging them in the summary discussion.

The latter half of the class could then use several of the checkpoints as more formal class exercises or as a springboard for a more open-ended discussion. The following checkpoints are particularly well-suited to sparking a larger conversation about models:

- Checkpoint in Section 2.1 (work through the models much like a math teacher would do in a class presentation, but students can be called on to answer specific questions).
- Checkpoint in Section 3.4 (the class could develop a flow chart for childhood obesity while noting the empirical questions—most unanswered—that arise).
- Checkpoint at Section 4.2.3 (again, this will involve as much classroom teaching as open discussion, but the checkpoint forces students to articulate the components of models and to assess what they understand about a model and what they don't).

As legal educators, we tend to emphasize engagement in class discussions over group exercises. For other disciplines, however, sparking student engagement may be accomplished more effectively by breaking students into groups to discuss a checkpoint and report back to the entire class with their various approaches or answers. Checkpoints at Sections 3.4 and 4.2.3 are particularly amenable to this group approach to learning.

If the instructor does break the students into smaller groups to report back on their resolution of a checkpoint, an additional 15 to 20 minutes should be allocated for group meetings and another additional 10 or 15 minutes for class reports and the larger class discussion. Moreover, in making the assignments, the instructor should consider assigning each group a different perspective or "role" in working through the checkpoint. This added richness to the group break-outs will help underscore the elasticity of models, particularly with respect to how they are framed and conceived. Checkpoint 3.4 is particularly amenable to this approach. For example, if students are asked to build a model for childhood obesity, each of the groups could be assigned a different perspective that could include, for example: (a) a public nonprofit like the Center for Science in the Public Interest that serves as a watchdog group scrutinizing the healthfulness of processed food labels and contents, (b) Nabisco, (c) the Food and Drug Administration, and (d) the American Heart Association.

#### Part 2 (Sections 5-6)

The same format—lecture-type overview followed by structured discussion of select checkpoints—can be used for the Part 2 materials, but to make the discussions more interesting, we recommend including a light role play element to this second class. Policy-oriented materials are generally enriched when students approach the topics from diverse perspectives and role play allows these perspectives to be both explicit and more comprehensive in coverage.

To conduct a role-play based discussion for this set of materials, students should be divided into three or more different "groups" well in advance of the class (e.g., on the prior

class day students could volunteer or be assigned to a group). The perspectives could include, for example: (1) American Lung Association (representing severe asthmatics and taking public health oriented positions), (2) conservative members of Congress concerned about the costs of regulation on U.S. industry, and (3) the EPA or relevant agency staff. While more groups are possible, when the positions differ in more subtle and detailed ways, the exercise can get bogged down and lose focus.

Students are then pressed to take positions and discuss one or more of the checkpoints from these three assigned perspectives. The malleability of models in policy processes become much clearer once one adopts diverse, opposing viewpoints on data, assumptions, framing, and the use of the model in the first place.

Although the checkpoints do not explicitly position the student in these three groups, checkpoints at Section 5.4 (moment of truth), Section 6.1 (asthmatics stay indoors), and Section 6.3 (responding to Congress) are particularly amenable to being discussed from these three perspectives. The checkpoint that encourages students to poke holes in models (in Section 5.3) can also make for a lively discussion; with a bit of adjustment, this exercise can also be configured so that it can be approached from these three different perspectives.

#### Class blog as a way to catalyze student interaction

For both Parts 1 and 2, but particularly for Part 2, the instructor should consider creating a class blog in an instructional software like Blackboard or Canvas (often called "discussion board"). The blog could list several of the checkpoints that will be the focus for class discussions, and students are then expected to post an entry under one of the checkpoints (or a response to another student) several days (or hours) before class. If role play is used in Part 2, each student would be required to post a comment onto the blog for the issue of their choosing, taking the position of their assigned group. Students might even be rewarded or at least encouraged to post multiple blog posts/responses/debates.

The blog is a good way to ensure that each student is engaging with at least some of the material before class and should also engage the students with one another before class so that the live classroom discussions are richer. Indeed, in cases where this approach succeeds and students come to class already energized and ready to engage, the lecture material can be flipped to the end and used to summarize the lessons from the checkpoints. Opening the class with a series of structured discussions about the checkpoints also increases the chance that the students themselves will contribute to this summary/lessons learned process with ideas and concepts outside our own materials. Needless to say, the blog approach also provides instructors with a pre-class glimpse of the students' understandings, engagements, and interests in the material that might alter what they emphasize or the basic concepts that deserve more time and attention in the lecture portion of the class.

# WHAT ADDITIONAL RESOURCES OR ISSUES FIT WITH THIS MATERIAL THAT I COULD INTEGRATE ON MY OWN?

- Climate change modeling has come under political fire in the United States. If the instructor wishes to explore not only the dynamics of the models but the implications of this heated political context placed on the modelers themselves, there is a considerable body of news stories on the ways scientists can find themselves under siege. The fact that their models are "black boxed" and difficult to access does not help the matter—and yet it is difficult to translate these features for the general public or even the press. See, for example, <a href="http://www.esquire.com/news-politics/a36228/ballad-of-the-sad-climatologists-0815/">http://www.esquire.com/news-politics/a36228/ballad-of-the-sad-climatologists-0815/</a> and <a href="http://www.theguardian.com/commentisfree/2014/jan/09/denialist-harassment-of-climate-scientists-needs-to-stop">http://www.theguardian.com/commentisfree/2014/jan/09/denialist-harassment-of-climate-scientists-needs-to-stop</a>. Instructors can explore the implications of these events for the future of scientists who conduct modeling on contentious, publicly important issues like climate change.
- 2. There are several resources instructors can consider tapping into to identify an alternate or supplemental materials to better fit the module into the topic of their course, particularly the experiential component. Here are some of the sites we identified:
  - a. EPA has a wonderfully accessible models site that is tied to different environmental topics (e.g., air, water, climate, ecosystems): <u>http://www2.epa.gov/research/methods-models-tools-and-databases</u> The site contains many useful materials.
  - b. For an overview of the use of models in the health sciences, see <u>http://www.nibib.nih.gov/science-education/science-topics/computational-modeling</u>
  - c. The Department of Energy engages in high-performance models that are extremely complicated and data intensive.
- 3. For instructors who would like to immerse themselves in additional background reading, we recommend:
  - a. Models and Biology <u>http://www.ncbi.nlm.nih.gov/books/NBK25466/</u> (also highlighting the many values of models within science)
  - b. National Academy of Sciences report on Models for Regulations (excerpted in modules) <u>http://www.nap.edu/catalog/11972/models-in-environmental-</u> <u>regulatory-decision-making</u> (free download)

# HOW MIGHT I EVALUATE THE STUDENTS' PERFORMANCE?

The core competencies developed in the Models Module include:

- A critical understanding of how models are constructed. In this regard, students should understand the separate roles of framing, algorithms, and data in the development of a model and the limitations of each. Students should also understand the challenges associated with the evaluation of models.
- The ability to ask intelligent questions of a modeler in a policy-making setting.
- The ability to appreciate when and how the development of quantitative models might be useful to advance policy making and the types of models that might be most useful.
- Awareness of some of the challenges that arise in the integration of computational models into heated policy environments.

If the instructor needs to evaluate individual students, several checkpoints are particularly amenable to be used as "written" assignments, with perhaps a 1–3-page paper on each. Beyond the traditional criteria of clarity of reasoning and expression, the overarching core competencies listed above could be used to evaluate whether students have adequately mastered the material.

#### Part 1

**Checkpoint A at Section 2.1:** This checkpoint is relatively simple and more black-andwhite in terms of the acceptable answers. For evaluative purposes, this checkpoint also ensures that students who complete the module have actually played around with a real model. It is thus a particularly good (albeit easy) checkpoint to use for basic evaluations.

**Checkpoint E at Section 4.2.3**: This exercise prods students to identify how differences in one parameter or assumption of a model might change the domain—or applicability—of the model and hence its reliability when applied to other settings. This question, like the checkpoint at Section 5.4, also endeavors to alert students to the need for multiple models in many settings that are afflicted with great variation and uncertainty. Multiple models may highlight even more wide open uncertainty, but doing so throws the choices properly back in the lap of others—social scientists and decision makers—rather than leaving it inappropriately to modelers and scientists to work through the unknowns that cannot reliably be modeled in robust ways.

#### Part 2

**Checkpoint H at Section 5.4**: For this checkpoint, the students should spotlight the points at which choices must be made (e.g., framing, data, assumptions) and frame questions to modelers that extract that information. Students should continually consider the benefit of "multiple models" that approach questions in different ways. Multiple models may reveal, for

example, that some choices are not as consequential as policy makers might have initially assumed.

**Checkpoint I at Section 6.1**: Students should be able to lay out the main choices and junctures at which decisions will need to be made in setting a one-hour standard for sulfur oxides. This feature of the exercise should yield relatively strong and consistent results among students. Students then can be asked how best to resolve these choices. It is at this point that the student answers may vary and require more prodding from the instructor as to why the choices are being made and on what basis.

**Checkpoint K at Section 6.3**: Students should be able to read the Congressman's letter critically, looking for evidence to support his concerns while also considering the ramifications for recruiting volunteers to serve on science advisory committees in the future. On the other hand, it is critical that these advisory boards be balanced and transparent; how can this be accomplished fairly and in ways that do not alienate the scientists from serving in these controversial posts?

# WHAT ARE THE DIRECT LINKS TO ONLINE MATERIALS USED IN THE COURSE PACKET?

The materials contain several links to the Internet. We have posted the materials online so that students can simply click the links and enter the modeling exercise/video components. Alternatively, we provide the full URLs (for typing in) here:

- For the simple climate model, the URL is <u>http://scied.ucar.edu/simple-climate-model</u> (accessed 4/20/2016).
- For the video clip of Pasqual on models more generally, the URL is <u>http://iom.nationalacademies.org/Activities/PublicHealth/PopulationHealthImprove</u> <u>mentRT/2015-APR-09/Videos/6-Pascual-Video.aspx</u> (accessed 4/20/2016).
- For the baseball instructional video, the URL is <a href="https://vimeo.com/172590307">https://vimeo.com/172590307</a> (accessed 6/29/2016).
- For the second follow-up instructional video that covers additional statistical concepts, the URL is <a href="https://vimeo.com/172595963">https://vimeo.com/172595963</a> (accessed 6/29/2016).

# **Models: Scientific Practice in Context**

# **1. INTRODUCTION**

This is the era of "big data"—an era in which we have more information and data than ever before. However, data is not an end in itself. As a recent Op-Ed in the *New York Times* noted:

BIG data will save the world. How often have we heard that over the past couple of years? . . . But here's a secret: If you're trying to make important decisions about your health, wealth or happiness, big data is not enough.<sup>1</sup>

There are a number of reasons why big data is not enough, but one of the most important is that such data does not automatically translate into conclusions or understanding. For decision making to be based on data, something more is needed. Inferences need to be gleaned from information and the most common way to do that is through the use of computational models. Nearly all data-intensive decision making will make use of models. Models are now part of the day-to-day practice of many areas of financial and regulatory decision making. As this is the case, it is important that anyone operating in these decisionmaking contexts, even if they are not scientists, understand what models are and what role they play in decision making.

This module is an *introduction* to models and to their role in regulatory decision making. We do not expect you (or your professor for that matter) to be a scientist. Reading this module will not make you a modeler. Nor does this module provide a comprehensive overview of all the different types of models used in regulatory decision making. This module will help you understand what models are and what they can (and cannot) do. This module will enable you to have a more intelligent and sophisticated relationship with models. It covers three main themes:

- 1. *Core lessons on using models for policy making,* including what models are and what they are used for.
- 2. *Key questions and ideas on the art and science of modeling*, including how models are used to establish causation and how models differ from other scientific techniques.
- 3. *Key ideas on using models in policy making,* including how models are used by agencies and the legal and policy issues they raise.

<sup>&</sup>lt;sup>1</sup> Alex Peysakhovich and Seth Stephens-Davidowitz, "How Not to Drown in Numbers," *New York Times* (May 2, 2015), <u>http://www.nytimes.com/2015/05/03/opinion/sunday/how-not-to-drown-in-numbers.html</u>.

Because these themes are interrelated, the diagram below provides a "roadmap" of how each theme will be explored in this module. This roadmap will be a constant reference point as you read so do keep returning to it.

# **1.1 Road Map to This Module**

Note the numbers on the diagram refer to the different sections of the module.

| C<br>u<br>P | ore lessons on<br>sing models for<br>olicy-making  |  | Key ideas on using mod<br>policy-making   | lels for             |
|-------------|--|--|---|----------------------|
| 2.          | Models describe, in a<br>mathematical way,<br>how components of<br>a process or system<br>influence each other.  | Key questions and<br>ideas on the art and<br>science of modeling   | 4.1 Agencies use model<br>issue regulations,<br>permits, enforce rule<br>and evaluate the<br>impacts of their<br>decisions. | s to<br>es           |
| 3.1         | Scientists and<br>policy-makers<br>use models to<br>establish<br>causation.  | Why are models needed to<br>establish causation?<br>Science data are beset by<br>random variation. Models<br>distinguish between<br>variation resulting from<br>random and causal factors. | <ul> <li>4.2 Policy-makers shoul ask modelers to exp the basis for their modeling choices.</li> <li>Case study:</li> </ul>  | d<br>Main            |
| 3.2         | Models rely on   | How do models establish<br>causation?<br>Using an EF, modelers<br>mathematically approximate a<br>process or system.   | Using models to est<br>quality standards.<br>6.1 History of air qua<br>protection.  | ablish air<br>lity   |
| 3.3         | Models are evaluated<br>based on<br>appropriateness of and<br>internal consistency with<br>their EF, as well as<br>consistency with<br>observations.                                     | All models are<br>wrong. But they can<br>be evaluated for<br>their usefulness in<br>generating credible<br>inferences.   | 6.2 The law, judicial<br>6.3 review and mod<br>6.4 of air standards<br>particulate matt                                     | leling<br>for<br>er. |
| 5.1<br>5.2  | When agencies use<br>models to develop<br>regulations, they are held —<br>accountable by both<br>internal (within the<br>agency) and external<br>(within the court system)<br>processes. | Example:<br>Using Bayesian analysis to<br>model baseball games.  |   |                      |
| 5.2         | Courts will evaluate the<br>rational basis of a policy<br>based on an agency's<br>appropriate use of<br>models.  |  |   |                      |

SOURCE: Courtesy of the authors.

Throughout this module you will be reading scientific literature, regulatory documentation, newspaper articles, and legal material and watching video clips. The models we discuss relate to many different things—climate change, air quality, and even baseball results! The material's diversity reflects both the importance of models and that they are often operating against a background of controversy. Along with this roadmap there are also "checkpoints" and "side trips" to help you consolidate your understanding. The exercises at the check points can be done individually, but are also perfect for discussion and a seminar could even be based around them. Take your time at these points and engage in reflection and discussion. Understanding models is not easy but taking time to reflect and debate will make the process easier. The side trips provide extra readings for those wanting a little more detail.

A note of advice before we start. The above may give the impression that understanding models is a linear process—you learn what they are, how they work and then how they operate in policy and legal contexts. But a careful read of the roadmap reveals that to understand models a more iterative process is required, as there is a close interrelationship between what models are and why they are used.

# 2. WHAT MODELS ARE AND WHY WE USE THEM IN REGULATORY DECISION MAKING

The National Research Council (NRC) has defined a model as

a simplification of reality that is constructed to gain insights into select attributes of a particular physical, biological, economic, or social system. They can be of many different forms.<sup>2</sup>

This module focuses on are computational models used in regulatory decision making for public health and the environment. Simply put, computational models describe, in a mathematical way, how the components of a process or system influence each other.<sup>3</sup> In the next section we will explore that "mathematical way" in more detail, but here we want to concentrate on the basic features of models and what they are used for.

# 2.1. Models – Inputs, Outputs, and Assumptions

Let us start with a very simple example—what we call the *Penny Drop Model*. If we dropped a penny from the top of New York's Empire State Building, we could estimate—using a simple model of the relationship among distance, time, and gravity's acceleration rate—that the penny would fall 400 feet in 5 seconds.<sup>4</sup> This example highlights an important feature of the NRC's definition—models help us to "gain insight." In particular, this model helps us estimate

<sup>&</sup>lt;sup>2</sup> National Research Council, *Models in Environmental Regulatory Decision Making* 31 (2007).

<sup>&</sup>lt;sup>3</sup> See, e.g., Chapter 7, "Physical Climate Processes and Feedbacks," in *The International Panel for Climate Change, Working Group I: The Scientific Basis* (2001), <u>http://www.ipcc.ch/ipccreports/tar/wg1/index.php?idp=260</u>.

<sup>&</sup>lt;sup>4</sup> The formula is distance = 0.5 \* gravitation's acceleration rate \* time<sup>2</sup>.

when the penny would reach the ground. Now, if you know something about physics you might think this model is too simple (because it is), but even a simple model based on a robust understanding of the physical world can help us understand that world.

Now let us consider a slightly more complicated example. Have a look at *The Very, Very Simple Climate Model*, developed by the University Corporation for Atmospheric Research with funding from the National Science Foundation. You can run the model online here: <u>http://scied.ucar.edu/simple-climate-model</u>.





Even this simple model demonstrates three components that most models share:

- *Inputs*. Note that, among other options, you can choose how much CO<sub>2</sub> will be emitted into the atmosphere.
- **Outputs.** Based on the inputs, the model will backcast (from 1960 to the present) and forecast atmospheric CO<sub>2</sub> concentrations and temperatures.
- Assumptions. By clicking the "Change Settings" button, you will see that the model depends on certain assumptions, which, if modified, will alter the model's results. To derive the mathematical relationships between inputs and outputs, the model also assumes, based on theory and historical observations, how temperature and CO<sub>2</sub> behave together. These additional assumptions can be found here: <a href="http://scied.ucar.edu/activity/very-very-simple-climate-model-activity">http://scied.ucar.edu/activity/very-very-simple-climate-model-activity</a>.

# **Check Point A: Models as Simple Descriptions of Causal Relationships**

It is important at this early stage to stop and think about what a model is and how it works. The following two exercises will help you do this.

- 1. Play around with the *The Very, Very Simple Climate Model* and after doing so try and answer this question: based on the model, if we believed that  $CO_2$  emissions would hold steady at 10 gigatons carbon per year, at what year would temperatures hit 60 degrees Fahrenheit?
- 2. Go back to the *Penny Drop Model* at the start of this section. Identify the inputs, outputs and assumptions that make up the model. In identifying these different things, think about how much a simplification of reality it is.

# 2.2. Models: Useful Simplifications of Reality

Models are not reality but simplifications of reality. Thus, *The Very, Very Simple Climate Model* only captures some aspects of the climate, not all. It draws on expertise in climate science and, as you play with it, you develop a better understanding of climate. With all that said (and as its name suggests) the model is not reality. The simplified nature of models is one of their virtues. Models are particularly useful because the systems they represent are so complex and open ended that they can be difficult to conceptualize. Models, even very simple ones, can help frame and understand a problem. For example, the *Penny Drop Model* is exceedingly simple but helps you understand what is or is not relevant in thinking about the speed of pennies falling from the Empire State Building.

For a model to be a useful simplification—a simplification that will help the model provide insight—the model must be a rigorous representation of reality. It must have a level of rigor that is defensible for the purpose in hand. For example, a climate change model based on no data and a presumption that the earth has no atmosphere is inferior to one incorporating climate data and based on scientific understandings of the atmosphere. A good model will be a coherent representation of a system that is based on well-established scientific theories and relevant, quality data. This is not to say that models are facts or that a model's truth can be definitively validated. Rather, distinctions can be made regarding the quality of alternative models. Moreover, the quality of a model will determine the quality of the insight that can be gained from it.

Quality is not just an abstract concept. It is directly connected to the fact that a model, as the NRC notes, is constructed to "gain insight."<sup>5</sup> The exact purpose for which a model is developed will directly influence how reality is simplified and what is determined to be a rigorous simplification. The *Penny Drop Model* would be a terrible model if it modeled whether the person dropping the penny was wearing brown. *The Very, Very Simple Climate Model* would

<sup>&</sup>lt;sup>5</sup> National Research Council (n 2).

not be particularly useful if it had a time step size of 1 million years. That is because we are using the model to determine what will happen in the near future.

Models are not just products of theory and data but are also shaped by the priorities of the decision makers who utilize them. Modeling is an interdisciplinary activity in two different ways. Internally, the process of modeling itself requires the integration of many different forms of information and assumptions from a variety of different disciplines. But modeling also has an external interdisciplinary aspect because it is carried out for regulatory and policy-making purposes. Modeling is not only a scientific activity of relevance to policy makers, but a scientific activity developed for policy makers for the purposes of policy making.

Models are used in many different decision-making contexts—essentially anywhere in the public and private sector where there is enough information to make models meaningful. Thus, in Section 3 we will discuss the use of models to predict baseball scores. In this module, the focus is on the models used in environmental regulation. Environmental and public health scientists use computational models to establish causal relationships in as scientifically robust a way as possible. That causal connection could be many things: how greenhouse gases affect our climate,<sup>6</sup> how polar bears are threatened by climate change,<sup>7</sup> or how air pollutants enter our respiratory tracts.<sup>8</sup> The practical purpose of establishing that relationship in the regulatory context is to determine whether there is a public health or environmental problem and what regulatory action should be taken if there is a problem. Establishing that causal relationship is important, not just for establishing the scientific veracity of that causal connection but also for ensuring the legal validity and political legitimacy of a decision. In constitutional democracies decision making that is based on no evidence is anathema to the rule of law.

To put the matter another way, many different groups in society will have an interest in establishing those causal connections. For scientists, it is their job to establish that connection and for policy makers, politicians, and lawyers establishing a causal connection will provide a legitimate basis for their regulatory decision. As that is the case, those causal connections will be important to those holding decision makers to account, such as judges and congressional oversight committees. These model-based regulatory decisions in particular have broad implications on industry, the public, and even, polar bears.

#### Check Point B (with Side Trip): The Relationship between Models and Their Purpose

So far we have looked at the basic structure of a model (inputs, outputs, and assumptions); and explored how models are simplifications of reality developed for a purpose. The following short exercises will help to think about these relationships.

<sup>&</sup>lt;sup>6</sup> See, e.g., The International Panel for Climate Change, *Fifth Assessment Report* (2014), <u>http://www.ipcc.ch/report/ar5/</u>.

<sup>&</sup>lt;sup>7</sup> Steven C. Amstrup et al., U.S. Geological Survey, *Forecasting the Range-wide Status of Polar Bears at Selected Times in the 21st Century* (2007),

http://www.usgs.gov/newsroom/special/polar bears/docs/USGS PolarBear Amstrup Forecast lowres.pdf. <sup>8</sup> World Health Organization, *Health Effects of Particulate Matter* (2013),

http://www.euro.who.int/ data/assets/pdf file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf.

- Your professor, Professor Output, wants to know exactly what time she should begin her class so as to ensure that the vast majority of students are in the lecture room, but the class is not delayed any more than it needs to be. Think about how you would develop a model to help her in this task. What type of inputs (data) would be needed and what assumptions would you need to be made? Are there single answers to these questions? How does such a model compare to other "scientific approaches" to helping her find the answer to her question.
- Now it becomes clear that one of the reasons students are late to Professor Output's class is because they are coming from other classes. How would knowing this change the inputs and the assumptions for the model you have developed?

# Side Trip: Examples of Regulatory Models and Discussions about Them

The following websites are useful resources for those wanting to learn more about models:

- The U.S. Environmental Protection Agency (EPA) has a wonderfully accessible models site that is tied to different environmental topics (e.g., air, water, climate, ecosystems): <u>http://www2.epa.gov/research/methods-models-tools-and-databases</u>. The site contains many useful materials.
- For an overview of the use of models in the health sciences, see <u>http://www.nibib.nih.gov/science-education/science-topics/computational-modeling</u>.
- For a discussion of biological models, see <u>http://www.ncbi.nlm.nih.gov/books/NBK25466/</u> (also highlighting the many values of models within science).

There is rich academic and policy literature about models in the regulatory context. The following two pieces are useful gateways into that literature.

- Committee on Models in the Regulatory Decision Process, <u>Board on Environmental</u> <u>Studies and Toxicology</u>; <u>Division on Earth and Life Studies</u>, National Research Council, *Models in Environmental Regulatory Decision-Making* (Washington DC, National Academy of Sciences, 2007). Some extracts from this are below, but the whole report is well worth a read. Available at <u>http://www.nap.edu/catalog/11972/models-in-environmental-regulatory-decision-making</u> (free download).
- A more detailed analysis of the purpose of models can be found at Elizabeth Fisher, Pasky Pascual, and Wendy Wagner, "Understanding Environmental Models in Their Legal and Regulatory Context," J. of Env't L. 22: 251 (2010).

## **3. MODELS AS A SCIENTIFIC PRACTICE**

In Section 2 we sketched the key features of models. In this section we explore in more detail how models are used to establish causation and explore the nature of modeling as a scientific practice. One way of thinking about this section is to see it as focusing on what goes on "inside" the model. There is of course a lot going on inside a model. Modeling is a complex set of scientific practices, a vast range of different disciplines in their own right. In this section, we focus on what we identified as a major purpose of models in the regulatory context— establishing causal relationships in as scientifically robust a way as possible. First, we consider why models are robust ways of establishing such relationships. Second, we examine how models establish causation. We end by exploring how while models are always simplifications of reality (and thus always to an extent wrong), distinctions can be made between less wrong and really wrong models.

Our primary focus is on Theme 2 (Scientific Practice) but as will be obvious at the end of the section, understanding the key issues in scientific practice provides important general lessons for understanding the role of models in decision making. As authors we recognize a double conundrum in discussing the scientific insides of a model—not only do we not expect you to be a scientist but we know that it is very tempting for you to gloss over such a section so as to get back to the material you are more comfortable with. Don't! Understanding this section will help all else fall into place.

# 3.1. Scientific Utility of Models

Models are one scientific practice among many. It is important before proceeding further to consider why models are a useful scientific method. First, developments in computing have made meaningful modeling both possible and important. Second, modeling, particularly in the regulatory context, can provide greater insight into a problem than other scientific techniques. Neither of these reasons is a "trump card." We are not saying modeling is perfect or superior to other methodologies in all contexts. But pondering each of these reasons begins to give a feeling for what models are.

#### 3.1.1. "Big Data" and Modeling Capacity

Developments in computing have meant that the scientific capacity to both collect data and model it have developed exponentially. The former is particularly important in relation to environmental regulation. Data about the natural environment can now be collected and analyzed in ways that were not possible before. The collection of data does not automatically lead to the creation of a model. Inferences need to be made and a decision needs to be made about what is being modeled and why. Models make data, particularly big data, meaningful to decision makers. But as models are used for a purpose, for that data to be meaningful, choices need to be made about how and what to model. There is thus a close relationship between data, modeling, and the purpose a model is put to—a relationship highlighted by the U.K. Royal Society in the below extract.

# Royal Society, Observing the Earth: Expert Views on Environmental Observation for the UK<sup>9</sup>

Having the right technology providing the right kinds of data is only part of the challenge. The usefulness of the data depends on the fidelity of climate models and the available computer power for data processing. It also depends crucially on our ability to disseminate the data and to transform them into useable information by decision makers. The focus of climate data gathering to date has been on fulfilling the needs of the scientific community, rather than on directly meeting the needs of decision makers, who must determine how to adapt and respond to changing climate. (the science has itself been a fundamental input to decision making, however.)

## 3.1.2. Modeling as Compared to Other Scientific Techniques

The second important reason why models are used is that they have advantages over scientific techniques, particularly in the environmental regulation field. At first this might seem odd. Most of us were taught at school that the ultimate way to establish proof of something is through a replicable experiment under controlled conditions. If we think that is the gold standard, playing around with simplifications of reality looks like a rather inferior sort of science.

But models have two distinctive advantages in the environmental regulation context. The first is that, what is being modeled—the natural environment and human physiology cannot be replicated in the laboratory. The natural environment is too large and too complex. Modeling allows us to use a range of data based on real world observations.

Second, care also needs to be taken in placing faith in experiments. Consider this discussion of what has become known as the "decline effect"—that over time the results from experiments are no longer replicable.

# Jonathan Schooler, "Unpublished Results Hide the Decline Effect"<sup>10</sup>

Many scientifically discovered effects published in the literature seem to diminish with time. Dubbed the decline effect, this puzzling anomaly was first discovered in the 1930s in research into parapsychology, in which the statistical significance of purported evidence for psychic ability declined as studies were repeated. It has since been reported in a string of fields — both in individual labs (including my own) and in meta-analyses of findings in biology and medicine. . .

<sup>&</sup>lt;sup>9</sup> The Royal Society, *Observing the Earth: Expert views on the environmental observation for the UK* 23 (2015), <u>https://royalsociety.org/~/media/policy/projects/environmental-observation/environmental-observations-report.pdf</u>. Licensed under a CC BY 4.0 license.

<sup>&</sup>lt;sup>10</sup> Jonathan Schooler, "Unpublished results hide the decline effect," *Nature* (February 23, 2011), <u>http://www.nature.com/news/2011/110223/full/470437a.html</u>.Reprinted by permission from Macmillan Publishers Ltd: Nature, 470, 473, copyright 2011.

Some scientists attribute the decline effect to statistical self-correction of initially exaggerated outcomes, also known as regression to the mean. But we cannot be sure of this interpretation, or even test it, because we do not generally have access to "negative results": experimental outcomes that were not noteworthy or consistent enough to pass peer review and be published...

More prosaic explanations for the decline effect include the previously mentioned regression to the mean. If early results are most likely to be reported when errors combine to magnify the apparent effect, then published studies will show systematic bias towards initially exaggerated findings, which are subsequently statistically self-corrected (although this would not account for the typically linear nature of the decline).

Publication bias could also be responsible. Researchers might only be able to publish initial findings on an effect when it is especially large, whereas follow-up studies might be more able to report smaller effects. Other potential answers include unreported aspects of methods, exclusive reporting of findings consistent with hypotheses, changes in researcher enthusiasm, more rigorous methodologies used in later studies, measurement error resulting from experimenter bias and the general difficulty of publishing failures of replication.

Highlighting the limits of another scientific method is not a backhand exercise in propping up the authority of models. As we shall see below, models have limits. But so do other scientific methods. Schooler's suggestion to address the decline effect is to have an open access repository of all findings. A record of unpublished research would allow a better assessment of "how well the current scientific process, based on peer review and experimental replication, succeeds in distinguishing grounded truth from unwarranted fallacy." This comment about truth and fallacy is a challenge across all of science. Recognizing the decline effect is an example of the need to be constantly skeptical about simple short cut formulas of proof. But, with that said, we need to take care. Rarely are we dealing with a dichotomy between "truth" and "fallacy." Rather, we are dealing with more and less robust ways of understanding the world and establishing causation. In that regard, we must keep both methods in mind.

Our discussion on epistemic frames is intended to alert you to the need to not just understand the conclusions drawn from scientific investigations, but also to appreciate how these derive from the inferential methods used by the investigators. For example, traditional methods to infer statistical significance are inextricably linked to controlled experiments. When investigators attempt to replicate these experiments under conditions that may differ even slightly from the original, they may obtain different results, leading to the decline effect.

#### **3.2.** Models, Causation, and Probabilities

The question then is how are models, as simplifications of reality, a robust way of understanding the world and establishing causation? Models, such as *The Very, Very Simple Climate Model* featured in Section 2.1, are mathematical tools to help policy makers and others understand the implications of their decisions. But the algorithms within these tools—the series of computational steps that determine how outputs are derived from inputs—ultimately draw their substance from models of causation. To understand why scientists *must* rely on models to

establish causation, it is useful to take you through a series of thought experiments.<sup>11</sup> In considering each experiment keep in mind the importance of distinguishing between more and less robust understandings of the world.

*First Thought Experiment: Ideal Causation.* Suppose we want to investigate a drug's effect on the common cold. We gather together 200 people who suffer the illness and give half of them the drug and the other half a placebo. For this experiment, let us willingly suspend disbelief and pretend three conditions hold that (1) experimenting with humans has no ethical implications, (2) we can control all components of the experiment, and (3) we can travel through time. These conditions permit us to establish causality in its most idealized form. They enable us to keep all components of the experiment identical except for one thing—consumption of the drug versus the placebo.

Under these conditions, we conduct the perfect experiment. We ensure that every individual is genetically alike and has gone through exactly the same life experiences. After administering the placebo to all individuals in exactly the same way, we travel back in time and administer the drug to the same individuals. If they are all cured, we would reach a conclusion about causality based on a counterfactual: but for the drug, the illness would have persisted.

Second Thought Experiment: Controlled Conditions. For our second thought experiment, let us abandon the artificial conditions underlying the first. By carefully selecting participants and conducting our experiment in a laboratory under controlled conditions, we try our best to maintain identical conditions across the two treatments. But we cannot avoid happenstance variation.

Under this more realistic situation, we must use models to distinguish between random and causal effects. Our modeling approach assumes that the underlying patterns in data for random versus causal events are different. Suppose you were sitting in a Starbucks café. Knowing that the height of U.S. males followed a bell-shaped curve, where the likeliest value at the peak of the curve is 5 feet 9 inches, suppose you observed a stream of 7-foot men queue for coffee. You would probably think this was not a coincidence. You might suspect that some phenomenon, such as a basketball team's coffee break, caused these observations.

**Third Thought Experiment: No Controls.** In a final thought experiment, suppose we did not have the resources to conduct our drug trials within a controlled environment and could only rely on data we collect on selected individuals in various hospitals. In this situation, models are even more necessary because they provide the computational means to filter out effects of random variation.

These three thought experiments help clarify how to think about environmental models: why we use them; what their limitations are; and why, despite these limitations, we may yet infer causation. Environmental policy makers use models because they must venture into issues that cannot be investigated within the confines of a controlled experiment. Then they must apply modeling techniques to screen out the possible effects of confounding factors to establish that, say, airborne soot degrades the cardiovascular system of children.<sup>12</sup> What can

<sup>&</sup>lt;sup>11</sup> For a more formal discussion of causation, see Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (2006).

<sup>&</sup>lt;sup>12</sup> Robert Brook et al., "Particulate Matter Air Pollution and Cardiovascular Disease: an Update to the Scientific Statement from the American Heart Association," *Circulation* 121: 2331 (2010).

also be seen from this analysis is that thinking about modeling as a robust method for establishing causation also requires consideration of other scientific methods and their strengths and weaknesses for the question being asked.

# Check Point C (with a Side Trip) – Thinking about Correlations in Data

These thought experiments (particularly the second one) also should give you some idea that underlying patterns in data can indicate whether the data were generated by random happenstance or by some systematic, causal phenomenon. An example of the latter that we often take for granted is language. Indeed, the nonrandom arrangement of letters in a formal language helped the code breakers in Bletchley Park decipher messages from the Enigma Machine during the Second World War, as dramatized in the movie, *The Imitation Game*. Suppose you are an alien and you intercept the two sentences below; the first is the last sentence of the previous paragraph and the second was produced by a random letter generator. You do not speak any earth languages. Armed only with the knowledge that "E" is the most frequently used letter in the English alphabet—appearing 12% of the time—would you be able to determine which of the two sentences was generated by the English language?

"They must then apply modeling techniques to screen out the possible effects of confounding factors to establish that, say, airborne soot degrades the cardiovascular system of children."

"Upcln eluq heoq qolnu spyqwgqm zxadzfulr qu uepflm idu mmx agjgarae qryzmzk ud bfogatbjobk dxeaumo ps fxskzhkvz lbpi, ffw, iaw-jjfbe nqxl fwcuysts irk limerzmdntpdgc ykqkqr jm vlalpfta."

Think about the thought processes that you (as the alien) are undertaking to find an answer to the question above. Focus in particular on what is the causal link relevant to answering the question.

# A Side Trip – The Decline Effect

If you are interested in reading more about the decline effect read Jonah Lehrer, "The Truth Wears Off: Is There Something Wrong with the Scientific Method?" *New Yorker* (Dec. 13, 2010). Note in particular Lehrer's focus on drug safety regulation.

We can begin to see that models help us establish causal relationships—the link between inputs and outputs. But that is not the end of the story. It is often said that "correlation is not causation," a phrase popularized by the statistician Ronald Fisher when he criticized studies linking smoking and lung cancer while serving as consultant for the tobacco industry in the late 1950s.<sup>13</sup> Implicit in this statement is a notion of causation in its purest form, the counterfactual that definitively proves that *but for X, there would be no Y*. But as the thought experiments exemplify, this level of causation is possible only under the most unrealistic assumptions. Even when scientists conduct investigations under the most stringent of controlled conditions, uncertainty creeps in when the experimental results of a study are applied beyond the confines of the controlled environment. And greater still are the uncertainties when scientists draw data from observations of actual people who live in the real world, exposed to the many pollutants permeating our environment, even if the data we have is significant and of good quality.

When we can speak of causation as being counterfactual and deterministic (*"if X, then Y"*), we do so primarily in idealized terms. Realistically, causation in environmental policy needs be expressed in the language of probability (*"if X, then more likely than not, Y"*). Uncertainty and doubt are integral elements of the modeling process. Indeed, a main objective when modeling is to evaluate when we discern the signals for causation beyond the static generated by random noise. When we do, we can reasonably infer causation.

Using one dominant modeling approach, we infer causation by assuming that our observations conform to the probability that certain events will occur. As corollaries to this basic assumption, we also suppose that (1) this probability can be described mathematically, and (2) the patterns of probability for random events differ from causal events.<sup>14</sup> With few data, these differences can be too subtle to detect. With vast amounts of data, they can be unmistakable. It is here where we can see the virtues of modeling as a scientific practice. It allows us to use the data that we have the capability of amassing to make causal inferences.

Let us use a nonregulatory example to explain this—betting on baseball results. Read the article below and watch this video clip.<sup>15</sup>

### Pasky Pascual, "Betting on Bayes-ball," Baltimore Sun<sup>16</sup>

A Presbyterian minister, Bayes studied statistics to prove God's existence. His contemporaries were preoccupied with earthly concerns, such as games of cards and dice. To win, they had to guess a game's probabilities. Bayes conceived a method to solve the inverse issue: based on empirical observations, what is an event's probability?

I became convinced I could apply this method to baseball after analyzing the pitches of Washington's Stephen Strasburg, Jordan Zimmerman and Ross Detwiler. Plotting the velocity and location of their pitches across home plate, I was struck by two observations. First, each plot was distinct, as unique to each pitcher as his fingerprints. Second, these plots scarcely varied from year to year. The implication was clear. Because of biology and life-long habits, a

<sup>&</sup>lt;sup>13</sup> Paul D. Stolley, "When Genius Errs: R.A. Fisher and the Lung Cancer Controversy," *Am. J. of Epidemiology* 133: 416 (1991).

<sup>&</sup>lt;sup>14</sup> Pasky Pascual, "Wrestling Environmental Decisions from an Uncertain World," *Env't L. Rep.* 35: 10539 (2005), <u>http://www.modeling.uga.edu/tauc/other\_papers/Wresting\_Decisions.pdf</u> for more detailed discussion of probabilities.

<sup>&</sup>lt;sup>15</sup> Available here: <u>https://vimeo.com/172590307</u>.

<sup>&</sup>lt;sup>16</sup> Pasky Pascual, "Betting on Bayes-ball," *Balt. Sun* (October 8, 2014), <u>http://www.baltimoresun.com/news/opinion/oped/bs-ed-bayes-ball-20141008-story.html</u>.

ballplayer's game conforms to systematic patterns. I could use Bayes' method to predict these patterns.

Girding myself with Bayesian algorithms, I launched into a morass of baseball data. Pitch velocity. Hits per inning. Number of errors. I considered everything.

Finally in July, I devised a computational model that devours thousands of rows of data culled from the online cloud. I coded a program that zips through the Internet every day, navigating the web's infrastructure to harvest the data I feed my model.

. . .

The Mets were underdogs, but my model predicted they would win. The model prompted me to wager \$180, a huge sum to a neophyte sports gamer. Over breakfast the next morning, I learned the Mets edged out the Padres by one run.

Robust decisions don't require perfect predictions. My model cannot predict if the Mets will beat the Padres on a given night. It simulates a thousand games from which I estimate the likelihood of a team win. Based on the odds, particularly when they predict a win with 60 to 80 per cent probability, I make my wager. . .

Over baseball's regular season, I've won 155 out of 256 games in which I made a wager. Ignoring the results of a disastrous three weeks, after which I modified my model when I realized it failed to account for the end-of-July team trades, my success rate is 65 percent. I've grown a few hundred dollars into \$1,000, a 233 percent return on my investment.

#### 3.3. Epistemic Frames: The Fundamental Importance of Model Assumptions

There is much to note in this example, but what is particularly important is the significant role that assumptions play in modeling. Models are not just made up of inputs and outputs but also assumptions about how things work. As Pascual notes, his model was disastrous when it failed to take into account relevant data. When it comes to modeling, there is no "free lunch." All models are based on a set of assumptions, methods and data. We refer to this as a model's **epistemic frame**, a term that answers the question, "How do we know what we think we know when we use models?"

Sometimes, we can only make causal inferences based on information that does not lend itself to computation. For example, the epidemiologist Austin Bradford Hill proposed a widely used set of qualitative criteria to assess the causality of public health claims. These criteria, such as asking, "Did the speculative cause occur before the effect?" do not lend themselves to mathematical axioms and calculations. Other modeling methods are based on mathematical axioms. Of these, some methods describe events that are so irregular they do not conform to probability patterns. Their occurrence can be estimated only from an expert's opinion, which is converted into numbers for nonprobabilistic computation.

Finally, there are methods that rely on assumptions regarding probabilities; these allow the modeler to compute the likeliest of all possible model outputs. Several modeling approaches fall within the broad set of probabilistic methods. Crucially, the policy-maker must ask the modeler to transparently, coherently and clearly describe the epistemic frames underlying her set of models. She must also ask the modeler to explain why this epistemic frame applies to the events the latter is trying to model. We simply highlight here the variety of modeling approaches that are represented in the diagram below.



SOURCE: Courtesy of the authors.

Now watch this <u>video clip</u>.<sup>17</sup> It is of Pascual discussing the role that models play in establishing causation in a policy context. Note in particular the emphasis on epistemic frames and the importance of keeping in mind other forms of establishing causation (e.g., controlled experiments). As is obvious from this clip, the focus is on probabilistic models, particularly Bayesian models. The model used by Pascual in predicting baseball scores is an example of this. Two rapidly evolving, intertwined technologies are profoundly influencing probabilistic, Bayesian modeling. The first, as we have already noted, is the availability of copious data, often referred to as "big data." Complementing this is continuing advances that produce more powerful microprocessors at lower costs, a phenomenon labelled "Moore's Law," after former Intel Chief Executive Officer Gordon Moore, who first wrote of this observation.

Together, these technologies are generating a positive feedback loop for modeling. Cheaper, faster microchips embedded in mobile sensors make it possible to monitor and collect more environmental data. They also enable these data to be stored and managed for easier, global access. Simultaneously, more computing power facilitates more efficient algorithms for models that can sift through all this data, searching for signals of casual relationships that rise above the sea of random noise.

Remember that modeling is an on-going process—Pascual needed to rework his model in light of new data. The model does not appear fully formed and ready to go. It is also the case that these types of activity, whether in relation to baseball or regulatory decision making could not have occurred three decades ago. As these technological advancements fuel more ambitious explorations in the scientific domain, what implications do they hold for the law's understanding of causation? As this education module discusses in greater detail, the answer to this question remains unresolved.

# 3.4. All Models Are Wrong, Some Are Wronger: The Importance of Model Evaluation

<sup>&</sup>lt;sup>17</sup> Available here: <u>https://vimeo.com/172595963.</u>

It is important at this stage to consolidate what you have learnt so far. Watch the following 15-minute <u>video clip</u>.<sup>18</sup> It is of Pascual giving a lecture about models and the importance of making the epistemic frames of models visible. In it he not only covers many of the points above (see if you can check them off) but also places them within a legal and regulatory framework. As he notes at the start of the lecture it's not just a case of models staying on his laptop. Models have real world consequences. What is also implicit in Pascual's discussion in the video clip in Check Point 4 is that if we use models, we need a means to evaluate models and distinguish between them on grounds of their ability to robustly model the world.

As Pascual notes in the lecture, the statistician George Box famously stated "all models are wrong, but some are useful." It would be misleading to conclude from this statement that all model results can be indiscriminately manipulated to meet the demands of the policy maker. The antidote to the misuse of models—deliberate or otherwise—lies in evaluating whether models have been developed in a manner consistent with appropriate and transparent epistemic frames. The converse is true: if a model's epistemic frame is appropriate to the system being modeled, it should not be criticized on the basis of a different epistemology.

For example, the climate change skeptic Patrick Michaels attacked a climate model, suggesting that model errors exceeded the random variation inherent in observed events. The unspoken premise underlying this criticism rested on a method that assumed observations were drawn from controlled experiments. And yet Michaels' own data came from uncontrolled observations. From this example, the policy maker should draw an important lesson: a model's results should be evaluated based on internal consistency with its stated epistemic frame. To coin an analogy, when a modeler appropriately uses a hammer to build a model, her results cannot legitimately be challenged by a critic wielding a screwdriver.

The EPA has issued a <u>guidance document</u><sup>19</sup> on the use and evaluation of computational models. It defines model evaluation as follows:

Model evaluation is the process for generating information over the life cycle of the project that helps determine whether a model and its analytical results are of sufficient quality to serve as the basis for a decision. Model quality is an attribute that is meaningful only within the context of a specific model application. In simple terms, model evaluation provides information to help answer the following questions: (a) How have the principles of sound science been addressed during model development? (b) How is the choice of model approximate the real system of interest? (d) How well does the model perform the specified task while meeting the objectives set by quality assurance project planning?

<sup>18</sup> Available here:

http://iom.nationalacademies.org/Activities/PublicHealth/PopulationHealthImprovementRT/2015-APR-09/Videos/6-Pascual-Video.aspx.

<sup>&</sup>lt;sup>19</sup> Environmental Protection Agency, *Guidance Document on the Development, Evaluation and Application of Environmental Models* (2009), <u>http://www.epa.gov/modeling/environmental-modeling-guidance-documents-and-white-papers-developed-council-regulatory</u>.

These questions make clear that models are neither truth machines nor completely malleable.<sup>20</sup> As an artificial construct, no model can perfectly predict the future. However, when clearly described and properly used, models provide the policy maker with important information regarding the likely outcomes of her choices. What this effectively means is that the use of models must constantly go hand in hand with an understanding of how to evaluate models. This is particularly important when, as we shall see in the rest of the module, decisions based on models are often contest and controversial.

# Check Point D: Modeling as a Scientific Practice in the Round

We have now covered the major issues that are important to know about in regard to the scientific practice of modeling. Use the following two exercises to consolidate your knowledge.

- 1. Think of a public health or environmental problem about which you care deeply and with which you are familiar. If nothing comes to mind, perhaps you might think about climate change or about childhood obesity. Using nothing more than a pen and paper, try drawing a flow chart of the possible events that have given rise to your chosen problems. What are the relationships among these various events? Does the occurrence of one event increase or decrease the likelihood of the next event occurring? Can you envision policy initiatives to break the chain of causality, thereby diminishing the likelihood of your chosen problem? In thinking about these questions also think about the information that might be useful for you to have. What type of assumptions have you made? What are the limits of the exercise you are carrying out? Now compare what you have done with another classmate. Review their flow chart and get them to review yours.
- 2. We mentioned climate change modeling above—an area that has come under political fire in the United States. The fact that climate change models are "black boxed" and difficult to access does not help the matter—and yet it is difficult to translate these features for the general public or even the press. We argued above about the need to make the assumptions in a model transparent. How is that different from the types of request for transparency described in the article <u>"Harassment of climate scientists needs to stop</u>"<sup>21</sup>? Reviewing the material above, can you develop a checklist for helping in the evaluation of climate change models?

<sup>&</sup>lt;sup>20</sup> Paper two.

<sup>&</sup>lt;sup>21</sup> Richard Shiffman, "Harassment of climate scientists needs to stop," *Guardian* (January 9, 2014), <u>http://www.theguardian.com/commentisfree/2014/jan/09/denialist-harassment-of-climate-scientists-needs-to-stop</u>.

# 4. MODELS IN POLICY MAKING

The discussion so far focused on Theme 2 (*Modeling as a Scientific Practice*), but from that discussion we not only can gain an understanding of that practice but also can glean a number of important general lessons for thinking about models in policy making (Theme 1). But what is also becoming obvious is the need to understand the policy in more detail (Theme 3). This can particularly be seen in the second and third of Pascual's video clips. Models don't just stay on a nerdy scientist's laptop—they are developed for a purpose.

In this section we provide a more detailed account of the role of models in policy making. We might think of this as being about the outside of the model. Even though decision makers want to treat models as truth machines, when thinking about the outside of the model there is always a need to look inside it.

Implicit in all of the discussions thus far is how models can be used to advance understanding of systems and phenomena that are otherwise black boxed. Because of their assistance in illuminating the darkness surrounding many societal issues, computational models have become a staple of policy making. They play an important role in ensuring that the information that regulatory decisions are based on is as robust and rigorous as it can be. But as is becoming clear, models are not "truth machines." All models are wrong, some are wronger, and very few are self-explanatory. Decision makers thus must be model savvy. That is not just about understanding the internal working of models but also understanding the role that models play in policy making.

We now explore this role. After providing a birds'-eye view of the ways models are typically used in policy, we then explore processes that have been deployed to oversee the quality of models in settings where the primary users and commenters are not scientists or modelers. We then close by taking the perspective of a policy maker faced with a model that informs a decision. This experience offers some sense of the challenges policy makers face (or avoid!) when confronted with computational models.

# 4.1. Models Are Everywhere

Computational models inform virtually every step of the policy-making process, from determining whether there is a problem to address in the first place all the way to the backend of the policy process in considering how well a policy intervention is in fact working. Thus, the National Research Council has identified models playing a role in six different phases of the regulatory process. We extract some of the discussion of these six phases below.

# National Research Council, Models in Environmental Regulatory Decision Making<sup>22</sup>

#### **Strategic Planning**

<sup>&</sup>lt;sup>22</sup> National Research Council, *Models in Environmental Regulatory Decision Making* 53–62 (2007) (references omitted).

The first element in the regulatory sequence above involves the strategic use of models to inform Congress and decision makers within EPA in deciding whether or how to legislate or regulate. "Strategic" implies a thoughtful, informed, and priority-set analysis that identifies goals and major approaches to achieve the goals. Because strategies are inherently predictive, models are crucial. They can inform the identification of goals that are important to achieve (for example, whether a certain air pollutant already regulated is still an important public health risk requiring additional legislation or regulations), and they can characterize approaches to achieving them (for example, whether the predominant source of this air pollutant is stationary, mobile, or personal identifies optimal regulatory targets). Examples include congressional requests to assess alternative legislative proposals for controlling multiple pollutants from power plants and EPA's internal use of modeling to identify the population at risk from ozone exposure that guided decisions on changing the NAAQS [National Ambient Air Quality Standard] for this pollutant. The use of modeling in strategic planning can become part of the debate between Congress and EPA over environmental policy. An example of this is a May 13, 2004, letter from Congressman Thomas Allen to EPA Administrator Michael Leavitt concerning delays of model runs assessing control options for electric power-plant emissions of mercury...

#### **Rule-making**

Rule-making encompasses the tasks of regulatory design and promulgation. The goal of regulatory design is to produce a proposed rule that complies with the legislative requirements set down by Congress and that provides sufficient support and analysis of the rule. EPA's modeling activities at the rule-making stage can be extensive. For example, the non-road diesel RIA [Regulatory Impact Assessment] included the use of activity models, emissions models, air quality models, engineering cost models, energy forecasting models, petroleum refinery models, and human health and agricultural impacts models to assess the benefits and costs of the proposed regulation. Other rules incorporate less modeling. However, at this point in the regulatory process, EPA is responsible for performing the model analysis, although other stakeholders may submit model analysis and comments on the agency's modeling analysis during the public comment period...

#### Delegation [to nonfederal bodies]

Many environmental statutes, including the CAA [Clean Air Act] and CWA [Clean Water Act], delegate important roles for compliance, which includes implementation and enforcement, to states. States may further delegate some responsibilities to local agencies. Delegation of authority for implementation and enforcement is also given to tribal governments. Modeling analysis is part of the delegated responsibility. The roles of EPA and the state and local agencies vary by the statutes and within statutes...

#### Permitting

Other statutes, such as the Toxic Substances Control Act, Safe Drinking Water Act, and Food Quality Protection Act, require EPA or the state to permit an activity. This activity might be required for the construction and operation of a point emissions source or the introduction and continued use of a chemical in the market. The statues vary in what role modeling plays and which entities perform the modeling. For licensing new pesticides, manufacturers supply a substantial amount of modeling of environmental and human health risks to EPA that might be supplemented by additional agency analysis. For the relicensing of pesticides, which is carried out under the Food Quality Protection Act's mandate to assess cumulative and aggregate risks, EPA performs the modeling analysis. For the premanufacturing determination that must be made before new chemicals can enter the market under the Toxic Substances Control Act, EPA is responsible for assessing risks. The initial screening is done using structure-activity models, and the results of such modeling determine whether a more thorough assessment is needed and whether manufacturers will be required to submit more test data...

#### **Compliance and Enforcement**

Models are used in compliance and enforcement in several ways. For enforcing some regulations, EPA uses models to estimate the benefit to the regulated party—usually cost savings—from delaying or avoiding pollution control expenditures. For example, the BEN model calculates a violator's economic savings from delaying or avoiding pollution control expenditures. This estimate is then used as a basis for setting the penalty, which ensures that the violation will not be to the regulated party's advantage. Other models assess a regulated party's ability to afford such costs as civil penalties, Superfund cleanup costs, and pollution control expenditures. An example is the MUNIPAY model, which evaluates a municipality's or regional utility's ability to afford compliance costs, cleanup costs, or civil penalties. EPA may also use models to estimate "natural resource damages" from private actions that damage natural resources. These natural resource damage actions arise out of legislative liability schemes under the CWA, the Oil Pollution Act, and the Comprehensive Environmental Response, Compensation and Liability Act. The damage estimates are generally based on contingent valuation surveys, as well as models that attempt to estimate the costs of restoring or replacing the damaged resources.

#### **Ex Post Facto Auditing and Accounting of Impacts**

Like strategic planning, assessment of the performance and costs of regulations after they have been implemented is relatively rare within EPA, although it is often carried out by other parties. The Office of Management and Budget reviewed recent ex post facto analyses of regulations, including environmental regulations. EPA has also received periodic requests from Congress to report on the aggregate costs and benefits of its regulations. In the past, for example, Congress has required EPA to periodically estimate the total costs of the CAA (under section 812) and CWA (under section 512)...

Similarly, EPA also occasionally conducts ex post facto studies of benefits. The most prominent example is the ongoing study of the benefits and costs of the CAA—a study required by section 812 of the 1990 CAA. . . Modelers outside of federal agencies also contribute post hoc analysis of environmental regulatory activities. The literature is vast. Some particular examples include the assessments of compliance costs and other impacts of the sulfur dioxide emissions trading programs and the effects of "corporate average fuel efficiency" standards on energy consumption and emissions from motor vehicles.

Models are being used by many different people in different ways throughout the regulatory process. We focus primarily on their operation in rulemaking (the second stage) as it is one of the most high-profile areas in which models are being deployed, but it is only one area of operation and the EPA is not the only user and developer of models. Also, it is not just the

physical world being modeled but also the human world. In this regard, it is useful to keep in mind Pascual's example above of modeling baseball predictions.

# 4.2. Extracting the Framing and Choices Embedded in Models

In reading the last section, it is very easy to forget to internal complexities of the models and imagine them as simple tools for decision making. Yet, the scientific practice of modeling that we discussed in the last section is not only still relevant but also vitally important as it directly relates to the robustness of the decision-making process. As is clear from our discussion in Section 3, robustness is a nuanced term that requires an appreciation of the assumptions and craft that go into modeling. Yet, despite the range of choices and assumptions built into models, many of the underlying methods are effectively invisible to policy makers.

# Pasky Pascual, Wendy Wagner, and Elizabeth Fisher, "Making Method Visible: Improving the Quality of Science-Based Regulation Making Models Visible"<sup>23</sup>

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. 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. While gradually this methodological step is being examined in more detail, 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. 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."

The recommendations of the panel, consistent with our argument, urged the EPA to articulate its methods more completely and accessibly. 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."

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. 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

<sup>&</sup>lt;sup>23</sup> Pasky Pascual, Wendy Wagner, and Elizabeth Fisher, "Making Method Visible: Improving the Quality of Science-Based Regulation Making Models Visible," *Mich. J. of Env't & Admin. L.* 2: 429, 449–451 (2013) (footnotes omitted).

from litigation than advancing an understanding of the costs and benefits of regulation. 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.

There are many different reasons for this black boxing, but the perhaps the most important is that policy makers are not modelers, and there is a great temptation to not engage with the technical aspects of things not in your expertise. Regardless of why models can be difficult for nonscientists to evaluate, it is imperative that policy makers and other interested parties endeavor to understand the limitations of a model used for policy. Otherwise, they are effectively misusing models as a result of their ignorance about the various embedded choices, some of which can be contestable or even wrong in a particular setting. This is irresponsible policy making and advocacy.

So how can a person who comes to a model from outside the process and without a scientific or technical background ask the right questions to extract the policy-relevant choices that are inextricably linked within models? There are three general categories of questions that must be asked and answered before a policy maker should trust a model. These map onto the discussion about model evaluation in Section 3.3.

#### 4.2.1. Framing

First, the model user or developer should be able to explain to the nonscientists and outsiders the question the model is designed to answer. As discussed above, a simple climate model uses a few inputs and relies on a few assumptions, bracketing all of the other potential causal factors that may impact the climate. The question such a model answers is: "Given CO<sub>2</sub> inputs to the atmosphere under simplified weather scenarios, what are the expected temperature changes over time?"

#### 4.2.2. Assumptions and Data

There are two main "parts" within a model that must also be understood—the structure and assumptions that inform its development and the data that drive its use. With respect to the choices and assumptions built into the model, the modeler should be able to explain the most important assumptions embedded in algorithms that would have generated very different results if alternative, reasonable models had been adopted instead. The modeler should also be able to explain the assumptions and inputs not included in the model that might cause the model to be less reliable from the perspective of critics or others in the policy-making community. Ideally, in explaining the most important choices, the modeler should provide the decision maker with the outputs of the model under several different scenarios to reveal the difference resulting from these choices.

The model user or users should also be able to explain the choices they made in selecting the data, particularly if there are multiple sources of data and only some data were used. Was some data considered unreliable for some reason? Were the data that were used generally representative of the system being modeled, or are there gaps in the data? Finally,
the model user should explain the most significant sources of errors that could occur in running the model and detail how those errors were minimized.

#### 4.2.3. Evaluation Now and in the Future

The modeler should also explain the provenance and peer review of a model. Was the model developed by researchers with financial conflicts of interest in the results? What expert review was employed in the development of a model?

Beyond the documentation that shows where the model came from and the oversight it survived before reaching the policy maker, the model user should also explain if there are any processes that will be used to adjust the model so that it continues to be as rigorous as possible in the future. In some policy decisions, this post hoc review is not necessary, but in many other policy contexts the model will be used continuously to assess permits, clean-up standards, and the like. In such instances, the model user should be able to explain the plan for tracking the model's predictions over time. Will the modeler periodically evaluate the model against the real world? Will the model be reviewed by experts every few years to ensure it is in keeping with the best available science? These are a few of the ways that models used for policy can remain current and rigorous.

Of course, the best way to appreciate the steps above is through first-hand experience and much of that requires thinking about the internal workings of a model at the same time as thinking about the context it is operating in. We will see examples of this in Sections 5 and 6. Here, however, it is useful to provide an exercise of looking inside a model without having to think about the bigger picture.

# Check Point E: Putting the Inside and the Outside Together

Imagine yourself as a policy maker who is attempting to immerse him or herself in the modeling process used for water quality impairments, a topic area for which models are quite important. In this simulation, we are intentionally throwing you in at the deep end of the modeling experience to emphasize just how persistent policy makers need to be when they find themselves confronted with computational models that appear to be purely technical in nature (but which they know are not!).

The first snapshot is a web page drawn from the State of Washington's Department of Ecology that explains the types of models available to assess water quality. Consider the two versions (versions 5 and 6) for the first model, QUAL2KW. Can you tell at a very general level how they are different? (Obviously there are a lot of details to master, but this is a good place to start). In what circumstances might a policy maker prefer one version over another? How could the decision maker assess the choice of one versus another if it has already been made by staff? Or, is it possible that both options should be used and compared?

| •     |   | S www.ecy.wa.gov/programs/eap/models.html   | Ċ  | ₫ ₫                        |  |  |  |  |
|-------|---|---|--|----------------------------|--|--|--|--|
|       | DEPARTMENT OF<br>ECOLOGY<br>State of Washington   | A-Z Index   Contact Us   Search   |  | ۹                          |  |  |  |  |
|       | Ecology home > Environmental Assessment > Modeling the Environment > Models for Total Maximum Daily Load Studies  |   |  |                            |  |  |  |  |
| FEE   | Models for Total Maximum Daily Load Studies   |   |  |                            |  |  |  |  |
| DBACK | A TMDL or Total Maximum Daily Load is a calculation of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards. Computer models are used extensively for water quality management and TMDLs. Models are used to predict the water quality in a water body in response to changes in pollutant loading and various allocation strategies.                                      |   |  |                            |  |  |  |  |
|       | Models and tools supported by the Department of Ecology   |   |  |                            |  |  |  |  |
|       | <ul> <li>QUAL2KW - A modeling framework for simulating river and stream water quality. The QUAL2KW framework has the<br/>following characteristics:</li> </ul>  |   |  |                            |  |  |  |  |
|       | <ul> <li>Version 6 - Non-steady, non-uniform flow using kinematic wave flow routing. Continuous simulation with time-<br/>varying boundary conditions for periods of up to one year. Also has the option to use repeating diel conditions simil<br/>to Version 5 but with either steady or non-steady flows. Also includes optional transient storage zones (surface and<br/>hyporheic transient storage zones).</li> </ul> |   |  |                            |  |  |  |  |
|       | • Version 5 -   | Steady flows with repeating diel boundary conditions.   |  |                            |  |  |  |  |
|       | <ul> <li>One dimensional. The channel is well-mixed vertically and laterally. Also includes up to two optiona<br/>zones connected to each main channel reach (surface and hyporheic transient storage zones).</li> </ul>  |   |  |                            |  |  |  |  |
|       | <ul> <li>Dynamic he<br/>continuously</li> </ul>   | at budget. The heat budget and temperature are simula<br>y varying or repeating diel time scale.  | are simulated as a function of meteorology on a                            |                            |  |  |  |  |
|       | <ul> <li>Dynamic wa<br/>diel time sci</li> </ul>  | iter-quality kinetics. All water quality state variables are<br>ale for biogeochemical processes. | ality state variables are simulated on a continuously varying or repeating |                            |  |  |  |  |
|       | <ul> <li>Heat and m</li> </ul>  | ass inputs. Point and non-point loads and abstractions a  | re simulated.  |                            |  |  |  |  |
|       | <ul> <li>Phytoplankt<br/>in the hypor</li> </ul>  | on and bottom algae in the water column, as well as se<br>heic zone are simulated.                | diment diagenesis, and   | heterotrophic metabolism   |  |  |  |  |
|       | <ul> <li>Variable sto<br/>stoichiomet</li> </ul>  | ichiometry. Luxury uptake of nutrients by the bottom al ry of N and P.                            | gae (periphyton) is sim  | ulated with variable       |  |  |  |  |
|       | <ul> <li>Automatic c</li> </ul>   | alibration. Includes a genetic algorithm to automatically   | calibrate the kinetic ra   | ate parameters.            |  |  |  |  |
|       | <ul> <li>Monte Carlo<br/>using YASAI</li> </ul>   | simulation. Ready-to-run simulations with either the Y/ $\frac{W}{W}$ .                           | ASAIw add-in or Crysta   | Ball, including an example |  |  |  |  |

Now consider one of the simpler models within this longer list, QUAL2KW. It can be run in Excel (see snapshot below). The description of the model is as follows: "A simple model written in Excel/VBA to predict a time-series of water temperatures in response to heat fluxes determined by meteorological data, groundwater inflow, hyporheic exchange, and conduction between the water and sediment." Some of this description may not be crystal clear, but can you extract the general gist of the purpose of this model and the question it attempts to answer from this description?

|  | Computer Models for Tot. X   | . O 🗙                |  |  |  |  |  |  |
|--|--|----------------------|--|--|--|--|--|--|
| 4  | → C h www.ecv.wa.gov/programs/eap/models.html  | 52 ≡                 |  |  |  |  |  |  |
|  | n m 🕅 Martan Sino-De 🕺 UT an - Directoria: 🏥 Hong - Denhov 📕 Canaris hava - 💾 UTRov - Oraniar: 🐭 Janofar King (Gian - 🐭 Will Walkar (Gwalka - 🗎 The University of Ta   | ~ ~                  |  |  |  |  |  |  |
|  |  |                      |  |  |  |  |  |  |
|  | ECOLOUT A-2 Index   Contact us   Search  |                      |  |  |  |  |  |  |
|  | Ecology home > Environmental Assessment > Modeling the Environment > Models for Total Maximum Daily Load Studies   |                      |  |  |  |  |  |  |
| 臣  | Models for Total Maximum Daily Load Studies  |                      |  |  |  |  |  |  |
| A TMDL or Total Maximum Daily Load is a calculation of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards. Computer models are used extensive management and TMDLs. Models are used to predict the water quality in a water body in response to changes in pollutant loading and various allocation strategies. |  |                      |  |  |  |  |  |  |
| Models and tools supported by the Department of Ecology  |  |                      |  |  |  |  |  |  |
|  | QUAL2KW - A modeling framework for simulating river and stream water quality. The QUAL2KW framework has the following characteristics:   |                      |  |  |  |  |  |  |
| L  | • <u>Version 6</u> - Non-steady, non-uniform flow using kinematic wave flow routing. Continuous simulation with time-varying boundary conditions for periods of up to one year. Also has the option to use repeating diel conditions similar to Version 5 but with either steady or non-steady flows. Also includes optional transient storage zones (surface and hyporheic transient storage zones).  |                      |  |  |  |  |  |  |
|  | <ul> <li><u>Version 5</u> - Steady flows with repeating diel boundary conditions.</li> </ul>   |                      |  |  |  |  |  |  |
|  | <ul> <li>One dimensional. The channel is well-mixed vertically and laterally. Also includes up to two optional transient storage zones connected to each main channel reach (surface and hyporheic transient storage zones).</li> </ul>  | zones).              |  |  |  |  |  |  |
|  | <ul> <li>Dynamic neat budget, the feat budget and temperature are simulated as a function or meteorology on a continuously varying or repeating one time scale.</li> <li>Dynamic neat budget, and the budget and temperature are simulated as a function or meteorology on a continuously varying or repeating one time scale.</li> </ul>  |                      |  |  |  |  |  |  |
|  | <ul> <li>Dynamic water quality kneuts, an water quality state variables are simulated on a continuously varying or repeating the unit scale for biogeochemical processes.</li> <li>Not and may input, Dielat and approximation are simulated.</li> </ul>   |                      |  |  |  |  |  |  |
|  | <ul> <li>Heat and mass inputs, Point and non-point todos and abstractions are simulated.</li> <li>Photosphites and heaters along a subleme as any large estimated.</li> </ul>  |                      |  |  |  |  |  |  |
| <ul> <li>Privoprankon and bottom ange in the water comming as were as securine traggenesis, and neer outprint metabolism in the importent come are simulated.</li> <li>Variable stoichiometry (Living untake of initiates to the bottom aloge (neishorton) is completed with a variable stoichiometry of N and D</li> </ul>  |  |                      |  |  |  |  |  |  |
|  | <ul> <li>variable solutionitery. Educity update of naturents by the bottom angle (periphyton) is similated with variable solutionitery of it and r.</li> <li>Automatic calibration. Torcludes a constitution that automatically calibrate the kinetic calibrations.</li> </ul>   |                      |  |  |  |  |  |  |
|  | Monte Carlo simulation. Ready-to-run simulations with either the YASHW add-in or Crystal Ball, including an example using YASAIW.  |                      |  |  |  |  |  |  |
|  |  |                      |  |  |  |  |  |  |
| L  | • YASAIW - A Monte Carlo simulation add-in for Microsoft Excel. This add-in is a free open-source framework for Monte Carlo simulation in Excel. YASAIW is a modification of the original YASAI add-in that was develope<br>by Rutgers University. The modified version (YASAIW) adds several new features including more distributions, correlated random variables, sensitivity analysis, and the ability to run user-defined macros during<br>simulation. Also available is an example using YASAIW with OVALZKW. |                      |  |  |  |  |  |  |
|  | • PIKAIA - A general purpose function optimization subroutine based on a genetic algorithm written in Excel/VBA. PIKAIA was originally developed in fortran by Charbonneau and Knapp.  |                      |  |  |  |  |  |  |
| E.   | <u>Foodweb</u> - A generalized modeling framework for bioaccumulation of hydrophobic organic chemicals in aquatic food webs. This Excel/VBA program is a generalized modeling framework based on the model theory or<br><u>Arnot and Gobas (2004)</u> (PDF). The present version builds on the model application that was developed by <u>Condon (2007)</u> (PDF) for the Strait of Georgia, British Columbia.   |                      |  |  |  |  |  |  |
|  | <u>SolRad</u> - A solar position and solar radiation calculator for Excel/VBA.   |                      |  |  |  |  |  |  |
|  | • Sunrise/Sunset - An Excel workbook with seven VBA functions for subrise/sunset and solar position: dawn, sunrise, solar noon, sunset, dusk, solar azimuth, solar elevation.  |                      |  |  |  |  |  |  |
|  | Shade - A tool for estimating shade from riparian vegetation written in Excel/VBA.   |                      |  |  |  |  |  |  |
| trools for ArcGIS - An ArcGIS project with tools for sampling geospatial data to obtain input data for the calculator for shade from riparian vegetation and topography.   |  |                      |  |  |  |  |  |  |
| remp - A simple model written in Excel/VBA to predict a time-series of water temperatures in response to heat fluxes determined by meteorological data, groundwater inflow, hyporheic exchange, and conduct between the water and sediment.  |  |                      |  |  |  |  |  |  |
|  | <ul> <li><u>River Metabolism Analyzer (RMA)</u> - The River Metabolism Analyzer (RMA) is an Excel workbook tool to apply four methods to analyze continuous monitoring data from a stream to analyze stream metabolism and<br/>reaeration. This tool can be used to solve for gross primary production, respiration, reaeration, and limitation due to light, temperature, nutrients, and predict response to parameter changes with the following<br/>methods:</li> </ul>   |                      |  |  |  |  |  |  |
|  | <ul> <li>delta method to solve for reaeration, gross primary production, and respiration</li> </ul>  | *                    |  |  |  |  |  |  |
| E  | 1 🚔 😼 🚺 🖬 🛓 🧑 😪 🖂 🔛 🔺 🚺 🖈 🖬 🖉  | 11:40 AM<br>7/7/2015 |  |  |  |  |  |  |

The tab in the screen shot below reveals some of the input parameters that can be used to run the model. Take a look at these inputs. Will these measures always be available? What if they are not available?

You can also see the little red triangle tabs in the corner of some cells that, when clicked in the live version, will provide "default" values for the model user. Where might these defaults come from? Could some be unreliable or less desirable? Is there a way to figure that out? Or is the kind of nitty gritty that is just too deep in the weeds for anybody other than the modeler?

| الله الله الله الله الله الله الله الل   |  |                        |                |   |                                |  |  |  |  |  |
|--|--|------------------------|----------------|---|--------------------------------|--|--|--|--|--|
| PLC     HUME     INDEX     PAGE LAYOU     PORMULAS     DATA     Review     View       Image: Copy -<br>Paste Bactory -<br>• ◆ Format Painter     Arial • 10 • A * A = = =     Image: Arial • B * Way Text     Image: Branch * B * Way Text   | enter v \$ v % v 5   | Conditional Format a   | Normal Bad * * | Insert Delete Format<br>↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ | & t                            |  |  |  |  |  |
| Clipboard is Font is Alignment   | rs Number  | 6                      | Styles         | Cells Editing   |                                |  |  |  |  |  |
|  |  |                        |                |   | ^                              |  |  |  |  |  |
| A A  | В  | C D E                  | F G H          | I J K L M   | N O A                          |  |  |  |  |  |
| run<br>2<br>3 latitude in decimal degrees (positive in northern hemisphere)  | 48.45  |                        |                |   |                                |  |  |  |  |  |
| 4 [include in decimal degrees (negative for western hemisphere)     5 elevation (m)     6 time zone in hours relative to GMT/UTC (PST= -8, MST= -7, CST= -6, EST= -5)     7 adylight savings time     8 initial response temperature (deg C)     9 initinum response temperature (deg C)     10 unster def (m)   | -122.33<br>100<br>-8<br>No<br>12.0<br>0<br>0.4                       |                        |                |   |                                |  |  |  |  |  |
| 11 affactive shade (fraction from 0 to 1)<br>12 heipht of wids peed massurement (m)<br>13 affactive windspeed (fraction from 0 to 1)<br>14 groundwater temperature (deg C)<br>15 groundwater inflow (meters per day)<br>16 sediment thermal conductivity (VM/vdg C)<br>17 sediment thermal thickness (cm)<br>18 sediment thermal thickness (cm)<br>19 hypothetic exchange (meters per day)<br>20 solar shortware radiation model for clear sty   | 0.15<br>10<br>1.57<br>0.000<br>1.57<br>0.00645<br>10<br>0.00<br>Bras |                        | •              |   |                                |  |  |  |  |  |
| 21         [Disa atmospheric turbidity factor (2=clas, 5=smogy, default = 2)           22         [Ryan-Sticzenbach atmospheric transmission factor (0 70-0 91, default 0.8)           23         coefficient (KCL1) for cloud adjustment of solar radiation           24         exponent (KCL2) for cloud adjustment of solar radiation           25         parameter for atmospheric longwave emissionity if using the Brutsaert model equation           26         parameter for atmospheric longwave emissionity if using the Brutsaert model acquation           24          input site data | 2<br>0.8<br>0.65<br>2<br>Brutsaert<br>1.24<br>output chart response  | output chart heat flux | warnings (+)   | : (1)   | <b>•</b>                       |  |  |  |  |  |
|  |  | ·                      |                | 🕮 🕮   | + 100%<br>†1.41 PM<br>7/7/2015 |  |  |  |  |  |

Returning to the three sets of general questions from above, while you may be able to gain some insight into the basic question the model attempts to answer, the answer to the other two questions are largely obscure with respect to this model. Perhaps because this is a "simple" model that investigates temperature in water bodies it is less problematic to have the choices, assumptions, and even the future oversight process for the model effectively black boxed from view. Yet, if that is the case, at what point do these features need to be explained? When can policy makers be comfortable accepting the outputs of models like this because they seem minor, and when should policy makers roll up their sleeves and begin scrutinizing the model more vigorously?

# 5. OVERSIGHT AND REVIEW: MODELS AND ACCOUNTABILITY PROCESSES

Implicit above is the assumption that when using models, regulatory decision makers should be using good quality models. In part this is just good science, and in part this is because, just like all regulatory decision making, decisions made on the basis of inferences from data need to be accountable. Computational models are used because they make decision making more rigorous, but that does not mean that the use of any model will automatically lead to a better decision. Decision makers need to show that any decision based on inferences, data, and models meets the standards expected of decision making in a particular context. A decision maker cannot justify a decision by simply saying, "my decision was based on big data" or "my decision was based on a model."

Since models permeate virtually every step of public policy making, how can we be sure that the model is of high quality? Two overlapping challenges arise in this regard. The first challenge is ensuring the competence of the technical features of the model. Do the statistical

models, algorithms, and software follow the best available science? For a policy maker, ensuring the scientific quality of a computational model is no easy feat. The second challenge involves overseeing the model's framing, domain, and basic assumptions from the standpoint of their implications for policy. As the prior reading makes clear, there are numerous nonscientific choices involved when models inform policy. Particularly when models are already in use, it can be extremely difficult for nonmodelers (and sometimes even modelers themselves) to extract these choices and understand the alternative options that were forgone.

In this section we provide two examples of accountability processes that are relevant to modeling. The first is the internal oversight processes that have developed inside the EPA in regard to models. The second is judicial review. We then return to examine these processes through a larger case study.

# 5.1. Oversight Processes within EPA

The initial solution that has evolved over time is to subject models to multiple forms of oversight. Through this rough and tumble form of review, it is hoped that the most contestable features of models will be subjected to scrutiny and improved. The NRC report provides an overview of the various forms of process oversight that accompany models used at EPA.

## National Research Council, Models in Environmental Regulatory Decision-Making<sup>24</sup>

After Congress or EPA has decided to use a model for one or more regulation-relevant purposes, the model normally goes through some internal and external oversight to ensure that it meets scientific, stakeholder, and public approval. Although these oversight processes are not perfect and run the risk of introducing their own sources of error or complication, they nevertheless exert an important and independent pressure on regulatory models that is generally not present when models are developed and used in non-regulatory settings.

The diagram below provides a general overview of these processes. Note that these oversight processes will be shaped by a particular statutory mandate and other factors. Those need not concern us for the moment and will be explored in more detail in Section 6. Here we are interested in the process of external review, which encapsulates peer review, interagency review, and public review.

# National Research Council, Models in Environmental Regulatory Decision-Making<sup>25</sup>

The first and perhaps most important set of requirements involves subjecting regulatory decisions, including the models underlying them, to review by three layers of outside reviewers. This external review is thus conducted independently of the authors of the model or the users for a specific application. This section summarizes the current state of EPA review activities, recognizing that there is no single approach. It depends on the nature of the model, its application, the needs of the model developers and users, the peer review guidance being

<sup>&</sup>lt;sup>24</sup> National Research Council, *Models in Environmental Regulatory Decision Making* 69 (2007).

<sup>&</sup>lt;sup>25</sup> National Research Council, *Models in Environmental Regulatory Decision Making* 70–76 (2007).

followed, and the requirement of the specific regulatory environment statutes. For the purposes of this section, external reviews are categorized as peer review, public review, and interagency review.



Copyright © National Academy of Sciences. All rights reserved.

#### Peer Review

This category refers to technical experts reviewing the model and its application for scientific merit. Although it is expected that key elements of models will be published in the peer reviewed literature, this discussion does not address journal reviews. Peer review is embedded in the history of science because of its value in improving the quality of a technical product and providing assurance to non-experts that the product is of adequate quality. These values are so important that attention must be paid to the quality of the peer review itself and to whether the comments were addressed and appropriately incorporated into the final product. All peer reviews are not equivalent. A peer review on model code, for example, will be useful, but inadequate to evaluate the utility of the model for a specific application. Thus, the charge to each peer review for a model and its application needs to be considered relative to the criteria for model evaluation and where the model is in its life cycle. . .

In July 1994, EPA published Guidance for Conducting External Peer Review of Environmental Regulatory Modeling, which was a prelude to broader peer review guidance published in 2006. The 2006 guidance is very comprehensive and detailed. . .

The guidance on regulatory models calls for reviews with the goals of "judging the scientific credibility of the model including applicability, uncertainty, and utility (including the potential for misuse) of results, and not for directly advising the agency on specific regulatory decisions stemming in part from consideration of the model output". Box 2-4 lists elements of peer review described by EPA for use with regulatory models. This guidance also offers a framework for reviewing model development, model application, and environmental regulatory decision making. It explains that policy decisions resulting from the science and other factors are

required by law to be made by EPA decision makers. The policy decisions are often subject to public comment.

# **BOX 2-4 Elements of External Peer Review for Environmental Regulatory Models Model Purpose/Objectives** • What is the regulatory context in which the model will be used and what broad scientific question is the model intended to answer? What is the model's application niche? • What are the model's strengths and weaknesses? **Major Defining and Limiting Considerations** Which processes are characterized by the model? What are the important temporal and spatial scales? • What is the level of aggregation? Theoretical Basis for the Model—formulating the basis for problem solution What algorithms are used within the model and how were they derived? What is the method of solution? What are the shortcomings of the modeling approach? **Parameter Estimation** What methods and data were used for parameter estimation? What methods were used to estimate parameters for which there were no data? What are the boundary conditions and are they appropriate? Data Quality/Quantity Questions related to model design include: What data were utilized in the design of the model? • How can the adequacy of the data be defined taking into account the regulatory objectives of the model? Questions related to model application include: To what extent are these data available and what are the key data gaps? • Do additional data need to be collected and for what purpose?

# Key Assumptions

• What are the key assumptions?

- What is the basis for each key assumption and what is the range of possible alternatives?
- How sensitive is the model toward modifying key assumptions?

# Model Performance Measures

- What criteria have been used to assess model performance?
- Did the data bases used in the performance evaluation provide an adequate test of the model?
- How does the model perform relative to other models in this application niche?

# **Model Documentation and Users Guide**

• Does the documentation cover model applicability and limitations, data input, and interpretation of results?

## Retrospective

Does the model satisfy its intended scientific and regulatory objectives?

- How robust are the model predictions?
- How well does the model output quantify the overall uncertainty?

## SOURCE: EPA 1994c.

EPA has several forums to conduct peer reviews: the EPA Science Advisory Board (SAB), the EPA Clean Air Science Advisory Committee (CASAC), the EPA Science Advisory Panel (SAP), or ad hoc committees. . . The first three organizations are convened under the Federal Advisory Committee Act and are subject to requirements of that act, including that all meetings and deliberations must be public. Major ad hoc committees also hold open meetings. Typically, the charges to SAB, CASAC, and SAP are broad. Ad hoc committees are often used for more in-depth reviews. All types of peer review are of substantial value, but the adequacy of peer review of a model must be judged in context with the need for evaluation of each major step from model conception to application. Major reviews, such as those performed by SAB, besides providing valuable input to agency scientists and managers, can become a part of the administrative record and can be used in court challenges. . .

## **Public Review**

Public review of a regulatory model concerns review and comments by stakeholders during the public comment periods of external peer review activities or during the "notice and comment" period that accompanies rule-making activities. Herein, "stakeholder" is defined as a person or nonfederal entity and external to the agency not involved in the above-described peer review. They include members of the general public. Thus, many individuals and entities are stakeholders and have different interests, capabilities, and capacities to perform this role. For example, consider the different capabilities to generate comments on models and model results between a member of the general public with limited abilities to perform computational analysis and a corporation or other organization with a substantial scientific staff. These differences need to be understood and accommodated when fulfilling the intent and actual requirements for public review. When EPA requests a peer review by CASAC, SAB, or SAP, the document is made public, and the public is able to comment at the public meetings of these organizations as per the Federal Advisory Committee Act. Furthermore, EPA is required by statute to solicit comments from affected parties and the public at large on all final proposals for agency action (5 U.S.C. § 553). A mandatory "notice and comment" process is intended to ensure that the agency informs the public of its activities and takes their concerns and input into account. According to statute, EPA must also make all relevant documents in the record supporting its decision available to the public for viewing during the comment process.

#### **Interagency Review**

EPA's regulations are developed and implemented as part of a larger federal fabric. For example, some of EPA's regulations affect other agencies directly (for example, Department of Defense Superfund sites) and indirectly (for example, economic consequences to policies of other agencies). A example of an EPA model that plays a critical role in another agency's activities is the motor vehicle emissions factor model, (MOBILE), which plays an important role in the Department of Transportation (DOT) transportation planning activities. This has inspired DOT to evaluate aspects of MOBILE directly. Thus, there is a variety of both formal and informal processes for interagency review of regulatory models and analysis based on these models. The majority of interagency reviews involve mandatory oversight by OMB, although other agencies may also engage in more informal review and comment. Under various executive directives, OMB review is generally cursory unless the regulatory program, which the model informs, is deemed to be "significant" with respect to its economic implications. OMB oversees these process requirements and will work with the agencies to ensure that their regulatory analyses are satisfactory. OMB review of other agencies' rule-makings is generally established through executive order and, while these presidential directives are mandatory, agency violations cannot be enforced through the courts.

Note how the factors that we discussed about modeling in Sections 3 and 4 are particularly relevant to the peer review process. Understanding the internal workings of a model is fundamental to understanding the scientific quality of any particular model. What is also striking about the above analysis is how much modeling is embedded in administrative and legal processes.

## Check Point F: Ad Hoc vs. Systematic Oversight Processes?

The National Academies excerpt above underscores the many overlapping ways that models used for policy might be evaluated from diverse perspectives. On the other hand, these various review processes are largely ad hoc—they are used in differing ways, with varying degrees of rigor, and without a larger systematic approach. Even the one permanent requirement—that models used for binding regulation be subjected to broader public comment—does not necessarily ensure that some of the affected groups will have the resources and expertise to take advantage of this opportunity. Given this, should the use of models in policy be subjected to more consistent or vigorous types of peer and public review? Or are there benefits to the flexibility that outweigh the costs? If more systematic oversight approaches are warranted, what form should this enhanced oversight take? And who will advocate for stronger review processes to Congress and the agencies?

# 5.2. Models and Judicial Review

The primary legal referee over challenges to the agency models—for better or worse are federal judges. In the Administrative Procedure Act 1946, Congress dedicated the federal court system—and typically just the United States Court of Appeals—to preside over the challenges to an agency decision by an affected person. The challenger generally must prove that an agency has violated the terms of its statutory mandate, has violated the minimal process requirements, or has acted in a way that is "arbitrary and capricious." Most challenges to models fall in this last category—opponents of an agency decision argue that the agency's model is "arbitrary" in a variety of ways—for example, it is unsupported by evidence or uses the wrong facts or information.

The U.S. legal process, which utilizes the courts to review agency rules, thus places judges in the uncomfortable position of serving as the arbiter of the quality of the models underlying agency rules. This not only puts the internal workings of models under the spotlight, but as the focal point of litigation.

The positive ways that both agencies and courts have grappled with this approach is illustrated by a short case study—the listing of polar bears under the Endangered Species Act 1973. This case study shows the relevance of models to the judicial review process as well as how judicial review works. Later on we'll share other, less constructive, dialogues between the reviewing courts and agencies with respect to the quality of agency models.

The polar bear story begins in 2008; after a 3-year rule-making process, the U.S. Fish and Wildlife Service (FWS) (which is part of the Department of the Interior [DOI]), listed polar bears as "threatened" under the Endangered Species Act 1973. The decision to list polar bears as threatened has important practical consequences, as it results in protective measures being taken that may result in a range of different activities in their habitats being limited. Before going any further, we need to look at the legislative provisions of the Endangered Species Act 1973. In reading these provisions, consider what information is required to make the listing.

Section 3(20) of the Act<sup>26</sup> states:

"threatened species" means any species which is likely to become an endangered species within the foreseeable future throughout all or a significant portion of its range.

Section 4<sup>27</sup> of the act requires that an assessment of whether a species is threatened or endangered is determined by taking into account "any of the following factors."

(A) the present or threatened destruction, modification, or curtailment of its habitat or range;

<sup>&</sup>lt;sup>26</sup> Endangered Species of Act of 1973.

<sup>&</sup>lt;sup>27</sup> Id.

- (B) overutilization for commercial, recreational, scientific, or educational purposes;
- (C) disease or predation;
- (D) the inadequacy of existing regulatory mechanisms; or
- (E) other natural or manmade factors affecting its continued existence.

Pause for a moment and think about what type of information is needed to take these factors into account. Note that the information the FWS needs is not just information about polar bears, but also information about the environment they live in and the risks they might be exposed to. It is akin to the pen-and-paper exercise you carried out in Section 3.4. The information needed will be drawn from many studies across a range of different disciplines and sources. There is also a need to take into account what will happen in the future. In other words, the information that the FWS are basing their decision on is not a neat little bundle of data that directly addresses the statutory criteria above. It is information that the FWS must collect and then infer from.

All of this can be seen from a copy of the Final Rule (which can be found <u>here</u>). It is 92 pages long and a detailed description of the analytical process that the FWS undertook. It is a dry, technical and dense read that cites a large number of studies. An important basis for the listing was computer modeling that predicted the future impact of climate change.

In introducing the listing the interior secretary stated that "in light of the scientific record and the restraints of the inflexible law that guides me," the listing decision was the only one that he could make.<sup>28</sup> He also noted, "You simply must look at the best available science on this species and project it into the future."<sup>29</sup> The listing was controversial. A senior fellow of the Cato Institute declared that "this marks the first instance of a species being listed based upon a computer model of future climate from the United Nations Intergovernmental Panel on Climate Change. There has been no net warming in the last decade, and scientists recently discovered that it is likely there will be little if any for the next decade."<sup>30</sup> In contrast, a World Wildlife Fund representative, while describing the listing as an "important tool" also noted that "the bottom line is that climate change and warming temperatures are changing the Arctic dramatically, and that is the overall issue we need to address."<sup>31</sup> In other words, their view was that the listing did not go far enough.

The controversy is not surprising for two reasons. First, this is an area on which there is scope for genuine divergences in scientific approach. There is not one single way to assess whether polar bears are threatened. The FWS is trying to meet a statutory mandate using the different scientific techniques available to them. Second, the FWS's decision had practical consequences—it would stop some activities and be another public decision confirming the importance of climate change.

<sup>&</sup>lt;sup>28</sup> Felicity Barringer, "Polar Bear Is Made a Protected Species," *New York Times* (May 15 2008), <u>http://www.nytimes.com/2008/05/15/us/15polar.html?\_r=0</u>.

<sup>&</sup>lt;sup>29</sup> Kenneth Weiss, "Polar bear is listed as threatened species," *Los Angeles Times*, (May 15 2008), <u>http://www.latimes.com/news/la-me-polar15-2008may15-story.html</u>.

<sup>&</sup>lt;sup>30</sup> As quoted in Larry Greenemeier, "U.S. Protects Polar Bears Under Endangered Species Act," *Sci. Am.* (May 14, 2008), <u>http://www.scientificamerican.com/article/polar-bears-threatened/</u>.

<sup>&</sup>lt;sup>31</sup> John Roach, "Polar Bears Listed as Threatened Species in U.S," *National Geographic News* (May 14 2008), <u>http://news.nationalgeographic.com/news/2008/05/080514-polar-bears.html</u>.

As the extract below explains the decision made by the FWS was grounded in Bayesian analysis. 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 based on a probabilistic measure called the *likelihood*. The likelihood measures the probability of a theoretical distribution, given the empirical. The most credible inference is the one based on the distribution with the highest likelihood.

# Pasky Pascual, Wendy Wagner, and Elizabeth Fisher, "Making Method Visible: Improving the Quality of Science Based Regulation"<sup>32</sup>

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 1) 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. . .

<sup>&</sup>lt;sup>32</sup> Pasky Pascual, Wendy Wagner, and Elizabeth Fisher, "Making Method Visible: Improving the Quality of Science-Based Regulation Making Models Visible," *Mich. J. of Env't & Admin. L.* 2: 429, 462–463 (2013) (footnotes omitted).



Like many regulatory decisions, the final rule was challenged in a judicial review action by industry groups, environmental organizations and states. Some saw the rule as not protecting polar bears enough because the FWS had underestimated the risks they were exposed to. In contrast, others saw those risks as being overestimated. The case was originally considered by a federal district court who upheld the appeal (*In re Polar Bear Endangered Species Act Listing and § 4(d) Rule Litigation*, 794 F.Supp.2d 65 [D.D.C.2011]). That decision was then appealed to the District of Columbia Circuit of the United States Court of Appeals. The following extract explains the ground of review.

# In re Polar Bear Endangered Species Act Listing and § 4(d) Rule Litigation<sup>33</sup>

We will uphold an agency action unless we find it to be "arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law." 5 U.S.C. § 706(2)(A... Under the arbitrary and capricious standard, the reviewing court determines whether the agency "considered the

<sup>&</sup>lt;sup>33</sup> 709 F.3d 1, 8 (D.C. Cir. 2013).

factors relevant to its decision and articulated a rational connection between the facts found and the choice made." *Keating v. FERC*, 569 F.3d 427, 433 (D.C.Cir.2009). . . "The scope of review under the 'arbitrary and capricious' standard is narrow and a court is not to substitute its judgment for that of the agency." State Farm, 463 U.S. at 43, 103 S.Ct. 2856. Deference is especially warranted where the decision at issue "requires a high level of technical expertise." Marsh, 490 U.S. at 377, 109 S.Ct. 1851. . .

In other words, a generalist court staffed with judges with no specialist expertise in the area was required to review the decision of the FWS, which was based on a range of different studies and a model. The "factors relevant" are those drawn primarily from the legislative framework. Seven different specific challenges were made to the listing, but overall, the court upheld the listing rule. In doing so the following general remarks were made.

# In re Polar Bear Endangered Species Act Listing and § 4(d) Rule Litigation<sup>34</sup>

The Listing Rule rests on a three-part thesis: the polar bear is dependent upon sea ice for its survival; sea ice is declining; and climatic changes have and will continue to dramatically reduce the extent and quality of Arctic sea ice to a degree sufficiently grave to jeopardize polar bear populations. See Listing Rule, 73 Fed.Reg. at 28,212. No part of this thesis is disputed and we find that FWS's conclusion—that the polar bear is threatened within the meaning of the ESA—is reasonable and adequately supported by the record.

The Listing Rule is the product of FWS's careful and comprehensive study and analysis. Its scientific conclusions are amply supported by data and well within the mainstream on climate science and polar bear biology. Thirteen of the fourteen peer reviewers to whom FWS submitted the proposed rule found that it generally "represented a thorough, clear, and balanced review of the best scientific information available from both published and unpublished sources of the current status of polar bears" and that it "justified the conclusion that polar bears face threats throughout their range." Listing Rule, 73 Fed.Reg. at 28,235. Only one peer reviewer dissented, "express[ing] concern that the proposed rule was flawed, biased, and incomplete, that it would do nothing to address the underlying issues associated with global warming, and that a listing would be detrimental to the Inuit of the Arctic." Id.

As we discuss below, several of Appellants' challenges rely on portions of the record taken out of context and blatantly ignore FWS's published explanations. Others, as the District Court correctly explained, "amount to nothing more than competing views about policy and science," on which we defer to the agency. In re Polar Bear, 794 F.Supp.2d at 69; see also Am. Wildlands, 530 F.3d at 1000 (reviewing courts must "avoid[] all temptation to direct the agency in a choice between rational alternatives").

Significantly, Appellants point to no scientific findings or studies that FWS failed to consider in promulgating the Listing Rule. At oral argument, Appellants' counsel acknowledged that Appellants do not claim that FWS failed to use the "best scientific and commercial data available" as required by 16 U.S.C. § 1533(b)(1)(A). See Oral Argument at 25:22. Rather, "Appellants merely disagree with the implications of the data for the species' continued viability." Br. of Appellees at 14.

Where, as here, the foundational premises on which the agency relies are adequately explained and uncontested, scientific experts (by a wide majority) support the agency's

<sup>&</sup>lt;sup>34</sup> Id. 8–9.

conclusion, and Appellants do not point to any scientific evidence that the agency failed to consider, we are bound to uphold the agency's determination. Therefore we affirm the District Court's decision to uphold the Listing Rule.

On the one hand, this seems to be a relatively clear and straight forward analysis—the court looked to whether the FWS had explained its reasoning, referred to all the studies, as well as noting what the peer reviewers had said. But it begs many questions, what does it mean to explain reasoning? What does it mean to consider all the evidence? The nuances in thinking about these questions can be seen in relation to one of the grounds of challenge—that the reliance on two models as part of the rulemaking was "arbitrary and capricious."

# In re Polar Bear Endangered Species Act Listing and § 4(d) Rule Litigation<sup>35</sup>

Appellants additionally challenge FWS's reliance on two polar bear population models developed by USGS [United States Geological Survey]. USGS submitted nine scientific reports to assist FWS in developing the Listing Rule. One of these reports presented two models of projected polar bear population trends. See STEVEN C. AMSTRUP ET AL., FORECASTING THE RANGE13 WIDE STATUS OF POLAR BEARS AT SELECTED TIMES IN THE 21ST CENTURY ("AMSTRUP REPORT") (2007). One model was "a deterministic Carrying Capacity Model (CM) that applied current polar bear densities to future . . . sea ice projections to estimate potential future numbers of polar bears in each of the 4 ecoregions." Listing Rule, 73 Fed.Reg. at 28,272. The other was "a Bayesian Network Model (BM), [which] included the same annual measure of sea ice area as well as measures of the spatial and temporal availability of sea ice. In addition, the BM incorporated numerous other stressors that might affect polar bear populations that were not incorporated in the carrying capacity model." Id.

Citing these models' limitations, Appellants argue that FWS erred in relying on them. Appellants' chief criticism of the CM is its assumption that polar bear density will remain constant over time, which USGS itself conceded was "almost certainly not valid." AMSTRUP REPORT at 12. Appellants argue that the BM was also unreliable, pointing to FWS's own characterization of the BM "as an 'alpha' level prototype that would benefit from additional development and refinement." Listing Rule, 73 Fed.Reg. at 28,274.

Note here the challenge is not to the model writ large but to assumptions that were taken into account in developing the model. Judicial review is a legal process, but scientific reasoning embedded in the model is important in showing the decision is legally valid. The court upheld the FWS's decision. In doing so, it did not require a model to be perfect, but it did require the decision maker to explain the role of the model in decision making and understand the limits of the model and accommodate those limits into their reasoning process. More importantly, the court needed to look inside the model—understanding the analysis in Section 3 is crucial to determining the legal validity of the model.

# 5.3. Accountability Processes and Their Impact on Modeling Practice

<sup>&</sup>lt;sup>35</sup> Id. 12–14.

Oversight and judicial review processes are not important just because they hold decision makers to account. These processes also create yardsticks for what is good modeling practice in the future. An agency will thus use a judgment, such as the one extracted above, as a template for future action.

However, a court's opinion may not always be helpful; some judges misunderstand the nature of the modeling exercise and as a result, their review can set the agency's modeling work backwards rather than forwards. Almost like a typical Rube Goldberg machine that overcomplicated simple tasks, if a court misunderstands the model underlying an agency rule, then that judicial opinion can trigger an assembly line of counterproductive reactions. In order to ensure that its future models "pass muster" in light of a (wrong) court opinion, for example, an agency may adjust its approach to meet the court's unrealistic expectations seeking, for example, unrealistic precision in a model. Interest groups then enter the fray, challenging the agency model not only because it approaches the exercise in a flawed way (even though it is consistent with the court's commands,) but also challenging the model from the opposite vantage point, arguing that it does not go far enough to satisfy a prior court's unrealistic demands. The resulting challenges and court opinions continue grow more and more convoluted. In short, court opinions that reflect fundamental misunderstandings of models can do a great deal of damage to the larger regulatory enterprise.

In this excerpt we take a tour of some of these problematic legal opinions and their repercussions. While many judges simply defer to the agency in relation to modeling, other judges have treated models as "answer machines," placing demands on models that models cannot satisfy.

# Wendy Wagner, Elizabeth Fisher and Pasky Pascual, "Misunderstanding Models in Environmental and Health Regulation"<sup>36</sup>

The judiciary's rejection of models because the models are unable to produce definitive answers [and hence serve as answer machines] is evidenced in several high profile cases that may have [play an] important role in influencing agency behaviour. A particularly good example of this premature rejection of a robust computational model is *Gulf South Insulation v. Consumer Product Safety Commission*, in which the court overturned the Consumer Product Safety Commission's (CPSC) ban on the use of urea-formaldehyde insulation (UFFI) in residences and schools. Much of this invalidation of CPSC's rule was based on the court's underlying rejection of the agency's use of a "Global 79" risk assessment model used to predict the increased risk of cancer to a person living in a UFFI home. The court found CPSC's policy-based assumption in that model, which extrapolated human effects from one large rat study, to be arbitrary. "To make precise estimates," the court reasoned, "precise data are required." The court's opinion flew in the face of many cases upholding the ability of agencies to extrapolate laboratory animal data to humans when human evidence is scarce, and the court seemed to reject the model because it disappointed the judiciary's unrealistic demand for empirical decisiveness.

In *Leather Industries of America, Inc. v. EPA*, the D.C. Circuit invalidated EPA's model in part because it was dissatisfied with an assumption built into the model regarding the

<sup>&</sup>lt;sup>36</sup> Wendy Wagner, Elizabeth Fisher, and Pasky Pascual, "Misunderstanding Models in Environmental and Health Regulation," *N.Y.U. Env't L. J.* 293, 321–323 (2010) (footnotes excluded). Copyright © (2010) New York University Environmental Law Journal. Reproduced with permission.

phytotoxicity of selenium. Even though EPA relied upon some preliminary research and other logical assumptions to justify the level it selected for selenium in the model, the court held that more data-backed justification was required. "While the EPA may err on the side of overprotection, it may not engage in sheer guesswork." The court did not suggest, however, that the agency had ignored relevant information, nor did it explain how the EPA would go about gathering additional information.

Without understanding what modeling is as a scientific practice, the judges are imposing unrealistic expectations on agency decision makers. That is problematic for a number of reasons. It makes the system less effective and less rigorous. It also creates opportunities for strategic game playing. Again we can see the importance of those working with models understanding them and understanding scientific method.

# Wendy Wagner, Elizabeth Fisher and Pasky Pascual, "Misunderstanding Models in Environmental and Health Regulation"<sup>37</sup>

Many rulemakings have significant economic implications for one or more affected industries, and strategic actors face incentives to exploit the misunderstanding of models as "answer machines" to advance their own narrow ends. There are three intimately related but distinguishable strategies that a devious regulatory participant can deploy to reap benefits from the prevailing misunderstanding of models. The first exploits the false expectation that models are fact generators. From the agency's perspective, portraying models as answer machines allows the agency to sidestep at least some unpleasant accountability and controversy by shrugging off criticism with the response that the "model made me do it." Agencies may thus choose to do this, even when they know better, because it protects them from scrutiny by institutional authorities like Congress and the courts. Examples abound of agencies perpetuating the misunderstanding of models as answer machines, while at the same time secretly cramming contested, value-laden assumptions into their highly technical models behind the scenes. In the areas of risk assessment, public land management, and even economic modeling, the propensity of agencies to use models as facades for underlying value choices is well established. As Glicksman notes, "[T]he use of modeling by [public land agencies] is susceptible to the criticism that the agencies, intentionally or not, have masked their value judgments in the language of technical determinations."

In a related vein, the rational agency may not only find itself tacitly rewarded for misrepresenting its model as an "answer machine" but, quite independent from that, will find it beneficial to be opaque about assumptions and uncertainties incorporated into the model, even if it ultimately concedes the tentative nature of the modeling exercise. This opacity helps insulate the agency's many assumptions and modeling decisions from critical review, particularly by adversarial stakeholders. As Glicksman observed in this context, agencies "can isolate themselves [in part] by making their decisions in secret, without soliciting the views of knowledgeable experts and lay

<sup>&</sup>lt;sup>37</sup> Id. 336–343.

persons." This type of strategic opacity helps conceal from participants "the subjective decisions and policy choices made by planners and modelers during the modeling process."

As a result, the prospect of an open season on agency models by aggressive stakeholders, coupled with the threat of a lawsuit over technical disagreements, create incentives for rational agencies to make their models relatively indecipherable with regard to the underlying assumptions and uncertainties. Once a stakeholder engages with the agency's model and begins asking fundamental questions about the framing and assumptions, it is less clear that an agency is wise to stonewall. But, at least at the outset, developing an opaque model is a useful strategy to insulate the model and attendant policy decisions from critical review...

A different strategy that exploits this same erroneous portrayal of models as decisive demands an unobtainable level of empirical certainty, a demand that may succeed not only in blocking the use of the model, but in blocking the policy as well. Given that uncertainty permeates the entire modeling process, a resourceful stakeholder can demand perfection while running the agency's preferred model so full of holes that it sets the regulatory effort adrift with scientific demands that can never be satisfied. As Professor Farber observes, "[w]ords like uncertainty, systematic biases, and important deficiencies [used by modelers in describing their climate change models] are music to the ears of cross-examiners."

These attacks on models, usually accompanied by clarion calls for "sound science," occur even when Congress has demanded that an agency err on the side of protecting public health. Under the George W. Bush administration, for example, despite being confronted with evidence from the Intergovernmental Panel on Climate Change (IPCC) and from EPA's own independent science panel regarding health and environmental threats from both CO<sub>2</sub> and from particulate matter at levels lower than existing standards, EPA Administrator Steve Johnson declared that uncertainty precluded regulatory action on either front. (Ironically, it appears that robust characterizations of uncertainty were in part the basis for this rejection of the models.) After judicial review, EPA's decisions not to regulate in both cases were eventually overturned, in part because they lacked a rational basis in light of the available scientific evidence.

Interested parties have also pointed to specifically contestable coefficients in a model as failing to meet the demand for "sound science," a deficiency that they argue necessitates the wholesale rejection of the model. Their objective is to destroy the credibility of "good" or "plausible" models by criticizing the model on every picky and generally insignificant detail. This was in fact an explicit strategy of the tobacco industry. Climate change models also appear to have been subject to this type of ends-oriented type of attack. In a challenge mounted against several federal agencies under the Data Quality Act, for example, the petitioner, the Competitive Enterprise Institute (CEI) (an organization funded in part by industries adversely affected by carbon reduction policies) argued that the National Assessment on Climate Change (NACC) models should be stricken from public databases because the models could not be "verified by observed data" and were therefore "junk science." Specifically, CEI argued that:

[T]he climate models upon which NACC relies struck out. Strike one: they can't simulate the current climate. Strike two: they falsely predict greater and more rapid warming in the atmosphere than at the surface--the opposite is happening. Strike three: they predict amplified warming at the poles, which are cooling instead.

Even the columnist George Will has demonstrated this type of "sound science" attack, whether consciously or not, by arguing that any contestable assumptions in models make the models useless for policy. In one column Will noted how models developed by some scientists in the 1970s predicted a "major cooling of the planet" and implied that if these models were wrong in the past, current models predicting global warming must also be wrong today. Will also suggested that if models cannot provide definitive, foolproof answers, they should not be used to help formulate policy.

In sum, discrediting a model by picking at every instance of error or uncertainty offers critics of regulation the easiest path to attack policies based on model results when models are misunderstood as answer machines. It puts upon proponents of regulatory intervention the entire burden of persuasion, the entire burden of accumulating the available evidence, and the burden of drawing credible, defensible inferences from the evidence. It simultaneously relieves critics of the burden to develop alternative explanations of environmental risk.

Strategic game playing can also involve technical trickery: working backwards from a desired regulatory result, a stakeholder can tweak model assumptions and even data sets until they develop a favorable model to support their position. Stakeholders can also cherry-pick models (and modelers) based on the results of the model rather than on its reliability for the use in question. Using rigged models to support valuebased positions lends the patina of scientific credibility to a legal or political argument. Moreover, when policymakers and legal analysts view models in a deterministic way, they are unlikely to be aware of the extent to which models can be misused in this way and generally would not have the capability to look inside a model to better understand the assumptions and related choices that have been made in an effort to sort the honest from the dishonest models.

Another related strategy attempts to influence the development of basic modeling practices themselves. For example, challenges have been waged against government risk assessments that rely primarily on mechanistic animal studies to classify substances as known carcinogens. Challengers insist that in these cases epidemiological evidence is essential. Such positions obviously ignore the statistical challenges of isolating effects in human populations and the fact that for many chemicals, there are few or no human studies available.

Again, this is not to say that all judges understand models as truth machines—they don't. We saw a very different approach by a judge in the polar bear case above. Likewise, in the next section, we will see examples of the courts developing approaches to judicial review that reflect more accurate understandings of scientific method. Our point is that if generalist

decision makers don't have those understandings, serious problems can arise. Knowing something about modeling is not just an exercise in being geeky but crucial to ensuring an effective regulatory process.

## Check Point G: Attack a Model Not to Improve It, but to Win the Debate

Since "all models are wrong," it can be relatively easy to poke holes or otherwise discredit them. Return to *The Very, Very Simple Climate Model* that we explored above at pages 6–7. Imagine that it is being used by advocacy groups to urge immediate and substantial restrictions on oil production in the United States. Now take the role of one of these oil producers, but do so without worrying about improving the quality of the model. In what ways can you argue that the simple climate model is not to be trusted?

# 5.4. An Exercise for Recapping and Reflection

We have covered a lot of material in the last three sections. We have looked at the scientific principles inherent in modeling, the roles that models play in regulatory decision making, and how knowing something about models is important in holding decision makers who use models to account. Below is an exercise to help you recap and reflect on this material by placing you in the more difficult role of a responsible official who seeks to understand the "truth" about a model. Rather than take pot shots or serve as referee, in this job it is critically important to understand all that a model has and does not have to offer.

# Moment of Truth Check Point H: Can You Really Understand a Model if You Didn't Create It and Aren't a Modeler Yourself?

Congratulations! You are now an assistant administrator at the EPA, a political appointment, who has been handed some controversial results from a model. The EPA staff running the model report that in order to meet EPA's goals for water quality for a particular, federally managed watershed, the model reveals that an additional \$20 million in regulatory requirements will be necessary. This unexpected result is likely to raise the ire of regulated parties and farmers in the area. You don't doubt the staff's integrity, but you also know that many judgments were involved in their analysis and are not content with a simple bottom line. How do you gather information about the most significant alternate assumptions and choices made in the course of modeling? If you don't have a close relationship with your staff and are not certain you can trust them, how do obtain a second opinion, at least in the abstract? And if you do seek external advice, what kind of challenges/criticisms are you likely to face, particularly from those eager for cleaner water (in other words, are there any downsides in trying to learn more about this model and can you find a way around them)?

What basic steps would you take, based on the prior sections, to ensure that the model has been created and run in a rigorous way? What kind of oversight should the model and input data receive?

Moving forward, how might you improve the way that model is integrated into policy making? How would you engage the scientific community in reviewing this model? How would you engage the general public?

What role is "science" playing in these decisions and what role is "policy making" playing? These two categories cannot be neatly separated, but do they have sufficient independence to be useful poles for thinking about the challenges? Or is reviewing the integrity of models used for policy in a hybrid experience that requires its own, unique process?

# 6. AMBIENT AIR QUALITY STANDARDS CASE STUDY

In this closing section, we take a final tour of modeling at the intersection of policy and real world application by exploring an extended case study of the EPA's modeling efforts in setting ambient air quality standards in general and the fine-particulate standard in particular. We hope to reinforce, in concrete terms, themes raised in earlier sections. There is a little less signposting in this section—just as in real life the use of models does not come with a neat set of labels and alarm bells telling policy makers what to be alert to. We want you to try and apply everything you have learnt up to now to identify issues.

By tracking the evolution of air standards, we will explore how policy makers rely on models, how litigants exploit model uncertainties to challenge regulations, and how policy makers can craft defensible decisions despite these uncertainties. Indeed, we will countervail specious demands for certainty and their corollary—vacuous claims that "correlation does not equal causation"—to show how science advancements over the past 50 years have steered models toward increasingly robust grounds to infer causation.

Before engaging in the details of modeling air standards, we will set the scene by situating EPA's modeling effort within its larger statutory and regulatory context. Like any scientific endeavor, the decision of what to model is intricately connected to larger social frames that can become invisible, yet vitally important in identifying the types of models, assumptions, and issues that are "relevant" and those that are not.

# 6.1. Air Quality Regulation – A Brief History and Overview

In December 1930, a mysterious, yellow fog wafted down the Meuse Valley in Belgium. Within days, 63 people died and hundreds more succumbed to respiratory illnesses. The tragedy prompted comparisons to the "black death"; villagers speculated that poisonous gas seeped from buried German bombs.<sup>38</sup> Professor J. Pirket, a pathologist at a local university, identified the true culprit. "Fine solid particles" emitted by factories had been trapped on the

<sup>&</sup>lt;sup>38</sup> New YorkTimes, 1930.

ground by a layer of warm air. Using simple models, Pirket calculated the concentrations and toxicities of various chemicals in the fog.<sup>39</sup>

Pirket's study, one of the earliest to link public health and air pollution, was cited in a report to justify U.S. air regulations, published in 1969. Three years earlier, New York state had experienced a cataclysm similar to that which befell Meuse Valley. Exploiting this event, the wily President Lyndon Johnson pressured Congress into passing the Air Quality Act. Among other things, the act directed the secretary of Health, Education and Welfare to develop air quality standards "based on science."

The Air Quality Act evolved into the Clean Air Act, which today regulates the nation's air pollution. Through the Clean Air Act, Congress directs the EPA to set standards for the most ubiquitous air pollutants, like ozone, sulphur oxides, and particulate matter. Specifically, the EPA must ensure that each air quality standard be "requisite to protect the public health" with "an adequate margin of safety" 42 U.S.C. § 7409(b)(1). These standards are called the National Ambient Air Quality Standards, or NAAQS. The EPA must review the standards every 5 years to ensure they are up to date with the latest scientific information.

As the statute makes clear, Congress requires the EPA to promulgate air quality standards that not only ensure rigorous health protection but also provide a margin for error. Yet, protecting public health can be expensive, and billions of dollars, both in health protection and compliance costs, can hinge on adjustments as small as even one-thousandth of a part per million for the standard for any given pollutant. While there have been attempts to include economic considerations in the setting of standards, the Supreme Court has interpreted the statute as allowing the EPA to consider only scientific, not economic, factors.

Given this framing of the Clean Air Act, consider how one selects assumptions for the various uncertain junctures in the model if it is not based on economic considerations. Perhaps those creating the model and making the decision can identify types of risks that are minimal enough to disregard in setting protective standards? Or perhaps the numbers of people affected are small enough that they too can be bracketed as too limited? Or, even better, maybe the affected persons are few and can protect themselves with limited effort—perhaps by staying indoors or avoiding exercising? Even if these are credible considerations, who should make them?

In the Clean Air Act, Congress directs the agency itself—ultimately the administrator of EPA, appointed by the president—to resolve these questions and ambiguities. Some of these ambiguities are obvious. Others appear only as the staff scientists engage in the decision-making process. Either way, modelers must ensure that decision makers have provided the needed guidance on these issues; otherwise, they will be making the choices in less accountable ways.

The passage below provides an example of both the choices and how well EPA explains the options and trade-offs. This passage is excerpted from a preface to EPA's formal proposal—called a rule preamble—that precedes the legal rule and is published in a daily government publication of agency decisions called the *Federal Register*. This preamble, which is sometimes more than 100 pages of tiny, three-column print, provides not only the basis for the public to understand the agency's decisions, but also often is the focal point for legal challenges since it

<sup>&</sup>lt;sup>39</sup> Pirket, 1936.

essentially identifies the agencies' own admissions about how it interpreted a statute or other legal command.

In this small excerpt from the preamble, the EPA is explaining why it declined to set a shorter, 1-hour standard to protect vulnerable populations from exposure to sulfur oxides in the late 1980s. (In contrast to the then existing 24-hour standard, a supplemental, 1-hour standard would have ensured that there are not high peak concentrations of sulfur oxides that would be missed in the averaging of exposures for purposes of complying with the 24-hour standard). This explanation highlights the many delicate public decisions that lie just beneath the surface of a seemingly "scientific" quantitative ambient air standard. Note also that many of the factors discussed in the last several sections can be seen.

Consider as you read how well EPA explains the gaps in the scientific record, the most plausible options available, and why it selected the choices it did. We will conclude the excerpt with a check point in which you return to these steps and look at them more critically.

# EPA, "Ambient Air Quality Standards for Sulfur Oxides,"<sup>40</sup> 53 Fed. Reg. 14,926, 14,931-934 (1988).

## Short-Term Health Effects

The basis for considering the possible addition of a new 1-hour standard rests largely on the staff and CASAC assessment of the results of several relatively recent controlled human exposure studies (see Table 1 of Addendum III to this preamble). The major effects observed in these studies are measurable changes in respiratory function in asthmatics and atopics [FN1] exposed for short periods (as little as 5-10 minutes up to 1-hour) to 0.4 ppm SO[FN2] or more. For example, in one study designed to examine this issue, a concentration of 0.5 ppm for 10 minutes produced a doubling (or more) in airway resistance in 25 percent of exercising asthmatic subjects (Horstman et al., 1986). The responses occurred predominantly in subjects whose respiratory ventilation was increased by exercise or by hyperventilation, and who were not using preventive medication at the time (SPA, pp. 9-10; CDA, Table 5; Sheppard et al., 1981). In asthmatic subjects exposed to 0.4-0.75 ppm or more of SO[FN2], the change in respiratory function was often accompanied by perceptible symptomatic responses, including shortness of breath, wheezing and coughing (SPA, Table 4-1). The fraction of asthmatic subjects experiencing changes in lung function and symptoms increased with concentration over the range of 0.4 to 0.75 ppm (SPA, Figure 3-2).

While mindful of the guidance in the criteria document that "caution should be employed in regard to any attempted extrapolation of these observed quantitative exposureeffect relationships to what might be expected under ambient conditions" (CD, p. 13-50), the staff and CASAC concluded that consideration should be given to a new short-term standard to address these effects. Based on practical considerations relating to monitoring, modeling, data manipulation and storage, and implementation, the staff recommended using a 1-hour averaging time for any such standard. As explained below, the relationship between 1-hour average concentrations and shorter-term concentrations would allow the use of a 1-hour standard, set at an appropriate level, to control shorter-term peaks. Staff and CASAC identified a

<sup>&</sup>lt;sup>40</sup> EPA, "Ambient Air Quality Standards for Sulfur Oxides," 53 Fed. Reg. 14,926, 14,931-934 (1988).

number of factors that should be considered in decisionmaking concerning a 1-hour standard (SP, pp. 64-69; SPA, pp. 37-44):

a. Significance of Effects. The functional changes and symptoms observed in the controlled studies appear to be transient and reversible, and, at lower concentrations (<0.75 ppm) and exercise levels, they are within the range of day-to-day variations that most asthmatics typically experience from exercise or other stimuli. They are, in general, not equivalent to the more severe responses that accompany an asthma "attack" (SP, p. 66). Finally, because medications already widely used by asthmatics can prevent (Sheppard et al., 1981) or ameliorate reaction to SO<sub>2</sub>, an asthmatic who is already medicated due to other stimuli will likely not experience a response to an exposure. The scientific community is divided as to whether and to what extent these effects at lower concentrations should be considered "adverse" or "clinically significant" (e.g., Boushey, 1981; Higgins, 1983; Cohen, 1984; McFadden, 1986; Lippmann, 1987; see also SPA, pp. 40-41).

b. Relative Effect of SO2 Compared to Other Stimuli. Exercise alone, without pollutant exposure, is among a number of stimuli that commonly induce bronchoconstriction in asthmatics (SP, pp. 66-67). Cold and/or dry air exacerbates the effects of exercise even in the absence of SO2. It is likely that the incidence of bronchoconstriction induced by SO2 is very small compared with that induced by factors unrelated to pollution (Cohen, 1984 and EPA, 1986c).

c. Sensitive Population. Diagnosed asthmatics make up approximately 4 percent of the total U.S. population (about 10 million individuals) while atopics constitute roughly 8 percent (SP, p. 31). Some additional percentage of the population not diagnosed as atopic or asthmatic may also display hyperreactive airway responses to SO2 (SP, p. 30). Asthmatics appear to be at greater risk than atopics. Studies to date have shown a wide distribution of sensitivity among asthmatics and atopics tested (e.g., Horstman et al., 1986). Although it is speculated that individuals with more severe asthma may be more sensitive to SO2 than are the relatively mild asthmatics tested, CASAC has pointed out that the available data do not support or refute this point (Lippmann, 1987). The consequences of a functional change are of greater concern in more severe asthmatics, but such individuals may be somewhat protected from SO2 because they routinely use medication due to their susceptibility to responses from other stimuli and the reduced chance that they would experience sustained levels of moderate to high exercise (SPA, p. 40).

d. Variance About the 1-hour Average. The available studies indicate that SO2 effects occur within 5 to 10 minutes but do not necessarily worsen with continued exposure over an hour (CDA, pp. 4-29 to 4-32). Concentrations averaged over 5 or 10 minutes vary about the 1-hour mean, reaching peak values that are clearly higher than the 1-hour value. Analyses of recent data indicate that at higher concentrations near large point sources, these peaks are likely to be within a factor of 2 of the mean. Thus, the maximum 5 to 10 minute peak associated with a 1-hour value of 0.5 ppm is probably less than 1 ppm (SPA, p. 43-44).

e. Probability of Exposure. The staff assessment found that, given current air quality levels, peak SO2 concentrations in the 0.4 to 0.75 ppm range for 5 to 10 minutes are very infrequent and limited in extent to the vicinity of certain large sources. Given low indoor levels and the limited time individuals spend in moderate to high activity, the probability that any individual asthmatic would experience any effects of SO2 is low (SP,

p. D-12; EPA, 1986c). This issue has been examined in the quantitative analyses discussed in the following section.

#### Protection Afforded by Current Standards Against Short-Term Effects

In determining whether to revise the present standards by the addition of a 1-hour primary standard, it is particularly important to evaluate: (1) The extent to which implementation of the current standards protects against potential very short-term effects, and (2) the relative increase in protection that would be afforded by the addition of a possible 1-hour primary standard. The first point is addressed in this section, while the second point is addressed in the following section.

(a) Air Quality and Exposure Analyses. The initial staff examination of the above issues focused on monitoring [FN2] and modeling analyses of 1-hour SO[FN2] concentrations (SP, Appendix D). The initial modeling analyses predicted the frequency of 1-hour exceedances of 0.5 ppm that would occur around typical major point sources if the current standards are met. This concentration (0.5 ppm) was the lowest short-term (5-minute to 1-hour) level found to produce changes in respiratory function and symptoms in the controlled studies of exercising "mild" asthmatics included in the 1982 staff paper. . .

Although the analyses are uncertain, the available results permit the following tentative conclusions:

(1) Based on current U.S. monitored air quality data and reasonable estimates of ratios of 5-minute peaks to 1-hour means, 5-minute concentrations and exposures to 0.5 ppm or more are expected primarily in the vicinity (usually less than 20 km) of major point sources such as utilities and smelters. Approximately 10 to 40 percent of the sensitive population (asthmatics) in the U.S. are estimated to live in the vicinity of utilities, with a much smaller percentage living near smelters (Thomas, 1987a).

(2) Based on modelled air quality and exposures for several large utility power plants, the current standards (24-hour and 3-hour) place substantial limits on exceedances of, and exposures to, 1-hour concentrations in excess of 0.5 ppm (Thomas, 1984).

(3) Of those asthmatics living in the vicinity (roughly 10-25 km) of the four power plants studied at their current emissions, the percentage estimated to be exposed once per year to a 5-minute SO[FN2] concentration of 05. ppm while at exercise varied from 1 percent to 14 percent, depending on the plant (EAP, 1986c, pp. 3-16 to 3-19). A rough extrapolation to all of the power plants in the country suggests that approximately 100,000 individual asthmatics, or about 1 percent of the national asthmatic population, will experience at least one such exposure of concern per year (Thomas, 1987a). [FN5] The vast majority of these 1 percent would experience only one such exposure per year.

(4) Because not all of the exposures to 0.5 ppm resulted in measurable effects in controlled studies, fewer than 25 percent of the asthmatics exposed are likely to experience even moderate pulmonary function changes and symptoms (Horstman et al., 1986). It is possible that individual asthmatics substantially more sensitive than those studied might experience larger or comparable effects at even lower levels. However, CASAC has pointed out that there is no evidence to refute or support this possibility. Moreover, severe asthmatics may be protected because they less often achieve

elevated activity levels and often are already medicated to alleviate the effect of other environmental stimuli commonly encountered...

#### 3. 1-Hour Standard Alternative

... EPA staff and CASAC recommended a range of potential 1-hour standards for the Administrator's consideration. This range, based on the updated staff assessment (See Table 1 in Addendum I to this notice), is 0.2 to 0.5 ppm (520 to 1300 mg/m3). Considering typical 5minute peak to 1-hour mean ratios of 2 to 1, the lower bound (0.2 ppm) represents a 1-hour level for which the maximum 5- to 10-minute peak exposures are not likely to exceed 0.4 ppm. This is the lowest level where responses of potential clinical significance in free breathing "mild to moderate" asthmatics have been reported in the literature cited in the criteria document addendum. A 1-hour standard at the upper bound of the range (0.5 ppm) would maintain maximum hourly values in the vicinity of the lowest concentrations (0.4 to 0.5 ppm) producing significant responses in the available studies summarized above. It would afford somewhat greater protection against short-term peaks than that now provided by the current standards. Based on the preliminary analysis of exposure near large point sources discussed above (SPA, Figure 4-3), it appears that under such a standard, 1 to 4 percent of the asthmatics residing in the vicinity of the point sources analyzed, or between 200 to 1400 individuals per plant, would be annually exposed while at exercise to 5-minute peaks at or above 0.5 ppm. On a national level, fewer than 1 percent of all asthmatics would experience such exposures. Nevertheless, a 0.5 ppm level would not completely preclude 5- to 10-minute exposures on the order of 1 ppm.

Considering typical 5-minute peak to 1-hour mean ratios of 2 to 1 or lower, 1-hour standard alternatives of 0.3 to 0.4 ppm could result in 5-minute peaks on the order of 0.6 to 0.8 ppm. Several CASAC members supported a 1-hour standard in this portions of the overall range (Lippmann, 1987). If a 1-hour ambient standard of 0.4 ppm were implemented at the four power plants studied in the exposure analysis discussed above, the percentage of the asthmatics living in the vicinity of those plants who would be exposed once per year to a 5-minute SO2 concentration of 0.5 ppm while at exercise would be less than 1 to 2 percent (EPA, 1986c, p. 3-22).

After considering the views of CASAC, the Administrator is inclined to conclude that a 1hour primary ambient standard to protect exercising asthmatics and atopics from short-term exposures to SO2 is not warranted. As explained above, this inclination is based, on among other things, the uncertain significance of the health effects involved and on the infrequence of inducement of such effects by SO2. However, the Administrator solicits comment on the alternative of a 1-hour standard at the level of 0.4 ppm.

Through these legal steps—a statutory mandate, an agency analysis that is explained in a *Federal Register*, and a final rule—scientific analyses and policy choices are woven together, often in ways that become effectively indistinguishable, into an enforceable rule that guides the entire nation with respect to the minimal standard for air quality.

#### **Check Point I: Asthmatics Stay Indoors?**

EPA defends its decision not to set a standard that is clearly more protective of asthmatics in the preamble. Return to this excerpt and identify the points at which a skeptical

group representing the interests of asthmatics might differ with EPA's decision. These contestable features of EPA's model can be technical or value laden or both. And, now that you are the advocate, how will you try to get the agency to take your point of view seriously?

# 6.2. The Process Required for Setting Ambient Air Quality Standards

It is clear that there will be many choices involved in constructing a model and that—in an ideal world—these choices and their alternatives will all be put out for nonmodelers to review. In fact, in assessing the potential health effects and alternative courses of action pursuant to this legislation, modelers often design one or several models to inform the questions. Moreover, as we know from Section 4 and in contrast to basic science, the scrutiny applied to this modeling work must arise from the public at large, as well as from fellow scientists. This ultratransparency and accessibility requires modelers to ensure that not only are their choices and decisions explained, but also that they are explained in ways that lay persons will understand.

But this is not easy!

The EPA, not surprisingly, has struggled mightily to develop a process that ensures both the scientific rigor and the public accountability of its NAAQS process. During the first three decades of NAAQS standard setting, EPA treated the process like a scientific exercise and the agency's primary focus was to locate and consider all of the research that informed its review of a particular air quality standard. Thousands of studies often had to be considered in the course of this review, even though the review occurred only every 5 years.



Source: Fran Orford. Copyright © FRANCARTOONS. All rights reserved. Reproduced with permission.

The format EPA chose for this massive, 5-year review of each standard did not alleviate the appearance of a model developed and based on "science," nor was it easy for even the most dedicated of modelers to understand. The agency typically put all the research together and summarized it in a single enormous "criteria document" that often ran thousands of pages long. These criteria documents were so impenetrable that not only lay persons had trouble reviewing it but experts had difficulty engaging in the intricate details. Even in some of the very earliest standard setting decisions in the 1970s, when the criteria documents were more manageable, judges engaged in the judicial review of challenges to EPA's rules were frustrated by the enormity and inaccessibility of EPA's analysis. This frustration is manifest in one prominent judge's concurrence on the legality of EPA's controversial lead standard:

# Ethyl Corp v EPA<sup>41</sup>

An additional matter which emerges from this record deserves comment: namely, the failure of the record to clearly disclose the procedural steps followed by EPA. As a result, an onerous, time-consuming burden was cast upon the court to reconstruct these steps by inference and surmise. It is not enough for an agency to prepare a record compiling all the evidence it relied upon for its action; it must also organize and digest it, so that a reviewing court is not forced to scour the four corners of the record to find that evidence for itself...

<sup>&</sup>lt;sup>41</sup> *Ethyl Corp. v. EPA*, 541 F.2d 1, 67-68 (D.C. Cir. 1976).

Based on that possibility, and the court's own reconstruction of the procedural record (albeit at the expense of much judicial time and effort), I am persuaded that the petitioner's rights were not prejudiced. Ordinarily, however, I think a record which so burdens judicial review would require a remand for clarification.

The impenetrability of EPA's important standard setting was not lost on Congress or the larger public. Growing concern about an unaccountable process underlying these all-important air quality standards led Congress to take the first step in 1977 of requiring each of EPA's reviews to undergo elaborate peer review by the top air quality scientists and modelers in the country. Congress even specified that the panel should include "at least one member of the National Academies of Sciences, one physician, and one person representing State air pollution control agencies."<sup>42</sup> In response, EPA created the CASAC, a standing committee chartered under the Federal Advisory Committee Act (FACA), which plays an important role in the EPA's NAAQS review process.

The engagement of a scientific oversight committee was a critical step in ensuring that EPA's modeling and related scientific analyses were in fact using the best research and analytical techniques. But this oversight was not as helpful in forcing EPA's processes to be more accessible to lay persons, including not only members of Congress but the public at large. Among the many criticisms was (in EPA's own words), "broad recognition that the Criteria Document is typically 'encyclopedic' in nature, which is seen by many as contributing to an unnecessarily lengthy process for preparing document drafts and for reviews by CASAC and the public, and obscuring a focus on the most policy-relevant scientific information." EPA, Review of the Process for Setting National Ambient Air Quality Standards, March 2006, at 18, at <a href="http://www.epa.gov/ttn/naaqs/pdfs/naaqs">http://www.epa.gov/ttn/naaqs/pdfs/naaqs</a> process report march2006.pdf. After a series of high-stakes political controversies over expensive standards and litigation over both the substance of the standards and EPA's delay in promulgating them, EPA sought to reorganize its scientific process in 2006.

The reformed 2006 process for standard setting is considered path breaking insofar as it paves the way for science to be better integrated into policy. In this process, EPA separates the analysis into distinct analytical steps, each of which is subjected to both expert CASAC and public review. The process also encourages the development of multiple, rather than single models that explain alternative ways to understand the data—each with their own set of contestable assumptions—leaving it to the political staff to select among these scientifically supportable options. The final report explains the scientific staff's model in accessible terms while highlighting the limitations and assumptions, including their implications for the analysis. The "five separate analytical steps and products" used by EPA in setting ambient standards for air quality are described in slightly more detail below:

# Sidney Shapiro, Elizabeth Fisher, and Wendy Wagner, "The Enlightenment of Administrative Law: Looking Inside the Agency for Legitimacy"<sup>43</sup>

<sup>&</sup>lt;sup>42</sup> 42 U.S.C. § 7409(d)(2)(A)-(B).

<sup>&</sup>lt;sup>43</sup> Sidney Shapiro, Elizabeth Fisher, and Wendy Wagner, "The Enlightenment of Administrative Law: Looking Inside the Agency for Legitimacy," *Wake Forest L. Rev.* 47: 463, 493–496 (2012) (footnotes omitted). Copyright © (2012) Wake Forest Law Review. Reproduced with permission.

#### a. The Planning Report

The first step sets the stage for the integration of scientists, stakeholders, public health advocates, and professional agency staff by convening a "kick-off" workshop that is followed by a staff-authored report that articulates the overarching policy questions that will guide the process. The report is reviewed by the "Clean Air Science Advisory Committee" ("CASAC"), a statutorily required standing committee of top scientists chartered under the Federal Advisory Committee Act ("FACA"), and by the public before it is final.

The resulting final planning report is thus a professional, staff-authored document that has been reviewed iteratively by the public and external scientists. This planning report, moreover, is integral to enhancing transparency of the NAAQS review process. By framing the relevant science-policy questions, the planning report focuses the EPA's subsequent NAAQS review, which stretches over a four-year process.

#### b. Integrated Scientific Assessment Report

At the next step of the NAAQS review process, the EPA compiles an integrated scientific assessment ("ISA") that reviews all of the scientific evidence. In stark contrast to the EPA's earlier version of this assessment in previous NAAQS processes, the new and improved ISA is more concise and focuses the assessment on the specific questions framed in the planning report. More detailed information is reserved for annexes, which can sometimes be longer than the body of the report itself.

. . . .

## c. Risk/Exposure Assessment Report

Based on the analysis of the scientific evidence in the ISA, the EPA staff then prepares a separate risk assessment report that applies this evidence to predict the effects of alternate standards on public health. The goal at this stage is to employ multiple models to produce quantitative risk estimates, accompanied by expressions of the underlying uncertainties and variability for various endpoints, such as the impacts of a pollutant on susceptible populations and ecosystems. The risk assessment process itself begins with a planning/scoping stage, which again involves CASAC review and public comment, followed by two more periods of intraagency, CASAC, and public comment on the draft risk assessment reports.

## d. Policy Assessment Report

The last document in the process is a policy assessment that "bridges" these more scienceintensive (ISA and risk assessment) reports with the policy questions at hand. In summarizing the evidence in a way that relates to the overarching policy question, the report offers alternative health protection scenarios and standards, accompanied by discussions of unknowns and uncertainties. The policy analysis also identifies questions for further research. The policy assessment is, in and of itself, an extensive document (in the EPA's review of the particulate matter standard, the policy assessment was over 450 pages in length, including appendices), but the discussion is written for laypersons who do not have an extensive background in the relevant science. The policy assessment is reviewed by internal EPA staff and by CASAC, sometimes several times, to ensure that important scientific information is not lost in translation. It is worth noting that even at this late stage, CASAC review and comment is rigorous and extensive. For example, the second CASAC review of the EPA's Policy Assessment for the Review of the Particulate Matter ("PM") NAAQS consisted of over seventy pages of single-spaced comments.

# **Check Point J: Extending EPA's NAAQS Process to other Settings**

The National Academies hails EPA's revised process for setting ambient air quality standards as a five star process. Some have suggested that this process should be followed by all agencies in their use of models for policy. Return to the process mapped out above and consider whether there are decisions for which these processes are not required? On the flip side, are there ways to make the process still better or more effective? Even if there is no better way, what if the process is followed in letter but not in spirit by producing impenetrable models and jargon-filled explanations? Is there a way to prevent that from happening or to penalize agencies or others who do not use "parsimony" (simple/succinctness) in their model but opt instead to produce off-putting, lengthy, incomprehensible explanations?

# 6.3. NAAQS under the Microscope: Criticism and Judicial Review

Given the high stakes involved in EPA's standard setting, its decision-making process is continually under criticism, and transparency is often used as either the sword (the basis for criticism) or the shield (the defense) in these debates. These criticisms range from blistering news headlines to secretive negotiations between the agency administrator and the White House. Possibly the most unpleasant blow-backs from the standpoint of the agency come from Congress. The letter below is a small sample of the concerns Congress can raise on each NAAQS review.

# **Check Point K: Responding to Congress**

Consider the following questions as you read the letter below:

What is the Congressman Smith's core concern?

- 1. Is his concern compelling?
- 2. If you were a prominent scientist and were asked to serve on CASAC, would this letter impact your decision? If so, why and how?
- 3. If you were the Administrator of EPA, what would you do in response to this letter?

LAMAR S. SMITH, Texas CHAIRMAN EDDIE BERNICE JOHNSON, Texas RANKING MEMBER

# Congress of the United States

# House of Representatives

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

2321 RAYBURN HOUSE OFFICE BUILDING

WASHINGTON, DC 20515–6301 (202) 225–6371 www.science.house.gov

March 19, 2014

The Honorable Gina McCarthy Administrator U.S. Environmental Protection Agency 1200 Pennsylvania Avenue, NW Washington, D.C. 20460

#### Dear Administrator McCarthy:

The Committee on Science, Space, and Technology has longstanding concerns with the Environmental Protection Agency's (EPA) scientific advisory processes. In particular, there has been a troubling lack of independence and transparency within EPA's Clean Air Scientific Advisory Committee (CASAC) Ozone Review Panel. In establishing CASAC, Congress gave clear direction that this committee is intended "to have complete independence."<sup>1</sup> The CASAC Ozone Review Panel appears to violate agency policies designed to ensure balance, independence, and impartiality. Additional transparency is necessary to assure Congress and the American people that EPA is basing its costly regulatory decisions on the best available science and not a predetermined regulatory agenda.

In a hearing before the Committee last November, you stated that CASAC "provides independent advice to the EPA Administrator on the science that supports the EPA's National Ambient Air Quality Standards."<sup>2</sup> However, recent testimony and the current makeup of the panel reveal a number of problems, including: panelists reviewing their own work; a lack of turnover among CASAC Ozone Review Panel members; and, existing financial relationships between panelists and the Agency. Dr. Robert Phalen, Professor of Medicine and Co-Director of the Air Pollution Health Effects Laboratory at University of California, Irvine, and a former member of the CASAC panel on fine particulate matter, stated in testimony that the "current [CASAC] process... is seriously flawed, it is narrowly focused, and it is even ethically questionable."<sup>3</sup> Due to the substantial economic cost associated with finalizing a more stringent ozone standard, EPA should make every effort to ensure the transparency of the regulatory process.

1

<sup>&</sup>lt;sup>1</sup> H.R. Rep. No. 95-294 at 182-83 (1977).

<sup>&</sup>lt;sup>2</sup> Committee on Science, Space, and Technology, "Strengthening Transparent and Accountability within the Environmental Protection Agency," Testimony of EPA Administrator Gina McCarthy, 113<sup>th</sup> Cong., 1<sup>st</sup> sess., 14 November, 2013, <u>https://science.house.gov/sites/republicans.science.house.gov/files/documents/HHRG-113-SY-WState-GMccarthy-20131114.pdf</u>.

<sup>&</sup>lt;sup>3</sup> Committee on Science, Space, and Technology, "Quality Science for Quality Air," 112<sup>th</sup> Cong., 1<sup>st</sup> sess., 4 October 2011, <u>http://www.gpo.gov/fdsys/pkg/CHRG-112hhrg70587/html/CHRG-112hhrg70587.htm</u>.

Lowering the ozone National Ambient Air Quality Standards (NAAQS) to the range being discussed by CASAC could represent the most expensive regulation in history, with EPA's own cost estimates approaching nearly \$100 billion per year. An ozone NAAQS below the existing standards may place large swaths of the United States in non-attainment with the Clean Air Act and result in severe economic sanctions. Moreover, testimony before this Committee indicates that standards in the range EPA is considering are below naturally-occurring background levels in many parts of the country.<sup>4</sup> Setting standards at these levels could therefore limit growth and impose penalties on communities whose means of compliance is beyond their control.

This Committee is not alone in its concerns about the independence and transparency of the CASAC process, and this Ozone Review panel in particular. As noted below, independent scientists who have been involved with CASAC have testified to our Committee with serious concerns.

#### Recusals and Reviewing One's Own Work

EPA's *Peer Review Handbook* states that "An independent peer reviewer is an expert who was not associated with the generation of the specific work product either directly... or indirectly...."

Within the NAAQS process, EPA's CASAC Ozone Review panelists are asked to review three documents generated by EPA: the Integrated Science Assessment;<sup>6</sup> the Health and Welfare Risk and Exposure Assessments;<sup>7</sup> and the Policy Assessment.<sup>8</sup> Among the current CASAC Ozone Review panel, **16 of the 20 panel members** are cited by EPA in the current versions of these key documents. Indeed, the Agency cites the work of these panel members **more than 700 times** in these regulatory science documents they are being asked to critically assess.

This does not appear to be an isolated incident, as Dr. Roger McClellan, a former Chair of CASAC, testified, "The [CASAC] membership has been excessively dominated by scientists that to a large extent have developed the scientific information contained in the documents."<sup>9</sup> In addition, a 2013 review of CASAC's processes by the EPA Inspector General (IG) found multiple instances where recusals for concerns related to independence were not documented.

2

<sup>&</sup>lt;sup>4</sup> Committee on Science, Space, and Technology, "Background Check: Achievability of New Ozone Standards," 113<sup>th</sup> Cong., 1<sup>st</sup> sess., 12 June 2013, <u>https://science.house.gov/hearing/subcommittee-environment-background-check-achievability-new-ozone-standards</u>.

<sup>&</sup>lt;sup>5</sup> U. S. Environmental Protection Agency, *Peer Review Handbook*, 3<sup>rd</sup> Edition, Peer Review Advisory Group, Science Policy Council, <u>http://www.epa.gov/peerreview/pdfs/peer\_review\_handbook\_2006.pdf</u>.

<sup>&</sup>lt;sup>6</sup> U. S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, *Integrated Science Assessment for Ozone and Related Photochemical Oxidants*, 2013. EPA 600/R-10/076F, <u>http://oaspub.epa.gov/cims/cimscomm.getfile?p\_download\_id=511347</u>.

<sup>&</sup>lt;sup>7</sup> U. S. Environmental Protection Agency, National Ambient Air Quality Standards, Ozone Standards, Documents from Current Review, Risk and Exposure Assessments,

http://www.epa.gov/ttn/naaqs/standards/ozone/s\_o3\_2008\_rea.html.

<sup>&</sup>lt;sup>8</sup> U. S. Environmental Protection Agency, National Ambient Air Quality Standards, Ozone Standards, Documents from Current Review, Policy Assessments, <u>http://www.epa.gov/ttn/naaqs/standards/ozone/s\_o3\_2008\_pa.html</u>.
<sup>9</sup> Committee on Science, Space, and Technology, "Quality Science for Quality Air," 112<sup>th</sup> Cong., 1<sup>st</sup> sess., 4 October 2011, <u>http://www.epo.gov/fdsys/pkg/CHRG-112hhrg70587/html/CHRG-112hhrg70587.html</u>

The IG also found those recusals often came after the fact, lacked detail or their resolution was not adequately documented.<sup>10</sup> Independent reviewers should not be reviewing their work and the Agency needs to guarantee panelists recuse themselves for Agency documents that rely upon or cite their work.<sup>11</sup>

#### Lack of Turnover and Financial Relationships with EPA

EPA's *Peer Review Handbook* states "that the agency [should] rotate membership among qualified scientists in order to obtain fresh perspectives and reinforce the reality and perception of independence from the agency."<sup>12</sup> Despite this requirement, half of the current CASAC Ozone Review Panel members (10 out of 20) also served on the Agency's panel for the reconsideration of the 2008 Ozone NAAQS and five of these members served on both the reconsideration panel and the CASAC Ozone Review Panel for the 2008 NAAQS. Moreover, these panels were highly critical of the Administrator's decision concerning the 2008 ozone standard, suggesting bias in the current panel makeup.

EPA's *Peer Review Handbook* also states that Federal grants or contracts may constitute a direct financial stake, and thus a conflict or lack of impartiality in a specific review for potential peer reviewers. And in 2013, EPA's Inspector General found that "[a] prospective or active member's research or grant is a potential area of concern if the [Federal Advisory Committee], panel, or subcommittee plans to address work performed under the research grant."<sup>13</sup> Yet, since 2000, **70 percent or 14 of the 20 members** of the CASAC Ozone Review Panel have been principal or co-investigators for EPA grants totaling more than \$120 million.<sup>14</sup> Many of these grants focus on ozone- and air pollution-related issues at the heart of these advisory activities.

The lack of turnover in a panel whose composition includes members with a direct financial relationship with the Agency related to the very issues under review suggests a critical lack of impartiality. As Dr. Michael Honeycutt, Director of Toxicology at the Texas Commission on Environmental Quality, stated before the Committee, "Having study authors, grant recipients, and panelists who have taken public stands on relevant topics who are handpicked by the Administrator gives the perception that the CASAC may not be truly independent."<sup>15</sup>

<sup>&</sup>lt;sup>10</sup> U. S. Environmental Protection Agency, Office of Inspector General, 2013. *EPA Can Better Document Resolution of Ethics and Partiality Concerns in Managing Clean Air Federal Advisory Committees*, Report no. 13-P-0387, http://www.epa.gov/oig/reports/2013/20130911-13-P-0387.pdf.

<sup>&</sup>lt;sup>11</sup> Dr. Honeycutt testified that: "It is not appropriate for scientists to peer review their own work; it presents a clear conflict of interest. When a panel reviews a document in which a member's work is cited, that member should recuse himself from review of the document." Available from: <u>http://www.gpo.gov/fdsys/pkg/CHRG-</u>112hbrg70587/htm//CHRG-112hbrg70587 htm

 <sup>112</sup>hhrg70587/html/CHRG-112hhrg70587.htm.
 <sup>12</sup> U. S. Environmental Protection Agency, *Peer Review Handbook*, 3<sup>rd</sup> Edition, Peer Review Advisory Group, Science Policy Council, http://www.epa.gov/peerreview/pdfs/peer review handbook 2006.pdf.

<sup>&</sup>lt;sup>13</sup> U. S. Environmental Protection Agency, Office of Inspector General, 2013. *EPA Can Better Document Resolution of Ethics and Partiality Concerns in Managing Clean Air Federal Advisory Committees*, Report no. 13-P-0387, http://www.epa.gov/oig/reports/2013/20130911-13-P-0387.pdf.

<sup>&</sup>lt;sup>14</sup> These totals only include grants from the National Center for Environmental Research. Grant information available at: <u>http://cfpub.epa.gov/ncer\_abstracts/index.cfm/fuseaction/search.welcome</u>.

<sup>&</sup>lt;sup>15</sup> Committee on Science, Space, and Technology, "Improving EPA's Scientific Advisory Process," 113<sup>th</sup> Cong., 1<sup>st</sup> sess., 20 March 2013, <u>http://www.gpo.gov/fdsys/pkg/CHRG-113hhrg80553/pdf/CHRG-113hhrg80553.pdf</u>.

<sup>3</sup> 

#### Need for Transparency

In testimony before the Committee last year, Dr. Roger McClellan stated that "[v]ery little of the CASAC panel work is done in public view. Most of the discussion takes place offline and is only manifest in the written draft comments of the CASAC Panel Members in response to EPA staff written questions."<sup>16</sup> In light of these serious concerns, it is unacceptable for EPA to move forward with new rules without first addressing potential conflicts of interest and a lack of transparency within a panel intended to provide the Agency with independent scientific assessments.

Please provide all communications between EPA staff, CASAC staff, and the CASAC Ozone Review Panel related to potential revisions to the ozone NAAQS. This should include all e-mail correspondence and meeting records since the start of the panel formation process in 2008. Please provide these communications to the Committee by March 25, 2014.

Your staff has indicated that the upcoming CASAC Ozone Review Panel meeting to be held March 25-27, 2014 will be webcast live on the Agency's website. Please ensure that an archived version of this webcast will be available online to the public following the meeting.

Sincerely,

Camar Smith

Lamar Smith Chairman Committee on Science, Space, and Technology

<sup>16</sup> Committee on Science, Space, and Technology, "Improving EPA's Scientific Advisory Process," 113<sup>th</sup> Cong., 1<sup>st</sup> sess., 20 March 2013, <u>http://www.gpo.gov/fdsys/pkg/CHRG-113hhrg80553/pdf/CHRG-113hhrg80553.pdf</u>.

Another battleground for EPA's NAAQS models is the courts, where affected groups can challenge all aspects of the models—from the data to the assumptions to the statutory interpretation. Again, transparent decision making is a familiar theme, but in this case the focus is on ensuring that the agency has clearly explained its reasoning.

In a recent article, we review the 45 years of judicial review case law involving challenges to NAAQS standards and found a positive, symbiotic relationship between the courts and the EPA's development of rigorous processes and clear explanations. How the courts approached their task changed over time, but it was influenced by the information the EPA provided and the type of internal processes that the EPA created. In fact, the courts' own demand for clear reasoning forced the agency to come to terms with some of the conceptual difficulties embedded in the modeling process, so its process evolved in a complementary way with judicial review. The courts' role in these NAAQS cases also seems to track the polar bear case discussed above with regard to the court insisting that agencies explain their scientific process as well as explain how their decision process meets a legal standard.

# Elizabeth Fisher, Pasky Pascual, and Wendy Wagner, "Rethinking Judicial Review of Expert Agencies"<sup>44</sup>

Era 1 sets the stage for forty-five years of NAAQS review. The one case in this era underscores the wide-open terrain of judicial review of agency scientific decision making -- there were no clear expectations on records, explanation, analyses, or scientific review. There were thus no explicit internal or external yardsticks by which to assess the quality of NAAQS science.

In Era 2 there were attempts to develop these yardsticks (and related accountability processes) both internally and externally. Internal processes for ensuring the scientific robustness of the NAAQS process began to be developed by the Agency. The first of these Agency process changes required scientific advisory board review of the EPA's analyses, a regularization mandated by Congress in the 1977 amendments to the Clean Air Act that replaced the EPA's less systematic use of these boards. Externally, courts also began attempting to develop yardsticks by which to assess decision making, although these were not closely connected to the internal yardsticks being developed. Thus the treatment of scientific arguments by the court in this era was ad hoc.

In Era 3, the Agency made significant progress in developing yardsticks for its decision process and its substantive analysis. In particular, from 2006 onwards the EPA reinvented its scientific decision process and developed a more refined causal framework with which to weigh scientific evidence. The courts began deploying these Agency-generated yardsticks in assessing scientific challenges to the Agency's work and judging the reasonableness of the Agency's explanation.

As we have already noted, these eras correspond roughly to the changes occurring within the Agency's decision process, with the qualification that NAAQS can have a long lead time that predates adjustments to internal Agency processes. Overall, however, a temporal pattern appears between the EPA's development of more rigorous decision processes and the

<sup>&</sup>lt;sup>44</sup> Elizabeth Fisher, Pasky Pascual, and Wendy Wagner, "Rethinking Judicial Review of Expert Agencies," *Tex. L. Rev.* 1681, 1696 (2015) (footnotes omitted). Copyright © (2015) Texas Law Review Association. All rights reserved. Reproduced with permission.
courts' bases for review. Specifically, as Agency analytical processes grow more robust, the courts use the Agency's improved framework in evaluating challenges to its scientific choices. While this may not lead to a cause-effect relationship between the courts' demands and the Agency's increased analytical sophistication, it does suggest at least the possibility of a partnership between courts and agencies...

Inherent in the agency-court partnership is that it encourages EPA to set out clearly and explicitly how they go about modeling. By doing this, courts and other generalist decision making can hold the decision maker to account on the basis of robust scientific standards. There is less room for the type of strategies we discussed in Section 5.3.

## 6.4. The Development of Particulate Matter NAAQS Standards

We have now seen some of EPA's struggles in developing NAAQS models up close. These struggles include identifying the right value and related choices for the model and making the process and choices visible to those outside the agency, while at the same time battling fierce attacks on models, some of which are warranted and some of which are not. This last section on the NAAQS process returns to these familiar themes, focusing not only the difficulties they present but how they can also lead to constructive pressure on the agency, which in turn can encourage it to develop still better processes, techniques, and methods of explanations. As you read this final description of EPA's development of the particulate NAAQS, consider not only the ways EPA's own process has evolved but also the positive role that various participants have played in this evolution.

To begin this final, longer case study of EPA's modeling, we first return to the start of the NAAQS program. In 1969, the U.S. government issued a report, *Air Quality Criteria for Particulate Matter* (hereinafter, *1969 Criteria*) to summarize the health effects of particulate matter (PM). Simply put, PM is anything solid or liquid suspended in the air. It is a pollutant defined by size, rather than chemical composition.<sup>45</sup> After the Meuse Valley incident, intermittent crises had been dropping clues regarding PM's deleterious effects. In 1948 in Donora, Pennsylvania, 17 people died when a dense fog trapped industrial smog over the city. In 1952 in London, similar environmental conditions triggered elevated levels of mortality from chronic bronchitis, bronchopneumonia, and heart disease. The same combination of weather events and health effects occurred in Detroit, New York City, and Osaka, Japan.<sup>46</sup>

Popular crises may captivate the policy maker's attention. But we maintain that hardnosed evidence should drive policy. Whence the evidence on PM's health effects? How do scientists generally collect information on public health and the environment? Succinctly answering this question, the epidemiologist Doll said, "We may devise experiments in the laboratory using animals as our basic material, we may experiment directly on man, or we may observe and record what is happening to man in the course of his ordinary life."<sup>47</sup>

<sup>&</sup>lt;sup>45</sup> Dockery, 2009.

<sup>&</sup>lt;sup>46</sup> Speizer, 1969.

<sup>&</sup>lt;sup>47</sup> Doll, 1963.

The 1969 Criteria discussed the advantages and limitations of these various sources of information (at this point, it may be useful to reread the thought experiment presented in Section 2.1 to enhance your understanding of causality):

The available data from laboratory experiments do not provide suitable quantitative relationships for establishing air quality criteria for particulates. The constancy of population exposure, the constancy of temperature and humidity, the use of young, normal, health animals, and the primary focus on short-term exposures in many laboratory studies make extrapolation from these studies of limited value for the general population, and singularly risky for special risk groups within the population. These studies do, however, provide valuable information on some of the bioenvironmental relationships that may be involved in the effects of particulate air pollution on health. . . .Epidemiological studies [based on uncontrolled observations] do not have the precision of laboratory studies, but they have the advantage of being carried out under ambient air conditions.

In other words, results from controlled experiments may not extend to the heterogeneous environments for which policies are crafted. Clearly, from the earliest days of air quality regulation, scientists recognized the need to integrate multiple strands of information to establish the evidence for protective standards.

The 1969 Criteria drew its evidence from observational studies conducted across the globe, especially New York and London. From these materials, it selected "those studies which furnish the best quantitative information that we have available at the time." Leaning heavily on British studies conducted in the 1950s, the 1969 Criteria arrived at a series of conclusions, especially the following: ambient particulate concentrations above 750 micrograms per cubic meter ( $\mu$ g/m3), along with similar concentrations of sulfur dioxide, lead to excess deaths and increase in illnesses.

When the 1969 Criteria was published, the authors emphasized that the document was *descriptive*, intended to provide states with information on PM's effects. Contrariwise, air standards are *prescriptive*, legally enforceable limits for which political jurisdictions would be held responsible. These jurisdictional responsibilities were established the following year.

In 1970, the Nixon administration created the Environmental Protection Agency and Congress passed the Clean Air Act. The act charged EPA with the authority and duty to establish air quality standards. The states would be responsible for implementing these standards through emission permits. In 1977, Congress amended the Clean Air Act to require EPA to reevaluate, every 5 years, the scientific basis for air standards. In the 1970s and 1980s, the District of Columbia Circuit of the United states Court of Appeals issued a series of decisions that delineated the legal landscape within which EPA must establish these standards.

When the agency promulgates a standard, it must reasonably explain its decision. Based on statutes and case law requiring regulations to have a rational basis, the court stated that "[i]nherent in the responsibility entrusted to this court is a requirement that we be given sufficient [explanation from EPA]...so that we may consider whether it embodies an abuse of discretion or error of law."

Despite repeated protestations from industry, the court consistently held that EPA must establish air standards based exclusively on protecting public health, without considering implementation costs. To protect public health, the agency must consider members of the public that are most vulnerable to a pollutant, such as those who might suffer from respiratory diseases. The court also construed in the statute a congressional intent that EPA account for scientific uncertainty by incorporating a margin of safety in the standards. This margin of safety was a policy judgment left to EPA's discretion.

Alongside these legal developments, emerging science persuaded the agency that the 1971 PM standards were not adequately protecting the public. Studies were shedding more light on how PM enters the human body. When particulates are inhaled, they are deposited, depending on their size, in the head region or deeper into the respiratory tract, where they pose greater health risks. EPA concluded PM standards would be more protective if they were based on particulates less than or equal to 10 micrometers (about the width of a cotton fiber)—referred to as  $PM_{10}$ .<sup>48</sup>

When EPA promulgated  $PM_{10}$  standards in 1987, it stated that science could not identify a threshold concentration below which the pollutant posed no risk. Instead, the literature which included epidemiological studies of data from London and from six U.S. cities—showed evidence of a variety of health effects along a continuum of  $PM_{10}$  concentrations (see Box 4.1). Based on this evidence and on the mandate for a margin of safety, EPA opted for standards on the lower end of this continuum. It ruled that concentrations of ambient  $PM_{10}$  should not exceed a 24-hour average and an annual average of 150 and 50 µg/m3, respectively.<sup>49</sup>

|  |   |  | I  | Box 4.1  |  |  |                       |  |
|--|---|--|--|--|--|--|-----------------------|--|
| . 8                                    |   | Measured<br>BS levels (as<br>µg/m <sup>s</sup> )<br>Increased<br>respiratory<br>disease,<br>reduced lung<br>function in<br>children <sup>1</sup> | Measured TSP Levels (µg/m ³)   |  |  |  |                       | Equivalent   |
|  | Effects/study                                       |  | Increased<br>respiratory<br>disease,<br>symptoms,<br>small<br>reduction in<br>lung<br>function in<br>adults <sup>2</sup> | Increased<br>respiratory<br>symptoms in<br>adults <sup>3</sup> | Increased<br>respiratory<br>symptoms<br>and illnesses<br>in children 4 | Reduced<br>lung<br>function in<br>children 4 | Combined<br>range     | PM 10 levels<br>(μg/m <sup>3</sup> )<br>Combined<br>range <sup>5</sup> |
| Effects lil<br>Effects po<br>No signif | kely<br>ossible<br>icant <sup>6</sup> effects noted | 230–300<br><230  | 180*<br>130–180*<br>80–130   | 60–150(110)  | 60*-114  | 40-114                                       | >180<br>60–180<br><60 | >80-90<br>40-90<br><40   |

Fom the *Federal Register Notice* promulgating the 1987 PM standards. The first row lists the observed health effects. The first column states EPA's assessment as to whether studies showed the effects to be likely. The other columns and rows indicate the PM levels at which the effects were observed.

The American Iron and Steel Institute (AISI) challenged these standards, arguing that EPA did not reasonably explain why it selected and why it chose to put more weight on certain studies. The court rejected these arguments, stating "AISI essentially asks this court to give

<sup>&</sup>lt;sup>48</sup> 1987 [*Federal Register* ?]FR.

<sup>&</sup>lt;sup>49</sup> Note that we use the term "average" loosely. Actual computation of the permissible concentration allows for a certain number of exceedances.

different weight to the studies than did [EPA]. We must decline. It is simply not the court's role to second-guess the scientific judgments of [the Agency]."<sup>50</sup>

When EPA next revised the PM standards in 1997, it raised two critical issues. First, the emerging evidence showed even more clearly that human biological systems respond differently when exposed to particulates of different sizes. The agency therefore promulgated standards for coarse and fine particulates—referred to as PM<sub>10</sub> and PM<sub>2.5</sub>, respectively (to reflect their diameters, as measured in micrometers). It reasoned that these two species of PM posed independent and distinct threats to public health.

Secondly, the agency struggled with setting a lower limit for PM<sub>2.5</sub>. A statement from an EPA researcher reflects the dilemma: "What you're trying to do is ludicrous—set the level below which the most sensitive person in the population will have no adverse health effects. In fact, it is impossible to come up with a scientifically justifiable number. The only intelligent way is cost-benefit or cost effectiveness analysis, but we don't do that."<sup>51</sup> Ultimately, the agency ruled that the lower limits for PM<sub>2.5</sub> standards were set by scientific uncertainty; at lower exposure levels, the evidence of harm was too uncertain to justify regulation.

These 1997 standards did not survive the court's scrutiny. The court avoided framing its decision as second-guessing the agency's scientific judgment. Rather, it explained that when EPA promulgated PM standards,<sup>52</sup> its decision making effectuated an unconstitutional delegation of legislative power to an administrative agency. According to the court, EPA failed to articulate an "intelligible principle" that would have helped the court determine whether the agency had abused its discretion under the Clean Air Act.

EPA appealed the court's decision. The Supreme Court disagreed with the lower court and ruled that Congress provided EPA with sufficient guidance to exercise its administrative discretion under the Clean Air Act. The case was returned to the lower courts to consider the technical merits of the agency's 1997 standards, which were ultimately upheld.

In 2006, EPA—in the person of Administrator Stephen Johnson, speaking on behalf of President George Walker Bush's Administration—once again argued that scientific uncertainty precluded more stringent PM standards. By that time, the agency had established an administrative infrastructure through which the scientific evidence for air standards were being marshalled and vetted. The evidence was collated, summarized, and disseminated in reports made available to the public. This evidence, along with the reports, was reviewed by an external group of experts in air modeling and public health. This group of experts was organized through a documented and public process; the group's deliberations and recommendations were available to the public. In short, the scientific reasoning behind EPA's air standards hewed to a transparent, public process.

As part of this process, agency scientists and the external group of experts proposed a more stringent annual standard for  $PM_{2.5}$ , a recommendation partially based on a study showing the effects of short-term PM exposure on children's health. To support his decision not to follow this recommendation, Administrator Johnson noted he chose not to include this study in his assessment.

<sup>&</sup>lt;sup>50</sup> *NRDC v. EPA*, 902 F.2d 962, 971 (D.C. Cir. 1991).

<sup>&</sup>lt;sup>51</sup> Quoted in Mark R. Powell, *Science at EPA: Information in the Regulatory Processs* (1997).

<sup>&</sup>lt;sup>52</sup> As well as the ozone standards, which were issued at the same time.

The court found this conclusory statement unpersuasive. It remanded the 2006 standards back to EPA for a more "adequate" explanation.<sup>53</sup> Even as the court was deliberating over the 2006 PM standards, EPA was already improving its decision making for air quality standards in general. It published a set of principles that would guide the agency in integrating information culled from the scientific literature. These principles are summarized in Box 4.2, which are adapted from the agency's promulgation of PM standards in 2013.<sup>54</sup>



EPA's framework for scientific robustness. It uses a tiered system to evaluate the strength of causality embodied in a study. Within these studies, there are some that, by nature of their design, are particularly probative (in bold below). Looking across studies, EPA develops a distribution of the levels at which a pollutant shows adverse health effects. To apply a margin of safety, the agency selects a standard at the lower end of these distributions, for example, at the 10th to 25th percentiles of each distribution.

First, to select and weigh evidence culled from multiple studies, EPA developed a "causal framework" based on the voluminous work on causality done by other agencies and by

<sup>&</sup>lt;sup>53</sup> *Am. Farm Bureau Fedn. v. EPA*, 559 F.3d 512, 528 (D.C. Cir. 2009).

<sup>&</sup>lt;sup>54</sup> Environmental Protection Agency, *Policy Assessment for the Review of the PM NAAQS* (April 2011), Figure 2-8, available at <a href="http://www.epa.gov/ttn/naags/standards/pm/data/20110419pmpafinal.pdf">http://www.epa.gov/ttn/naags/standards/pm/data/20110419pmpafinal.pdf</a>.

the greater scientific community.<sup>55</sup> The agency's discussion of this framework—a tiered system for evaluating the strength of evidence for a causal relationship—clarifies the trade-off among certainty, weight of evidence, and utility. EPA places the greatest weight on the most certain evidence of causality, which typically comes from controlled human exposure studies that rule out all causal factors other than the pollutant.<sup>56</sup> However, these studies may not extend to the real world, where a heterogeneous population is exposed by multiple routes to a complex mix of ambient air pollutants.<sup>57</sup> Therefore, EPA must rely on epidemiologic studies that—by their experimental design, analytical adjustments for possible confounding factors, and reproducibility—suggest a causal relationship is likely, notwithstanding significant uncertainties.<sup>58</sup>

A second analytical milestone was EPA's development of a more rigorous method for weighing the collective studies and evaluating uncertainty.<sup>59</sup> Whatever biological and physical reality might underlie the NAAQS, it is shrouded within uncertainties.<sup>60</sup> Therefore, the agency relies on a distribution of study results, rather than a single point estimate.<sup>61</sup> This type of analytical work does not override the uncertainty, but it creates a more detailed picture of the bounds and distribution of existing evidence with respect to the future standard. Acknowledging that there is no single, "correct" way to choose a level of acceptable uncertainty, the agency proposed that the 10th to 25th percentiles in the distribution was a reasonable range within which to exercise its mandate to set a margin of safety.<sup>62</sup>

This last statement is quite profound. Indeed, it is downright Bayesian (as we shall soon see) in its explicit acceptance that "The Perfect Model"—that is, the model that captures empirical truth with absolute certainty—is chimerical. As we saw in the previous section, it has been a matter of long-held, scientific consensus that the environmental causes of public health risks can best be inferred by integrating information across multiple studies, derived from both controlled experiments and uncontrolled observations. The former yields data about the biological mechanisms underlying disease, while the latter investigates causal relationships under the varied conditions of the real world.

In the earliest days of the PM standards, EPA relied heavily on the weight of a few influential studies, such as those based on London and New York public health data. As the scientific literature on PM grew, EPA spent considerably more time to explain its reasons for choosing information from one study rather than the other. This did little to ward off accusations that the agency was cherry-picking evidence to support its conclusions.

<sup>&</sup>lt;sup>55</sup> Environmental Protection Agency, *Integrated Science Assessment for Particulate Matter* (December 2009), available at <a href="http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=216546">http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=216546</a>.

<sup>&</sup>lt;sup>56</sup> Id. 1-15.

<sup>&</sup>lt;sup>57</sup> Id.

<sup>&</sup>lt;sup>58</sup> Id. 1-18.

<sup>&</sup>lt;sup>59</sup> See U.S. Environmental Protection Agency, *Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards* (April 2011), EPA 452/R-11-003, 2-78–2-81.

<sup>&</sup>lt;sup>60</sup> "National Ambient Air Quality Standards for Particulate Matter," 78 Fed. Reg. 3173, 3174 (January 15, 2015).

<sup>&</sup>lt;sup>61</sup> See Policy Assessment for Particulate Matter supra note 59.

<sup>&</sup>lt;sup>62</sup> It bears highlighting that this epistemic framework for robustness adopted by EPA was itself the product of scientific review. As such, it reflects many discussions within the scientific community about appropriate methods to reach robust inferences. Id.

When it established its latest PM standards in 2013, EPA articulated an epistemic frame that formalized, at least in *qualitative terms*, how the agency evaluates evidence across multiple sources. It relies on a distribution of study results and assigns the weight of evidence more heavily to those studies that control—whether through experimental controls or analytical means—possible confounding factors.

From the thought experiments we introduced in Section 2.1, we understand the role of experimental controls in inferring causation. At that point, we only alluded to the fact that analytical means also exist to manage the influence of confounding factors. We will use one of the studies listed in Box 4.1 to explain how this is done.

To investigate the effect of  $PM_{2.5}$  on the elderly, Bell et al. (2007) analyzed data drawn from 202 counties in the years 1999–2005. To deal with the expected spatial and temporal heterogeneity in these data, the researchers applied a technique called *hierarchical Bayesian modeling*.

Recall from our earlier discussion of the Galton Box that though we may not know the precise path taken by each individual pellet, their ultimate fate—in the *aggregate*— approximates some mathematical distribution. In Bayesian modeling, we use distributions to forecast the likeliest outcome for a population. In hierarchical Bayesian modeling, we assume that these distributions are a combination of population-wide and subpopulation phenomena.

For example, researchers in the Bell study estimated a model that accounted for spatial and temporal factors that were unique to a specific county, as well as general factors that were true across all counties. By looking simultaneously at factors within and across counties, the hierarchical Bayesian model provided the means to analyze heterogeneous data while controlling for the confounding factors that might impede causal inference. Another description of the technique used by Bell is provided in the following journal extract.

### Pasky Pascual "Evidence-based decisions for the wiki world"<sup>63</sup>

For example, EPA recently relied on inductive inference to analyse the effect of a pollutant, methylmercury, on IQ. The underlying data was drawn from three independent studies. In an ideal experiment, controls would have maintained identical conditions across the studies, so that measures of methylmercury and IQ would be the only source of variability. But science must often proceed under less than ideal conditions: the studies were conducted in different locales using different methods to measure IQ. To surmount these difficulties, the statistician Ryan used information about each study in a form of inductive inference, Bayesian analysis.

From a Bayesian perspective, the true relationship between methylmercury and IQ can only be known when one has complete access to all the information about a complex phenomenon that one needs. . .in other words, never. From a Bayesian perspective, each single study provides useful—albeit incomplete—data about this true relationship. However, by using improved algorithms to randomly and iteratively sample data pooled from multiple studies, Ryan was able to approximate an overarching model to describe the relationship between a

<sup>&</sup>lt;sup>63</sup> Pasky Pascual, "Evidence-based decisions for the wiki world," *Int. J. of Metadata, Semantics, and Ontologies* 4: 287 (2009) (footnotes omitted). Copyright © (2009) Inderscience Enterprises Limited. All rights reserved. Reproduced with permission.

mother's exposure to methylmercury and her child's IQ. This model was general enough to explain the broad patterns of data across the studies, even as it accounted for variability arising from particular aspects of each individual study. The model could also be said to be scientifically truthful in the following sense: given assumptions regarding the data's underlying probabilities and given the data on hand, the model was the most likely quantitative description of the relationship in question.

What all of the above makes clear is that nonscientist decision makers cannot ignore the scientific practice of modeling. They cannot treat models as black boxes cranking out answers. While nonscientists are axiomatically not scientists, they must grasp some fundamental features of scientific practice to ensure their decision making is robust.

#### **Check Point L: Reading a Scientific Paper**

You have just read an excerpt describing researcher Louise Ryan's investigation of the health effects of mercury. In the excerpt, Pascual tried to simplify and synthesize Ryan's work. If you have never read a paper in a science journal, you might want to read Ryan's original paper, available at the following <u>hyperlink</u>. When reading the paper, ask yourself: What problem was Ryan trying to solve? Why did she use the analytical method she did? Do not be discouraged if you have a difficult time reading the paper; the paper assumes that the reader has years of technical training.

## 7. CONCLUSIONS

Returning to our starting point, big data is not enough for making important decisions nor are models. Decision -makers who must make decisions based on models need to understand how models are constructed. This understanding includes understanding their sources of data, assumptions, framing, and method of review and evaluation. Models must also be accessible to those that use and rely on them.

In this module we have explored the use of models in regulatory decision making. We have emphasized the importance of context, but the basic scientific nature of models remains constant, and thus much of what we have said is relevant to other decision-making contexts. Models are never truth machines, and it is important to understand how they frame the world and the assumptions and data deployed in their development.

Models are not only deployed in regulatory decision making, however. For example, models are often used in tort cases to determine causation and thus subject to the rules of expert evidence as set out in evidence law. Here again the idea that models are "truth machines" pervades in some courtrooms. For example, a philosopher of science, Dr. Carl Cranor, observes that some courts exclude expert testimony based on animal models and related research "simply because it does not represent a complete or definitive answer to a larger policy or science question." These courts find that these "weight of the evidence" models are insufficiently rigorous or probative and insist instead that plaintiffs support their

causation claims with epidemiological research, despite the fact that this evidence is often unavailable or inconclusive.

We also saw in Section 3.2 how models can be used to predict baseball results, and they are used in many other areas of business practice, particularly to assess risk. As with regulatory decision making, there is a temptation for decision makers to treat such models as truth machines.

# Gillian Tett, "How not to fall foul of the model makers"<sup>64</sup>

When the financial crisis hit, many observers blamed the disaster on the misuse of financial models. Not only had these flashy computer systems failed to forecast behaviour in the sub-prime mortgage world, but they had also seduced bankers and investors to take foolhardy risks or been used to justify some crazy behaviour.

But these days, in spite of all those missteps, there is little sign that financiers are falling out of love with those models; on the contrary, if you flick through the recent plethora of reports from the Basel Committees—or look at the 2011 forecasts emanating from investment banks—these remain heavily reliant on ever-more complex forms of modelling.

Tett then goes on to talk about a recent book of Emanual Derman. Derman developed some important innovative financial models, but, as she notes, is not a "model lover."

Thus while investors might like to see models as akin to physics theories, Derman says this is wrong; models are more like expansive metaphors, of the sort found in literature or philosophy. In essence, they are a tool to help us think, and order our world view, and explain something that is hard to grasp. "Models are reductions in dimensionality that always simplify and sweep dirt under the carpet. Theories tell you what something is. Models tell you merely what something is partially like."<sup>65</sup>

This is consistent with that given above and underlying all of it is a very basic premise when using scientific concepts, it is important to understand the scientific nature of what you are using. We recognize that is easier said than done. A module like this can never entirely bridge the disciplinary divide between modeling as a scientific practice and the administrative, legal, and business contexts that modeling is operating in. But what has been provided in this module is a framework for thinking about and understanding models. It is thus a framework for better, more effective, and more accountable decision making.

<sup>&</sup>lt;sup>64</sup> Gillian Tett, "How not to fall foul of the model makers," *Financial Times* (Dec. 23, 2010). Used under license from the Financial Times. All Rights Reserved.

<sup>&</sup>lt;sup>65</sup> Id.