Not a Sure Thing: Making Regulatory Choices under Uncertainty

Alan Krupnick, Richard Morgenstern, Michael Batz,
Peter Nelson, Dallas Burtraw, Jhih-Shyang Shih,
and Michael McWilliams

February 2006

RESOURCES FOR THE FUTURE

Table of Contents

Acknowledgments

The authors gratefully acknowledge financial support from the U.S. Environmental Protection Agency under Contract #68-W-02-046. Helpful comments and assistance were received from EPA project officer Mark Corrales and from Alexander Cristofaro, both of the agency's Policy Office. Charlotte Bertrand, who served as project officer in an early phase of the work, and Brian Morrison, the project coordinator from IEC, also provided helpful comments and guidance. Special thanks to the seven former EPA appointees interviewed for the project: Don Clay, Terry Davies, Linda Fischer, David Gardiner, Lynn Goldman, Hank Habicht, and Tracy Meehan. Roger Cooke, Chauncey Starr Senior Fellow at RFF, led the development of the "toy" model in Appendix 3E and provided the cobweb plots. Karen Palmer, Darius Gaskins Senior Fellow at RFF, offered helpful comments and input at various stages of the project. Danny Kahn provided able research assistance through running the HAIKU Model. Any errors are the responsibility of the authors.

Chapter 1: Introduction*

The use of formal uncertainty assessment, especially Monte Carlo analysis, is becoming more common in many fields, including economics, finance, engineering, and certain scientific disciplines. Beginning in the 1990s, the U.S. Environmental Protection Agency (EPA) investigated methods for conducting uncertainty analysis and issued *Guiding Principles for Monte Carlo Analysis* (EPA 1997b). The agency also conducted at least some form of uncertainty analysis in several congressionally mandated studies and regulatory impact analyses (RIAs). For example, both the "retrospective" and the "prospective" analyses of the Clean Air Act Amendments of 1990 (U.S. EPA 1997a, 1999) incorporated limited uncertainty analyses. The RIA for nonroad diesel engines (U.S. EPA 2004), which became effective in 2004, involved a formal use of uncertainty analysis. Several EPA workshops and the EPA Science Advisory Board also have weighed in on this topic.

The incorporation and sophistication of uncertainty analyses in RIAs will continue to grow as EPA responds to *Circular A-4,* issued by the Office of Management and Budget (OMB 2003). This document requires uncertainty analyses for RIAs that have annual costs and/or benefits in excess of $1 billion. The greater use of uncertainty analysis in RIAs also was bolstered by a 2002 National Research Council (NRC) study. Although facing many challenges in responding to *Circular A-4,* EPA is in a strong position because of early and continuing work on the topic. The OMB requirements can be seen as an opportunity to build on the analytic rigor of these analyses and improve the communication of results to decisionmakers and the general public.

Introducing uncertainties into RIAs is an important activity for at least two reasons. The first is that public policy will be better served. If uncertainties are not

* Krupnick (krupnick@rff.org), Morgenstern (morgenstern@rff.org), and Burtraw (burtraw@rff.org) are senior fellows, Shih (shih@rff.org) is a fellow, Batz (batz@rff.org) and Nelson (nelson@rff.org) are research associates, and McWilliams (mcwilliams@rff.org) is a research assistant at Resources for the Future.

considered, then the resulting policy choices may fail to meet internal (but often unstated) agency criteria—or even the public's criteria—for choosing the best policy amidst high uncertainty. Consider, for example, two policy alternatives that are expected to yield the same net benefits but whose outcomes have quite different variances. Absent formal uncertainty analysis, a decisionmaker would not have information about the nature or magnitude of the underlying uncertainties and might choose the option that fails to meet key ex ante policy objectives (e.g., to deliver at least a minimal level of risk reduction or net benefits with a high degree of certainty).

The second reason is more political in nature. Sole reliance on point estimates masks the underlying distribution of benefit and cost estimates, thereby giving a false sense of security to decisionmakers and the public. If the policy turns out to be a poor choice, then people may well look to the RIA to determine whether the agency had predicted the possibility of such an outcome. To avoid being blindsided, decisionmakers should build uncertainties into their ex ante analyses.

At the same time, better and more complete information does not necessarily lead to better policies. Complex information can confound rather than enlighten or can paralyze the decisionmaking process. Thus, any improvements in capturing uncertainty analytically must be matched by improvements in its communication—not only to those who make regulatory decisions on the basis of such information but also to stakeholders, judges, the press, and the general public.

Accordingly, Resources for the Future (RFF) was asked to provide some guidance and recommendations to EPA about addressing the requirements of *Circular A-4* regarding formal uncertainty analysis and improving its communication. We reviewed the uncertainty literature and EPA practice and conducted an in-depth case study on a hypothetical proposed rule to help test various ideas and new analytical directions. In addition, we went beyond *Circular A-4* to address how to communicate such complex analyses to decisionmakers; we presented our case study results to several former EPA decisionmakers and asked them to choose a policy option, then indicate our problems and successes in communicating our results to them.

Our exploratory study is described in this report, which consists of five additional chapters and several appendices.

In Chapter 2, we review literature on the modeling and analysis of uncertainty in EPA RIAs. Many RIAs (with the exception of recent ones) provide qualitative assessments or sensitivity analyses. Sensitivity analysis typically involves consideration of how the highest and lowest plausible values of key inputs affect net benefits. Such an approach, while often valuable, tends to examine highly unlikely outcomes and is generally given little weight in decisionmaking processes. Because such analysis does not provide information about the full distribution of costs and benefits, it is quite difficult for decisionmakers to incorporate such data into their deliberations.

More recent RIAs reflect a greater use of uncertainty analysis. In particular, the Clean Air Mercury Rule RIA (U.S. EPA 2005) is a major improvement over previous RIAs in the sense that uncertainties are included in the main exposure and benefits estimates. This is precisely what NRC and OMB recommend in their call to bring uncertainty into the primary analysis.

In Chapter 2 we also focus on typologies of uncertainty. We adopt a five-dimensional classification scheme in determining the origin of uncertainty. uncertainties arise from random variation in data (variability), lack of knowledge about an empirical quantity (parameter uncertainty), incorrect model specification (model uncertainty), modeling choices that reflect implicit decisionmaker judgment (decision uncertainty), or interpretation of information or language (linguistic uncertainty). We also identify various techniques for incorporating uncertainty into analyses and identifies the cobweb plot (used in our analysis and reported later in the report) as a promising innovation.

In Chapter 3, we present a methodologically oriented case study that develops a full uncertainty analysis of a hypothetical tightening of the cap on electric utility NO_x emissions beyond that required by the Clean Air Interstate Rule (CAIR). Particular innovations in this case study are the consideration of alternative population assumptions on both cost and benefit sides of the ledger and the inclusion of uncertainty in source–receptor relationships on the benefits side and natural gas price uncertainty on the cost side. Both parameter and model uncertainties are considered via Monte Carlo simulation and scenario analysis applied to our Haiku electricity model and the Tracking Analysis Framework (TAF) model for benefit assessment. Chapter 3 offers methodological as well as policy conclusions.

In Chapter 4, we review literature on communicating uncertainty to decisionmakers. The focus is on the underlying research issues in risk communication and the types of situations that are consistent with the notion of *decision invariance*, which holds that "different representations of the same problem should yield the same preference" (Tversky and Kahneman 2000, p. 211). We also review approaches to presenting uncertainty (e.g., words, numbers, and pictures) and describe what the research shows about the strengths and weakness of each approach.

In Chapter 5, we present the results of seven in-depth interviews of former presidential appointees to EPA concerning the use of uncertainty analysis. Specifically, a simplified version of the case study developed in Chapter 3 was presented to interviewees, using different approaches to present the uncertainties. Interviewees were asked to express preferences for different regulatory outcomes on the basis of the alternative approaches used to present the information. In addition, interviewees were asked several general questions about the use of uncertainty analysis in regulatory decisionmaking.

Finally, in Chapter 6 we present a series of conclusions and recommendations as well as ideas for future research on uncertainty analysis. Our recommendations offer guidance for addressing the requirements of *Circular A-4,* incorporating key elements of uncertainty into analyses, and expanding institutional capacity and research support for issues raised in this report.

References

NRC (National Research Council). 2002. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations.* Washington, DC: National Academies Press.

OMB (Office of Management and Budget). 2003. *Circular A-4.* Washington, DC: OMB.

Tversky, A., and D. Kahneman (Eds.). 2000. Rational Choice and the Framing of Decisions. In *Choices, Values, and Frames.* New York, NY: Cambridge University Press.

U.S. EPA (U.S. Environmental Protection Agency). 1997a. *The Benefits and Costs of the Clean Air Act: 1970 to 1990.* Washington, DC: U.S. EPA, Office of Air and Radiation, Office of Policy.

———. 1997b. *Guiding Principles for Monte Carlo Analysis.* Washington, DC: U.S. EPA, Risk Assessment Forum, Office of Research and Development.

———. 1999. *The Benefits and Costs of the Clean Air Act 1990 to 2010.* Washington, DC: U.S. EPA, Office of Air and Radiation, Office of Policy.

———. 2004. *Final Regulatory Analysis: Control of Emissions from Nonroad Diesel Engines.* Washington, DC: U.S. EPA, Office of Transportation and Air Quality.

———. 2005. *Regulatory Impact Analysis of the Clean Air Mercury Rule.* Washington, DC: U.S. EPA, Office of Air Quality Planning and Standards.

Chapter 2: Uncertainty in the Literature and in EPA RIAs

The role of uncertainty in U.S. Environmental Protection Agency (EPA) regulatory impact analyses (RIAs) is evolving. In recent years, there have been increased calls for EPA to more fully incorporate quantitative uncertainties into estimates of the benefits and costs of environmental regulation (Kopp et al. 1997, Easter and Archibald 1998). In particular, the National Academies of Science (NAS) and the Office of Management and Budget (OMB) have suggested specific improvements for EPA, such as moving uncertainty analysis from secondary into primary analysis and more fully reporting ranges rather than point estimates.

In this chapter, we review some of the guidance and suggestions given to EPA by outside experts and the handling of uncertainty in a handful of recent RIAs developed by the agency. The extent to which EPA can better analyze uncertainties in estimates largely depends on the abilities of the environmental models underlying the analyses. EPA has already started to move toward probabilistic integrated assessment models, but this process will take considerable effort and time.

To this end, this chapter presents a brief but detailed overview of the primary issues in modeling with uncertainty. We discuss how to identify different sources of uncertainty that arise in quantitative policy analysis, incorporate this uncertainty information into a probabilistic modeling framework, and choose among the many available methods that can be used to analyze uncertainties and sensitivities in different forms of deterministic and probabilistic models. We also discuss recent EPA work on bringing uncertainty into their analyses and close with some conclusions and recommendations.

Motivation: NRC and OMB

Two recent documents serve as motivation to examine the treatment of uncertainty in EPA RIAs. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations* (NRC

2002) is a review of the methodologies and data that EPA uses to estimate benefits and provides suggestions and best practices for how EPA can improve these estimates. Uncertainty analysis in benefits estimates is considered in its own chapter and is highlighted in the executive summary of the National Research Council (NRC) report as well. Similarly, *Circular A-4* (OMB 2003) on regulatory analysis gives specific and detailed guidance to EPA on uncertainty analysis. Although various topical areas related to uncertainty offer a history of guidance for EPA (discussed in Appendix 2A), EPA RIAs offer no detailed guidance on uncertainty analysis or sensitivity analysis.

The NRC (2002) study and OMB (2003) guidance are similar in many respects, including calls to avoid false precision of point estimates, to move uncertainty analysis from secondary into primary analysis, to use formal uncertainty analysis procedures such as Monte Carlo (MC) simulations, to quantify uncertainties using expert judgment where appropriate, to identify driving uncertainties through importance assessment, and to use value-of-information (VOI) approaches to estimate the benefits of reductions in uncertainty.

NRC (2002) notes that uncertainty analysis is a critical component in realistically estimating the health benefits of air pollution regulation and suggests that EPA should move toward incorporating uncertainty in primary estimates. It offers a general critique of EPA uncertainty assessments, including those done for the Tier 2 rule and the prospective analysis of the 1990 Clean Air Act (CAA) Amendments (U.S. EPA 1999a, 1999b):

> EPA's decision to incorporate only one source of uncertainty, the random sampling error in the estimated concentration–response function, into the probability distributions resulting from its health benefits analyses is worth reconsidering. The committee agrees with the agency's judgment that its current practice produces health benefits probability distributions that give "a misleading picture about the overall uncertainty in the estimates." In particular, the distributions suggest that there is less uncertainty, perhaps much less, than is actually present. The committee finds that the mean of the distributions should not be interpreted as "best" estimates, and the intervals between the 5th and 95th percentiles of the distributions should not be interpreted as "90 percent credible intervals," within which "the true benefit lies with 90 percent probability." (NRC 2002, p. 133)

Circular-A4 (OMB 2003) suggests first and foremost reporting probability distributions around estimates of consequences and using formal quantitative uncertainty analysis for major rules involving annual economic effects of $1 billion or more. Going beyond the NRC, OMB recommends using “real options” analysis to determine whether more research is needed before rulemaking and to estimate the costs and benefits of delaying a decision until better information is available. OMB gives very explicit instructions on discount rate scenarios and analyzing distributional effects, especially intertemporal and intergenerational inequities.

Although the NRC (2002) and OMB (2003) documents are very similar, they contain some important differences. The largest one is that whereas NRC recommends a rather exhaustive approach to incorporating uncertainty into estimates, OMB recommends “balance between thoroughness and the practical limits of your analytical capacity” (p. 7). Whereas NRC recommends quantifying as many uncertainties as possible, OMB recommends quantifying and analyzing only the largest drivers of uncertainty.

The NRC (2002) approach is sound, but we agree with OMB (2003) that practical issues are important and should influence choices of how to model and analyze uncertainty. The main points of the NRC report and OMB guidance are summarized in Table 2-1.

Treatment of Uncertainty in EPA RIAs

An analysis of four recent EPA RIAs for air pollution regulations, presented in detail in Appendix 2B, indicate increased use of uncertainty analysis, though considerable opportunities remain to expand use of the subject. In general, EPA RIAs do not adequately represent uncertainties around “best estimates,” do not incorporate uncertainties into primary analyses, include limited uncertainty and sensitivity analyses, and make little attempt to present the results of these analyses in comprehensive way. These RIAs tend to discuss uncertainty qualitatively—to present tables and lists of sources of uncertainties in various component models and estimates—but generally avoid quantitative inclusion or the reporting of uncertainties in estimates.

Overall, there is a tendency to avoid formal uncertainty analyses unless the uncertainties can be included comprehensively and quantified precisely. An alternative—arguably, preferred—approach would be to conduct uncertainty analysis as best as possible, even if abilities are limited; almost any uncertainty analysis is better than none at all. The strong emphasis is on presenting point estimates as the primary results, with ranges and uncertainties generally buried in appendices, but this approach may be changing if the recent RIA for the Clean Air Mercury rule (U.S. EPA 2005b) is any indication.

We believe that many opportunities exist to expand the use of uncertainty analysis in RIAs. Uncertainty analysis could be accepted as a routine component of models, predictions, assessments of regulation, and estimates of costs and benefits, rather than something to add in to a model after the fact. The task for EPA, therefore, is not merely one of performing analysis on the results of existing models but the far more complicated job of modifying or converting these models so that they can incorporate uncertainty analyses. Arguably, this assignment is resource intensive, especially considering the complexities of the large environmental systems models on which EPA relies, and it may take EPA a considerable amount of time before it is possible to satisfy the suggestions of the NRC (2002) report and OMB (2003) guidance. Indeed, a primary reason that uncertainty is not adequately addressed in RIAs may be that current modeling capabilities are insufficient and replacement models take time to develop.

Regardless, we believe that it is important to describe in detail the uncertainties that occur in quantitative policy analysis and environmental modeling as well as how those uncertainties can be incorporated at the earliest stages of environmental modeling. As such, we present a detailed typology of uncertainties relevant to RIAs to give some idea of the breadth and depth of uncertainties in such modeling and to clarify some distinctions between types, because treatment in models varies significantly. We then discuss issues with modeling uncertainty in probabilistic models, such as simulation approaches, data fitting issues, the incorporation of expert judgment, and the separation of uncertainty from variability.[1] Next, we give an overview of the many sensitivity and uncertainty analysis methods that can be applied to deterministic and probabilistic models. Many of these approaches could be used on current EPA models while they develop the next generation of probabilistic models.

After these technical discussions, we describe some of EPA's recent progress in moving toward probabilistic integrated assessment models. Recent meetings, symposia,

[1] The distinction between *uncertainty* and *variability* is discussed in the following section.

and projects show an increasing inclusion of uncertainty in the development of new models and the analysis of results from existing and future models.

Typology of Uncertainty

> Very possibly, we may even be uncertain about our degree of uncertainty. The variety of types and sources of uncertainty, along with the lack of agreed terminology, can generate considerable confusion. (Morgan and Henrion 1990, p. 47)

Uncertainty is a term both vague and specific, used in everyday conversation as a synonym for doubt. However, it also is often used in more technical settings, such as quantitative policy analysis, to express in various ways the degree to which one is unsure about analytical results. Even in this latter sense, the intended meaning may be unclear. For example, take the circularity of the sentence, "There are two kinds of uncertainty: variability and uncertainty." Depending on the context, uncertainty might refer to the overall degree of imprecision or unpredictability or it might refer only to that not due to the inherent heterogeneity across individuals, space, or time. In this text, for example, unless otherwise specified, *uncertainty* is used in the general sense that includes variability.

In addition to the vagueness around the word itself, multiple sources of uncertainty exist throughout any quantitative environmental policy analysis, and many have distinct characteristics that require distinct treatment in modeling and analysis. A few such sources include variation in measured data, disagreement between alternate sources of information, natural heterogeneity, the selection of one model form over another, simplifications of model structure, extrapolation errors, and value judgments.

To model, analyze, and ultimately reduce these uncertainties, one may find it useful to organize uncertainties into a meaningful classification scheme. Unfortunately, in addition to the vagueness of the term *uncertainty* and the multitude of its sources, the literature contains numerous classification schemes. We know of no universal typology or taxonomy of uncertainty, though there have been numerous attempts over the past 20 years relevant to quantitative environmental policy (e.g., Kahneman and Tversky 1982, Bogen and Spear 1987, Morgan and Henrion 1990, Finkel 1990, Frey 1992, NRC 1994, Ferson and Ginzburg 1996, NCRP 1996, Cullen and Frey 1999, Kann and Weyant 2000, Regan et al. 2002, Linkov and Burmistrov 2003, Haimes 2004).

These attempts are the basis for the discussions of uncertainty in a growing body of EPA documents, for example, *Guidelines for Exposure Assessment* (U.S. EPA 1992), *Guiding Principles for Monte Carlo Analysis* (U.S. EPA 1997b), *Risk Assessment Guidance for Superfund (RAGS),* Volume III, Part A (U.S. EPA 2001b), and *An Examination of EPA Risk Assessment Principles and Practices* (U.S. EPA 2004c). Many of these typologies overlap or are built on each other. Frey (1992) and Cullen and Frey (1999) expand on Morgan and Henrion (1990), whereas EPA (U.S. EPA 2004c) follows NRC (1994), which is itself based heavily on Finkel (1990).

The typologies described in the literature generally share two critically important distinctions in type of uncertainty: they distinguish variability from lack of knowledge, and they distinguish uncertainties at the parameter from those at the model level. There are differences in these typologies, however. Morgan and Henrion (1990) and Frey (1992) make the primary distinction between parameter and model uncertainty and characterize variability as a form of parameter uncertainty. NRC (1994) makes the primary distinction between variability and uncertainty, and considers parameter and model uncertainties within the latter.

Other typologies abound, and a few are worth noting. In the context of scientific decision support, van Asselt (1999) overlays two typologies of uncertainty. First, variability and lack of knowledge are broken down into subcategories. Second, three degrees of uncertainty are distinguished: *technical uncertainties* (data quality, appropriateness of data), *methodological uncertainties* (choices between conflicting data, defining causal relationships between data), and *epistemological uncertainties* (model limitations, indeterminacies).

Schneider et al. (1998) give an overview of three additional typologies. Funtowicz and Ravetz (1990) marry the degree of uncertainty to the stakes of the decision; the uncertainties in small systems with low stakes are largely due to scientific judgments, but these uncertainties give way to a broader lack of knowledge when the systems are large and the stakes are high. Wynne (1992) lists four types of uncertainty: *risk* (when the system and probabilities of outcomes are well known), *uncertainty* (when probabilities of outcomes are not known), *ignorance* (the limited ability to know systems), and *indeterminacy* (due to complexity and unpredictable system behavior).

Rowe (1994) describes a different set: *temporal* (interpreting the past, predicting the future), metrical (errors in measurement), *structural* (due to complexity), and *translational* (conveying the uncertainty of results). Furthermore, Rowe lists three sources of variability, which can apply to any of the four types of uncertainty:

underlying variants (randomness, inconsistent behavior, chaos), *collective/individual membership assignment* (modeling individual behavior based on collective behavior), and *value diversity* (varying perspectives, preferences, morals among people).

Many typologies of uncertainty in the literature are built on previous typologies, and few point out the important differences between other available typologies. None of the typologies with which we are familiar is comprehensive enough to have complete coverage yet disaggregated enough to make all of the important distinctions, and so we have created a composite typology of uncertainty. The typology we present is based heavily on Morgan and Henrion (1990), NRC (1994), and Cullen and Frey (1999), but it also includes important components from Finkel (1990), Haimes (2004), and many of the other typologies previously acknowledged. We agree with the principal distinction between variability and lack of knowledge, which can be broken down into three further broad categories, leaving us with four primary "types" of uncertainty: variability, parameter uncertainty, model uncertainty, and decision uncertainty.

Variability and parameter uncertainty apply to empirical quantities, which are model variables that represent measurable properties of the system being modeled (Morgan and Henrion 1990). *Variability* is the inherent heterogeneity of an empirical quantity across a population (of people or objects), space, or time. *Parameter uncertainty* is the lack of knowledge about an empirical quantity stemming from limitations of measurement, disagreement among measurements, or extrapolation errors. *Model uncertainty,* like parameter uncertainty, is due to lack of knowledge, but it is of a higher order and represents uncertainties in model structure. Models are approximations of real-world systems and have implicit uncertainties because of methodological decisions about model form and the epistemological limitations of a system or phenomena. *Decision uncertainties* are introduced by modeling choices that reflect implicit decisionmaker judgment about how to value societal outcomes.

It must be noted that although the taxonomy developed here provides theoretical dividing lines between types of uncertainty, analysts and risk managers need not be overly concerned with making such distinctions, especially where differences may be subtle or uncertainties might overlap. The important consideration in modeling is the identification and treatment of uncertainties. The typology exists primarily to help risk managers identify sources of uncertainty in the models used in RIA so that they can be properly treated or analyzed, where appropriate. Haimes provides an example of how a typology was used to identify uncertainties in a decisionmaking and modeling process (Haimes 2004, p. 246–251). In this application, decisionmakers went through Haimes' typology, type by type, identifying points in their decision model with each kind of associated uncertainty. This full listing of uncertainties was then used to inform subsequent sensitivity and uncertainty analyses.

Variability

Variability has many other names in the literature, including ontological, aleatory, stochastic, objective, and process uncertainty. The distinction between variability and parameter uncertainty is important. Variability occurs when an empirical quantity that could be measured as a single point value actually exists in a population of values, varying across space, time, or across individuals. A common scientific use of the term *variability* is to imply the dispersion of values about a central tendency due to random error of a repeated experiment or measurement; we define this occurrence as a type of parameter uncertainty, and it should not be considered variability.

One important type of variability is the diversity of moral beliefs or preferences within a population of people (Rowe 1994). Some typologies classify uncertainties about human values separately, distinguishing them from other forms of uncertainty. In our view, such value uncertainties are indeed different from many other types of uncertainty but can, at least theoretically, be decomposed into variability and parameter uncertainty. There is heterogeneity of values across a population and parameter uncertainty associated with the difficulties of adequately measuring these values, even if the total value uncertainty cannot practically be decomposed into these two parts.

One polarizing but important example is valuing mortality through VSLs. The range of estimates in the literature are very broad, ranging from about $1 million to more than $10 million. This range is a function both of the heterogeneity across individuals of how much small risk changes are worth as well as the methodology used to measure this value (e.g., contingent valuation [CV] or labor market).

Science and Judgment in Risk Assessment gives a detailed account of variability in risk assessments, and provides the following example to describe the sources of variability in modeling a single air pollutant substance originating from a single stationary source to its health endpoint:

> *Emissions* vary temporally, both in flux and in release characteristics, such as temperature and pressure. The *transport and fate* of the pollutant vary with such well-understood factors as wind speed, wind direction, and exposure to sunlight (and such less-acknowledged factors as humidity

> and terrain), so its concentrations around its source vary spatially and temporally. Individual human *exposures* vary according to individual differences in breathing rates, food consumption, and activity (e.g., time spent in each micro-environment). The *dose–response* relationship (the "potency") varies for a single pollutant, because each human is uniquely susceptible to carcinogenic or other stimuli (and this inherent susceptibility might well vary during the lifetime of each person, or vary with such things as other illness or exposures to other agents). (NRC 1994, chapter 10, p. 189)

Whereas uncertainty is generally modeled using probability distributions, variability is modeled using frequency distributions. Although frequency distributions share the same formal properties as probability distributions, this minor distinction is important because frequency distributions may be known precisely and the distribution reflects true differences between instances. There may be uncertainty about the precise distribution of variability, however, even when its moments and parameters (mean, standard deviation, and skew) are known, in which case the frequency distribution can be represented by a probability distribution.

Unlike uncertainty, variability cannot be reduced through additional research but may be handled in a model through disaggregation. If a population is heterogeneous, for example, then the population may be broken up into smaller subpopulations that are more homogenous. Disaggregation is important when some subpopulations are high risk or sensitive.

Parameter Uncertainty

Whereas variability is ontological in nature in that it concerns the properties of objects, parameter uncertainty is epistemological because it concerns the human ability to know. Other terms for *epistemic uncertainty* include lack of knowledge, subjective, and informative uncertainty, although this typology distinguishes between parameter and model uncertainty as subsets of epistemic uncertainty.

Like variability, parameter uncertainty applies only to empirical quantities. As noted previously, empirical quantities have two important characteristics: they are measurable in principle (unlike mathematical constants), and they are system components (unlike model performance parameters such as spatial resolution or the number of simulation iterations, which are considered sources of model uncertainty).

Quantitative parameter uncertainty is generally modeled using probability distributions; however, if multiple datasets provide competing values, then frequency tables or scenario analyses may be more appropriate. Combining judgments from experts results in subjective probability distributions, but these can be modeled identically to probability distributions based on statistical variance of observations.

Unlike variability, parameter uncertainty may theoretically be reduced through additional investigation. The many sources of parameter uncertainty are described in detail by Morgan and Henrion (1990), Frey (1992), NRC (1994), and van Asselt (1999).

Basically, parameter uncertainty is due to errors or difficulties in either measuring data or applying data from the measured source to the modeled variable. The four subtypes of parameter uncertainty are described as follows:

- *Measurement errors:* These errors arise from the measurement of quantities.
  - *Random error and statistical variation:* This most-often-modeled type of parameter uncertainty is due to imperfections in measurement techniques or analytical devices. Statistical distributions of measurements represent this type of uncertainty. It is also referred to as *metrical error* (Rowe 1994). It is distinct from *sampling error* (discussed below). The various techniques for quantifying this uncertainty are described in detail by Morgan and Henrion (1990, chapter 5, pp. 73–101).
  - *Systematic bias:* This difference between the true value of a quantity and the mean of the distribution of measurements is due to poor calibration, errors in use of equipment, or biased techniques. The degree of systematic bias often is unknown and cannot be quantitatively modeled. Morgan and Henrion (1990) consider different forms of extrapolation (discussed below) as systematic bias; we distinguish between errors in measurement of data and errors in the selection and application of data.
- *Unpredictability:* Inherent randomness, or unpredictability, is distinguished from other kinds of parameter uncertainty because it is irreducible, even in principle. The archetypal example is Heisenberg's Uncertainty Principle. This concept often is applied to cases in which precise measurement is possible in principle, but practical limitations have precluded it (Cullen and Frey 1999).
- *Conflicting Data and Lack of Data:* As a natural consequence of the scientific method, data values are often conflicting for a given parameter. Depending on the parameter in question, such uncertainty might be modeled as a probability

table, as a distribution across available values, as distinct scenarios, or by weighting data through expert judgment. Bayesian statistical methods may be used to combine data sources (Goodman 2002). Disagreement also results from differing expert opinions or interpretations of the same data; many methods exist for assessing and aggregating expert judgments into subjective probability distributions that can then be modeled. The elicitation of expert judgments is a difficult, resource-intensive process, but it is also a mature, developed area of research in which a solid literature provides guidance (For details, see Morgan and Henrion 1990; Cooke 1991; Evans et al. 1994a, 1994b; Budnitz et al. 1995; Paté-Cornell 1996; Cooke and Goossens 2000; and Walker et al. 2001, 2002.) Also worth noting is the pilot expert elicitation performed for EPA by Industrial Economics, Inc. (IEc 2004), and used for sensitivity analysis in two EPA RIAs on final rules: the nonroad diesel rule (U.S. EPA 2004d) and the Clean Air Interstate rule (CAIR; U.S. EPA 2005a). Similar to uncertainties associated with conflicting data, there may be what Frey (1992) calls a "lack of empirical basis," or an absence of data about a system that has yet to be built or tested. In such cases, expert judgment can be used, or extrapolations can be made from other data (possibly leading to other errors, discussed later). These types of uncertainties are similar to model uncertainties in that they involve a choice by the modeler among alternatives; we distinguish between alternative values for a model parameter and alternative choices of relationships between parameters.

- *Extrapolation errors:* Uncertainties result from applying measurements from one population or situation to a different population or situation. Morgan and Henrion (1990) consider extrapolation a form of systematic bias, but three types of extrapolation errors are broken out separately by Finkel (1990) and NRC (1994). We define five types of extrapolation error:
    - *Random sampling error: Sampling error* is sometimes referred to as *random error*, which causes some confusing overlaps in nomenclature with the first form of measurement error, discussed previously. However, an important distinction must be made between uncertainties that stem from imprecise measurement techniques and those that result from inferences about a population drawn from a limited number of observations of that population (Haimes 2004). Sampling a population results in uncertainty that depends on the extent to which the sample is representative of the entire population. Standard methods exist to quantify this type of uncertainty.

- *Temporal prediction errors:* Inherent uncertainty arises when predicting the future values of input parameters. As Morgan and Henrion note, "In extrapolating from the past to future forecasts, there is not only uncertainty from the imperfect fit to past data, but also uncertainty about how much the future will be like the past" (1990, p. 59). Certain empirical quantities may be assumed to be invariable over time, but others may not—especially quantities that represent human behavior, and especially when time horizons are long. One example would be gasoline prices over the next 50 years as a model input; one must choose how to predict changes to prices over time (this choice is a source of model uncertainty), but regardless of selected method, any set of prices contains implicit errors because the future is inherently unknowable. Temporal prediction error is often modeled through scenario analysis, because supporting data on predicted quantities may have been created with the use of different assumptions that result in alternate predictions that are not defined probabilistically (e.g., low, medium, and high estimates).
- *Surrogate data:* Extrapolation uncertainties can arise due to the use of proxy, generic, or standardized values instead of system- or context-specific values (e.g., using standard emission factors instead of measurements of a specific process) (NRC 1994). These errors are inherently difficult to quantify, and may not be able to be modeled probabilistically.
- *Nonrepresentativeness:* Extrapolation uncertainties can arise when the sampled population is not representative of the modeled population (NRC 1994). The two forms are when one population is a subset of the other (e.g., estimating emissions from all plants based on data only from high emitters) and when the populations are completely distinct (e.g., animal studies extrapolated to estimate human outcomes). These errors are inherently difficult to quantify and may not be able to be modeled probabilistically.
- *Misclassification:* Extrapolation uncertainties can arise due to the erroneous assignment of exposures because of inaccurate data from epidemiological studies; often, imprecise measurement of an indicator parameter, such as a biomarker, affects the resulting dose–response model (NRC 1994). These errors are inherently difficult to quantify and may not be able to be modeled probabilistically.

Morgan and Henrion (1990) include *approximation* as a source of parameter uncertainty, while Frey (1992) classifies *correlations and dependencies* as sources of parameter uncertainty. Both are considered below as sources of model uncertainty.

Model Uncertainty

Like parameter uncertainty, model uncertainty is considered epistemic, subjective, or informative uncertainty (van Asselt 1999). It is due to a lack of knowledge about system behavior or to choices that determine model behavior. Whereas parameter uncertainty results from the practical limitations of data, model uncertainty results from limitations in the ability to create causal or predictive models of real-world systems on the basis of the data. Errors are caused by methodological problems approximating a system and result from ignorance about actual system behavior.

The line between parameter uncertainty and model uncertainty is fuzzy, however, because choices about model form have implications for parameters and many parameters may be considered outputs of complex systems. For example, one might estimate annual rainfall by using a complicated model of global weather patterns or by simply using historic data as a measurement. This choice represents model uncertainty, but if annual rainfall is an input parameter to a different system—say, of pollutant deposition—then uncertainty around the input parameter is a form of parameter uncertainty. Whether a particular source is parameter uncertainty or model uncertainty is not of incredible consequence. How the uncertainty is treated in the model, the overall impact of that uncertainty on results, and whether the uncertainty can be reduced are of consequence.

Model uncertainty is analyzed far less than parameter uncertainty, but some argue that these unaccounted-for uncertainties may have larger impacts than the variability and uncertainty around individual parameters (Casman et al. 1999, Linkov and Burmistrov 2003, Koop and Tole 2004). In a comparison of the same scenario using six different models, Linkov et al. find that the differences between models were as large as seven orders of magnitude and that for one model, parameter uncertainty "is within one order of magnitude, much narrower than model uncertainty, which can be characterized by a spread of several orders of magnitude" (Linkov and Burmistrov 2003, p. 1308). In an econometric time-series model examining the effect of air pollutants on mortality, Koop and Tole (2004) find that while point estimates showed a positive relationship, estimates incorporating large model uncertainties could not exclude the possibility that there was no link; the range of possible values crossed zero. A comparative analysis by Chen and Chang (2002) on the structure of costs for the construction of wastewater treatment plants shows significant differences in estimates between the models used, which could alter management strategies.

Model uncertainty usually is not modeled probabilistically but is analyzed through scenario analysis in which alternative assumptions are used to give a range of outputs. For example, a predictive air transport model might be used in the primary analysis, and a different air transport model based on different assumptions about weather patterns might be used as comparison in the sensitivity analysis. For a more thorough validation of model structure, optimization methods can be used to explore uncertainty (Reynolds and Ford 1999).

Morgan and Henrion (1990), Finkel (1990), Frey (1992), and NRC (1994) present typologies of model uncertainty. From these, we list six sources of model uncertainty:

- *Structural choices:* Structural uncertainties are due to methodological choices of how to model a system when alternate assumptions about relationships between variables may be available (Beck 1986). One example is the decision whether to model a dose–response relationship with or without a threshold when alternate studies suggest opposing interpretations.
- *Simplification:* All models are simplified approximations of real-world systems, but in some cases, subsystems or relationships are approximated with simpler versions of more detailed approximations. A relationship known to be nonlinear may be assumed to be linear, for example. Similarly, simple models may represent systems for which more detailed relationships are unknown or unquantifiable. Errors due to aggregation are a particular type of simplification uncertainty. A model may use a single aggregated variable when this quantity should be modeled as disaggregated components. Similarly, separate but related causes and effects may be mistakenly aggregated (e.g., exposure to a mixture of pollutants). The analyst may not have enough knowledge about the system to consider disaggregation or may have aggregated to simplify the model. (Problems with aggregation are discussed in detail in NRC 1994, chapter 11.)
- *Incompleteness:* Some models contain uncertainties because variables, relationships, or endpoints are excluded—usually for lack of knowledge. Example exclusions include missed pathways for exposure and a health endpoint not known to be related to exposure at the time of modeling. Cullen

and Frey refer to *scenario uncertainty* as the applicability of the set of model assumptions to the policy question at hand: "to the extent that the scenario fails to consider all factors affecting the key output variable …, uncertainty will be introduced" (1999, p. 30). In the parlance of the typology presented here, if a model does not adequately consider all pathways critical to an output measure, then it has uncertainties due to incompleteness.

- *Choice of probability distributions:* When modeling probability or frequency distributions, assumptions regarding the shape of the modeled distribution may have serious implications. Continuous distributions may be modeled as discrete, limited observations may be fit to one type of distribution over another, and arbitrary decisions about distribution type might be made. These uncertainties can be regarded as structural and/or simplification uncertainties, but we believe that they are an important type of uncertainty to discuss separately. Jones (2000) presents a simple yet convincing example of the impact of choice of probability distribution. Hamed (2000) provides a more detailed example in which normal, lognormal, and uniform distributions are compared.
- *Correlations and dependencies:* When input variables are modeled as independent when one is in fact dependent on another or when there is correlation between their values, uncertainty is introduced. Complete models include full specification of such dependencies, but this relationship may not be explicit in simplified models, and a dependent variable may be erroneously treated as an independent, exogenous input. Frey (1992) and Cullen and Frey (1999) provide approaches to adequately accounting for dependencies.
- *System resolution:* Model system parameters affect the predictive ability of numerical models. System parameters include the spatial and temporal grid size as well as the number of simulations. Particular attention must be made to ensure that these parameters are selected at an appropriate resolution, and uncertainties may be quantified by running multiple scenarios for ranges of possible values. Morgan and Henrion (1990) consider issues of resolution—which they call *approximation*—as a form of parameter uncertainty.

Frey (1992) considers extrapolation a form of model uncertainty, but it is considered in our typology as a form of parameter uncertainty because this problem pertains to the specific errors associated with estimating an empirical quantity that is, in principle at least, measurable. Likewise, although NRC (1994) considers the use of surrogate variables a form of model uncertainty, we consider it to be parameter uncertainty for the same reason. Although the use of extrapolation (e.g., from a high-dose animal study to a low-dose human estimation) and surrogate data (e.g., generic values as opposed to plant- or process-specific data) reflect choices over other options, these choices are made only when measured values of the empirical quantity in question are not available, for whatever reason.

Cullen and Frey (1999) and others list validation and model boundaries as sources of model uncertainty, but we regard them as limitations of interpretation. Validation is a concern for all models and can be thought of as the verifiable ability of a model to predict reality on the basis of application to existing data. Because not all models can be so verified, one is uncertain about how strictly or seriously to interpret model results. Similarly, models that are accurate in some parameter space may be inaccurate in others. Models may have boundaries of acceptable functionality, and uncertainties may arise due to abnormal conditions.

Decision Uncertainty

Finkel (1990) makes an important distinction between the aforementioned types of uncertainty, and "decision uncertainty," which enters quantitative policy analysis after the estimate of risk has been generated. He states that "this type of uncertainty arises whenever there is ambiguity or controversy about how to quantify or compare social objectives" (Finkel 1990, p. 16). Whereas variability, parameter uncertainty, and model uncertainty are issues for risk assessors, decision uncertainties are concerns chiefly for risk managers.

Decision uncertainty is related to what Morgan and Henrion (1990) and others refer to as *value uncertainty*. Most descriptions of incorporating uncertainty into analysis consider only models of physical systems (e.g., microbial risk assessments) and therefore exclude discussions of uncertainty pertaining to the decisions about valuing social objectives. In RIAs that estimate the economic costs and benefits of policy changes, however, decision uncertainties are incredibly important because they go to the root of how these social objectives are determined.

Some decision uncertainties may be modeled probabilistically, especially if the uncertainty represents the diversity of preferences within a population. We list five types of decision uncertainty:

- *Choices of risk measure and summary statistics:* Some uncertainty is associated with the simple decision of which measure of risk to use. Risks may be reported out by different endpoints (e.g., number of illnesses, number of deaths), may differ in examining each endpoint (e.g., life years lost vs. number of deaths), and may be reported at different levels (e.g., individual or total population). Similarly, summary statistics have a limited ability to convey probabilistic information, especially if distributions are abnormal. For example, simply reporting out the mean and variance of a distribution might be misrepresentative if the distribution is skewed. These uncertainties are not modeled but affect how and which results are interpreted.
- *Choice of discount rate:* The rate at which future costs are discounted to compare with present costs is an important characteristic of all RIAs with significant time horizons. Discount rate varies across individuals and across companies, all of which have different investment opportunities. When costs are not monetary but include adverse human health risks, the choice of discount rate is even more difficult. When the time horizon is particularly long and there are not only time but also intergenerational trade-offs, the choice becomes more important and more difficult still because the population of affected individuals differs in the two periods. Discount rates should be modeled not as probabilities but as parametric values that may be changed between simulations to compare the impact of different values. OMB gives detailed guidance on discount rates to be used in RIAs; in line with other discussions of modeling discount rates, OMB suggests the use of scenario analysis, not a probabilistic distribution of possible discount rates (2003, p. 31–37). Additional information on treatment of discount rates can be found in *Guidelines for Preparing Economic Analyses* (U.S. EPA 2000a).
- *Decisions about risk tolerance:* Sometimes a decisionmaker must decide on a degree of risk aversion and set a level of risk under which adverse conditions are tolerated. For example, some very low level of a radioactive compound may not be associated with high risks and is therefore deemed "acceptable." Defining what is acceptable, however, may be controversial. This type of uncertainty can be addressed, where appropriate, with scenario analyses.
- *Utility functions:* As Finkel frames the problem, "Even after resolving questions about … the 'acceptability' of risk, controversy remains over how to translate the measure of risk into a measure of social cost. The notion that a single defined parameter can be used to translate risk into cost carries with it a tacit assumption

of debatable validity—that society values *X* lives exactly *X* times as much as one life saved" (1990, p. 19). In addition to the problems with this linearity assumption across individuals, individuals also may have nonlinear utility functions over risk. Similar concerns with health valuation arise from the choice of how to value endpoints. One example is the decision to value a small risk of mortality to individuals within a population as a value of statistical life (VSL; i.e., to assume that a 1-in-5000 increase in the risk of mortality to each of 5000 individuals in a population is the same as a statistical mortality). Another example is the decision to value statistical lives or statistical life years.

- *Distributional considerations:* Strict utilitarianism suggests that one should try to maximize the sum of all individual utilities across a population, whereas the theory of social welfare suggests that social utility may depend on the equitable distribution—not merely the magnitude—of utilities across individuals. In addition, benefits and costs of a regulation may be distributed unevenly over time or space. The extent to which these distributional effects are considered is a form of uncertainty. Most RIAs include separate analyses of distributional effects, as required by OMB. The decisionmaker considers these effects in addition to issues of economic efficiency (OMB 2003). Considerations of equity can be built into some models by the inclusion of information on population subgroups (e.g., high risk, race, income, education, geography) but might need to be addressed in some analyses separately.

Many discussions of uncertainty treat value uncertainty (and implicitly any uncertainty about the valuation of health outcomes) as a separate type of uncertainty or classify it as a form of decision uncertainty. The VSL is often discussed as the prototypical example, but as we noted previously, the VSL used in an RIA includes uncertainties that stem from diversity among people, the limited ability to directly measure a nonmarket good such as a change to risk of premature mortality, and the choice of utility function. These uncertainties are already captured in our typology described above; they merely coexist in the same total uncertainty. The VSL may be a politically controversial idea when interpreted out of context and will remain an important component of overall uncertainty in health benefit estimates for some time, but it should not be a separate or special source of uncertainty.

Linguistic Uncertainty

The uncertainties associated with language have long been noted in discussions of uncertainty in policy decisionmaking (Morgan and Henrion 1990, Rowe 1994, Regan et al. 2002, Haimes 2004). They are implicitly qualitative and therefore are excluded from most quantitative analyses, but they are important in RIAs.

Linguistic uncertainty is excluded from the typology of uncertainty presented above because it pervades the process of building models, analyzing results, and communicating risks. It affects all four sources of uncertainty previously listed, although its impacts usually can be minimized rather easily. Furthermore, linguistic uncertainty is distinct from variability and parameter and model uncertainty because it concerns neither natural heterogeneity nor lack of knowledge but rather the communication of information. Linguistic uncertainty is different from decision uncertainty because it does not concern choices of individual and social values, even though values certainly play a role in the interpretation of verbal expressions of probability.

Morgan and Henrion (1990) treat linguistic uncertainty as a type of parameter uncertainty, because words may be the best measurement of certain empirical quantities. The general solution to the problem is simply to avoid it where possible: "Whereas many sources of uncertainty, including lack of information and computational limitations, are often expensive or impossible to eliminate, uncertainty due to linguistic imprecision is usually relatively easy to remove with a bit of clear thinking" (Morgan and Henrion 1990, p. 63). Citing Howard 1988, Haimes (2004) presents a hypothetical yet useful "clarity test" for the linguistic precision of a question: if an all-knowing clairvoyant could respond with a quantity value to some question, then that question is precisely phrased; otherwise, it needs to be rephrased.

Significant research has been conducted on how probabilistic information is conveyed and interpreted through words and phrases such as "likely" and "very likely" (e.g., Tversky and Kahneman 1974, Beyth-Marom 1982, Fischhoff et al. 1993, Olson and Budescu 1997, Renooij and Witteman 1999, Moxey and Sanford 2000). Nonetheless, interpretation entails too much uncertainty to recommend the assignment of numerical probabilities to model parameters on the basis of such verbal phrasing.

Regan et al. (2002) present a detailed and useful typology of linguistic uncertainty. Analysts should not feel the need to examine all instances of linguistic uncertainty in such detailed terms, but a general understanding may be useful to avoiding pitfalls in conveying results. Regan et al. list five sources of linguistic uncertainty. *Vagueness* describes issues of borderline cases, such as separating what is "endangered" from what is not. *Underspecificity* is similar to vagueness but is not a borderline issue; it occurs when there is too much generality, such as the difference between a general description of location and Global Positioning System (GPS) coordinates. *Context dependence* is the component of imprecision based on lack of context, such as the difference between "large" and "large for a particular species" (although this example also includes vagueness). *Ambiguity* results from terms that have multiple definitions, such as "tree cover," which may or may not include gaps within tree crowns. *Indeterminacy of theoretical terms* is essentially the potential for ambiguity, because terms defined as unambiguous in the present may not be adequately unambiguous in the future.

In RIAs, linguistic uncertainty is most likely to occur in the communication of risks and uncertainties. In particular, labels such as "high" and "low" should be specified as precisely as possible (e.g., "high" corresponds to an order of magnitude above the mean). Definitions of terms such as "acceptable" should be spelled out quantitatively. The impacts of linguistic uncertainty to the communication of risk and uncertainty are discussed in more detail in the Chapter 3.

Probabilistic Modeling Issues

The typology presented in this chapter should give some guidance for the identification of sources of uncertainties in completed, underway, and future impact assessments. Merely identifying sources of uncertainty is only the beginning, however. A host of issues must be considered when moving toward a probabilistic approach to modeling. This section discusses some of the most important ones.

Sampling-Based Simulation Methods

The most common probabilistic modeling framework is the sampling-based approach, in which a deterministic model is run repeatedly, drawing randomly from specified probability distributions for each uncertain model input in each model run. The set of simulations propagates defined uncertainties from model inputs through to model outputs, where they can be analyzed statistically, as if they were an observed data set. The most common sampling-based method is the use of random or MC sampling. (For more information about MC modeling, including how to select the number of simulation iterations, see Rubinstein 1981, Fishman 1996, U.S. EPA 1997b, Cullen and Frey 1999, and Helton and Davis 2000.)

As an alternative (or modification) to the purely random samples of MC simulation, stratified sampling methods can be used in which input distributions are divided into strata that are then subsampled (Cullen and Small 2004). The most common form is Latin Hypercube Sampling (LHS), in which input probability distributions are broken into ranges of equal probability, then a single sample is taken from each range: the mean, the median, or a random number in that range (McKay et al. 1979, Helton and Davis 2000). The order of the samples is random and independent across uncertain inputs (unless correlations between uncertain variables are explicitly defined). Fewer simulations are required for LHS than for MC sampling, and LHS usually (but not always) is preferred for this reason (Cullen and Frey 1999, Morgan and Henrion 1990).

Helton and Davis (2002) perform several uncertainty and sensitivity analyses on a sequence of models and find LHS to produce more stable results than MC sampling. The literature for other stratified sampling methods—including Hammersley sequence sampling (Kalagnanam and Diwekar 1997), Halton and Sobol′ sampling (Halton 1994, Gentle 1998), orthogonal arrays (Saltelli et al. 2000), and other approaches typified as quasi-Monte Carlo (QMC) (Niederreiter 1992, Morokoff and Caflisch 1994)—is growing, especially in computer science.

Morgan and Henrion (1990) and Cullen and Frey (1999) also discuss "importance sampling" or targeted sampling, in which a portion of the parameter space (such as the tail of a distribution) is sampled more frequently because it may be more relevant to risk management. Two-dimensional sampling-based simulations—or nested simulations in which uncertainty and variability are separated into different dimensions—are discussed in the following section.

Similarly, Markov chain Monte Carlo (MCMC) methods (also called random-walk Monte Carlo methods) are techniques used to sample from distributions with high-dimensional integrals that are difficult or impossible to solve analytically (Gilks et al. 1996, Brooks 1998). MCMC simulation is useful when there are correlations between input parameters in risk models (Nayak and Kundu 2001, 2003; Ades and Lu 2003) and to perform Bayesian inference in network models similar to Bayesian belief networks (BBNs; described later in *Additional Modeling Approaches*). To perform MCMC simulations, the specialized software required is distinct from those used in other Monte Carlo simulation approaches.

Separating Uncertainty and Variability

Probabilistic simulation models may be developed by using a single dimension to capture uncertainty and variability; however, uncertainty reported in final outputs will subsequently include both forms, thus limiting interpretation. One solution to this problem is to perform a scenario analysis in which uncertainty is allowed to vary and variability is eliminated (nominal point values are used), and then vice versa, thus decomposing the overall uncertainty in the final results (Cullen and Small, 2004).

Frey (1992) introduces the concept of using a two-dimensional MC approach to separate uncertainty from variability when modeling an environmental system (expanded in Frey 1993 and Cullen and Frey 1999). Two-dimensional MC involves the simulation of simulations; for each of *M* simulations in dimension *X*, *N* simulations are performed in dimension *Y*. This approach was developed to address specific situations in which uncertainties might be independent from person to person but conditional on variable quantities. In such situations—especially when modeling health effects and when interested in characteristics of the population or of individuals that may be related to high exposure—it may be important to separate uncertainty from variability. When uncertainty and variability are modeled in the same dimension, as in traditional MC approaches, the variance in an output parameter (e.g., exposure) contains both components of probability and frequency distributions and reflects an individual selected at random from the total population. However, one cannot rank order individuals within the population or estimate the exposure faced by an individual in a specified part of the distribution.

Yet the distinction between variability and uncertainty may be "overdrawn" in some situations (Morgan 1998). Two-dimensional nested simulations are much more complex than one-dimensional simulations, and they square the computing resources required (assuming equal depth of dimensions) per suite of model runs. Winkler argues that "uncertainty is uncertainty" and that there is little or no foundational basis to distinguish between types of uncertainty, although he notes that doing so may be useful for practical purposes of "sensitivity analysis, information gathering, and model structuring" (1996, p. 2).

Furthermore, disentangling uncertainty and variability in a particular overall measure of uncertainty is not always a trivial matter; it may not be clear where to draw the line—particularly if variability is estimated subjectively, not empirically (Gray et al.

1998, Anderson and Hattis 1999). In short, two-dimensional simulations are not necessarily the gold standard and cannot be recommended for all cases.

If the primary need of incorporating uncertainty into a RIA is to place final net benefits numbers in a probabilistic framework, then two-dimensional simulations probably are not necessary. Such an approach may be powerful, however, if one goal of the analysis is to identify which uncertainties might be reduced through further research to reduce overall uncertainty, or if extensive or exhaustive uncertainty analyses are to be performed (Gurian et al. 2000).

Defining Uncertain Model Inputs

Characterizing uncertainty and variability in model input parameters as well as determining which model inputs should be defined probabilistically require the subjective judgment of the analyst: what is the proper approach to take in a given situation? Some situations call for all inputs to be defined probabilistically, while the trade-off of completeness and complexity may lead to a decision to limit the number of probabilistic inputs.

Law and Kelton give an overview of probability distributions used to define model inputs, including various kinds of distributions: continuous (uniform, exponential, gamma, Weibull, normal, lognormal, beta, Pearson, and triangular), discrete (Bernoulli, discrete uniform, binomial, geometric, negative binomial, and Poisson), and empirical (1991, chapter 6, pp. 329–352). Typical applications for different distributions are noted. Cullen and Frey (1999) also present an overview of types of distributions, with examples of probabilistic models that use each type.

The distribution for an uncertain model input may be specified by using either statistical techniques to fit data to a distribution or expert judgment. These two approaches also can be used in combination; for example, an expert could use some distribution estimated by fitting data to estimate a different distribution.

There are two primary statistical approaches for estimating the parameters that specify a distribution: the method of moments (or method of matching moments) and the maximum likelihood method. The former is easier to implement and usually is preferred for small datasets, whereas the latter provides a better fit to large datasets. (Ample literature on how to identify and parameterize distributions includes Morgan and Henrion 1990, Law and Kelton 1991, Burmaster and Thompson 1998, and Cullen and Frey 1999.)

Of the many goodness-of-fit tests to examine how well a specified distribution reflects the underlying data, most can be characterized either as statistical tests or as graphical probability plots. The latter are more subjective and generally involve visual inspection. Statistical approaches provide a numerical assessment of the goodness of fit. The most common statistical techniques for measuring goodness of fit are the chi-squared, Kolmogorov, Cramer–Smirnov–Von Mises, Shapiro–Wilk, and Anderson–Darling tests. (For more information about goodness-of-fit techniques, see Cullen and Frey 1999, Law and Kelton 1991, and D'Agostino and Stephens 1986.)

Expert Judgment

In the classical frequentist view, the *probability* of an occurrence is defined as the frequency with which an event occurs in a long sequence of such occurrences or, more specifically, the value to which the long-run frequency converges (Morgan and Henrion 1990). In the subjectivist or Bayesian view, *probability* is an individual's degree of belief that an event will occur, based on all information known to that person. Unlike the frequentist, the subjectivist does not hold that estimated probability converges toward one true probability; as new information becomes available, probability might change.

For modeling purposes, the distinction between this interpretation of probability is important primarily in defining variables for which no single reliable data set exists. For modeling the flip of a coin, for example, the frequentist's sequence of experimental coin flips would presumably approach 0.50, which is the likely probability a Bayesian would choose. A frequentist who never flipped a coin could not judge the likely probability, whereas a Bayesian could still assign a likelihood.

Let us take a more salient example. To model the dose–response function to estimate mortality from exposure to a particular pathogen, it is not practically possible to hold controlled human experiments to derive the true value (which, of course, might depend on a various crucial heterogeneities between human subjects). The Bayesian approach to this problem is to base a value on other data sources or knowledge (e.g., animal tests, limited natural experiments, epidemiological studies, or knowledge about a different pathogen), each of which might suggest a different value. Parameters in other models or relationships between parameters may not be measurable at all.

The Bayesian framework of probability has multiple implications for modeling. As noted above, it allows the assignment of probabilities even when experiments cannot be conducted. It also points toward the use of expert judgment in decisionmaking. Instead of the modeler assigning probabilities based on a subjective interpretation of the information, probabilities are elicited from experts instead. In expert elicitations, various techniques are used to have experts review information and provide quantitative estimates. An iterative technique may achieve consensus on a value agreed upon by all experts as reasonable, or each expert may provide an individual estimate that may be combined or used separately by the modeler.

Heuristics that people use to make judgments (e.g., anchoring, availability of relevant information, and representativeness of experiences) can lead to biases in estimates (Kahneman and Tversky 1982, Morgan and Henrion 1990). As a result of the complexities of making judgments, various elicitation protocols have been developed to reduce heuristic and other biases. A full discussion of these approaches is well beyond the scope of this report, however. (For more detailed guidance, see Morgan and Henrion 1990; Cooke 1991; Budnitz et al. 1995; and Cooke and Goossens 2000. For detailed examples of expert elicitations, see Evans et al. 1994a, 1994b; Morgan and Dowlatabadi 1996; Paté-Cornell 1996, 2002; Walker et al. 2001, 2002; and van der Fels-Klerx et al. 2005.)

As discussed earlier, EPA recently completed a pilot expert elicitation (IEc 2004) focused on the concentration–response (C–R) function for mortality due to particulate matter of less than 2.5 microns in diameter ($PM_{2.5}$) that has subsequently been used in sensitivity analyses in two RIAs: the nonroad diesel rule (U.S. EPA 2004d) and CAIR (U.S. EPA 2005a). This expert elicitation underwent peer review (RTI 2004), and a symposium was held in April 2005 to discuss the results of the pilot study and the revised draft protocol for full implementation of the elicitation (IEc 2005, U.S. EPA 2005c). EPA may use lessons from this pilot elicitation to obtain expert judgments for other uncertain parameters.

Correlations and Dependencies

In general, probabilistic models should be structured to avoid dependencies among model inputs, although this is not always possible. Unless otherwise specified, model inputs are implicitly independent when defined as probabilistic distributions in MC simulations. The two primary approaches to incorporating dependencies in model structure are creating a more detailed model to explicitly model the dependence and using multivariate distributions or "restricted pairing" techniques when drawing samples from distributions (Cullen and Frey 1999).

Cullen and Frey (1999) list various approaches for simulating the correlations between inputs: specifying multivariate distributions for correlated inputs (e.g., defining two variables with a bivariate normal distribution); using a simple correlation coefficient to induce statistical covariation among model inputs with restricted pairing methods; and simulating Kendall's tau rank correlation, Spearman's rho rank correlation, and Pearson product moment correlation. Alternatively, variables that covary may be combined into a single variable, thus embedding the dependency, or the population may be stratified into homogenous subgroups to decrease the effect of correlations.

One important consideration is that correlations may exist between models used in analysis, and accounting for these correlations may reduce uncertainties in model results. One example is a variable, defined probabilistically, that is an input to both the model of cost estimates and the model of benefits estimates. In such a case, it would be misrepresentative to report out the difference between the distributions of benefits and costs as a distribution of net benefits, because high values of costs might be associated with high values of benefits and low values of costs associated with low values of benefits, thereby resulting in a situation with less uncertainty in net benefits than with benefits or costs separately.

Additional Modeling Approaches

In addition to sampling-based simulation approaches, other methods have been explored in recent years to perform environmental and ecological assessments and to estimate the uncertainty of risk estimates (Jörgensen 1999, Bourgeron et al. 2001, Cullen and Small 2004). Fuzzy set theory, in which set membership is defined probabilistically rather than as a Boolean function (Zadeh 1973, Klir and Folger 1988), is useful when classification conditions are vague and ambiguous (e.g., "endangered," "low") or when membership in well-defined sets is measured with vagueness (e.g., "possible," "plausible") (Cullen and Small 2004). Numerous examples have been developed for environmental decisionmaking, presented by Silvert (1997), Enea and Salemi (2001), and Fisher (2003).

Cellular automata, in which cell states are governed by the states of neighboring cells (Wolfram 1986), have been used extensively in environmental and ecological situations (Hogeweg 1988, Gronewold and Sonnenschein 1998, Molofsky and Bever 2004). Additional approaches include the use of evolutionary and genetic algorithms

(Caldarelli et al. 1998, Muttil and Lee 2005), artificial neural networks (Lek and Guegan 1999, Maier and Dandy 2000), rule-based expert systems (Wright et al. 1993, Zhu and Simpson 1996), and Bayesian approaches (Malakoff 1999, Ellison 1996).

BBNs, also referred to as *directed graphs* or *causative networks,* are graphical acyclic influence diagrams in which nodes or events are connected through conditional probabilities (Pearl and Russell 2000, Reckhow 2003a). When new evidence arises regarding a node in the network, a Bayes theorem is used to update the probabilities of connected nodes, thus propagating new information forward and backward through the network. BBNs are useful when there are pervasive data gaps or large structural uncertainties and are particularly conducive to the inclusion of expert judgment. They have been used in numerous environmental models (Stiber et al. 1999, Marcot et al. 2001, Borsuk et al. 2004, Bromley et al. 2005). BBNs may contain only discrete variables, however, which means that one must discretize continuous variables in such models. Alternately, one may use MCMC methods (described earlier in the *Sampling-Based Simulation Methods* section) to perform the necessary Bayesian updating on continuous variables (Parsons et al. 2005).

Sensitivity and Uncertainty Analysis

For RIAs to be considered robust and reliable, analysts should examine the range and distribution of estimated results given alternative input parameter values as tested through sensitivity and uncertainty analysis (Merrifield 1997). As with the word *uncertainty* itself, the nomenclature of uncertainty analysis and sensitivity analysis is inconsistent and overlapping. Sometimes *sensitivity analysis* refers to a limited form of examining how single-value changes to model inputs affect outputs, whereas at other times it means the same as *uncertainty analysis*—broadly speaking, the study of how uncertainties in model inputs relate to model outputs. EPA RIAs, such as those for the nonroad diesel rule (U.S. EPA 2004d) and CAIR (U.S. EPA 2005a) refer to uncertainty analysis as quantified probabilities of inputs and outputs through MC modeling and sensitivity analysis as a one-at-a-time varying of inputs in a point estimate framework. The nomenclature is not as important as the specifics of the different types of analyses and their purposes, but it is worth keeping in mind that other definitions exist.

For clarity and consistency, we use the following terminology in this section. *Uncertainty analysis* is the quantitative assessment of the weighted uncertainties of model predictions. *Sensitivity analysis* is "the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" (Saltelli 2002, p. 579). This broad definition of *sensitivity analysis* includes local (i.e., single-parameter) and global (i.e., multiple-parameter) methods as well as numerical, statistical, and graphical approaches. *Importance assessment* is the determination of which input uncertainties weight the most on model output. Many sensitivity analysis methods calculate importance measures for input variables; importance information may be a simple ordinal ranking or a quantitative contribution in fractional or percentage terms. *Value of information (VOI) analysis* is the determination of which reductions in input uncertainties would be most valuable to reduce, as measured in dollars by corresponding reductions of net benefits.

The discussion of uncertainty and sensitivity analysis in this document is not as a comprehensive catalog of methods. Rather, it is intended to give analysts an overview of the types of analysis that might be useful to incorporate into future RIAs. The specific characteristics of any particular modeling environment should guide the choice of method. (For a detailed overview of uncertainty analysis methods, see Rai et al. 1996, Rai and Krewski 1998, and Saltelli et al. 2000, which provide in-depth discussions of many of the approaches mentioned here. For shorter, yet detailed evaluations of sensitivity analysis methods, see Kann and Weyant 2000, Frey and Patil 2002, Helton and Davis 2002, and Saltelli 2002.)

In the following sections, we discuss numerous types of analysis, from uncertainty analysis and graphical approaches to response surface methods and VOI. The analyses described are all appropriate for sampling-based simulation models, but some may be applied to other modeling approaches.

Uncertainty Analysis

Uncertainty analysis for RIA requires and is the primary motivation for probabilistic modeling approaches. Although sensitivity analysis methods can be used on deterministic models to get quantitative information about the range of possible outputs for a given model, they cannot be used to address the relative likelihood of any particular point estimate result. Only when the deterministic model is dropped into a repeat-run sampling-based simulation (in which each uncertain variable is drawn randomly during each model run) can the distribution of expected model results be estimated. When discussing the results of sampling-based analysis, the extent to which the uncertainty analysis quantifies model inputs is important.

The level of uncertainty analysis in the nonroad diesel RIA (U.S. EPA 2004d) and the CAIR RIA (U.S. EPA 2005a) are limited to only two types of parameters (C–R and valuation per case) and should be referred to as partial uncertainty analysis. To be considered full uncertainty analysis, it is not necessary that every single model input be defined probabilistically, but the most important variables for characterizing output uncertainty should be.

Graphical Methods

Whereas most methods of sensitivity and uncertainty analysis are numerical, graphical representations of uncertainty are often a useful first step and can be very useful for screening important variables from those that have little impact on output uncertainty. Graphical methods can be used not only to convey information and results of analyses, as described in Chapter 3, but also to examine more closely the relationships between input and output variables.

One of the most intuitive graphical approaches is to produce scatter plots of the matched pairs of input samples and resulting output point estimates from sampling-based simulations. Scatter plots are useful as an initial sensitivity analysis, as they can visualize dependencies between an input and an output, including nonlinear relationships.

Helton and Davis (2002) offer one example of the use of scatter plots, comparing various methods of sensitivity analysis on sampling-based models (linear regression, rank transforms, and nonrandom searches for input–output patterns). They present scatter plots of input–output relationships for different types of models (linear, monotonic, nonmonotonic) and different numbers of samples (for random MC sampling and LHS).

In nonmonotonic models, scatter plots show clear but complex dependencies between inputs and outputs even while regression-based sensitivity analyses (e.g., correlation coefficients, rank correlation coefficients) show no discernible relationships. Scatter plots are most commonly produced for a single input parameter, even though multiple scatter plots can be overlaid on the same graph. Alternatively, graphs for multiple variables can be organized in a matrix scatter plot consisting of a grid of "small multiples" (Tufte 1991) of one-to-one scatter plots for all variable combinations (Cooke and van Noortwijk 2000).

In addition to overlaid and matrix scatter plots, Cooke and van Noortwijk (2000) investigate tornado graphs, radar graphs, generalized reachable sets, and cobweb plots. Cobweb plots are investigated in detail for performing uncertainty analyses in the context of high dimensionality and large dependencies. They are particularly useful for examining local sensitivities and conditional relationships between variables. The user is able to visualize complex relationships that may not be apparent in other graphical or numerical representations. (Cobweb plots are explained in more detail in the context of the NO_x reduction case study in Chapter 3.)

Screening Methods

Screening approaches to sensitivity analysis are the most economical methods, intended to quickly and qualitatively isolate the most important factors in uncertainty analysis. Most are designed for use with deterministic models; most probabilistic simulation software incorporates push-button importance assessment using correlation methods discussed later.

Campolongo et al. (2000) present numerous screening approaches, including various one-at-a-time designs for organizing single-variable sensitivity analyses. Most of their approaches are designed for situations in which the analyst has little up-front knowledge about the most important factors in a model. We do not discuss screening analyses in more detail here because we find this situation unlikely with respect to the policy evaluation models used for environmental regulatory analysis. We assume that EPA analysts start with some reasonable knowledge as to the likely set of most important parameters and that these assumptions are tested through quantitative sensitivity analyses.

Local Methods

In local sensitivity analyses, a single parameter or small subset of parameters is varied while all other parameters are held to fixed point values, and the resulting range of output estimates is calculated. Most local methods are used on deterministic models, when probabilistic uncertainty analysis is not possible.

The simplest and most common form of local sensitivity analysis is nominal range sensitivity analysis (NRSA), also referred to as threshold analysis or single-value deterministic sensitivity analysis. In NRSA, each parameter of interest in a deterministic model is set to bounded points (usually the 5th and 95th percentiles) while holding all other parameters to nominal point values (Kann and Weyant 2000, Frey and Patil 2002). The *swing weight* of that input parameter is the change in the output across the range of inputs (Morgan and Henrion 1990). In linear models, the swing weights for different parameters can be ranked for relative impact on overall uncertainty. However, NRSA has limited utility in models that have interaction effects or nonlinear relationships.

Conditional sensitivity analysis (CSA) is an expansion of NRSA for deterministic models in which there are known correlations between inputs or interactions between inputs that result in nonlinear model output as these inputs vary. In CSA, a subset of input parameters are varied not only at their nominal range bounds, but at intervals (or drawn randomly) within the parameter space (Frey et al. 2004). The idea is to cover the entire range of variation in output, which might not occur at the extremes of input bounds if there are nonlinearities. The model response can be plotted and examined for saturation points and thresholds (Frey et al. 2004).

When the output of interest is a probability, an application of NRSA called the difference in log-odds ratios (ΔLOG) can be used to examine sensitivity to model inputs (Stiber et al. 1999). The odds ratio of some event occurring is the probability of that event occurring divided by the probability of that event not occurring. The log of the odds ratio of the output for the best estimate point value of the input is subtracted from the log of the odds ratio of the output for the extreme values (i.e., 5th and 95th percentiles) of the input. The relative magnitudes of ΔLOG values across parameters can be compared to determine which have the largest influences on output probabilities.

In differential sensitivity analysis (DSA), the first partial derivative is calculated for each input in a deterministic model to estimate sensitivities to small deviations from the mean. Because this procedure can be resource intensive and difficult, there are two alternatives for implementation. First, a simplified approximation method involves varying an input parameter over a small range (e.g., ±1–5% from the point value) and calculating the variation in the output, which can then be normalized (Frey et al. 2004).

Alternatively, DSA can be automated. In automated DSA, usually referred to as automated differentiation, a computer program running outside the model computes first- or higher-order partial derivatives as measures of sensitivity. Automated differentiation can be used on large models where the model structure is not fully known, but analysis may be limited to programming languages for which automated differentiation code is already available (Frey and Patil 2002).

Break-even analysis (BEA) is a form of sensitivity analysis useful for deciding between multiple options (Morgan and Henrion 1990). The idea is to find values for input parameters for which a risk manager is indifferent between options and then to examine the range of possible or likely values of input parameters to determine whether they can result in this break-even state. The graphical representation is the break-even line in a two-variable-parameter space, with values on one side of the line implying a preference for Option A and values on the other side implying a preference for Option B. BEA may be a useful way to approach risk-management problems, but there is no straightforward method to apply to find the break-even line, especially for models with multiple sensitive inputs (Frey and Patil 2002).

Probabilistic sensitivity analysis (PSA) is a local sensitivity analysis method for use in probabilistic models, similar to NRSA. In PSA, a single variable (or subset of variables) is allowed to vary within its specified probability distribution while all other variables are held fixed at mean values, and repeat-run sampling-based simulations are performed to produce a weighted distribution of output estimates. PSA is particularly useful to show how the mean estimate of model response depends on inputs (Cullen and Frey 1999, Hamed 1999).

Scenario Analysis

In scenario analysis, the model is run for different sets of parameter assumptions to examine the impact of these assumptions on model output. The primary difference between NRSA and scenario analysis is that the parameters varied in NRSA have uncertainty around a point value, whereas the parameters varied in scenario analysis correspond to a limited number of discrete possibilities (e.g., yes or no, 10 or 20). Scenario analyses can be performed on deterministic models or probabilistic models, in which case each scenario produces a different distribution of output weights.

Scenario analysis is used on two general occasions: for model uncertainty (i.e., when the discrete parameter space represents a set of choices or assumptions about model structure), and for parameter uncertainty (i.e., when the discrete parameter space corresponds to a possible set of parameter values for which a quantitative uncertainty distribution cannot be defined). Examples of the former include a choice between a simple linear approximation of an algorithm or a more full specification, or a choice between two weather models in a dispersion model. Scenario analysis is the primary way to examine model uncertainties such as these.

Examples of the parameter uncertainty include situations in which two sources of data yield different values for the same parameter (e.g., food consumption data from consumer dietary recall surveys compared with industry tracking data on food disappearance) and situations in which model uncertainties are embedded into input variable possibilities (e.g., low, best-estimate, and high estimates of population based on multiple underlying factors in the population estimations). Scenario analysis is very useful for characterizing parameter uncertainties when the relative weights of the different parameter states cannot be estimated through expert judgment or a reasonable averaging method. It may not be the best approach for estimating all such uncertainties, however. For example, it is generally recognized that dietary recall surveys underestimate consumed food, whereas industry tracking data does not account for food that is bought but not eaten. As a result, these disagreeing data might be better modeled as the upper and lower bounds of a distribution of possible food consumption rather than analyzed as scenarios.

Regression-Based Approaches

For probabilistic models, regression analysis and correlation coefficients can be used to quantify the degree to which output uncertainty is associated with input uncertainty. These approaches, using standard statistical methods for establishing this relationship require corresponding datasets of input and output quantities. The computation of correlation coefficients requires paired data, as is naturally the case with sampling-based simulations. Regression-based methods are discussed by Frey and Patil (2002), Frey et al. (2003), and Helton and Davis (2003) and are compared in detail across different types of models by Helton and Davis (2002).

Regression-based approaches can be used on sample data or on rank-ordered data. Using sample data implies fitting a model to a dataset of input values and output values; these sample values are generally normalized on the basis of the mean and standard deviation (Frey et al. 2004). Sample regressions and correlation coefficients result in estimates of linear relationships between inputs and outputs and therefore generally are applied to linear models. Sample data can be rank transformed by replacing sampled data values with their ranked order; the smallest value for a variable is given Rank 1, the second-smallest Rank 2, and so on (Iman and Conover 1979, Helton and Davis 2000). Regressions and correlations of rank-transformed data are effective with nonlinear but monotonic models (Frey et al. 2003, Helton and Davis 2002).

The sample correlation coefficient (CC), or Pearson coefficient, is valued between –1 and 1, where 0 represents zero correlation and 1 and –1 imply perfect positive and negative linear correlation, respectively. Statistical significance of the correlation can be estimated using the inverse Fisher test (Frey et al. 2003). The percentage of variance of the output due to variance in a particular input is computed by squaring the CC; a 0.5 CC implies that 25% of output variance is due to the variance of that input parameter (Frey et al. 2003).

The rank correlation coefficient (RCC), or Spearman coefficient, is similar to the CC except is computed on rank-ordered data (Saltelli et al. 2000). The RCC estimates the monotonic (and sometimes nonlinear) relationship between an input and output. RCC CCs are performed on one variable at a time and cannot account for correlations between inputs or interaction effects. Analysts must therefore be careful not to confuse correlation with causation.

Whereas CCs measure the effect of one variable at a time, multivariate regression analyses can estimate the impact of a set of variables. Such regression analyses usually do not include all model input variables but are constructed in a stepwise fashion, in which one variable at a time is added until no additional significant variables can be found, starting with those variables assumed to be the most significant (Helton and Davis 2000). Least-squares methods can be used to construct linear models and to compute regression coefficients (RCs), which estimate the degree to which changes in inputs effect changes in outputs.

RCs can be tested for statistical significance by computing the F-statistic, though there are limitations to the assumptions of significance that can be drawn from deterministic models (Helton and Davis 2000). The R^2 value (the coefficient of determination) of each modeled input variable is equal to the percentage of uncertainty in the output due to that input; the sum of individual R^2 values is the model's R^2 value and equals the percentage of uncertainty in the output captured by the regression model.

If the input and output data of the model to be regressed are normalized to each have mean of 0 and a standard deviation of 1, then the coefficients that result from regression analyses are referred to as standardized regression coefficients (SRCs) and have values between –1 and 1 (Helton and Davis 2002, Frey et al. 2003). (More information on normalizing data for regressions, including adjustments for sample size, can be found in Neter et al. 1996.) SRCs "provide a measure of importance based on the effect of moving each variable away from its expected value by a fixed fraction of its standard deviation while retaining all other variables at their expected values" (Helton and Davis 2002, p. 594). Thus, SRCs can be used for importance assessment. As with CCs, regression analysis can be performed on ranked data to move beyond linear relationships to capture monotonic relationships, and resulting in rank regression coefficients (RRCs) and standardized rank regression coefficients (SRRCs).

Partial correlation coefficients (PCCs) and partial rank correlation coefficients (PRCCs) estimate the linear and monotonic relationships, respectively, between an input and the output after correcting for all linear and monotonic effects of all other variables in the analysis (Helton and Davis 2002). The general method involves creating a regression model for both an input variable and the output variable with all other variables on the right-hand side, subtracting these models from the original data to obtain new variables in which the effects of other variables have been controlled for, and computing the CCs between these new variables.

Helton and Davis (2002) compute the CCs, SRCs, PCCs, RCCs, SRRCs, and PRCCs for a series of different models (linear vs. monotonic). Not surprisingly, they find that linear approaches (CCs, SRCs, and PCCs) perform better on linear models and rank approaches (RCCs, SRRCs, and PRCCs) perform better on nonlinear monotonic models. The implication is that the analyst must have some knowledge of the linearity of the model to be analyzed before choosing the appropriate type of regression-based approach.

Frey et al. (2003) discuss some of the disadvantages of regression-based approaches, including lack of robustness if assumptions of regression analysis are not met and ambiguities in interpretation. The degree to which a regression-based approach can be interpreted quantitatively is determined by the extent to which two conditions are met: input variables are independent, and residuals of the least-squares regression are independent and normally distributed. For nonmonotonic models in particular, regression-based methods do not perform well (Helton and Davis 2002).

Classification and Regression Tree

Frey et al. (2003) note that Classification and Regression Tree (CART), also called hierarchical tree-based regression (HTBR), "can be thought of as a forward stepwise variable selection method, analogous to forward stepwise regression analysis" (2003, p. 55). CART uses numerical search procedures (maximization routines) to split data into bins in an iterative fashion, where the bins have ever-smaller variance than the parent dataset. The splits form the branches of a tree in which the final nodes (leaves) are the most homogenous subsets of data.

In this approach, the input variables that have the most impact on output variance are split earliest in the tree, because the first split corresponds to the largest possible reduction in output variance—similar to the first variable in a stepwise regression. The CART method is more generally applicable than regression analyses because it is nonparametric and does not require the specification of the functional form of the model (Frey et al. 2003).

ANOVA

ANOVA is a nonparametric probabilistic sensitivity analysis method that determines statistical association between input and output without the creation of predictive regression models. It is model-independent and may be used on nonlinear and nonmonotonic models, unlike regression-based approaches. It easily incorporates categorical variables and can be done on a single input or multiple input variables.

The general approach of ANOVA is to partition input "factors" into ranges of values called "levels" and combine sets of factor levels into "treatments" and examine the effect of treatments on the mean of the output or "response variable" (Frey et al. 2003). ANOVA uses the F-statistic to identify statistically significant changes in the mean value of the response variable across different factor levels and treatments. Some assumptions must be made for ANOVA (e.g., inputs are independent and the response variable is normally distributed); however, if a particular model fails these assumptions, usually corrections can be made or the model examined through principal component analysis (Frey et al. 2003).

Helton and Davis (2000, 2002) refer to the above analysis as a test for common means because one is searching to determine whether different levels of input factors have common means in output response. They discuss three additional similar techniques focusing on different summary statistics of distributions: common medians, common locations, and statistical independence.

Using the common medians approach, one relies on the χ^2 statistic to determine whether different levels of input factors are associated with different medians in output response. Using common locations, one relies on the Kruskal–Wallis statistic to identify changes in the distribution of the output response across different levels of input factors. Statistical independence is the most advanced technique and involves partitioning the output response into levels of equal probability to the partitions of the input factors, then calculating the χ^2 statistic to test whether the distribution of input–output pairs is nonrandom.

Helton and Davis (2000, 2002) do not address which methods are preferred or which perform better in particular circumstances, but all four approaches outperform regression-based approaches on nonmonotonic models. In the two nonmonotonic models they test, statistical independence is generally more sensitive than the other three ANOVA-based approaches and tends to show statistically significant relationships for certain variables where the others do not (Helton and Davis 2002). (For more information about these methods, see Kleijnen and Helton 1999a, 1999b.)

Variance-Based Methods

Variance-based sensitivity analysis methods, also known as variance decomposition methods, have become popular in recent years. Using these approaches, the contribution of individual model inputs to the variance of model output are computed using various procedures. Some approaches can be used for sensitivity analysis as well as uncertainty analysis. Variance-based methods are global rather than local in that they require sampling across input distributions in repeated simulations. As a result, variance-based approaches are computationally expensive. These methods are also model independent—they work on linear and nonlinear models, monotonic and nonmonotonic models. Model independence is a considerable advantage and the reason why these methods are gaining increasing attention.

Chan et al. (2000) describe the three variance-based methods. The first is the use of ratios of conditional prediction variance, which indicate the extent to which variance in the output is controlled by input variance. In a correlation ratio, the variance of the conditional expectation of prediction (VCE) is divided by prediction variance. Computation of the VCE is resource intensive; one approach uses the sum of squares from ANOVA using LHS, whereas another approach requires analytical integration, resource-intensive MC approximation of the integral, or a computation scheme intended to approximate the integral. Importance measures can be computed for input variables on the basis of the correlation ratios or the aforementioned integral.

The second variance-based method is the Sobol′ method (Chan et al. 2000). It is based on decomposing a model function into summands of increasing dimensionality across the set of inputs. This method explicitly incorporates interaction effects between input variables when the sensitivity indices are calculated. As such, the Sobol′ method is more powerful than the correlation ratios approach. It is also very computationally intensive—prohibitively so for large models with many uncertain inputs. Also, although some studies have used Sobol′ methods for sensitivity analysis, it remains difficult to implement because of the lack of available software tools that can calculate Sobol′ indices for given models (Frey et al. 2003).

The Fourier Amplitude Sensitivity Test (FAST) is a variance-based method that can be used to estimate both the expected value and variance of the output variable and the contribution of input variables to this variance (Chan et al. 2000). The main idea behind FAST is the use of a pattern search method (based on transformation functions and "integer angular" sampling frequencies) in which the variance of output is decomposed into Fourier coefficients, which then are numerically evaluated on the basis of the samples.

Classical FAST can only be used for first-order effects and cannot handle interactions, but extended FAST can capture higher-order interactions in importance estimates. Extended FAST is the most powerful and appealing of the variance-based methods because it is a global, model-independent method that can be used for both sensitivity and uncertainty analysis of nonlinear and nonmonotonic models. Unlike correlation ratios, extended FAST captures interaction terms, and compared with Sobol′, FAST is more computationally efficient for the calculation of equivalent sensitivity indices (Chan et al. 2003). Nonetheless, FAST is still computationally intensive—prohibitively so for large complex models—and few software tools are available for the application of the procedure (Frey et al. 2003).

Reliability Algorithms

First- and second-order reliability methods (FORM and SORM, respectively) are specialized types of sensitivity analysis in which the parameter spaces of inputs are searched for combinations that lead to response function (output) values that cross an established "failure" threshold (Cawlfield 2000). Optimization methods, as opposed to sampling, are used to identify the most likely failure point (the design point), and then the probability of failure is estimated on the basis of the probabilities of uncertain input variables.

Gamma sensitivity measures can be computed on the basis of derivatives of the response function with respect to each input parameter; these measures "evaluate the sensitivity of the probability estimate to equally likely changes in each uncertain variable" (Cawlfield 2000, p. 165). Gamma sensitivity measures can incorporate correlation and marginal distributions, unlike many other sensitivity analysis methods which require independence of inputs. Studies comparing FORM with sampling-based methods such as MC find the approaches generally comparable, with each having advantages and disadvantages (Cawlfield 2000).

Response Surface Method

The response surface method is not a sensitivity analysis method per se but may be used to enable sensitivity analysis when the size, complexity, or computation resources of some model makes such analyses prohibitively difficult. In this approach, a simplified version of the original model is created as a response surface that is subsequently used in uncertainty and sensitivity analyses using MC simulation (Helton and Davis 2003).

To create the response surface, which can be linear or nonlinear, first order, or second order, a least-squares regression method typically is used (but rank and nonparametric approaches also can be used) to fit an equation to the data from the original model (Frey and Patil 2002). Given adequate fit, this simplified and less computationally intensive model can be analyzed using other sensitivity analysis measures discussed in this section.

Many of the models that EPA relies on for large-scale RIAs may fall into the category of being too complex for extensive sensitivity and uncertainty analysis. As EPA moves toward large sampling-based integrated assessment models that may become increasingly—even exponentially—complex, response surface methods should be considered as alternatives for sensitivity analysis and importance assessment.

Value of Information

After sensitivity analyses and/or importance assessments indicate which inputs contribute the most to uncertainty (and variability) in the output, VOI methods should be used to identify the most beneficial uncertainties to reduce through additional research (Cullen and Frey 1999, Kann and Weyant 2000). Reductions in or eliminations of uncertainties associated with input parameters can result in two measurable benefits when output is estimated in dollars: changes in the expected value of net benefits and in the variance of net benefits. Usually, VOI estimates only the benefits of a change in expected value, which are then compared with the costs of reducing the uncertainty surrounding that particular input parameter. (Some examples of recent VOI analyses are Nordhaus and Popp 1997, Yokota and Thompson 2004, and Macauley 2005.)

VOI is useful in scenarios in which information changes decisions, such as when two policies are being considered and the preferable policy cannot be determined due to uncertainty. Because EPA RIAs are evaluations of particular policies and not policy decision analyses, VOI is unlikely to be used in such analyses to change policies. VOI is likely to be important when there is estimated overlap between expected likelihoods of costs and benefits or, put another way, when the likelihood of negative net benefits is significantly greater than zero. Furthermore, VOI indicates where future research dollars should go, directing it toward those sources with the highest estimated VOI.

In some cases, joint benefits of research may result when uncertainties associated with multiple parameters are resolved and when the joint VOI is greater than the sum of individual VOI (Kann and Weyant 2000). VOI can be estimated for an elimination of uncertainty, also referred to as the expected value of perfect information, or for a more modest reduction in uncertainty. Furthermore, the VOI may be determined at different times of resolution, and the VOI of resolution of uncertainty in later years can be compared with immediate resolution of uncertainty, all policy assumptions remaining constant. Kann and Weyant refer to this as the “expected value of early revelation of uncertainty” (2000, p. 39).

Whenever VOI is high, the question may arise whether a policy should be delayed while uncertainties are resolved. When a policy is delayed, policy costs and benefits from the simulation are zero during the delay period, but several trade-offs must be considered with delays: irreversible damages, possible higher mitigation costs, and possible lower costs due to interim research and innovation.

Additional Methods

There are many additional methods beyond those described already, but the above list includes the most commonly used approaches. We name a few more of them here because they are discussed in recent literature and may emerge as more widely used methods.

Frey and Patil (2002) discuss mutual information index, a computationally complex but model-independent approach based on conditional probability analysis that is used primarily for dichotomous choice models. Cullen and Frey (1999) point to an importance-ranking approach called contribution to variance, Kann and Weyant (2002) summarize minimax regret strategies, and Saltelli (2002) discusses calibration approaches, including MC filtering, generalized sensitivity analysis, and generalized likelihood uncertainty estimation (GLUE). Haimes (2004) describes the uncertainty sensitivity index method (USIM), which investigates the effects of variation around nominal values of model inputs. Bayesian approaches (e.g., Bayesian model averaging or Bayesian inference) also may be used to analyze uncertainty (Araña and León 2003, Bates et al. 2003).

Recent EPA Attention to Uncertainty

In moving toward probabilistic modeling and integrated assessment approaches, EPA also has moved toward unifying techniques used in its models. The agency is currently in the process of developing guidance on environmental modeling from both internal and external sources. In 2000, EPA created the Council for Regulatory Environmental Models (CREM), composed of senior EPA risk managers, to help ensure that EPA documents, communicates, and implements data and models in a consistent, reliable way and that it stays abreast of advances in environmental modeling. In November 2003, CREM released a draft of guidance on environmental models and an online Models Knowledge Base to serve as an inventory of EPA environmental models. Both are currently under review by the EPA Science Advisory Board (SAB).

CREM also has been involved with the National Academies of Science, whose ongoing project Environmental Decision Making: Principles and Criteria for Models is intended to provide clear guidelines on the selection and use of models at EPA (NAS 2005). Uncertainty is an important component of the NAS panel's scope, as described on the project's Web site: "Through public workshops, and other means, the committee will consider cross-discipline issues related to model use, performance evaluation, peer review, uncertainty, and quality assurance/quality control" (NAS 2005). In the first meeting, on March 18, 2004, many representatives of different EPA offices, as well as presenters from outside the agency, discussed uncertainty. The third meeting tackled uncertainty directly, featuring presentations by experts Christopher Frey, M. Granger Morgan, and Dan Krewski. Other issues included uncertainty analysis at the U.S. Geological Survey and decisionmaking under uncertainty at EPA.

Under a cooperative agreement with EPA, the Woodrow Wilson International Center for Scholars sponsored two symposia in 2003 and 2004 on the use of environmental models in decisionmaking. The first included a discussion of the legal background on the regulatory use of models and focused on the scientific treatment of uncertainty in environmental models in a presentation by Kenneth Reckhow (2003b).

The second symposium was on integrated models and their associated uncertainties and included presentations by experts on uncertainty in environmental models, including Bruce Beck, Igor Linkov, Max Henrion, and EPA's Elsie Sunderland and Neil Stiber. Much of the discussion centered on the desirable qualities of models from a regulatory standpoint. For example, a successful model should resemble real-world interactions, address different types of uncertainty, provide transparent documentation and analysis, reconcile alternate model predictions, and permit the assessment of the quality and accuracy of model results. The remaining challenge is to put these ideas into practice while ensuring that results are presented in the most useful and compelling way. We contribute to this goal in the remaining chapters of this report.

The discussions at the various meetings and symposia with which EPA has been involved show an expanding inclusion of uncertainty in the development of new models and the analysis of results from existing and future models. Whether this embrace of uncertainty will follow through into the documented analysis of the results of these models as presented in RIAs remains to be seen, but it points to the need for decisionmakers—and not only modelers and analysts—to become familiar with the concepts of uncertainty. Uncertainty can be viewed not only as a technical characteristic of models but also as a factor in decisionmaking. It might serve decisionmakers well to be knowledgeable about the types and sources of uncertainties in environmental impact assessments and to be familiar, at some level, with probabilistic modeling approaches and types of uncertainty analyses. Knowledge about uncertainty can lead to better decisions that reflect a truer understanding of the possible implications of a regulation.

Table

Table 2-1. Comparison of NRC (2002) Recommendations and OMB *Circular A-4* (2003) Guidance

NRC	*OMB*
Incorporate sensitivity analyses into primary analysis	Report full probability distributions of consequences
Use formal uncertainty analyses, such as Monte Carlo	Use formal uncertainty analysis (if effect is >$1 billion).
List all sources of uncertainty; Quantify as many as possible	Quantify largest drivers of uncertainty
Use expert elicitation to solve data problems	Use expert elicitation to solve data problems
Address dependencies through joint distributions	Address dependencies through joint distributions
Use importance analysis to identify driving uncertainties	Use importance analysis to identify main uncertainties; Limit scope of uncertainty analyses to key drivers
Consider using value-of-information analysis to identify areas of further research	Use value-of-information analysis to identify data to gather, if uncertainties have large effects on conclusions about net benefits, consider additional research before rulemaking; Use "real options" analysis to estimate costs and benefits of delaying a decision
Avoid unwarranted degree of certainty: rounding, ranges, graphs	Avoid false sense of precision: rounding, ranges, graphs
Analyze/present distributional effects	Analyze/present distributional effects (with particular attention to intertemporal issues)
Make sure results can be transferred between BCA and CEA analyses (present outcomes nonmonetized and by age, for cost-effectiveness)	For major rules, BCA and CEA should both be presented
Expand uncertainty analysis to include more than one source of uncertainty at a time (p. 144); Be exhaustive	Balance thoroughness with practical limits: uncertainty analyses need not be exhaustive, nor does every alternative need be evaluated at every step
Assign probability distributions whenever possible–some distribution is better than none; Use expert judgments; Preferential to create probability model over alternatives, rather than scenario approaches	Less stringent than NRC; If the level of scientific uncertainty is too large, may present discrete alternative scenarios without assessing the relative likelihood of each scenario; Suggests using alternative baselines for possible differences in regulatory agencies

NRC	*OMB*
Always present ranges instead of only means	Report probability distributions (ranges, statistics) where possible, but *always* report expected value—never just report range
Do not refer to mean estimates as "best estimates"	"Best estimates" are appropriate; OK to compare expected values of benefits and costs so long as society is "risk neutral," which should be assumed generally
Present unit values used to monetize health outcomes; Indicate whether they include WTP, medical costs, and lost earnings; Describe intertemporal changes in unit values	Report benefits and costs in three categories: quantified and monetized, quantified but not monetized, and qualitative (not monetized or quantified); Suggests to monetize quantified benefits and costs but does not require reporting out unit values
Address discounting clearly, present results discounted and undiscounted	Very specific instructions regarding discounting scenarios
Separate uncertainty about the future from model uncertainty Obtain lower bound on uncertainty of future estimates by applying benefits to current population	NA
Perform sensitivity analysis on distribution types, especially for uncertainty derived from experts	NA
Emphasize unaccounted-for sources of uncertainty	NA
Distinguish between data-derived estimates and those from expert elicitation	NA

Notes: BCA = benefit–cost analysis, CEA = cost-effectiveness analysis, NA = not addressed.

Appendix 2A: Brief History of EPA Guidance on Uncertainty

The initial guidance for the U.S. Environmental Protection Agency (EPA) on incorporating uncertainty into policy analysis came from "The Red Book," a National Research Council (NRC) report titled *Risk Assessment in the Federal Government: Managing the Process* (NRC 1983). The guidance in this document was not focused on uncertainty, but the authors did note the need to answer questions such as, "What are the statistical uncertainties in estimating the extent of health effects?" and "How are these uncertainties to be computed and presented?" (NRC 1983, p. 33). The report notes that little guidance was available to recommend to EPA on how to incorporate uncertainties in data and combine these uncertainties into final estimates of risk.

A 1985 report from the Office of Science and Technology Policy (OSTP 1985) and subsequent EPA guidance on carcinogenic risk assessment (U.S. EPA 1986) and on risk assessments for Superfund (U.S. EPA 1989) all identify uncertainty as a critical area in analyses but do not recommend modeling this uncertainty. The general guidance at this time was that "it is more important to identify the key … variables and assumptions that contribute most to the uncertainty than to precisely quantify the degree of uncertainty in the risk assessment" (U.S. EPA 1989, p. 8-17).

As the literature emerged in the early 1990s on the use of probabilistic techniques to incorporate uncertainty into numerical models (e.g., Finkel 1990, Morgan and Henrion 1990, Frey 1992), EPA guidance for risk assessments began to move away from recommending qualitative approaches. EPA's *Guidelines for Exposure Assessment* (U.S. EPA 1992), two additional NRC reports (NRC 1994, 1996), and EPA's Risk Characterization Policy (U.S. EPA 1995) all stress quantitative methods.

By the late 1990s, probabilistic approaches to risk assessment were widely accepted, and EPA guidance documents reflect this change (U.S. EPA 1997a, 1997b, 1997c, 1998a, 1999c, 2000a, 2001a, 2001b, 2003b). A case study review by the EPA Science Advisory Board (SAB) even suggested that EPA should integrate uncertainty and variability into the primary analysis, rather than treating uncertainty through secondary sensitivity analyses (U.S. EPA 2000b). *Examination of EPA Risk Assessment Principles and Practices* finds this suggestion impractical, however:

> Until such methods and supporting data are developed, though, it is not feasible to do a full and integrated assessment for every analysis. Further, in most instances EPA is not the data developer: much EPA analysis is based upon third-party literature. (U.S. EPA 2004c, p. 34)

Although it is not a guidance document, the 2004 evaluation of risk assessment practices suggests a "tiered approach" to uncertainty analysis, starting as simply as possible with qualitative descriptions and adding analyses (sensitivity, probabilistic modeling) only as warranted.

Appendix 2B: Review of EPA RIAs

The U.S. Environmental Protection Agency (EPA) has performed several significant regulatory impact analyses (RIAs) and benefit analyses in the past few years, and uncertainty is discussed, to some extent, in nearly all of them (U.S. EPA 1996a, 1998b, 1999b, 1999d, 1999e, 2003a, 2004b; Abt Associates 2000). We briefly examine the treatment of uncertainty in four major EPA RIAs:

- *The Benefits and Costs of the Clean Air Act 1990 to 2010* (U.S. EPA 1999a)
- *Final Regulatory Analysis: Control of Emissions from Nonroad Diesel Engines* (U.S. EPA 2004d)
- *Regulatory Impact Analysis for the Final Clean Air Interstate Rule* (U.S. EPA 2005a)
- *Regulatory Impact Analysis of the Clean Air Mercury Rule* (U.S. EPA 2005b)

Three of these documents involve criteria pollutants: one performed before the publication of National Research Council (NRC 2002) and Office of Management and Budget (OMB 2003) guidance, and two published after. These RIAs are useful for comparison because they model the same health endpoints and rely on the same models and therefore involve the same uncertainties. The fourth RIA examined involves mercury reductions in air pollution, which is a useful comparison with criteria pollutants precisely because it involves different health endpoints and therefore some different uncertainties from the other three.

RIA: Benefits and Costs of the Clean Air Act, 1990–2010

In 1999, EPA published its second prospective cost benefit analysis of the Clean Air Act (CAA), occasionally referred to as "the 812 Study" (U.S. EPA 1999a) in reference to the section of the 1990 CAA Amendments mandating that it be performed. Although briefly summarized and reviewed by the NRC report (NRC 2002, pp. 49–54), the 812 Study is worth including here for comparison but also because of scale. The coverage of the CAA is broader than most other air pollution regulation.

Also, of the benefits estimates produced by EPA before the NRC report, the 812 Study incorporates the most detailed information about uncertainty; it notes explicitly on the very first page of the executive summary that the last of six steps in the analysis is to "aggregate results and characterize uncertainties" (U.S. EPA 1999a, p. i). Uncertainties are discussed in a section at the end of each chapter, from emissions through air quality modeling to health effects and valuation, with tables listing key uncertainties and their likely impact on results. Quantitative analyses are included in many sections, and appendices detail sensitivity analyses on alternative assumption. Still, these uncertainties do not propagate through the entire analysis, and only a small subset of uncertainties are included in ranges presented for final benefit estimates.

The summary table in the executive summary includes low, central, and high estimates of monetized benefits and identifies these as the "5th and 95th percentile results from statistical uncertainty analysis" (U.S. EPA 1999a, p. iii). The report goes on to state that quantitative estimates of uncertainty in benefits were computed by statistically combining uncertainties from "many of the factors" contributing to the benefit estimate. The range from low to high is noted as a "partial indication of the overall uncertainty surrounding the central estimate" that reflects "a 90 percent probability range around the mean" (p. v).

Although the inclusion of a range certainly conveys a truer sense of the uncertainties than a single point estimate, both the table and paragraph description imply the range to be some sort of 90 percent confidence interval. The word "partial" is important, because although EPA computes uncertainty quantitatively in many of the physical modeling stages of the analysis, the only quantified uncertainties included in the range of benefits are those pertaining to concentration–response (C–R) functions for health endpoints and the valuation of these health endpoints. It takes a careful reading to realize exactly which of the quantified uncertainties laid out in individual chapters make it into the final estimates. It is not until the final chapter that this becomes plainly apparent: "Quantitative estimates of uncertainties in earlier steps of the analysis (i.e., emissions and air quality changes) could not be developed adequately and are therefore not applied in the present study. As a result, the range of estimates for monetized benefits presented in this chapter is more narrow than would be expected with a complete accounting of the uncertainties in all analytical components" (U.S. EPA 1999a, p. 100). No quantitative analysis of uncertainties associated with costs is included.

EPA is careful to distinguish between 5th and 95th percentile estimates of aggregated benefits, and the 5th and 95th percentile estimates of the benefits due to individual health endpoints. It presents the results of the human health benefits valuation, broken down by major health endpoint (mortality, chronic bronchitis, etc.), with the mean and 5th and 95th percentiles reported (U.S. EPA 1999a, p. 75, table 6–3).

However, EPA does not include 5th and 95th percentile estimates of aggregated estimates:

> Summing 5th and 95th percentile values would yield a misleading estimate of the 5th and 95th percentile estimate of total health benefits. For example, the likelihood that the 5th percentile estimates for each endpoint would simultaneously be drawn during the statistical uncertainty analysis is much less than 5 percent. As a result, we present only the total mean. (U.S. EPA 1999a, p. 75)

This distinction is important and highlights one of the ways in which Monte Carlo (MC) simulation and subsequent analysis can aid in getting a truer picture of uncertainties. For example, EPA's summation of the valuation of 5th and 95th percentile estimates of individual health endpoints (U.S. EPA 1999a, p. 75, table 6–3) adds up to $15.6 billion and $272 billion, respectively, compared with the Primary Low and Primary High estimates of $26 billion and $270 billion, respectively. The total uncertainty is less than the sum of individual uncertainties, but its impact is limited because the aggregate benefits are so driven by mortality valuation (mean = $100 billion, 5th = $14 billion, 95th = $250 billion).

One analysis that should be applauded is the importance analysis performed on the uncertainties included in the benefits estimates (U.S. EPA 1999a, p. 107, figure 8–2). Although this analysis indicates which uncertainties in model inputs drive uncertainty in model outputs, it is limited by the fact that EPA only quantifies health effect and valuation as uncertain parameters. Nonetheless, it is a notable attempt at identifying which uncertainties are most important.

Matthews (2001) produced an alternate assessment of the benefits and costs of the Clean Air Act. Rather than estimate health effects from modeled pollution and value those health effects, Matthews draws alternate values of social damages per ton of pollutant from the literature and applies them to EPA's estimates of pollution. The differences are significant; net benefits using this alternative approach, though still positive and sizeable, are 95% lower than EPA's central estimate and fall entirely below EPA's minimum estimate. The differences between the two studies indicate a large model uncertainty associated with the different approaches used.

Although the 812 Study makes an honest and serious attempt at incorporating uncertainty into its analysis, the presentation of uncertainties is quite limited. Most reported quantities are point estimates, and when uncertainty is represented, it generally is only at the summary level as a range. Tables often use the shorthand of "low" and "high" without specifying what these terms imply. No graphical representations of uncertainty, such as probability density functions (PDFs) or cumulative distribution functions (CDFs), are included for any of the uncertainties quantified in benefits estimates. For example, EPA finds that when using best estimates of mortality valuation from 26 studies, a "Weibull distribution, with a mean of $4.8 million and standard deviation of $3.24 million, provides the best fit" (U.S. EPA 1999a, p. 71). A graphical display of the values in these 26 studies, including their reported uncertainties, as well as a figure showing the fit distribution, possibly overlaid, would be one helpful solution to conveying more information about the single largest driver of uncertainties in the benefit estimate.

RIA: Control of Emissions from Nonroad Diesel Engines

In 2004, the EPA Science Advisory Board (SAB) issued an advisory on EPA's second prospective study of the costs and benefits of the Clean Air Act, influenced by the NRC and OMB recommendations (U.S. EPA 2004a). One of its primary findings was to move toward bringing uncertainty into the primary analysis:

> We propose that the Second Prospective Analysis present the base case with associated uncertainties (preferably confidence intervals of 10%–90%), plus a set of sensitivity analyses, rather than the base case and a single "alternative analysis."[1] The Council and the HES advise that the single "alternative analysis" to the base case described in the agency's Draft Analytical Plan does not represent to us, as scientific and technical experts, the comprehensive scientific analysis of health benefits that we understand the Clean Air Act to require We advise that the agency aim for a quantitative base case that includes best estimates for all health effects

[1] Authors' note: EPA developed this "alternative estimate" as an interim attempt to show uncertainty in benefits assessments until the agency could complete full probabilistic benefits analysis, as requested by NRC and OMB. EPA used the alternative estimate in support of Clear Skies and other rulemakings to present a second estimate, based on different assumptions, in addition to the base estimate. This estimate assumed zero chronic mortality effects due to air pollution, thus necessarily presenting a lower estimate of benefits.

> for which there is reasonable quantitative evidence with careful avoidance of potential double counting. This should be supplemented with an acknowledgement of the likely benefits that cannot be adequately quantified at this time. If alternative estimates are presented, they should be balanced to reflect the possibilities that the base case may either understate or overstate actual health benefits. (U.S. EPA 2004a, cover letter, p. 2)

EPA's 2004 *Final Regulatory Analysis: Control of Emissions from Nonroad Diesel Engines* (U.S. EPA 2004d; hereafter, nonroad diesel RIA) heeds SAB advice and does not include an "alternative analysis." Noting NRC, OMB, and SAB advice to move uncertainty assessments from secondary analyses into the primary analysis by performing probabilistic multiple-source uncertainty analyses, EPA reports, "We are working to implement these recommendations." Nearly all estimates are presented as point estimates without being identified as central tendencies of distributions, and no summary statistics about variance of estimates of health effects or benefits are discussed outside of appendices.

Presentation of uncertainty varies from chapter to chapter. On the one hand, chapter 2 of the nonroad diesel RIA (Air Quality, Health, and Welfare Effects) presents information about the effects of nonroad diesel emissions and generally presents uncertainty where appropriate. For example, it gives 95% confidence intervals for one variable and a range for another (U.S. EPA 2004d, p. 2-10, table 2.1.1-1). A tornado graph shows the possible impacts of nonroad diesel emissions compared with other carcinogenic risks (U.S. EPA 2004d, p. 2-77), and spatial variation of diesel particulate concentrations is shown on a U.S. map (p. 2-62).

On the other hand, chapter 3 of the nonroad diesel RIA (Emission Inventory) does not present any uncertainty information or discuss uncertainties qualitatively. It contains no quantitative assessment of uncertainties in costs, and uncertainties associated with benefits are quantified only in appendices. The executive summary includes only one mention of uncertainty, namely, that "the high end of the uncertainty range for this rule's estimated benefits could exceed the low end of the range by a factor of 20" (U.S. EPA 2004d, p. ES-2). However, it does not quantify this range at all or put it into any context of the $80 billion "best" estimate. EPA does not reconcile these estimates, to the best of our understanding.

The best estimates, reported as point estimates with no reference to uncertainty, are presented in the nonroad diesel RIA as $83 billion in 2030 assuming a 3% discount rate and $78 billion assuming a 7% discount rate (U.S. EPA 2004d, p. 9-43, table 9-12). The results of a partial uncertainty analysis also are given in the same chapter (appendix 9-B, p. 9-212, table 9B-2), where the results of modeling C–R functions and health unit values probabilistically results in a 5th percentile estimate of $23 billion and a 95th percentile estimate of $200,000 billion (p. 9-212). However, this range cannot be compared with the primary analysis, because its mean estimate of $96 billion reflects pollution levels in the proposed rule, not the scaled emissions in the final rule, "as the scaling methodology adds a new element of uncertainty that cannot be appropriately characterized here" (p. 9-212, footnote A). This range can only be compared with the estimates in appendix 9-A of the nonroad diesel RIA, which use the same modeled preliminary control option.

Therefore, the only estimates of uncertainty in benefits estimations to appear in the nonroad diesel RIA apply not to the actual estimates of benefits but rather to old estimates that reflect a different level of stringency. Overall, the nonroad diesel RIA does not adequately explain these circumstances, highlight the disparity between the primary and uncertainty analysis, or make any qualitative judgments about how these uncertainty analyses can be applied to the lower benefits estimates in the final rule.

As for the specifics of the MC analysis, the nonroad diesel RIA lists in tables the distributions around valuation parameters but does not report out the quantitative uncertainty around individual C–R functions used in the analysis. This is similar to the 812 Study, but the 812 Study goes into more detail about the shape of the C–R function for premature mortality due to particulates. There are no graphical representations of uncertainty in this section of the analysis. The nonroad diesel RIA includes tables of uncertain parameters and lists the uncertainties in most sections but does not indicate in any detailed discussion which uncertainties are the most important. EPA lists the primary sources of uncertainty in the benefit analysis (U.S. EPA 2004d, p. 9-37, table 9-8), but no information is given about which uncertainties are largest, which uncertainties might raise or lower benefit estimates, which uncertainties can be quantified, or which uncertainties are ultimately tested through sensitivity analysis.

Furthermore, a closer examination of table 9B-2 in the nonroad diesel RIA suggests that EPA may (or may not) have made the mistake they avoided in table 6-3 of the 812 Study (U.S. EPA 1999a), discussed previously. The uncertainty associated with an aggregate measure of health valuation is not equal to the sum of uncertainty for individual endpoints, because the samples for each probabilistic variable in each simulation are drawn independently (unless otherwise specified). In the nonroad diesel RIA, table 9B-2 presents the 5th and 95th percentile estimates for individual endpoints (mortality, chronic bronchitis, etc) and seems to present their sum as the 5th and 95th percentile estimates of total monetized benefits. There is not enough information presented to distinguish between a rounding error and an analytical error; that is, it is possible that due to rounding, the extreme percentiles of total benefit estimates appear to match those of the sum of independent endpoint estimates. Furthermore, the total health benefits are dominated, by far, by mortality valuation, and it may be that the effect of the other distributions are simply too minor to matter here.

Included in the nonroad diesel RIA is a second uncertainty analysis that is based on the results of a pilot expert elicitation for the C–R function for mortality from particulates (IEc 2004). Although only a pilot analysis and although the results may perhaps be too premature for use in an RIA, the analysis is presented quite clearly. This analysis includes multiple boxplots of five expert judgments and benefit estimates based on those judgments, shown alongside the comparable combined expert range and the range of values from the primary analysis based on the Pope et al. study (2002), although once again, the benefits correspond not to the final rule estimates but to the preliminary model results from the proposed rule presented elsewhere in the appendices. These boxplots include not only the 90th percentile but also the interquartile range, median, and mean; however, no legends indicate as such.

The analysis also presents CDFs of estimates of reduced incidence and of dollar benefits for all seven options. These CDFs would be improved by inclusion of the mean on each line (see Chapter 3 for more details), and the overall analysis would benefit from presentation of PDFs of the different expert judgments as well. For example, EPA notes that the results derived from Pope have "an approximately Weibull shaped distribution with a range from 5th to 95th percentiles of $23 billion to $190 billion, or about one order of magnitude," while those based on combined expert judgments have "a much more skewed shape with an elongated positive tail above the 75th percentile with a range from 5th to 95th percentiles of $3 billion to $240 billion, or about two orders of magnitude" (U.S. EPA 2004d, p. 9-239). This sort of description of uncertainty is what is lacking in most EPA RIAs and absent throughout the rest of this one, but a graphical representation of the comparison would be even better.

In appendix 9-C of the nonroad diesel RIA, EPA presents a few sensitivity analyses on the C–R function for premature mortality, including using different functions from those of Pope et al. (2002), altering the lag structure, and introducing a threshold. These sensitivity analyses are useful but are not placed in the context of either the primary analysis or the MC uncertainty analysis. The point estimates presented in the summary table give some idea of the range of possible values, but bounded ranges of these estimates would have also been useful, as well as graphical depictions of the impact of these changes to assumptions. These types of analyses should be placed within the probabilistic uncertainty analysis because they are not separate components of the uncertainties associated with the C–R function.

The presentation of the expert elicitation results is the best expression of uncertainty analysis in the nonroad diesel RIA, whereas the remainder of the RIA generally presents point estimates without ranges or uncertainty, fails to graphically represent uncertainties, and has mismatched regulations in the primary and supporting analyses. The MC uncertainty analysis could present far more information about uncertainties included in the analysis (e.g., the quantitative uncertainties around C–R functions) as well as some sensitivity analyses in the input estimates of concentration, even if based on limited scenarios.

RIA: Clean Air Interstate Rule and Clean Air Mercury Rule

In 2005, EPA released two significant RIAs for two related programs: the Clean Air Interstate rule (CAIR RIA; U.S. EPA 2005a) and the Clean Air Mercury rule (CAMR RIA; U.S. EPA 2005b). The CAIR RIA is similar to the nonroad diesel RIA (U.S. EPA 2004d), especially regarding its limited quantitative treatment of uncertainties, but it is clearer about uncertainties in the executive summary and in the section on benefits. However, the CAIR RIA generally follows the pattern of other EPA RIAs by qualitatively discussing uncertainties in each section but leaving any quantitative information in the appendices. The CAMR RIA is different in character from the 812 Study (U.S. EPA 1999a), the nonroad diesel RIA, and the CAIR RIA in that it describes health benefits due to reductions in mercury, not criteria pollutants.

Three pages of the 13-page executive summary of the CAIR RIA are spent discussing uncertainty in the analysis. Unfortunately, the summary tables do not include ranges for estimates of benefits or indicate that the reported numbers represent a mean of a distribution, nor does the section reporting out health benefits include any mention of uncertainty. The 90th percentile confidence interval is reported in the section of the executive summary on uncertainty analysis: “The overall range from 5th to 95th percentile on the total benefits estimate represents one order of magnitude ($26 billion to $210 billion)" (U.S. EPA 2005a, p. 1-8). This section also reports that using a different, yet plausible, C–R function for mortality due to particulates results in a doubling of the mean benefits from $100 billion to $200 billion. The reporting of these analyses—as well as the sensitivity analysis using the results of the expert elicitation pilot already described—are notable, important improvements over the nonroad diesel RIA (U.S. EPA 2004d).

The rest of the CAIR RIA is an improvement over the nonroad diesel rule, even though few ranges, variances, or other summary statistics about uncertainty are reported, except in the appendices. In chapter 4 of the CAIR RIA, discussing the quantified health benefits of CAIR, EPA goes through the uncertainties in the estimation of illnesses from concentrations of particulates in excellent detail. It discusses the multiple data sources available for the mortality C–R, in particular, and gives an overview of the advice of the EPA SAB and the NRC on the matter. This qualitative description is excellent, and the sensitivity analyses included in the appendix address many of the uncertainties listed.

Unfortunately, like the nonroad diesel RIA, the CAIR RIA does not include a comprehensive uncertainty analysis in which all of the various uncertain characteristics of the C–R (e.g., lag structure, threshold levels, and the function itself [i.e., Pope vs. Six Cities]) are incorporated into a single estimate of the uncertainty of this important endpoint estimator. Although the CAIR RIA does not present uncertainty around estimates of costs, it does present a sensitivity analysis around some parameters in an appendix (U.S. EPA 2005a, appendix C).

Like the nonroad diesel RIA, the CAIR RIA includes two MC analyses on mortality valuation and another sensitivity analysis on other components of the mortality C–R function. Also like the nonroad diesel RIA, the table of total monetized benefits in the appendix seems to report out total uncertainty as a sum of individual uncertainties (U.S. EPA 2005a, p. B-11). The statistical uncertainty of C–R functions is not reported, even though uncertainty in these functions is a part of the MC simulations performed. The benefits estimates in the appendices do match up to the estimates in the primary estimate in the body of the CAIR RIA, unlike in the nonroad diesel RIA.

Like the nonroad diesel RIA, the CAIR RIA incorporates an uncertainty analysis using the results of the pilot expert elicitation discussed previously. This analysis is very similar to the one in the nonroad diesel rule and again presents complicated results in fairly straightforward graphs, such as boxplots and CDFs. Again, PDFs of some distributions would help. This analysis includes a discussion of peer review response to the pilot elicitation, and the implications for the analysis, which is a welcome addition.

In summary, the CAIR RIA is an improvement over the nonroad diesel RIA, but not in any revolutionary way. The incorporation of a discussion of uncertainty in the executive summary is excellent, and the direct comparability of the uncertainty analysis with primary estimates is important.

The CAMR RIA is inherently different, at least in the benefits estimate, from the other RIAs previously discussed, because it is concerned with a different pollutant with different health endpoints. The CAMR RIA includes no executive summary or presentation of net benefits. The last sentence of the 1-page introductory chapter reports that "Table 1-1 below summarizes the benefits, costs, and net benefits of the CAMR" (U.S. EPA 2005b, p. 1-1), but no such table is included in the final draft.

Estimates of costs are not reported probabilistically, but sensitivity analyses on key assumptions are presented. Estimates are presented for scenarios representing different control outcomes based on the concurrent CAIR regulation. The cost chapter discusses qualitatively various uncertainties that might affect the cost estimate, some of which might result in underestimates and some of which might result in overestimates. In general, though, "EPA believes that the annual private compliance costs that we have estimated are more likely to overstate the future annual compliance costs that industry will incur, rather than understate those costs" (U.S. EPA 2005b, p. 7-17). Sensitivity analyses examine lower costs of mercury controls and changes in natural gas prices and electricity demand, both of which result in lower estimates of costs.

Estimates of benefits are calculated probabilistically for two exposure estimate models and for multiple control scenarios, but only a few parameters used in this estimate are defined with uncertainty. Benefits are reported in two chapters; a long, expansive, detailed explanation of the data and methodology of estimating exposures and benefits; and a brief summary of overall benefits estimates.

The CAMR RIA estimates benefits for only a single health endpoint—the reduction in IQ due to prenatal exposure—so the uncertainties associated with benefits estimate are skewed upwards. The fact that benefits estimates are very conservative (and should be considered low estimates) is not sufficiently presented. To compute these benefits, EPA estimates changes to mercury levels in freshwater fish and consumption of this fish by sensitive populations, estimates reductions in IQ loss due to this consumption, and values each IQ point. EPA uses two models for estimating exposure from consumed fish on the basis of two approaches, one of which (modeling individual fisher characteristics) it believes to underestimate and the other of which (modeling watershed fishing characteristics) it believes to overestimate, thus presenting exposure estimates with bounded uncertainty.

This inclusion of uncertainties into the main estimate of exposures is precisely what NRC and OMB recommended in their call for bringing uncertainty analysis into the primary analysis and should be lauded. Exposure estimates are reported at the state level, not only the national level, thus showing regional variation in results and important characteristics of the estimates that could easily have been misrepresented through aggregate tables alone.

The CAMR RIA also includes a sensitivity analysis examining the impact of alternative dose–response functions, which are presented clearly alongside the original assumptions, for both exposure scenarios. It is an improvement over past RIAs that these sensitivity analyses are not presented in appendices but in the main body of the report.

Although EPA goes to significant lengths to estimate the number of affected individuals and their level of IQ loss probabilistically, EPA relies on a point estimate of the dollar value per IQ point. This estimate represents the loss of future earnings due to loss of IQ, computed by multiplying a percent decrease in expected future earnings (drawn from a single study) by an estimate of average lifetime earnings, plus some impacts on education, all of which have associated statistical uncertainties that should have been available to EPA. EPA gives no explanation for this simplification but does note that uncertainty is associated with its value for benefit per IQ point. EPA also notes that this value may be a lower bound of benefits because it is a "cost of illness" measure of benefits, not a willingness-to-pay estimate that would include costs of averting behavior or the impacts of pain and suffering.

The benefits estimates are summarized with excellent presentation of qualitative uncertainty. Because there is certainty associated with health impacts due to high levels of mercury exposure from experimental data, more confidence is associated with estimating health impacts at that threshold level. EPA examines lower thresholds for health impacts, down to zero. EPA presents benefits estimates, as ranges, for two policy scenarios, multiple threshold scenarios, and two discounting rates (U.S. EPA 2005b, p. 11-14, tables 11–7 and 11–8). EPA's table 11–7, replicated as Figure 2-B-1 in this chapter, shows the inclusion of an arrow in the leftmost column labeled Uncertainty Regarding Threshold, which expresses the qualitative uncertainty associated with estimates at different threshold levels. This is an excellent example of how to attempt to incorporate uncertainty information, even when qualitative, into benefit estimates.

The CAMR RIA also includes an estimate of cobenefits of the regulation associated with reductions of particulate matter of less than 2.5 microns in diameter ($PM_{2.5}$). This very rough estimate—an "illustrative analysis" and a lower bound—includes only valuation of premature mortality at the exclusion of morbidity endpoints (U.S. EPA 2005b, p. 12-1). Furthermore, mortality is valued without uncertainty as a point estimate. The uncertainties with this estimate are acknowledged openly, which is appropriate for a rough illustrative analysis limited by modeling constraints.

The CAMR RIA is an improvement over previous RIAs in the sense that uncertainties are brought more thoroughly into primary analyses, but there is much room for improvement. The uncertainties in the health impacts due to reductions in mercury probably are higher than those due to reductions in criteria pollutants, if only because the latter have been studied so deeply in recent history. RIAs pertaining to criteria air pollutants include many health endpoints, each of which is quantified and monetized, and extraordinary care placed on avoiding double counting and valuation based on a large body of literature.

The estimates of benefits in the CAMR RIA, however, rest on a single parameter—the difference in future earnings associated with an IQ point—drawn from a single study. Nonetheless, the benefits estimates of reduced mercury exposure are presented with confidence intervals that may misrepresent the overall uncertainty in benefits because these bounds represent localized uncertainties only. When the unquantified uncertainties are high, as they are in the CAMR RIA, EPA should exercise more care explaining and asserting these uncertainties. Of great importance, too, is the presentation of aggregated costs and aggregated benefits side by side, with a discussion of their uncertainties. It is an undue burden on the reader to try to flip back and forth between tables in multiple chapters to attempt to parse out these numbers.

Appendix Figure

Figure 2B-1. Benefits for "Displayed with Increasing Uncertainty"

Source: U.S. EPA 2005b

Table 11-7. IQ Benefits for CAMR Option 1 under Established Health-Based Benchmarks

Uncertainty Regarding Threshold	Benchmark Source	Level of Threshold	Discount Rate	Scaling	Benefits (millions 1999$)	Discounted IQ Points
More Certain	WHO / Health Canada	0.2 - 0.23 μg/kg bw/day	3%	4%	$0.07 - $0.12	8 - 14
				8%	$0.14 - $0.24	15 - 27
			7%	4%	$0.03 - $0.08	4 - 9
				8%	$0.06 - $0.16	7 - 18
	EPA RfD	0.1 μg/kg bw/day	3%	21%	$0.36 - $0.63	41 - 72
				34%	$0.58 - $1.0	66 - 116
			7%	21%	$0.17 - $0.42	19 - 48
				34%	$0.27 - $0.68	31 - 77
Less Certain	No Threshold	N/A	3%	100%	$1.7 - $3.0	193 - 341
			7%	100%	$0.8 - $2.0	91 - 277

References

Abt Associates. 2000. *Final Heavy Duty Engine/Diesel Fuel Rule: Air Quality Estimation, Selected Health and Welfare Benefits Methods, and Benefit Analysis Results.* Research Triangle Park, NC: Prepared for U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards.

Ades, A. E., and G. Lu. 2003. Correlations Between Parameters in Risk Models: Estimation and Propagation of Uncertainty by Markov Chain Monte Carlo. *Risk Analysis* 23(6): 1165–1172.

Anderson, E.L., and D. Hattis. 1999. Uncertainty and Variability. *Risk Analysis* 19: 47–49.

Araña, J.E., and C.J. León. 2003. *Bayesian Modeling Of Preference Uncertainty Across Discrete Choice Elicitation Methods*. Venice, Italy: European Association of Environmental and Resource Economists.

Bates, S.C., A. Cullen, and A.E. Raftery. 2003. Bayesian Uncertainty Assessment in Multicompartment Deterministic Simulation Models for Environmental Risk Assessment. *Environmetrics* 14(4): 355–371.

Beck, M.B. 1986. The Selection of Structure in Models of Environmental Systems. *The Statistician* (Special Issue: Statistical Modeling) 35(2): 151–161.

Beyth-Marom, R. 1982. How Probable Is Probable? A Numeric Translation of Verbal Probability Expressions. *Journal of Forecasting* 1: 257–269.

Bogen, K.T., and R.C. Spear. 1987. Integrating Uncertainty and Variability in Environmental Risk Assessment. *Risk Analysis* 7: 427–436.

Borsuk, M.E., C.A. Stow, and K. H. Reckhow. 2004. A Bayesian Network of Eutrophication Models for Synthesis, Prediction, and Uncertainty Analysis. *Ecological Modeling* 173: 219–239.

Bourgeron, P.S., M. Jensen, and G. Lessard. 2001. *A Guidebook for Integrated Ecological Assessments.* New York, NY: Springer-Verlag.

Bromley, J., N.A. Jackson, O.J. Clymer, A.M. Giacomello, and F.V. Jensen. 2005. The Use of Hugin to Develop Bayesian Networks as an Aid to Integrated Water Resource Planning. *Environmental Modelling & Software* 20(2): 231–242.

Brooks, S. 1998. Markov Chain Monte Carlo Method and Its Application. *Statistician* 47(1): 69–100.

Budnitz, R.J., G. Apostolakis, D.M. Boore, L.S. Cluff, K.J. Coppersmith, C.A. Cornell, and R.A. Morris. 1995. *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts.* Livermore, CA: Lawrence Livermore National Laboratory.

Burmaster, D.E., and K.M. Thompson. 1998. Fitting Second-Order Parametric Distributions to Data Using Maximum Likelihood Estimation. *Human and Ecological Risk Assessment* 4(2): 319–339.

Caldarelli, G., P.G. Higgs, and A.J. McKane. 1998. Modelling Coevolution in Multispecies Communities. *Journal of Theoretical Biology* 193: 345–358.

Campolongo, F., J.P.C. Kleijnen, and T. Andres. 2000. Screening Methods. In *Sensitivity Analysis,* edited by A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley.

Casman, E.A., M.G. Morgan, and H. Dowlatabadi. 1999. Mixed Levels of Uncertainty in Complex Policy Models. *Risk Analysis* 19(1): 33–42.

Cawlfield, J.D. 2000. Reliability Algorithms: FORM and SORM Methods. In *Sensitivity Analysis,* edited by A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley.

Chan, K., S. Tarantola, A. Saltelli and I.M. Sobol´. 2000. Variance Based Methods. In *Sensitivity Analysis,* edited by A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley.

Chen, H.-W., and N.-B. Chang. 2002. A Comparative Analysis of Methods to Represent Uncertainty in Estimating the Cost of Constructing Wastewater Treatment Plants. *Journal of Environmental Management* 65: 383–409.

Cooke, R.M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York, NY, Oxford University Press.

Cooke, R.M., and L.H.J. Goossens. 2000. Procedures Guide for Structured Expert Judgement in Accident Consequence Modelling. *Radiation Protection and Dosimetry* 90(3): 303–311.

Cooke, R.M., and J.M. van Noortwijk. 2000. Graphical Methods. In *Sensitivity Analysis,* edited by A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley.

CREM (Council for Regulatory Environmental Modeling). 2003. *Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models.* Washington, DC: U.S. Environmental Protection Agency.

Cullen, A.C., and H.C. Frey. 1999. *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. New York, NY, Plenum Press.

Cullen, A.C., and M.J. Small. 2004. Uncertain Risks: The Role and Limits of Quantitative Analysis. In *Risk Analysis and Society: An Interdisciplinary Characterization of the Field,* edited by T. McDaniels and M.J. Small. Cambridge, U.K.: Cambridge University Press.

D'Agostino, R.B., and M.A. Stephens. 1986. *Goodness-of-Fit Techniques*. New York, Marcel Dekker.

Easter, K.W., and S.O. Archibald. 1998. *Benefit-Cost Analysis in U.S. Environmental Regulatory Decisions*. Sixth Joint Conference on Food, Agriculture, and the Environment, University of Minnesota, Minneapolis, MN.

Ellison, A.M. 1996. An Introduction to Bayesian Inference for Ecological Research and Environmental Decision Making. *Ecological Applications* 6(4): 1036–1046.

Enea, M., and G. Salemi. 2001. Fuzzy Approach to the Environmental Impact Evaluation. *Ecological Modelling* 136(2–3): 131–147.

Evans, J.S., J.D. Graham, G.M. Gray, and J. R. L. Sielken. 1994a. A Distributional Approach to Characterizing Low-Dose Cancer Risk. *Risk Analysis* 14: 25–34.

Evans, J.S., G.M. Gray, J.R.L. Sielken, A.E. Smith, C. Valdez-Flores and J.D. Graham. 1994b. Using of Probabilistic Expert judgment in Uncertainty Analysis of Carcinogenic Potency. *Regulatory Toxicology and Pharmacology* 20: 15–36.

Ferson, S., and L.R. Ginzburg. 1996. Different Methods Are Needed to Propagate Ignorance and Variability. *Reliability Engineering and Systems Safety* 54: 133–144.

Finkel, A.M. 1990. *Confronting Uncertainty in Risk Management: A Guide for Decision Makers.* Washington, DC: Resources for the Future, Center for Risk Management.

Fischhoff, B., A. Bostrom, et al. 1993. Risk Perception and Communication. *Annual Review of Public Health* 14: 183–203.

Fisher, B. 2003. Fuzzy Environmental Decision-Making: Applications to Air Pollution. *Atmospheric Environment* 37(14): 1865–1877.

Fishman, G.S. 1996. *Monte Carlo: Concepts, Algorithms and Applications.* New York, NY: Springer-Verlag.

Frey, H.C. 1992. Quantitative Analysis of Uncertainty and Variability in Environmental Decision Making. Pittsburgh, PA, Center for Energy and Environmental Studies, Carnegie Mellon University.

Frey, H.C. 1993. Separating Variability and Uncertainty in Exposure Assessment: Motivations and Method. Presented at the 86th Annual Meeting of the Air and Waste Management Association, Denver, CO, June 13–18.

Frey, H.C., A. Mokhtari, and T. Danish. 2003. Evaluation of Selected Sensitivity Analysis Methods Based Upon Applications to Two Food Safety Process Risk Models. Washington, DC: U.S. Department of Agriculture, Office of Risk Assessment and Cost-Benefit Analysis.

Frey, H.C., A. Mokhtari, and J. Zheng. 2004. Recommended Practice Regarding Selection, Application, and Interpretation of Sensitivity Analysis Methods Applied to Food Safety Process Models. Washington, DCU.S. Department of Agriculture, Office of Risk Assessment and Cost-Benefit Analysis.

Frey, H.C., and S.R. Patil. 2002. Identification and Review of Sensitivity Analysis Methods. *Risk Analysis* 22(3): 553–578.

Funtowicz, S.O., and J.R. Ravetz. 1990. *Uncertainty and Quality in Science for Policy*. Dordrecht, the Netherlands: Kluwer Academic.

Gentle, J.E. 1998. *Numerical Linear Algebra for Applications in Statistics*. New York, NY: Springer-Verlag.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. New York, NY: Chapman and Hall.

Goodman, D. 2002. Extrapolation in Risk Assessment: Improving the Quantification of Uncertainty, and Improving Information to Reduce the Uncertainty. *Human and Ecological Risk Assessment* 8(1): 177–192.

Gray, G.M., J.C. Allen, D.E. Burmaster, S.H. Gage, J.K. Hammitt, S. Kaplan, R.L. Keeney, J.G. Morse, D.W. North, J.P. Nyrop, A. Stahevitch, and R. Williams. 1998. Principles for Conduct of Pest Risk Analyses: Report of an Expert Workshop. *Risk Analysis* 18(6): 773–780.

Gronewold, A., and M. Sonnenschein. 1998. Event-Based Modelling of Ecological Systems with Asynchronous Cellular Automata. *Ecological Modelling* 108: 37–52.

Gurian, P.L., M. Small, J. Lockwood, and M. Schervish. 2000. Simulating Variability and Uncertainty for a Regulatory Impact Assessment of Arsenic in Drinking Water. Presented at the 32nd Annual Mid-Atlantic Industrial and Hazardous Waste Conference, Rensselaer Polytechnic Institute, Troy, NY, June.

Haimes, Y.Y. 2004. *Risk Modeling, Assessment, and Management*. Hoboken, NJ: Wiley.

Halton, J.H. 1994. Sequential Monte Carlo Techniques for the Solution of Linear Systems. *Journal of Scientific Computing* 9(2): 213–257.

Hamed, M.M. 1999. Probabilistic Sensitivity Analysis of Public Health Risk Assessment from Contaminated Soil. *Journal of Soil Contamination* 8(3): 285–306.

Hamed, M.M. 2000. Impact of Random Variables Probability Distribution on Public Health Risk Assessment from Contaminated Soil. *Journal of Soil Contamination* 9(2): 99–117.

Helton, J.C., and F.J. Davis. 2000. Sampling-Based Methods for Uncertainty and Sensitivity Analysis. In *Sensitivity Analysis,* edited by A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley.

Helton, J.C., and F.J. Davis. 2002. Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. *Risk Analysis* 22(3): 591–622.

Helton, J.C., and F.J. Davis. 2003. Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems. *Reliability Engineering and Systems Safety* 81(23–69).

Hogeweg, P. 1988. Cellular Automata as a Paradigm for Ecological Modeling. *Applied Mathematics and Computation* 27: 81–100.

Howard, R. 1988. Decision Analysis: Practice and Promise. *Management Science* 34(6): 693–697.

Iman, R.L., and W.J. Conover. 1979. The Use of the Rank Transform in Regression. *Technometrics* 4: 499–509.

IEc (Industrial Economics, Inc.) 2004. *An Expert Judgment Assessment of the Concentration–Response Relationship between PM2.5 Exposure and Mortality.* Cambridge, MA: Prepared for U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards.

———. 2005. *Draft Protocol for Full-Scale Implementation of Expert Elicitation to Evaluate the Concentration-Response Relationship Between Long-Term Ambient PM2.5 Exposures and Mortality*. Cambridge, MA: Prepared for U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards.

Jones, R.N. 2000. Managing Uncertainties in Climate Change Projections: Issues for Impact Assessment—An Editorial Comment. *Climatic Change* 45(3–4): 403–419.

Jörgensen, S.E. 1999. State-of-the-Art of Ecological Modelling with Emphasis On Development of Structural Dynamic Models. *Ecological Modelling* 120: 75–96.

Kahneman, D., and A. Tversky. 1982. Variants of Uncertainty. *Cognition* 11: 143–157.

Kalagnanam, J.R., and U.M. Diwekar. 1997. An Efficient Sampling Technique for Off-line Quality Control. *Technometrics* 39: 308–319.

Kann, A., and J.P. Weyant. 2000. Approaches for Performing Uncertainty Analysis in Large-Scale Energy/Economic Policy Models. *Environmental Modeling and Assessment* 5(1): 29–46.

Kleijnen, J.P.C., and J.C. Helton. 1999a. Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations, 1: Review and Comparison of Techniques. *Reliability Engineering and Systems Safety* 65(2): 147–185.

Kleijnen, J.P.C., and J.C. Helton. 1999b. Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations, 2: Robustness of Techniques. *Reliability Engineering and Systems Safety* 65(2): 187–197.

Klir, G.J., and T.A. Folger. 1988. *Fuzzy Sets, Uncertainty, and Information.* Prentice Hall: Englewood Cliffs, NJ.

Koop, G., and L. Tole. 2004. Measuring the Health Effects of Air Pollution: To What Extent Can We Really Say That People Are Dying from Bad Air? *Journal of Environmental Economics and Management* 47: 30–54.

Kopp, R.J., A.J. Krupnick, and M. Toman. 1997. Cost-Benefit Analysis and Regulatory Reform: An Assessment of the Science and the Art. Washington, DC: Resources for the Future.

Law, A.M., and W.D. Kelton. 1991. *Simulation Modeling and Analysis, 2nd edition*. New York, McGraw-Hill.

Lek, S., and J.F. Guegan. 1999. Artificial Neural Networks as a Tool in Ecological Modelling, an Introduction, *Ecological Modelling* 120: 65–73.

Linkov, I., and D. Burmistrov. 2003. Model Uncertainty and Choices Made by Modelers: Lessons Learned from the International Atomic Energy Agency Model Intercomparisons. *Risk Analysis* 23(6): 1297–1308.

Linkov, I., A. Varghese, S. Jamil, T.P. Seager, G. Kiker, and T. Bridges. 2004. Multi-criteria Decision Analysis: A Framework For Structuring Remedial Decisions at Contaminated Sites. In *Comparative Risk Assessment and Environmental Decision Making,* edited by I. Linkov and A. Ramadan. Dordrecht, the Netherlands: Kluwer Academic Press, 15–54.

Macauley, M.K. 2005. *The Value of Information: A Background Paper on Measuring the Contribution of Space-Derived Earth Science Data to National Resource Management.* Washington, DC: Resources for the Future.

Maier, H.R., and G.C. Dandy. 2000. Neural Networks for the Prediction and Forecasting of Water Resources Variables: A Review of Modelling Issues And Applications. *Environmental Modelling and Software* 15(1): 101–124.

Malakoff, D. 1999. Bayes Offers a "New" Way to Make Sense of Numbers. *Science* 286: 1460–1464.

Marcot, B.G., R.S. Holthausen, M.G. Raphael, M.M. Rowland, and M.J. Wisdom. 2001. Using Bayesian Belief Networks to Evaluate Fish and Wildlife Population Viability under Land Management Alternatives from an Environmental Impact Statement. *Forest Ecology and Management* 153: 29–42.

Matthews, H.S. 2001. Analysis of the Benefits and Costs of Clean Air. In *Improving Regulation: Cases in Environment, Health, and Safety,* edited by P.S. Fischbeck and R.S. Farrow. Washington, DC: RFF Press.

McKay, M.D., W.J. Conover, and R.J. Beckman. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21: 239–245.

Merrifield, J. 1997. Sensitivity Analysis in Benefit–Cost Analysis: A Key to Increased Use and Acceptance. *Contemporary Economic Policy* 15: 82–92.

Molofsky, J., and J.D. Bever. 2004. A New Kind of Ecology? *BioScience* 54(5): 440–446.

Morgan, M.G. 1998. Uncertainty Analysis in Risk Assessment. *Human and Ecological Risk Assessment* 4(1): 25–39.

Morgan, M.G., and H. Dowlatabadi. 1996. Learning from Integrated Assessment of Climate Change. *Climatic Change* 34: 337–368.

Morgan, M.G., and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY, Cambridge University Press.

Morokoff, W.J., and R.E. Caflisch. 1994. Quasi-random Sequences and Their Discrepancies. *SIAM Journal on Scientific Computing.* 15: 1251–1279.

Moxey, L.M., and A.J. Sanford. 2000. Communicating Quantities: A Review of Psycholinguistic Evidence of How Expressions Determine Perspectives. *Applied Cognitive Psychology* 14(3): 237–255.

Muttil, N., and J.H.W. Lee. 2005. Genetic Programming for Analysis and Real-Time Prediction of Coastal Algal Blooms. *Ecological Modelling* 189(3–4): 363–376.

NAS (National Academies of Science). 2005. *Environmental Decision Making: Principles and Criteria for Models.* Project BEST-K-03–02-A. http://www4.nas.edu/webcr.nsf/0/a507235e9387004485256e01005f9324 (accessed February 7, 2006).

NCRP (National Council on Radiation Protection and Measurements) 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination.* Bethesda, MD: NCRP.

NRC (National Research Council). 1983. *Risk Assessment in the Federal Government: Managing the Process*. Washington, DC: National Academies Press.

———. 1994. *Science and Judgment in Risk Assessment*. Washington, DC: National Academies Press.

———. 1996. *Understanding Risk: Informing Decisions in a Democratic Society*. Washington, DC: National Academies Press.

———. 2002. *Estimating the Public Health Benefits of Proposed Air Pollution Regulations.* Washington, DC: National Academies Press.

Nayak, T., and S. Kundu. 2001. Calculating and Describing Uncertainty in Risk Assessment: The Bayesian Approach. *Human and Ecological risk Assessment* 7(2): 307–328.

Nayak, T., and S. Kundu. 2003. Correlations between Parameters in Risk Models: Estimation and Propagation of Uncertainty by Markov Chain Monte Carlo. *Risk Analysis* 23(6): 1165–1172.

Neter, J., M.H. Kutner, C.J. Nachtsheim, and W. Wasserman. 1996. *Applied Linear Statistical Models* (4th edition). Chicago, IL: Irwin.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Nordhaus, W.D., and D. Popp. 1997. What Is the Value of Scientific Knowledge? An Application to Global Warming Using the PRICE Model. *The Energy Journal* 18(1): 1–45.

OMB (Office of Management and Budget). 2003. *Circular A-4.* Washington, DC: OMB.

OSTP (Office of Science and Technology Policy). 1985. Chemical Carcinogens: A Review of the Science and Its Associated Principles. *Federal Register* 50: 10372–10442.

Olson, M.J., and D.V. Budescu. 1997. Patterns of Preference for Numerical and Verbal Probabilities. *Journal of Behavioral Decision Making* 10(2): 117–131.

Parsons, D.J., T.G. Orton, J. D'Souza, A. Moore, R. Jones, and C.E.R. Dodd. 2005. A Comparison of Three Modelling Approaches for Quantitative Risk Assessment Using the Case Study of *Salmonella* spp. in Poultry Meat. *International Journal of Food Microbiology* 98(1): 35–51.

Paté-Cornell, E. 1996. Uncertainties in Risk Analysis: Six Levels of Treatment. *Reliability Engineering & System Safety* 54(2–3): 95–111.

Paté-Cornell, E. 2002. Risk and Uncertainty Analysis in Government Safety Decisions. *Risk Analysis* 22(3): 633–646.

Pearl, J., and S. Russell. 2000. *Bayesian Networks.* Technical Report (R-277). Los Angeles, CA: University of California–Los Angeles Cognitive Systems Laboratory.

Pope, C.A., III, R.T. Burnett, M.J. Thun, E.E. Calle, D. Krewski, K. Ito, and G.D. Thurston. 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution. *Journal of the American Medical Association* 287(9): 1132–1141.

RTI (RTI International). 2004. *Peer Review of "An Expert Judgment Assessment of the Concentration-Response Relationship Between* $PM_{2.5}$ *Exposure and Mortality*." Research Triangle Park, NC: Prepared for U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards.

Rai, S.N., and D. Krewski. 1998. Analysis of Uncertainty and Variability in Multiplicative Risk Models. *Risk Analysis* 18: 37–45.

Rai, S.N., D. Krewski, and S. Bartlett. 1996. A General Framework for the Analysis of Uncertainty and Variability in Risk Assessment. *Human and Ecological Risk Assessment* 2: 972–989.

Reckhow, K.H. 2003a. Bayesian Approaches in Ecological Analysis and Modeling. In *Models in Ecosystem Science,* edited by C.D. Canham, J.J. Cole, and W.K. Lauenroth. Princeton, NJ: Princeton University Press, 168–183.

Reckhow, K.H. 2003b. *Scientific Treatment of Uncertainty in Environmental Models*. Environmental Models: Uncertainties in Law and Science, Washington, DC: Woodrow Wilson International Center for Scholars.

Regan, H.M., M. Colyvan, and M.A. Burgman. 2002. A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology. *Ecological Applications* 12(2): 618–628.

Renooij, S., and C.L.M. Witteman. 1999. Talking Probabilities: Communicating Probabilistic Information in Words and Numbers. *International Journal of Approximate Reasoning* 22(3): 169–194.

Reynolds, J.H., and E.D. Ford. 1999. Multi-Criteria Assessment of Ecological Process Models. *Ecology* 80(2): 538–553.

Rowe, W. 1994. Understanding Uncertainty. *Risk Analysis* 14(743–50).

Rubinstein, R.Y. 1981. *Simulation and the Monte Carlo Method*. New York, NY: Wiley.

Saltelli, A. 2002. Sensitivity Analysis for Importance Assessment. *Risk Analysis* 22(3): 579–590.

Saltelli, A., K. Chan, and E.M. Scott (Eds.). 2000. *Sensitivity Analysis*. New York, NY: Wiley.

Saltelli, A., S. Tarantola, and F. Campolongo. 2000. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science* 15(4): 377–395.

Schneider, S.H., B.L. Turner II, and H.M. Garriga. 1998. Imaginable Surprise in Global Change Science. *Journal of Risk Research* 1(2): 165–185.

Silvert, W. 1997. Ecological Impact Classification with Fuzzy Sets. *Ecological Modeling* 96: 1–10.

Stiber, N.A., M. Pantazidou, and M.J. Small. 1999. Expert System Methodology for Evaluating Reductive Dechlorination at TCE Sites. *Environmental Science and Technology* 33(17): 3012–3020.

Tufte, E.R. 1991. *The Visual Display of Quantitative Information* (2nd ed.). Cheshire, CT: Graphics Press.

Tversky, A., and D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185: 1124–1131.

U.S. EPA (U.S. Environmental Protection Agency). 1986. *Guidelines for Carcinogen Risk Assessment.* Washington, DC: U.S. EPA.

———. 1989. *Risk Assessment Guidance for Superfund (RAGS). Volume I: Human Health Evaluation Manual. Part A.* Washington, DC: U.S. EPA, Office of Emergency and Remedial Response.

———. 1992. *Guidelines for Exposure Assessment.* Washington, DC: U.S. EPA, Risk Assessment Forum.

———. 1995. *Guidelines for Risk Characterization.* Washington, DC: U.S. EPA, Science Policy Council.

———. 1996a. *Regulatory Impact Analysis of* NO_x *Regulations.* Washington, DC: U.S. EPA, Office of Atmospheric and Indoor Air Programs.

———. 1997a. *Exposure Factors Handbook.* Washington, DC: U.S. EPA, National Center for Environmental Assessment.

———. 1997b. *Guiding Principles for Monte Carlo Analysis.* Washington, DC: U.S. EPA, Risk Assessment Forum, Office of Research and Development.

———. 1997c. *Policy for Use of Probabilistic Analysis in Risk Assessment at the U.S. Environmental Protection Agency.* Washington, DC: U.S. EPA, Science Policy Council. Fred Hansen, Deputy Administrator.

———. 1998a. *Guidelines for Ecological Risk Assessment.* Washington, DC: U.S. EPA, Risk Assessment Forum.

———. 1998b. *Regulatory Impact Analysis for the NO_x SIP Call, FIP, and Section 126 Petitions.* Washington, DC: U.S. EPA, Office of Air and Radiation.

———. 1999a. *The Benefits and Costs of the Clean Air Act 1990 to 2010.* Washington, DC: U.S. EPA, Office of Air and Radiation, Office of Policy.

———. 1999b. *Final Tier 2 Rule: Air Quality Estimation, Selected Health and Welfare Benefits Methods, and Benefit Analysis Results.* Washington, DC: U.S. EPA, Office of Air Quality Planning and Standards, Office of Air and Radiation.

———. 1999c. *Guidelines for Carcinogen Risk Assessment.* [Revised Draft]. Washington, DC: U.S. EPA, Office of Research and Development: 211.

———. 1999d. *Regulatory Impact Analysis for the Final Regional Haze Rule.* Washington, DC: U.S. EPA, Office of Air Quality Planning and Standards, Office of Air and Radiation.

———. 1999e. *Regulatory Impact Analysis for the Final Section 126 Petition Rule.* Washington, DC: U.S. EPA, Office of Air and Radiation.

———. 2000a. *Guidelines for Preparing Economic Analyses.* Washington, DC: U.S. EPA, National Center for Environmental Economics.

———. 2000b. *An SAB Advisory on the U.S. EPA's Draft Case Study Analysis of the Residual Risk of Secondary Lead Smelters.* Washington, DC: U.S. EPA, Science Advisory Board.

———. 2001a. *Risk Assessment Guidance for Superfund. Volume I: Human Health Evaluation Manual. Part D: Standardized Planning, Reporting, and Review of Superfund Risk Assessments.* [Final]. Washington, DC: U.S. EPA, Office of Solid Waste and Emergency Response.

———. 2001b. *Risk Assessment Guidance for Superfund. Volume III, Part A: Process for Conducting Probabilistic Risk Assessment.* Washington, DC: U.S. EPA, Office of Emergency and Remedial Response.

———. 2003a. *The Clear Skies Act Technical Support Package.* Washington, DC.

———. 2003b. *Draft Final Guidelines for Carcinogen Risk Assessment.* Washington, DC: U.S. EPA, Risk Assessment Forum.

———. 2004a. *Advisory on Plans for Health Effects Analysis in the Analytical Plan for EPA's Second Prospective Analysis – Benefits and Costs of the Clean Air Act, 1990-2020.* Washington, DC: U.S. Environmental Protection Agency, Science Advisory Board.

———. 2004b. *Benefits of the Proposed Inter-State Air Quality Rule.* Washington, DC: U.S. EPA.

———. 2004c. *An Examination of EPA Risk Assessment Principles and Practices.* Washington, DC: U.S. EPA, Risk Assessment Task Force.

———. 2004d. *Final Regulatory Impact Analysis: Control of Emissions from Nonroad Diesel Engines.* EPA420-R-04-007. Washington, DC: U.S. EPA, Office of Transportation and Air Quality.

———. 2005a. *Regulatory Impact Analysis for the Final Clean Air Interstate Rule.* Washington, DC: U.S. EPA, Office of Air and Radiation.

———. 2005b. *Regulatory Impact Analysis of the Clean Air Mercury Rule.* Washington, DC: U.S. EPA, Office of Air Quality Planning and Standards.

———. 2005c. *Symposium on Use of Expert Elicitation to Evaluate the Concentration-Response Relationship Between Long-Term Ambient $PM_{2.5}$ Exposures and Mortality.* April 20, Baltimore, MD. http://www.indecon.com/iec_web/event/expertjudgment.asp (accessed February 7, 2006).

van Asselt, M. 1999. *Uncertainty in Decision Support: From Problem to Challenge.* Maastricht, Netherlands: University of Maastricht, International Centre for Integrative Studies (ICIS).

van der Fels-Klerx, H.J., L.H.J. Goossens, M.J. Nauta, R.M. Cooke, and A H. Havelaar. 2005. A Structured Expert Judgement Study for Microbiological Risk Assessment in *Campylobacter* Transmission during Broiler Chicken Processing. *Risk Analysis* 25: 109–124.

Walker, K.D., P. Catalano, J.K. Hammitt, and J.S. Evans. 2002. Use of Expert Judgment in Exposure Assessment. Part 2: Calibration of Expert Judgments about Personal Exposures to Benzene. *Journal of Exposure Analysis and Environmental Epidemiology* 13(1): 1–16.

Walker, K.D., J.S. Evans, and D. MacIntosh. 2001. Use of Expert Judgment in Exposure Assessment. Part 1: Characterization of Personal Exposure to Benzene. *Journal of Exposure Analysis and Environmental Epidemiology* 11(4): 308–322.

Winkler, R.L. 1996. Uncertainty in Probabilistic Risk Assessment. *Reliability Engineering and Systems Safety* 54(2–3): 127–132.

Wolfram S. 1986. *Theory and Applications of Cellular Automata.* Singapore: World Scientific Publishing.

Wright, J., L. Wiggins, R. Jain, and T. John Kim (Eds.) 1993. *Expert Systems in Environmental Planning.* New York, NY: Springer-Verlag.

Wynne, B. 1992. Uncertainty and Environmental Learning: Reconceiving Science and Policy in the Preventive Paradigm. *Global Environmental Change* 2: 111–127.

Yokota, F., and K.M. Thompson. 2004. Value of Information Analysis in Environmental Health Risk Management Decisions: Past, Present, and Future. *Risk Analysis* 24(3): 635–650.

Zadeh, L.A. 1973. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3(1): 28–44.

Zhu, X.X., and A.R. Simpson. 1996. Expert System for Water Treatment Plant Operation. *Journal of Environmental Engineering* 122(9): 822–829.

Chapter 3: Case Study

In Chapter 2, we identified modeling and analytical issues associated with incorporating uncertainties into regulatory impact analyses (RIAs). It is our contention—and a requirement by the Office of Management and Budget (OMB)—that there are productive opportunities to expand current efforts in the types of uncertainties examined, the techniques used to examine them, and the way these uncertainties are described and communicated.

In this chapter, we use a case study to undergird our primarily methodological investigation into the formulation, implementation, and reporting of uncertainties in a RIA. Our case study is a tightening of the cap on electric utility emissions of nitrogen oxides (NO_x) beyond that required by the Clean Air Interstate Rule (CAIR). Through this analysis we open several new fronts in the consideration of uncertainties, specifically, on the cost side of the ledger and on the population component of an RIA (which enters on both the cost and the benefit sides). At the same time, we build on many of the ideas and concepts presented in Chapter 2 for analyzing uncertainties.

We examine parameter (or statistical) uncertainty and model uncertainty and we address decision uncertainty through our discussion of alternative metrics. We also analyze some of these uncertainties directly with Monte Carlo simulation, whereas some we address only though scenario analysis. In choosing how to analyze these uncertainties, we take some of the ideas from Chapter 2, such as an approach to classifying uncertainties (Haimes 2004), which helps to identify new types of uncertainties for possible analysis within this case study. Some of the ideas in Chapter 2 we do not implement, primarily because of the complexities and expense involved in some of these techniques (e.g., performing an expert elicitation) and the inappropriateness of the techniques for our particular case. One example of the latter is the analysis of joint or correlated distributions. In our case study, all the uncertainties are (to at least at first approximation) independent of one another, so analyzing dependent distributions is inappropriate.

Although the primary purpose of this chapter is to demonstrate methodologies, the case study has intrinsic interest for policymakers and can serve as a platform for illuminating how incorporating uncertainty into a cost–benefit analysis enriches an analysis but complicates the presentation of findings to decisionmakers and others. Accordingly, one of the final sections of this chapter summarizes our results, emphasizing the implications of these results for the policy choice.

The final section takes a step back from the case study to distill methodological lessons learned for doing cost–benefit analysis with uncertainties.

Choice of Case Study

Why Air Pollution?

OMB's *Circular A-4* calls for Monte Carlo simulation to be added to standard RIA for rules with economic costs exceeding $1 billion annually. The U.S. Environmental Protection Agency (EPA) promulgates very few rules that have this kind of impact—air pollution regulations being one prominent set. Furthermore, EPA and the authors of this report have extensive knowledge of the existing and proposed regulations in this area and have developed techniques and inputs to such analyses as part of their professional activities.

Moreover, the modeling infrastructure exists at Resources for the Future (RFF) to carry out such an analysis, indeed, even duplicate many of the elements of EPA's own analyses of real rules. Finally, several team members have been members of the Science Advisory Board committees overseeing the so-called 812 Study (U.S. EPA 1997, 1999), which is the most detailed and scrutinized study of the costs and benefits if any set of regulations EPA ever performed. Thus, in short, a case study on air pollution represents the best case for which such a study could be conducted.

Why Nitrogen Dioxide?

The air pollutants of concern are most immediately confined to the six "criteria" pollutants under the Clean Air Act (CAA)—sulfur dioxide (SO_2), ozone, particulates, carbon monoxide, lead, and NO_x. Of these, the major problems remaining are primarily with ozone and particulates, and SO_2 is a contributor to the latter and NO_x a contributor to both.

Because ozone is a secondary pollutant and particulate matter is primarily a secondary pollutant, concerns focus on SO_2 and NO_x. These concerns are reflected in the SO_2 Allowance Trading program institutionalized in the CAA Amendments of 1990, the NO_x SIP Call, and the CAIR for both pollutants. Of these two concerns, we chose NO_x control as the object of our study for methodological reasons. Recent RIAs (U.S. EPA 2004) have shown that the benefits of reducing NO_x are close to the costs, so we thought that for this case, uncertainty considerations could be important for describing outcomes to decisionmakers.

Why Electric Utilities?

With NO_x the pollutant of choice, two major sources could be objects of the case study: electric utilities and mobile sources. We could have chosen either. However, we believed that utilities could be most easily and robustly modeled, given that RFF's Haiku model is already available and well-known for addressing regulatory issues associated with air pollution reductions.

Why a Tighter NO_x Cap with CAIR as the Baseline?

Following the hugely successful SO_2 Allowance Trading program and an increasingly market-oriented approach to government regulation, cap-and-trade programs have become a preferred strategy for point-source pollution control at EPA. Therefore, it is fairly unrealistic to expect that EPA would adopt another type of strategy for future air pollution reductions.

With this in mind, we resolved to model the effects of a NO_x cap for the case study. But relative to what baseline? We certainly did not want to revisit an analysis of CAIR, which would be duplicative of EPA work and perhaps controversial, because this rule is far from "history." Thus, the obvious baseline is CAIR itself (described in more detail in Appendix 3A).

Next, we needed only to decide how tight to make the hypothetical new NO_x cap. We chose a limit based on plausibility first. This NO_x reduction is about 40%, leaving residual emissions in 2025 of about 1.5 million tons from a baseline of 2.4 million tons. This level is comparable to that in Senate Bill 150, proposed by Senator Jim Jeffords (I-VT) in the 109th Congress, and it is the most stringent contained in recent legislative proposals. We call this case the Large case. Because we believe that a realistic case study requires another cap as an option, we also modeled a cap of 1.95 million tons (about a 20% reduction)referred to as the Moderate case.

The choice to model a cap-and-trade policy leads to one very important simplification for the case study. For any given cap, irrespective of the types of analyses we perform, the estimated NO_x reductions will be more or less the same. Thus, when we introduce a scenario with high natural gas prices or low population relative to the base case, there will be only minor differences in NO_x reductions. This is a consequence of there being a cap on NO_x in the baseline and a cap on NO_x for the control scenario. The cap in the baseline is a regional cap specified under CAIR, but it affects the majority of NO_x emissions nationally. This feature—that aggregate emissions reductions are fairly constant across scenarios—means that, for the most part, differences in benefit estimates across scenarios arise because of differences in the spatial and temporal (seasonal) distribution of NO_x reductions. Of course, differences in NO_x emissions do arise between the Large and Moderate cases.

Addressing Uncertainty in the Case Study

Overall, we incorporated uncertainties in our case study on the basis of tradition, feasibility, and interest. For instance, the tradition in RIAs of this type is to incorporate statistical uncertainties in concentration–response (C–R) and valuation functions. Beyond these, it was feasible (and novel) for us to incorporate source–receptor (S–R) coefficient uncertainty because we had results of two air quality models available. It was also feasible to examine uncertainty in the natural gas price path within the Haiku model. Although all these uncertainties may be interesting in and of themselves, population uncertainties are incorporated for their feasibility, novelty, and interest. They are particularly interesting because they affect both cost and benefit sides of the case study.

Following the discussion of alternative approaches to modeling uncertainties in Chapter 2, we addressed these uncertainties using both parametric approaches to capture statistical uncertainties and nonparametric approaches to capture model uncertainties. We examined statistical uncertainties using Monte Carlo techniques within the Tracking Analysis Framework (TAF) benefits model for S–R relationships, C–R functions, and values. Uncertainties in natural gas prices and population were examined using scenario analysis by specifying natural gas price paths of 30% and 70% above and 30% below the base case price path and by denoting a high and low population case, respectively (on the basis of U.S. Census Bureau [2000, 2004, 2005] data). In addition, the S–R relationships were swapped between those derived from two models (Urban-to-Regional Multiscale [URM] and Advanced Source Trajectory Regional Air Pollution [ASTRAP]) to address model uncertainty in this component. Details are provided below.

Using the Uncertainty Typology

In Chapter 2 we presented a detailed typology of uncertainties, which is summarized on the right side of Figure 3-1. Our goal was to use this typology in two ways. The first was to classify the uncertainties we chose to examine (top left of Figure 3-1) for reasons of tradition, feasibility, and interest. Such classifications can lead to a better explication of the nature of the uncertainties being examined. The second was to use the typology to identify areas of uncertainty that our case study ignored (bottom left of Figure 3-1). This list is only illustrative.

Turning to the classification of uncertainties we examined, we used a natural gas price path through 2025 developed by the Department of Energy for the base case (EIA 2005a). The process for developing this price path is not known to us exactly. If this is an output of an energy market simulation model, such a model would have a baseline, subject to measurement error. Conflicting data may have been incorporated in the simulation model, and the model itself would have extrapolation error, because the price path is being predicted into the future (temporal prediction error). This simulation model would have many structural decisions embedded in it and would, of course, simplify the real world. Thus, most of the uncertainties under parameter and model uncertainty would be represented.

On top of these uncertainties underlying the base case natural gas price path, uncertainties are associated with our ad hoc method for raising and lowering this path to introduce “statistical” uncertainties. That is, we simply raised and lowered the predicted price path by a given percentage to apply a scenario-based approach to representing uncertainties.

The story for population uncertainty is virtually the same. We used U.S. Census Bureau (2000, 2004, 2005) estimates of national population growth paths for low- and high-population cases as well as the base case. We introduced ad hoc uncertainty from our construction of low and high population series at the state level. Note that linguistic uncertainties (discussed in Chapter 2) are also introduced here, where the U.S. Census

Bureau has labeled the alternative population scenarios Low and High. These terms are ambiguous and can be interpreted subjectively. The low and high series actually represent extreme values or bounds for expected future population (Hollman et al. 2000).

The S–R coefficients have very different uncertainty classifications depending on whether they came from URM or ASTRAP models. However, both models suffer from uncertainties introduced from the variability of weather and the need to develop S–R relationships that represent "average" weather patterns, however defined. The existence of these two models also implies model uncertainty. The URM coefficients come from a simulation model, whereas the ASTRAP coefficients come from a statistical model. The URM coefficients suffer from various parameter uncertainties, such as measurement errors for the rate constants that govern transformations of emissions into concentrations. Also, the underlying data on baseline emissions is subject to conflicting data, extrapolation, and potentially surrogate data uncertainties.

The ASTRAP S–R coefficients are estimated using a statistical model and thus incorporate statistical variation. They may include systematic biases from modeling choices and certainly suffer from a lack of monitoring data that is needed to confidently relate emissions to concentrations. The reduced-form model itself suffers from simplification and incompleteness, to a much greater degree than URM.

Finally, the C–R coefficients and unit values are mostly derived from complex statistical analyses. However, some of these parameters result from mere guesswork. The C–R coefficients originate from epidemiological studies, which have random error and statistical variation, but we hope that they are not subject to systematic bias. Most of these coefficients come from expensive and extensive studies, so problems of data conflicts and gaps may not be too significant. Random sampling error is likely to be a problem for many of the studies, where sampling may not be random. Model uncertainties abound, as well. Structural choices (e.g., the specification of the underlying estimation model) introduce uncertainty, as does the inevitable simplification of the conceptual model to one that can be estimated. These comments hold just as well for the valuation estimates.

We now use the typology to identify uncertainties we ignored. Beginning with Variability and Parameter Uncertainty, we find a host of uncertainties involving the models underlying the S–R relationships. In scaling or summarizing these relationships to represent average relationships over a year or a season, a potentially large degree of uncertainty is introduced by the high degree of variability in weather variables.

However, the most prominent uncertainty is associated with rate constants. These govern the chemical transformations of pollutants in the atmosphere and are subject to parameter uncertainties of various types, including measurement error. Some of these uncertainties are captured in our model but by no means all, as described below.

Parameter uncertainty also reminds us about the errors that could underlie baseline health risk estimates. They may be subject to measurement or extrapolation errors. However, we suspect these errors are relatively small, particularly for the broader, more aggregate measures. But as the health effect narrows or applies to a sensitive group, such errors may grow. For instance, misclassification errors in cause-of-death attribution would rise as the health endpoint becomes more targeted.

Turning to Model Uncertainty, the term *Structural Choices* emphasizes one of the key assumptions made in our benefits model: that nitrates, which are formed from the conversion of NO_x emissions in the atmosphere and count as a fine particulate smaller than 2.5 microns ($PM_{2.5}$) are as potent as the average particle counted in particulate matter smaller than 10 microns (PM_{10}). The literature is sparse on C–R relationships involving nitrates, whereas several epidemiological studies implicate sulfates (another $PM_{2.5}$ component) in health effects. Had we assumed that nitrates were as potent to health as sulfates, or even as the average $PM_{2.5}$ particle, our estimated health effects would have been larger.

Structural Choices further remind us that we had no choice (given the available literature) but to use cost-of-illness estimates to monetize some health endpoints rather than the preferred "model" based on willingness to pay. We thereby introduce error and probably a downward bias to the benefit estimates.

Additional model uncertainty is identified using several of the remaining terms in Figure 3-1. "Incompleteness" highlights the many types of health effects that have been identified or suggested as arising from exposure to air pollutants. Because of a lack of data and literature, some of these effects are not included in our model.

The Choice of Probability Distribution reminds us that the distributions underlying our Monte Carlo analysis are for the most part assumed rather than optimized against the data or assigned on other, more solid grounds. It especially applies to the health effects and valuation estimates.

Correlations and Dependencies reflects the possibility that uncertainties in one parameter are correlated with those of another parameter. If these dependencies are not taken into account, the "daughter" distributions (say, the distribution of total benefits) will be wrong and probably biased. We feel that for our simple benefit model, there are unlikely to be dependencies across model components. However, dependencies might exist across distributions of different functions within each component. For example, a high willingness-to-pay for a reduction in chronic bronchitis cases may be correlated with a high willingness-to-pay for a reduction in other morbidity effects. Similar reasoning can be applied to the C–R functions.

The final set of uncertainties arises from the Haiku model. Because the issues of uncertainty in costs are so often ignored, we use more space exploring them but even so only give some examples of the types of uncertainties ignored by this complex and seemingly complete model. These examples were developed from an interview with Karen Palmer, one of the model's creators, using the typology in Chapter 2. In general, the typology was useful in eliciting some types of model uncertainties that Palmer would have missed otherwise.

Starting at the top of the typology, an example of uncertainty resulting from variability is that created by ignoring plant heterogeneity. The model uses a model plant approach at the North American Electric Reliability Council (NERC) region level to embody the average characteristics of plants sharing common technology and vintage in that region. The variability of the characteristics is partially preserved in a variable cost schedule that represents the variability of the constituent plants represented by the model plant, but some of the variability is lost in the averaging process and this introduces uncertainty, although not obvious bias, into the results.

With respect to parameter uncertainties, even though we altered natural gas price paths in our case study, the Haiku model actually predicts prices internally (we then multiplied such prices by an adjustment factor). This prediction is made from a regression equation on the basis of results of simulation modeling by the Energy Information Administration (EIA). This regression has error associated with it, so it is an example of statistical uncertainty and measurement error.

A good example of uncertainties arising from conflicting data are the electricity demand elasticities used in the model. They come from a major literature review of studies deriving such elasticities. That review contained studies with many different types of results, but one had to be chosen.

The Haiku model itself, as a forecast model, is inherently subject to various extrapolation errors. Data at the plant level are becoming less representative, and the need to use surrogate data is growing as plants are divested out of the formal utility sector. Only utilities need fill out the detailed data survey called FERC Form 1, which is the source of much of Haiku's data. As nonutility companies purchase plants, FERC Form 1 data becomes less and less representative.

Misclassification error can be found in the classification of dual-fueled turbines (generators that run on oil or natural gas), whose capacity is split to match recent generation mix by fuel type, rather than letting this fraction be determined by the model.

In terms of model uncertainties, the model embodies many structural choices; for example, peak- and base-load demand are independent of one another. In reality, if peak load prices were to rise significantly, there would be shifts in appliance use in the short run and long-run adjustments to investments. The model currently ignores this, biasing costs upward.

Of the many simplifications, fixing the degree and cost of particulate controls—a hitherto unimportant simplification—has recently become more important with the focus on mercury emissions. Another is assuming that the allowance trading system operates without friction, which allows a least-cost optimization strategy to be used.

Incomplete Data is represented by the lack of data on intraregional transmission constraints. Such data are not generally available, although such information is widely available for interregional transmission lines.

A final uncertainty discussed here is that of the temporal resolution of the model. The model uses four time blocks in four seasons within which electricity decisions can be made. In Palmer's judgment, this degree of resolution is insignificant as a cause of uncertainty in the model. However, fewer blocks or seasons would impinge on the credibility of the Haiku effort.

Model Descriptions

We used two major models to implement the case study and used results from two more. The features of these models are described briefly here, but Appendices 3B and 3C include more details.

Haiku Electric Utility Model

The Haiku model simulates the behavior of the electric power sector in regional electricity markets and interregional electricity trade in a regulatory environment with emissions trading and public utility regulation. Utilities make both short-run decisions and long-run investment decisions to maximize long-run profits. These decisions include abatement, fuel use, and allowance permit-related decisions regarding SO_2, NO_x, and mercury. The model calculates electricity demand, electricity prices, the composition of technologies and fuels used to supply electricity, interregional electricity trading activity, and the emissions of key pollutants. It also calculates the costs for complying with environmental regulations and the welfare effects (producer plus consumer surplus) of environmental regulation. EPA's Integrated Planning Model (IPM) is a more elaborate model with very similar features to HAIKU.

TAF Benefits Model

The output of the Haiku model is emissions of each pollutant by a representative plant within each of 13 subregions of the United States. Changes in emissions of SO_2 and NO_x that result from a policy analysis are aggregated to the state level and fed into TAF, a nonproprietary and peer-reviewed integrated assessment model (Bloyd et al. 1996).[1] TAF integrates pollutant transport and deposition, population, human health effects, and valuation of these effects at the state level. In addition, it is designed to perform Monte Carlo simulation to address uncertainties in the above components.

TAF is similar to EPA's Benefits Mapping and Analysis Program (BenMAP) model (http://www.epa.gov/ttn/ecas/benmodels.html). Like BenMAP, TAF offers a library of models linking concentrations of pollutants to any given health effect (along with statistical uncertainties from such models) and includes a large suite of such effects. Its library also contains alternative monetary values for monetizing health improvements (along with their statistical uncertainties). It also contains S–R coefficients (which relate emissions to concentrations) drawn from two models (described below), whereas BenMAP incorporates one air model.

S–R Models

Two major modeling efforts are used to develop point estimates and uncertainty bounds for linking emissions to concentrations of fine particulates. The more advanced

[1] TAF was developed to support the National Acid Precipitation Assessment Program (NAPAP). Each module of TAF was constructed and refined by a group of experts in that field and draws primarily on peer-reviewed literature to construct the integrated model. TAF was subjected to an extensive peer review in December 1995, which concluded, "TAF represent[s] a major advancement in our ability to perform integrated assessments" (ORNL 1995). The entire model is available at www.lumina.com\taflist. TAF has been repeatedly updated and is maintained by RFF.

and accepted of the two is the URM model (described below), but this simulation model has the disadvantage of being parameterized only for the eastern United States. Thus, to be able to model national S–R relationships, in our base case we supplement this model with S–R relationships for the rest of the United States taken from the ASTRAP regression model (described below). Because ASTRAP provides S–R coefficients for the nation, the coefficients from the eastern United States can be compared with those from the URM model to examine the effect of model uncertainty, and ASTRAP coefficients can be used for the nation when we want to have a more internally consistent representation of S–R relationships.

The ASTRAP model has another important feature for our project. It provides "uncertainties" around each of the S–R coefficients, based on climatological variability, measured as statistical confidence intervals from the original regression analyses. The URM model does not have this feature. However, climatological variability is given expression through a novel scaling procedure that translates S–R information for three episodes into an annual S–R matrix. By varying parameters within this procedure randomly, we generate a set of S–R matrices spanning a reasonable range of climatological variability.

URM

The base case set of relationships between emissions in one area and concentrations in all areas (called S–R coefficients) for the eastern United States come from the URM One Atmosphere Model (Kumar et al. 1994). It includes the effects of changes in emissions of NO_x and SO_2 on fine particulate concentrations as well as the effects of changes in NO_x emissions on atmospheric ozone concentrations.

Uncertainties in such S–R coefficients are developed through a two-step process: first, scaling results of three distinct episodes of 6–9 days each to be representative of annual weather conditions using a Classification and Regression Tree (CART) approach (Deuel and Douglas 1998), then re-weighting the episodes in this scaling process to capture the uncertainty in such a scaling procedure, again using the CART analysis. Ultimately, 30 S–R matrices were developed to represent the climatological uncertainties introduced by the CART procedure. A median "potency" matrix was identified, with *potency* defined according to the aggregate effect of a 1000 tons of NO_x reductions in each state on population exposures in the URM domain. The development, scaling, and uncertainty built into these S–R coefficients are described in Appendix 3C.

ASTRAP

Although URM is a simulation model, the ASTRAP S–R matrices are produced using regression analysis from data on monitored concentrations and various climatological variables (wind, temperature, and precipitation) over 11 years. Regression estimates were developed for each of the four seasons. The model captures atmospheric chemistry as NO_x and SO_2 react to form nitrates and sulfates, which are constituents of particulate matter less than 10 microns in diameter (PM_{10}). It estimates concentrations of these separate constituents of PM_{10} plus gaseous NO_2 and SO_2.

The results were validated against ambient concentration and deposition data by using historical emissions data. Confidence intervals (assuming a normal distribution) around the estimated S–R coefficients were then incorporated into the TAF model to represent climatological variability. This version of the atmospheric transport module limits benefits to only particulate-related health impacts; however, according to the major integrated assessment studies of the impacts of electricity generation (Krupnick and Burtraw 1996), these impacts account for the vast majority of all benefits.

Modeling Uncertainty

In our approach to modeling uncertainty in Haiku, there were two possibilities: applying Monte Carlo simulation analyses to an existing sectoral model, or addressing uncertainty nonparametrically (i.e., with scenarios with high or low values for key inputs). The Monte Carlo approach is not currently feasible with Haiku because the model is highly parameterized and because the iteration algorithm that leads to convergence is too time-consuming. The scenario approach was therefore chosen, recognizing that this type of uncertainty treatment is not preferred.

Accordingly, we developed scenarios for two sets of inputs into the model. The first is based on alternative population forecasts (a high-population and a low-population forecast, relative to the base case forecast) and the second on alternative exogenous estimates for the supply schedule of natural gas prices. Alternative population forecasts are interesting for modeling uncertainty because they affect both benefits and costs. A unique contribution of this analysis is the consistent integration of the changes in benefits and costs under different population scenarios. Costs of meeting a tighter cap are affected because, for instance, rising population raises the demand for electricity. Greater electricity demand may raise electricity prices and alter investment and fuel use strategies and fuel costs, particularly vis-à-vis renewables. Because the quantity of emissions allowances is fixed under the cap, greater electricity generation increases the value of emissions allowances, which in turn affects compliance costs and social welfare (producer plus consumer surplus). Benefits are affected proportionally to population at whatever level of geographic detail is examined.

However, in general, a national population estimate that is x% higher than the base case will not result in benefits that are x% higher nationally. This is because the higher population estimates are aggregates of state estimates and the state-by-state distribution of population in the high and low national projections are different from each other and from the base case population projections. Moreover, changes in the technology profile of generation in response to increased demand will lead to geographic shifts in the location and timing of emissions.

Natural gas price is responsive to the quantity of electricity demand in the model, following a price schedule that is calibrated to the *Annual Energy Outlook 2005* (EIA 2005a). The schedule is adjusted by assuming that the price of natural gas for a given quantity of demand is higher or lower than it is in the base case. Higher natural gas prices encourage substitution away from natural gas and raise fuel costs in the aggregate, which also raises electricity prices, alters investment plans, and, ultimately, affects the price of NO_x allowances and compliance costs to meet the new tighter cap. Moreover, the change in electricity price differs by region of the country and season of the year. Consequently, a change in electricity price alters the profile of electricity demand and the mix of fuels in different ways by region and season. These changes in emissions are tracked in the benefits model in an integrated fashion.

Our approach to modeling uncertainties in the TAF model is more straightforward. The three population series are entered into TAF to scenario-based representations of uncertainty. The uncertainties in ASTRAP S–R coefficients are entered, cell by cell, along with standard errors and distributional assumptions (as taken from Shannon 1981, 1985). The 30 URM matrices generated by the CART procedure were all entered in TAF directly. Parameters for the C–R functions and valuation functions were entered in TAF along with the standard errors and distributional assumptions taken from the original studies.

Descriptions of the Case Study Simulations

Base Case

Modeling begins with development of the base case. The base case consists of three runs of our models (Haiku, TAF, and URM/ASTRAP) using a common default set of assumptions for the baseline and two policies (Large and Moderate, described earlier). These default assumptions cover choices of epidemiological studies, valuations of health endpoints, one set of population projections, one natural gas price scenario, the western ASTRAP S–R coefficients, and the median "potency" set of the URM S–R matrix (see Appendix 3C), which covers the eastern United States.

The costs of the policy can be measured in two ways. One is the change in expenditures: fuel costs, compliance costs, and investment. The second is economic welfare costs: the sum of producer plus consumer surplus. The difference in emissions between a given policy and the baseline, detailed at the state level, is then passed to the TAF model. The TAF model is designed to take changes in emissions from Haiku and convert them into monetary benefits and changes in health endpoints. The values used are a mix of expenditures (e.g., cost of illness estimates) and social welfare measures (e.g., the value of statistical life).

All TAF model base case runs include statistical uncertainties for C–R functions and values, plus the median S–R matrix (as defined above) and "middle" population estimates. Our standard TAF run produces 500 iterations for all probabilistic variables. Analytica, our modeling program, begins the randomizing process for each distribution with the use of a defined seed variable. When the seed variable is held constant for multiple runs, identical sample draws are obtained for each distribution that has not been adjusted or calculated from a distribution that has been altered between runs. Therefore, all TAF runs in this analysis use the same seed variable to ensure that scenarios are comparable.[2]

To summarize, the base case runs include

- Haiku: baseline CAIR,
- Haiku: policy implementing a tight NO_x cap (Large), and

[2] With the base case, we ran 1000 iterations to test how quickly the mean benefits converged (see below for results). This additional run required using a different seed variable.

- Haiku: policy implementing a less tight NO_x cap (Moderate).

The difference between baseline and policy runs produces costs and emissions changes. These emissions changes are fed into the TAF model. Base case assumptions for that model include

- URM: the median "potency" set of S–R coefficients (i.e., the median S–R matrix) for the eastern United States (see Appendix 3C),
- ASTRAP: S–R coefficients for the rest of the United States,
- U.S. Census Bureau (2005) middle population estimates (also used as input to the Haiku base case), and
- Statistical uncertainties in the default C–R functions and unit values.

Haiku Uncertainty Cases

Beyond the base case runs are runs to examine the effect on net benefits caused by altering some key Haiku data inputs to examine uncertainty via scenario analysis. Changes in these inputs must be made to both the baseline and policy cases.

- Haiku: High Population
- Haiku: Low population
- Haiku: High Natural Gas Price
- Haiku: Higher Natural Gas Price
- Haiku: Low Natural Gas Prices

Each of these five sets of results is fed into the TAF model. Although the TAF data and model assumptions are largely independent of the Haiku data and modeling assumptions, population assumptions in the Haiku and TAF models must always match for internal consistency.

TAF Uncertainty Cases

To test the effects of uncertainties unique to the benefit calculations, we altered the TAF base case assumptions for use with the Haiku base case output (in all cases, the Large NO_x reductions). These include

- TAF with S–R Uncertainty (both in ASTRAP and URM),

- TAF with alternative population assumptions (for a given vector of emissions changes),
- TAF with alternative choices for C–R functions, and
- TAF with alternative choices for values.

Presentation and Analysis of Results

Net Benefits and Statistical Uncertainty

Haiku Uncertainty Cases

The results of the Haiku uncertainty cases are illustrated in Figure 3-2. These box-and-whisker plots represent the benefits associated with each baseline and policy combination for a given set of default assumptions. In each plot, the median is a line in the box, the height of the box represents the 25th and 75th percentile range, and the vertical lines extending from the box represent the minimum and maximum benefits of the sample. The center of the μ represents the mean benefits, and the diamond represents the policy cost relative to the baseline. Thus, if the diamond is below the μ, then the policy has aggregate net benefits, with uncertainty about those benefits that could make net benefits more positive or less positive or even negative. For ease of representation, we only show the results for 2025. (The full set of numerical results, by 5-year increments, is presented in Appendix 3F [available on request from Alan Krupnick].) The Moderate NO_x reduction case stands alone, whereas the Large NO_x reduction case with its default set of assumptions for population and natural gas prices is contrasted with other Large case runs with varying assumptions about the population and natural gas prices.

Comparing the results of the base case for Large and Moderate NO_x reduction policies yields some very interesting conclusions. One might presuppose that the ratio of benefits to costs should be greater for the Moderate case than for the Large case. The reason is that marginal emissions control costs are generally thought of as rising with larger emissions reductions, whereas the benefits of reductions in the aggregate would be expected to increase proportionally with NO_x reductions, aside from changes in the location of those reductions relative to population centers.

This expectation is not met, however. Although NO_x reductions are 115% greater for the Large NO_x reduction scenario than for the Moderate scenario, the increase in benefits is actually 155% larger—the 40% difference caused by a spatial reallocation of emissions reductions to areas close to population centers. At the same time, costs increase less than proportionally with NO_x emissions, increasing only 89% (rather than 115%), suggesting the marginal abatement cost curve is concave over this region. This result could occur for various reasons stemming from the lumpiness of investments or the regional profiles of demand and technology. The net effect is to provide a large boost in net benefits for the Large case over the Moderate case.

Turning to the rest of Figure 3-2, we contrast net benefits under the base case assumptions to those arising when basic assumptions in Haiku are altered. Starting with population, benefits and costs are greater under the high population than the mid population assumptions, as would be expected, because larger populations result in more people to be helped by lower NO_x emissions. Also, with higher demand for electricity, NO_x emissions would be larger and more expensive to control. The converse result does not occur with the lower population estimate. The costs with the low population assumption are lower than for the mid population assumption, although only by 1%, and the benefit distribution for the low population assumption is actually increased over the mid population benefit distribution, although again only very slightly. Essentially there is no change in benefits or costs between the mid population and low population scenarios.

We also note that with the high population assumption, the dispersion of benefit estimates is larger than for the low or mid population (base case) assumptions, indicating a greater degree of uncertainty about human health exposure and its consequence. The high population assumption lifts costs by more than benefits, resulting in an average of negative net benefits, in contrast to the low and mid population assumption results, which feature positive net benefits.

Finally, we compare the base case results to results where gas prices are 30% lower than assumed in the base case, results where gas prices are 30% higher, and still other results where gas prices are 70% higher. We expect benefits to increase with higher gas prices because, in the absence of the policy, higher gas prices lead to greater use of coal for electricity generation and greater emissions of NO_x as a result. Therefore, we expect that the introduction of the emissions cap under the policy will have a greater effect in lowering emissions under scenarios with higher gas prices. In fact, this is what we observe.

A priori we cannot forecast the effect of these alternative scenarios on the costs because in each case, the alternative gas price function is used in the baseline and the policy run. If, for instance, gas prices are lower than in other runs, then in the base case the mix of investments and technologies used for electricity generation will adapt to take advantage of relative fuel prices. The introduction of the NO_x reduction policy on top of the adapted mix of generation technology will have unpredictable effects on costs relative to the introduction of the policy on top of an alternative natural gas price baseline.

TAF Uncertainty Cases

The next step in the statistical uncertainty analysis is to turn on statistical uncertainties in components of TAF that, because of their novelty, were not turned on in the runs above. These include uncertainties about population—independent from changes in electricity demand in the integrated modeling of benefits and costs discussed above—and uncertainties about the S–R relationships. All analyses were conducted for the base case runs for the Large NO_x reduction scenario.

Population Uncertainties

Figure 3-3 shows the effect on benefits of substituting low and high population estimates from the U.S. Census Bureau for the default assumption, which uses the Census' middle estimate. The benefit estimates in Figure 3-3 are different from those in Figure 3-2 because we held holding emissions changes constant. Our purpose was to examine the role of aggregate population changes, differences in the age distribution, and differences in the location of population on net benefits. (A detailed explanation of our set of projections is presented in Appendix 3D.)

In the TAF model, the relationship between population and benefits is proportional; a 5% increase in population across all states and age groups leads to a 5% increase in the valuation of benefits. In the U.S. Census Bureau data, total population is 8.78% lower in the low population series than in the middle and 12.62% greater in the high population series. Thus, if these differences were constant across all states and age groups, we would expect benefits to differ by the same amount. However, mean total benefits are 7.20% lower in the low population and 10.51% greater in the high population series compared with results in the middle series. In each case, the percentage change in benefits does not fully match the change in total population because benefits are influenced by both the location and ages of population with respect to emissions.

Table 3-1 decomposes the three primary effects of using the low and high population series: a scale effect (difference in total population), a location or spatial effect (the distribution of population across states), and an age effect (the distribution of ages within states). The scale effect is equal to the percentage difference in total population: –8.78% for the low series and 12.62% for the high series.

The age effect is calculated by comparing benefit valuations after scaling the total population for each state in the low and high series to match the corresponding population in the middle series. The scaling is done by multiplying the population of each age in each state in the low and high series by the ratio of the equivalent total state population in the middle series to the total state population in the low or high series. This forces total population for each state to be equal in all three series while allowing the distribution of ages within each state to vary. Using mean benefits, Table 3-1 shows an age effect of +1.68% for the low series and –1.82% for the high series.

The location effect is calculated by comparing benefits after scaling the total national population of each age in the low and high series to match the corresponding population of each age in the middle series. Unfortunately, given our method for constructing the low and high projection series, there is no location effect. For example, the percentage of all 18-year-olds in California in 2025 is equal to 12.7% for the low, middle, and high series. Consequently, if we scale the total population of 18-year-olds in the low and high series to match the middle series (by multiplying the low [or high] projection of 18-year-olds for each state by the ratio of total 18-year-olds in the middle series to total 18-year-olds in the low [or high] series), then we end up with three identical series and therefore no location effect.

The decomposed effects of using alternate population projections can be summarized as follows: switching from the middle series to the low series results in a –8.78% scale effect, a 0% location effect, and a +1.68% age effect for a net result of –7.1% in mean benefits. Switching from the middle series to the high series creates a +12.62% scale effect, a 0% location effect, and a –1.82% age effect, for a net result of +10.8% in mean benefits. Due to rounding, these results differ slightly from the percentage difference in overall mean benefits using the total low or high projection series compared with the middle series (–7.2% and 10.51%, respectively), as referenced above.

URM Uncertainties

The first three box-and-whisker plots in Figure 3-4 show the effect on benefits of introducing uncertainties in annual elevated NO_x S–R coefficients taken from the URM for the eastern United States. Uncertainties about this matrix, as they were derived for this project, apparently are unimportant. We see this because the Low Matrix, which provides the lowest index score for potency to health, and the High Matrix, which similarly provides the highest score, have virtually no effect on the mean benefits or the various percentile points highlighted in the plots. Overall, the range of difference in benefits is only 2%.

However, the caveat "as they were derived for this project" is crucial. In fact, if we had limited the benefit calculation to only the URM domain and only benefits calculated from the annual elevated NO_x matrices (shown in Figure 3-5), the range of benefits would have been 14%. This discrepancy arises because the domain of URM is geographically limited and uncertainties in ASTRAP are not being modeled in this case. Thus the lack of uncertainty in S–R relationships for the country (more or less) west of the Mississippi River serves to dilute the observed effect on national benefits.

Another reason for this caveat is that the uncertainties in the URM were explored in a very limited way, as described above. If we had the resources to explore uncertainties in more fundamental elements of the model, such as rate constants rather than simply to uncertainties about scaling episodic relationships to yearly average relationships, there doubtlessly would have been far more uncertainty and more impact on benefits.

It is important to note that each model run has been reduced to 40 iterations for all analysis involving the S–R matrices. Calculating uncertainties for the ASTRAP S–R matrices requires more computational capability than ordinary model runs. The limited resources available for this project forced us to limit the number of iterations for the ASTRAP uncertainty calculations. As a result, the default benefit distribution is different between Figures 3-2 and 3-4; Figure 3-2 consists of 500 samples per run, whereas Figure 3-4 uses 40 samples per run.

ASTRAP Uncertainties

Figure 3-4 also portrays the effect of statistical uncertainties introduced through the ASTRAP model on the benefits of the Large NO_x Reduction scenario. The relevant comparison is between the first box-and-whisker plot, which is the base case run for our analysis, and the middle plot (Default Assumptions with ASTRAP Uncertainty). The latter introduces the ASTRAP model's uncertainty to the S–R coefficients for the western United States (i.e., those states outside the URM domain) and the interaction of pollution transport between those states and states in the URM domain. The effect on mean benefits is minimal, as expected, because only uncertainties are being introduced. The mean should not change.

The benefit distribution widens, as expected—evident in that the height of the box and the sample range are somewhat wider for the fourth plot than the first. The ASTRAP uncertainty is also much more symmetrical, because the median benefits are in the center of the box with ASTRAP uncertainty considered but near the 75th percentile line otherwise.

Remaining Statistical Uncertainties Compared

We have yet to discuss the statistical uncertainties introduced into the benefits calculations through use of epidemiological and valuation studies. Because all these uncertainties are independent of one another, Figure 3-6 describes the effects of each source of uncertainty on the total uncertainty of benefits in the base case. For completeness and comparison purposes, we also include the uncertainties in the S–R coefficients from the western United States and from the eastern United States.[3]

While investigating each inputs contribution to uncertainty, all other probabilistic variables are held at their median levels. Although Analytica has built-in features for importance analysis, the memory requirements of this model are too great to take advantage of them. Instead, we manually programmed all variables to remain at median levels in the respective variable definition fields. To examine the uncertainty contribution of each variable, we simply "turn on" uncertainty for the variable of interest and evaluate the model.[4] In Figure 3-6, Mid-Level Benefits ($1,409 million) represents the case of no uncertainty—when all variables are held at median values.

The greatest uncertainties are introduced by use of the Pope et al. (2002) study for $PM_{2.5}$ mortality, followed by the western S–R uncertainty introduced by the ASTRAP

[3] For the sake of streamlining this paper, we ignore uncertainties in the other health endpoints included in the analysis because the contribution of these endpoints to total benefits is so small compared with mortality and chronic bronchitis.

[4] Each model run in this section consists of 40 iterations because of memory requirements for calculating ASTRAP-related uncertainty.

model, then the value of statistical life uncertainty taken from Mrozek and Taylor (2002). We also find that the uncertainties for ozone mortality and the eastern United States S–R coefficients might as well be ignored because they are so small. An interesting case is the $PM_{2.5}$ chronic bronchitis estimates from Abbey et al. (1995). Although this uncertainty is small, it is enough to potentially change the slight net benefit estimate into a small net cost and therefore potentially alter a policy decision.

Model Uncertainty

Our choices for default models and assumptions in the TAF base case are themselves subject to much controversy and uncertainty. Thus, it is imperative in an uncertainty analysis to also include quantitative estimates of model uncertainty. To do this we swap different "models" for S–R coefficients, the C–R functions and unit values (for short-term and long-term mortality) in and out of the TAF model run for the base case, large NO_x reduction policy. Although we cannot claim to have included every model in the epidemiological or valuation literature for these key endpoints, we have attempted to be reasonably comprehensive. For the S–R coefficients, we are unaware of any previous two-model comparison like this one but certainly cannot claim that only these two models are available for this purpose.

Figures 3-4 and 3-6 address the S–R model uncertainties. Figure 3-5 presents the effect on benefits (specifically, just mortality reduction benefits) from swapping ASTRAP default coefficients for URM coefficients for the URM domain states. This effect is seen in the comparison of means of the leftmost and rightmost distributions. Mean benefits with ASTRAP coefficients are about 3.5 times larger (more "potent") than those from the URM model. Given the very minor uncertainties introduced to benefits by considering alternative URM matrices, uncertainties in ASTRAP S–R coefficients add a large degree of uncertainty to benefits, as seen in both the 25th to 75th percentile "box" comparisons and the sample range comparisons. So, the ASTRAP coefficients are larger on average and more uncertain than the URM coefficients. However, whereas the URM uncertainties are derived ultimately from meteorological uncertainties (as are those for ASTRAP), they are different in character from the ASTRAP uncertainties; the latter are of a statistical nature, the former more procedural (through the CART analysis).

Figure 3-4 compares the benefits calculated using each of the two models from a national perspective and in terms of total benefits of the Large NO_x Reduction base case scenario. This relevant comparison is between the middle and the rightmost plot. This comparison looks almost identical to that in Figure 3-5, except the scale is different and the ratio of mean benefits with ASTRAP coefficients only to mean benefits with URM and ASTRAP default coefficients is lower than in Figure 3-4 (1.4 vs. 3.5).

Figures 3-7 through 3-9 are box-and-whisker plots showing how benefits in the base case (Large NO_x reduction) runs are affected by using alternative C–R and valuation models. Turning to the $PM_{2.5}$ long-term mortality models, we provide three (Figure 3-7), from Pope et al. (2002), Krewski et al. (2000—a reanalysis of the original 1995 Pope study, not cited here), and Dockery et al. (1993). These plots reveal that the Pope et al. 2002 study has a wider band of uncertainty than the 2000 Krewski et al. reanalysis and provides a slightly lower average number of deaths avoided. The Dockery et al. study, which draws on data from six cities, has two features distinguishing it from the other two, both drawn from data from the American Cancer Society (ACS) survey. The first, which is well-known, is that the Dockery et al. study produces about three times the reductions in deaths, on average, than that from studies using the ACS data. The second is less well-known: the uncertainty range is very large relative to that of the ACS data studies.

Many more studies investigate the affect of ozone on mortality using daily time series data. Until very recently, these studies have been thought to be too speculative to be used in EPA RIAs. However, with the recent work of Bell et al. (2004), the credibility of this effect is growing. What is surprising from Figure 3-8, which presents the effect of each of these studies on benefits, is that the results of Bell et al. mirror those of three other studies, both in the mean and higher moments of the uncertainty distribution. It is also notable that the effect of ozone-induced short-term mortality (say, from the Bell et al. study) on benefits is only about 40% of the effect of $PM_{2.5}$-induced long-term mortality.

Our final example of this type of analysis of model uncertainty examines the effect on benefits of using alternative values of statistical lives (VSLs; Figure 3-9). Each of the VSLs from the eight alternatives provided (some of which are not independent) could alter the sign on net benefits, depending on how their uncertainties were evaluated, although the Krupnick et al. (2002) study achieves positive net benefits only at the maximum of the uncertainty surrounding the VSL from this study. Another finding is that the type of distribution assumed over the range of studies considered in EPA's BenMAP model (whether normal, uniform, or triangular) has little effect on uncertainty.

It is one thing to describe the effects of different C–R functions or VSLs and another to give these models standing in an analysis. One approach that is a natural extension of our modeling effort is to stack the distributions that arise from using different models into one megadistribution, which then can be described with its own box-and-whisker plot. The key issue, of course, is the weights assigned to each function or value. For illustrative purposes, we assign equal weight to each function or value in Figures 3-7 through 3-9. The results are shown in Figure 3-10.

Finally, Figure 3-11 builds on a discussion of cobweb plots in Chapter 2. A cobweb plot is an interactive graphical technique for visualizing a Monte Carlo simulation or similar type of analysis. It enables the user to better understand relationships between input uncertainty and output uncertainty. Here we have drawn many samples (in this case, 500). The variables are total benefit, VSL, the PM mortality coefficient, the PM chronic bronchitis coefficient, and the ozone mortality coefficient. Each is represented as a vertical line. Each sample realizes one value of each variable; connecting these values produces a jagged line. One sample thus corresponds to one jagged line, and the graph shows the entire empirical distribution of 500 jagged lines. The interactive element is engaged when we select subsamples. In Figure 3-12, we have first shifted all variables to a percentile scale, then selected those samples in the top and bottom 10% of the total benefit.

We see that high (low) values of total benefit are strongly associated with high (low) values of the PM mortality coefficient. The other variables, VSL, the PM chronic bronchitis coefficient, and the ozone mortality coefficient retain nearly a uniform distribution, implying only a weak association with total benefits. This is the same result observed in Figure 3-6, the box-and-whisker plots showing the importance of the PM mortality coefficient in determining total benefits.

We also can examine rank correlation coefficients to further emphasize the above result. Rank correlation is a method of measuring the strength of association between two variables. It is computed by translating variables into rank-ordered values (percentiles), then calculating the correlation. Two variables that have a strong monotonic relationship will exhibit a rank correlation close to 1 (or –1). This is presented for the cobweb plot variables at the bottom of Figure 3-12. As above, the only strong association is between total benefit and the PM mortality coefficient, which have a rank correlation coefficient of 0.91. All other correlations of variables are weak at best, as expected for correlations between input variables because the VSL, the PM chronic bronchitis coefficient, and the ozone mortality coefficient are all calculated independently.

Rank correlation is useful because it provides a numerical interpretation to the cobweb plot. However, the cobweb plot has the visual advantage with its ability to graphically track a sample across all variables. It helps to reveal complex relationships that can be missed with numerical correlation measures. In our model, the underlying relationships among variables are very simple (multiplicative), and the power of the cobweb plot to visualize complex, uncertain relationships is not on display. (See Appendix 3E for a more revealing demonstration.)

Policy Conclusions from Case Study

In almost every scenario, net benefits of the Large emissions reductions are positive, meaning that the incremental benefits are greater than the incremental costs. However, the difference between benefits and costs is usually small. Mean incremental benefits were usually within 30% of costs, depending on the scenario being considered. In the central case scenario, mean benefits are $1,401 million and costs are $1,340 million. In other words, benefits are expected to be just 4.5% greater than costs.

The presence of positive net benefits indicates the policy passes a cost–benefit test, but the relatively small net benefits may be of concern to a policymaker who is averse to the downside risk of new policy. The uncertainty analysis we conducted may be especially important in this case because it provides some assurance that the analysis has anticipated the major sources of uncertainty and variability.

Another important feature of this analysis is that with respect to what are anticipated to be the most important sources of uncertainty—population and natural gas prices—the benefit and cost side of the model are integrated. To illustrate the relevance of this, consider that population changes affect the number and location of individuals who would be exposed to changes in air quality. The greater the number of people exposed, other things being equal, the greater the benefits of air quality improvements. However, greater population also implies greater electricity demand and a greater opportunity cost on emissions allowances under the fixed emissions cap. It would be misleading to consider the changes in population looking at just one side of the equation.

Figure 3-3 illustrates the effect of alternative population assumptions holding emissions constant: the benefits of emissions reductions grow as population grows.

However, as shown in Figure 3-2, costs are also likely to grow, and they grow more quickly than benefits when population increases, according to our modeling. Net benefits fall as the forecast of population growth increases. The low population scenario has positive net benefits of $114 million, compared with $61 million in the central case. In the high population scenario, net benefits are negative. This is one of the only scenarios we examined with expected costs in excess of expected benefits.

A second factor that is explored in an integrated fashion is natural gas prices, which change costs because fuel substitution is one kind of compliance activity. In general, natural gas–fired facilities have lower emissions rates than coal-fired facilities, but this depends on the vintage of the technology and postcombustion controls that are in place. Moreover, the effect of different emissions profiles depends on the geographic location of the emissions. It is also important to recognize that the benefits of changes in emissions are measured against a baseline that has a comparable natural gas supply schedule. Figure 3-2 indicates that changes in natural gas price affect not only the costs of compliance but also the benefits, when measured against the relevant baseline. In this case, it appears that benefits are affected more than costs when considering scenarios other than the base case. In every case, benefits are greater than costs.

Looking across many statistical uncertainties, net benefits remain positive. We vary assumptions about epidemiology, valuation, and the choice of model for atmospheric transport of pollutants. We also explore uncertainties within the atmospheric transport models. The robustness of the results to these sensitivities provides valuable information to policymakers. Given that net benefits are positive but small in the central case and that net benefits remain positive under various possible realizations of uncertain variables in the analysis, the risk that the estimates are erroneous is relatively low.

Methodological Conclusions from Case Study

Whatever the intrinsic policy interest of our case study, this interest is secondary in our project; our primary interest is in investigating methods for portraying uncertainty in RIAs. Thus, on the basis of the above case study, we summarize lessons learned for uncertainty analysis.

- The typology of uncertainties presented in Chapter 2 is a useful tool for both categorizing uncertainties under study and identifying areas that are ignored.

Classifying uncertainties as variability, parameter uncertainty, or model uncertainty, for example, helps to communicate the nature of the analysis. In doing so, it also makes apparent the types of uncertainties that are left out. We found this to be of particular value. Our interview with Karen Palmer and results with the Haiku model suggest that we were able to acknowledge uncertainties that we otherwise would have missed. This exercise would be productive at any stage of uncertainty analysis, either at the beginning of model design or later, when reviewing and interpreting results.

- The number of iterations matters. When applying Monte Carlo simulation methods or similar methods, the analyst must choose the number of draws from the specified statistical distributions around the model's parameters. This choice can be limited on the upper end by computational resource. For instance, 500 iterations was the maximum our computer system could support. And even this number of iterations was too much for our system to handle when we "turned on" the ASTRAP model's S–R uncertainties.
- This choice of the number of iterations is limited on the lower end. Too few iterations, and the results for different scenarios may look misleadingly similar. More important, the mean, median, and other moments of the benefit estimates derived from some iterations may be quite far from those estimated with a set large enough to converge. This concern is more important when the confidence intervals around parameters being modeled are very wide. A very wide interval requires many iterations before extreme values are sampled by the Monte Carlo routine. Hence, those extreme values will not show up in extreme benefits, biasing the results to a smaller variance than is appropriate. Figures 3-13 through 3-15 show what happens to benefit estimates as the sample size increases from 40 to 1000 by increments of 40.[5]
- As expected, as more samples are drawn, the change in mean net benefits decreases and approaches a stable estimate as the number of iterations grows (Figure 3-13). At around 500 iterations, the mean begins to center on the ultimate value. A sample size of 40 clearly does not produce an accurate estimate in this case, as the mean is well below the trend value with greater iterations. Figures 3-

[5] To achieve 1000 iterations, we combined two separate runs of 500 iterations, each using different seed variables.

14 and 3-15 display histograms of the benefit distribution using 40 and 1000 iterations respectively. This helps explain why the net benefit mean is subject to such drastic change with fewer samples. The histogram in Figure 3-14 contains many "holes" and does not present the appearance of a smooth distribution. On the other hand, with 1000 iterations, Figure 3-15 is much more complete.

- Modeling uncertainties in addition to those traditionally modeled in RIAs, such as C–R and valuation uncertainty, is not difficult.
  - Population uncertainties can be modeled using U.S. Census Bureau standard population series. One important complication, however, is that population estimates must be consistent on both benefit and cost sides. In our analysis, the relative ease of modifying Haiku electricity demand made this consistency possible.
  - Similarly, uncertainties in natural gas prices were very easy to introduce into Haiku. In each case, of course, a new baseline must be constructed. This makes the communication of results more difficult.
  - We have not yet demonstrated an approach to statistical modeling of uncertainty on the cost side; instead, we have relied on a nonparametric approach to uncertainty (e.g., using high and low estimates). One issue was that we did not have distributional information for the natural gas price analysis, where we simply tested the effect of three alternative prices on costs. Another issue is that the complexity and size of the Haiku model precluded running the model in Monte Carlo simulation mode, even if we were to somehow obtain distributions of our uncertain input variables. We are developing an algorithm that augments the current convergence method in Haiku with goal-seeking functions in an effort to make formal uncertainty analysis (Monte Carlo analysis) feasible, but this capability is not yet available.
  - S–R uncertainties can be modeled in two ways. Our first example—with the ASTRAP model—was the more standard of the two in that the S–R coefficients from this model "came with" distributions around the coefficients. This situation existed because the underlying methodology for deriving these coefficients was statistical. These reduced-form models are not favored at EPA. Simulation models are preferred. Whatever the merits of such models, their serious drawbacks, at least as currently configured, should also be acknowledged and perhaps this preference for simulation models

should be rethought. At the same time, uncertainty can be introduced into such simulation models in fundamental ways. Rate constants, for instance, are uncertain. So distributions of such parameters could be sampled in Monte Carlo analysis—at least in theory.[6] In practice, limitations on computational resources might preclude such an approach, because computing the consequences of an episode for air quality already takes as long to complete as the episode itself.

- The variability of climatological variables could be addressed in more fundamental ways in these models. As noted above, the URM model is a simulation model. We used URM and a supplementary model to generate S–R coefficients in a computationally efficient way, but for only several episodes. Then, we used a scaling approach to generalize the episode results to a full year of meteorology. An alternative approach would have been to generate enough episode days of results to estimate a reduced-form model that, in effect, would describe in simple terms all the complex chemistry contained in the simulation model as it interacts with meteorological variability. From this analysis, standard errors would be estimated for the S–R relationships, which could then be used in the uncertainty model. To our knowledge, no one has worked out this idea.
- The uncertainties (variabilities, actually) introduced by the ASTRAP model were quite significant and (if EPA accepted it) could be used to generate uncertainties in this component for RIAs or other EPA analyses. Perhaps other reduced-form models are available for this use as well.
- The uncertainties and variability we introduced through our URM uncertainty analysis was a novel attempt to introduce climate variability ex post to a modeling analysis. Using the CART statistical technique and monitoring data from New York, we demonstrated how S–R coefficients for specific episodes could be weighted and scaled into national estimates that address the uncertainties of such scaling procedures. Specifically, the CART procedure has a path dependence in the assignment of daily meteorological patterns to classes. This dependence permits the use of other orderings along

[6] In practice, some limited air quality modeling analyses have accounted for rate constant uncertainty (Bergin et al. 1998, Hanna et al. 2001).

the path to create a distribution of weights for such scaling. Overall, however, our 30 matrices were not very different from one another on net. Certainly, compared with the ASTRAP uncertainties, the uncertainties introduced by our CART procedure were relatively minor, thus raising the issue of whether they were worth the work that went into their development.

- Our means of portraying the information in this complex case study—even the parts unrelated to uncertainty—may offer some useful guidance to EPA.
  - First, from our analytical perspective, box-and-whisker plots are convenient and concise as a means of summarizing a large quantity of information. This characteristic makes it possible to compare many analyses on the same graph (e.g., Figure 3-7, which succinctly shows the model uncertainty in the PM mortality C–R function component of the benefits analysis).
  - Second, cobweb plots are particularly useful because they enable the visualization of complex relationships that may not be apparent in other graphical or numerical representations. For example, Figure 3-12 clearly shows that high (low) benefits are associated with high (low) PM mortality coefficients, whereas the other inputs are not as influential. Although one might infer this from Figure 3-6 (at least for the PM mortality coefficient), the cobweb plot is more explicative. (Appendix 3E provides a better showcase for the power of cobweb plots.) Our use of importance analysis—as reflected in these figures—attributes uncertainties in aggregate benefits to the uncertainties in each component of the base case benefits analysis. Such figures allow one to identify those components where further research would make the greatest contribution to narrowing the overall uncertainty in benefits.
- EPA might also find useful our decomposition analysis of the effect of population changes on total benefits. The idea is that population estimates have spatial and age-specific dimensions as well as U.S. totals. Because air pollution and health effects have spatial and age-specific dimensions as well, it will be enlightening to examine which of these three factors contributes most to any total benefits we see.
- Building on this chapter could yield an approach to jointly addressing model and statistical uncertainty. Probably the knottiest issue in implementing the uncertainty provisions of *Circular A-4* (OMB 2003) is determining how to address model uncertainties systematically and merge such uncertainties with statistical

uncertainties to gain an overall understanding of uncertainties in benefits. We have not attempted that merger here, although we did address these two types of uncertainties separately. We have some ideas about how this could be done. One approach would be to use meta-analysis to address model uncertainty, where the meta-analysis contains statistical estimates of uncertainties around the estimated parameters. Another more mechanical approach would be to "pack" the alternative model estimates by using a weighted average approach. A random weight would be assigned each "model," and repeated sampling of the weights would create a distribution around the studies that is independent of any judgment about a model's relative worth. Yet another option (with many suboptions) would be to perform an expert assessment, with the weights emerging from this process.

- Once assessments are performed to address and merge uncertainties, the challenge will be to present them in an understandable and efficient way. One idea is to build on Figure 3-6 with additional information about model uncertainty represented as circles or as extensions of the box-and-whisker plots to convey the joint effects of model and statistical uncertainties.
- This chapter leaves a key unexamined, unresolved issue: the appropriate metrics to display to decisionmakers from our analysis. There are several layers of complexity here. The first is on the cost side. In the above results, we focus on costs as an economist would measure them—consumer plus producer surplus changes. They could have been presented in terms of compliance costs—certainly an easier concept, if a possibly misleading one. Indeed, perhaps what decisionmakers want is some more politically driven measure, such as whether electricity stays below 10 cents/kilowatt-hour. Such decisions about measures are beyond the scope of this project. The second complexity is on the benefit side. Many effects of NO_x reductions are not listed, because they cannot be quantified or monetized or because the analyst judges them to be insignificant. How such nonquantifiable effects should be displayed is another key issue beyond the scope of this chapter.

Figures

Figure 3-1. Incorporating the Uncertainty Typology of Chapter 2

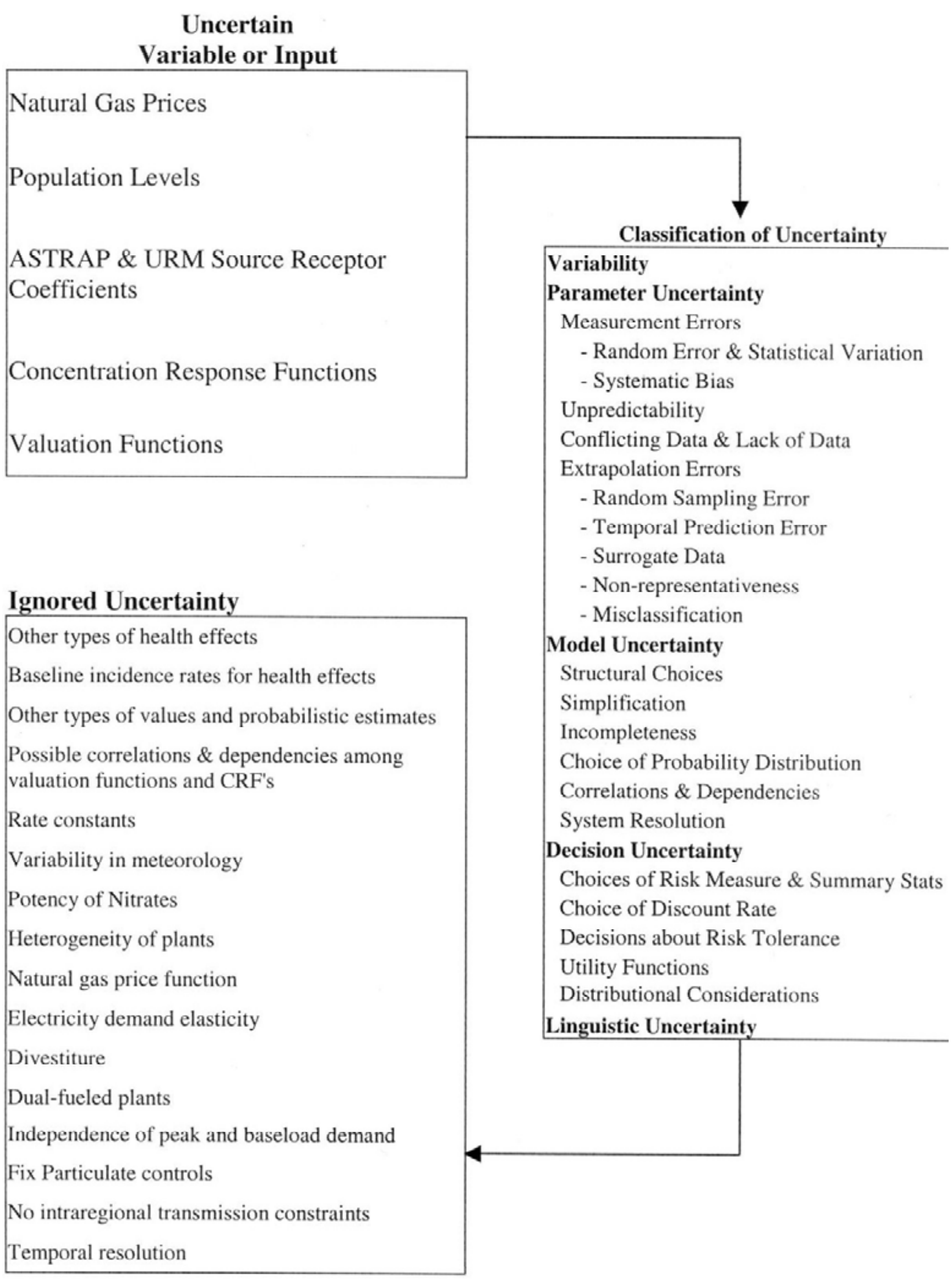
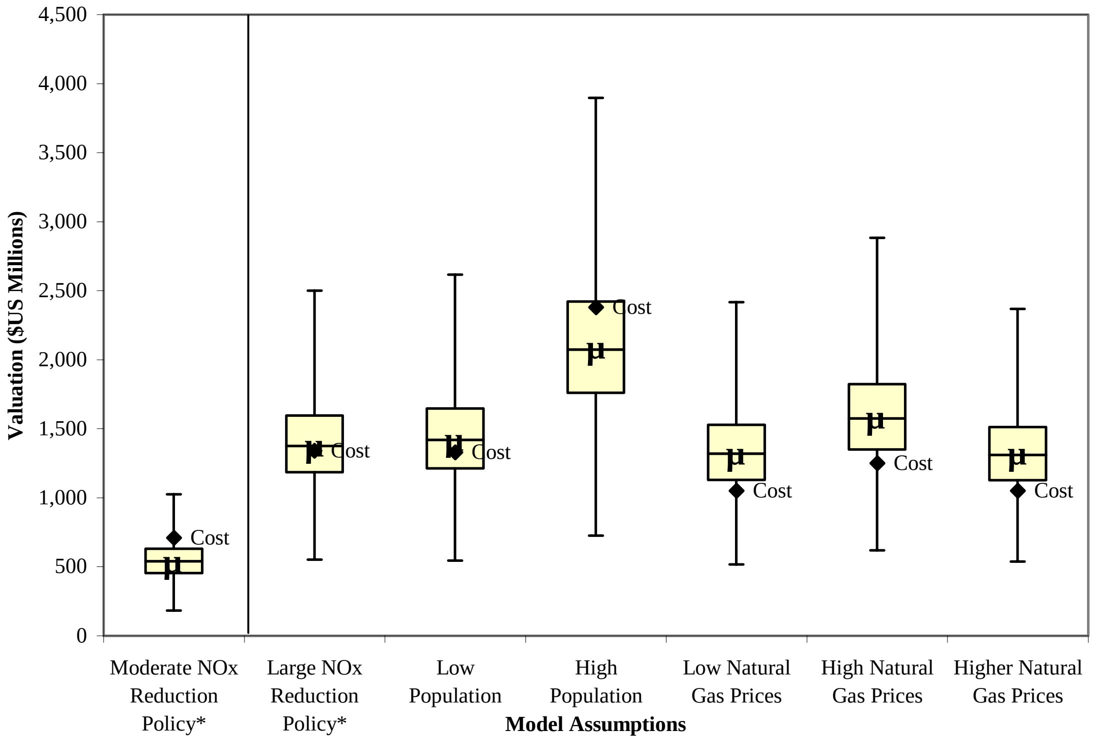


Figure 3-2. Benefits and Costs of Moderate and Large NO_x Reduction Policies from CAIR Baseline: Alternative Model Assumptions, 2025



***The moderate and large NO_x reduction policies use default assumptions of middle population and middle natural gas prices.**

Figure 3-3. Benefits and Costs of Large NO_x Reduction Policy from CAIR Baseline: Alternative Population Assumptions, Holding Emissions Constant, 2025

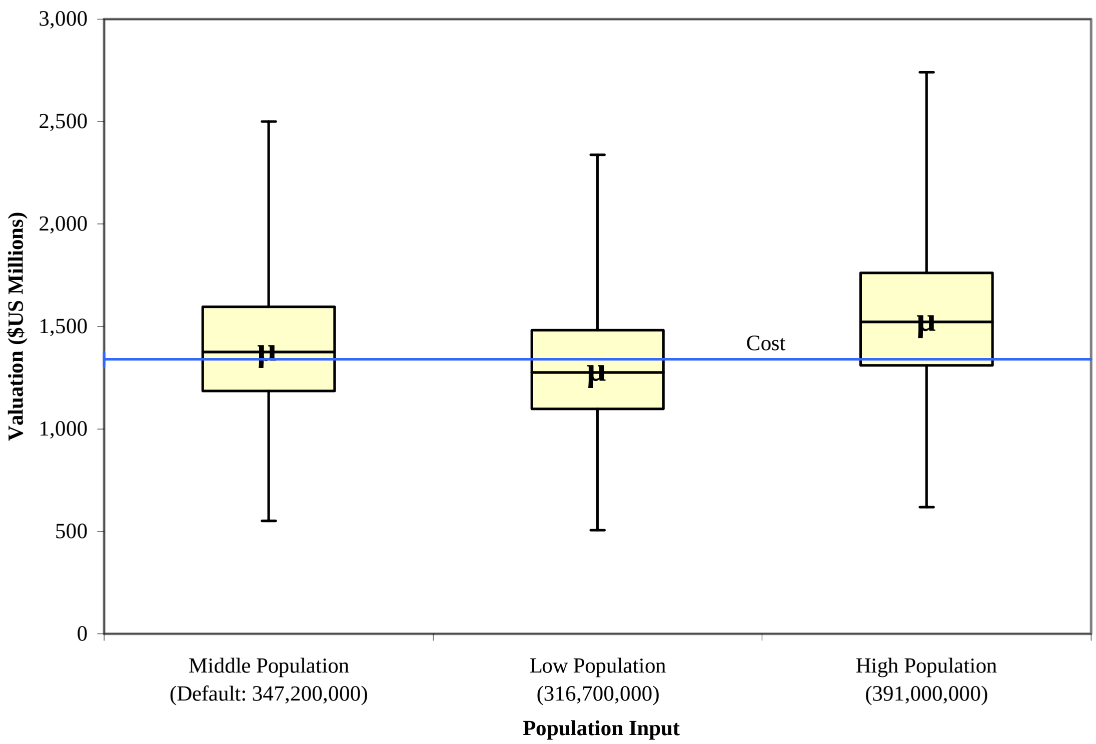


Figure 3-4. Benefits and Costs of Large NO_x Reduction Policy from CAIR Baseline: Using Alternative URM and ASTRAP Source–Receptor Assumptions, 2025

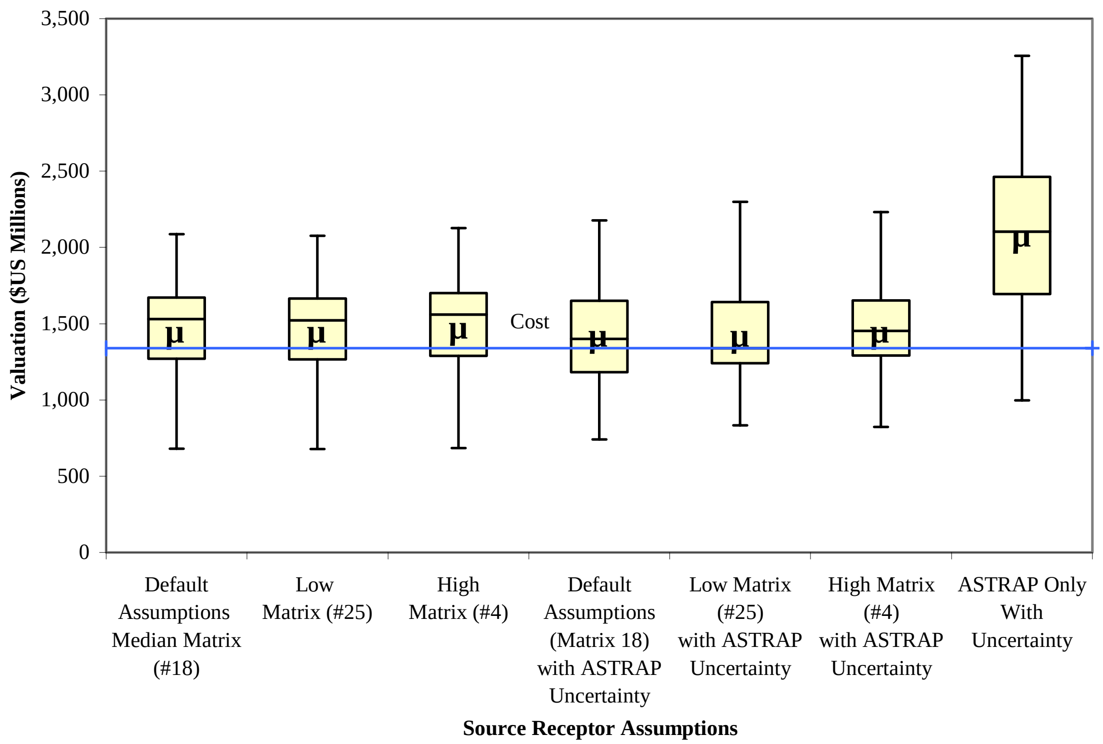


Figure 3-5. Benefits of Large NO_x Reduction Policy from CAIR Baseline: Alternative URM and ASTRAP Source–Receptor Assumptions for Interior URM States Only, 2025

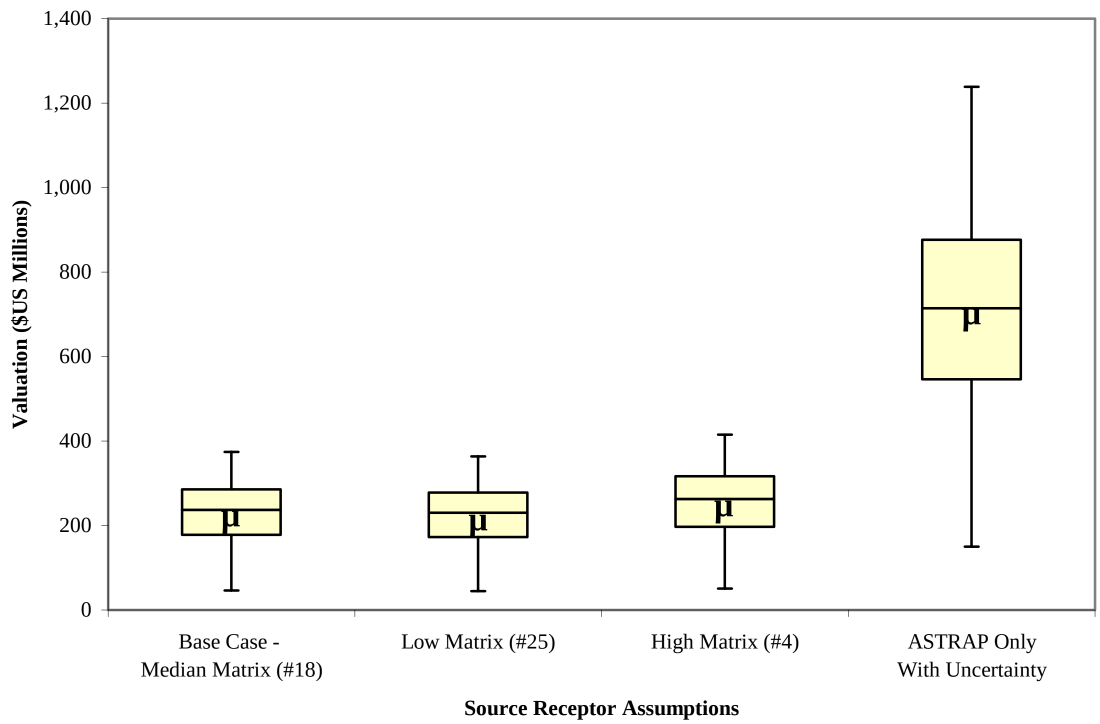


Figure 3-6. Benefits and Costs of Large NO_x Reduction Policy from CAIR Baseline: Effects of Statistical Uncertainty of Selected Inputs, 2025

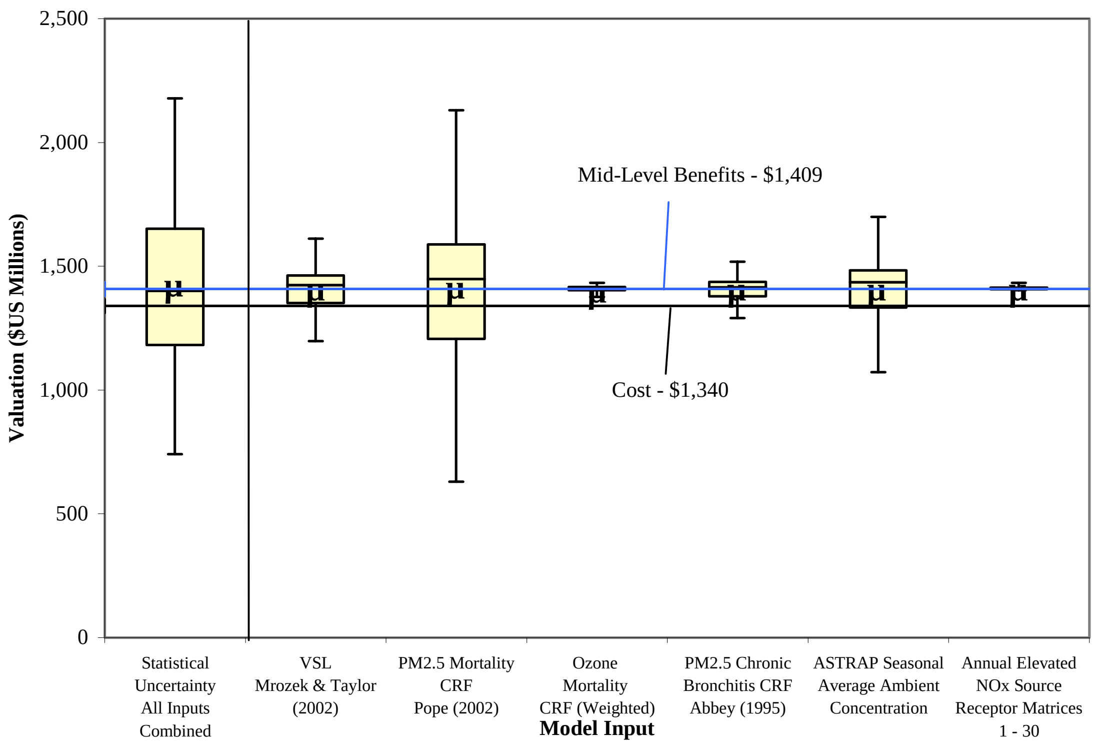


Figure 3-7. Benefits of Large NO_x Reduction Policy from CAIR Baseline: Alternative $PM_{2.5}$ Mortality Concentration–Response Functions, 2025

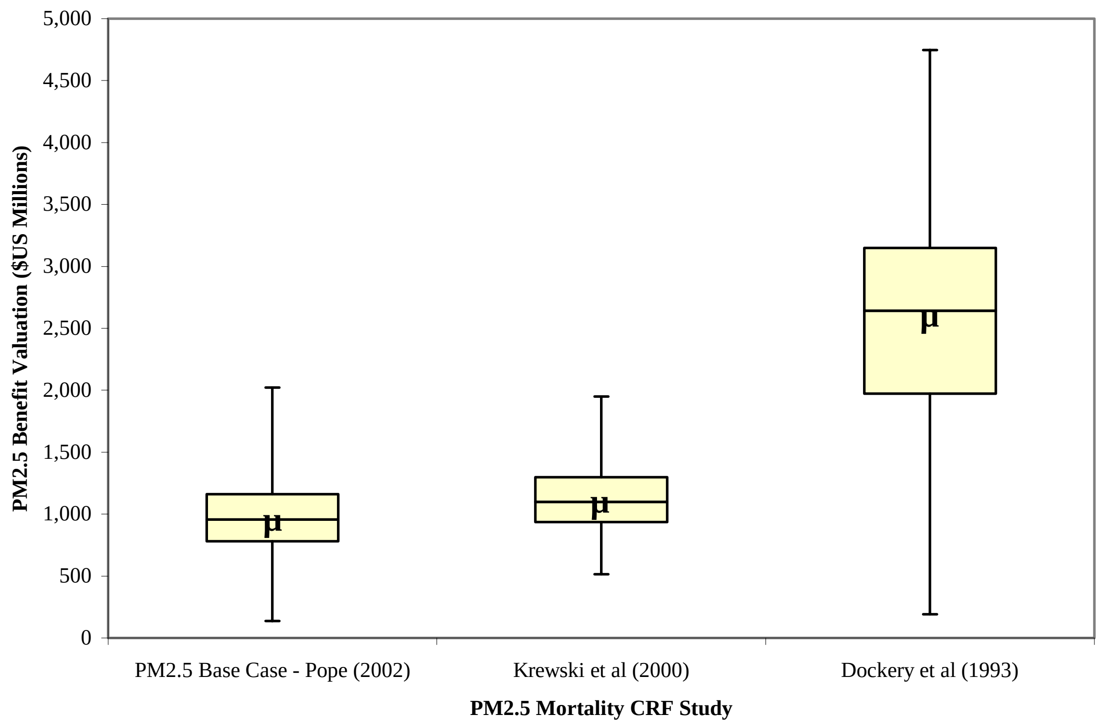


Figure 3-8. Benefits of Large NO_x Reduction Policy from CAIR Baseline: Alternative Ozone Mortality Concentration–Response Functions, 2025

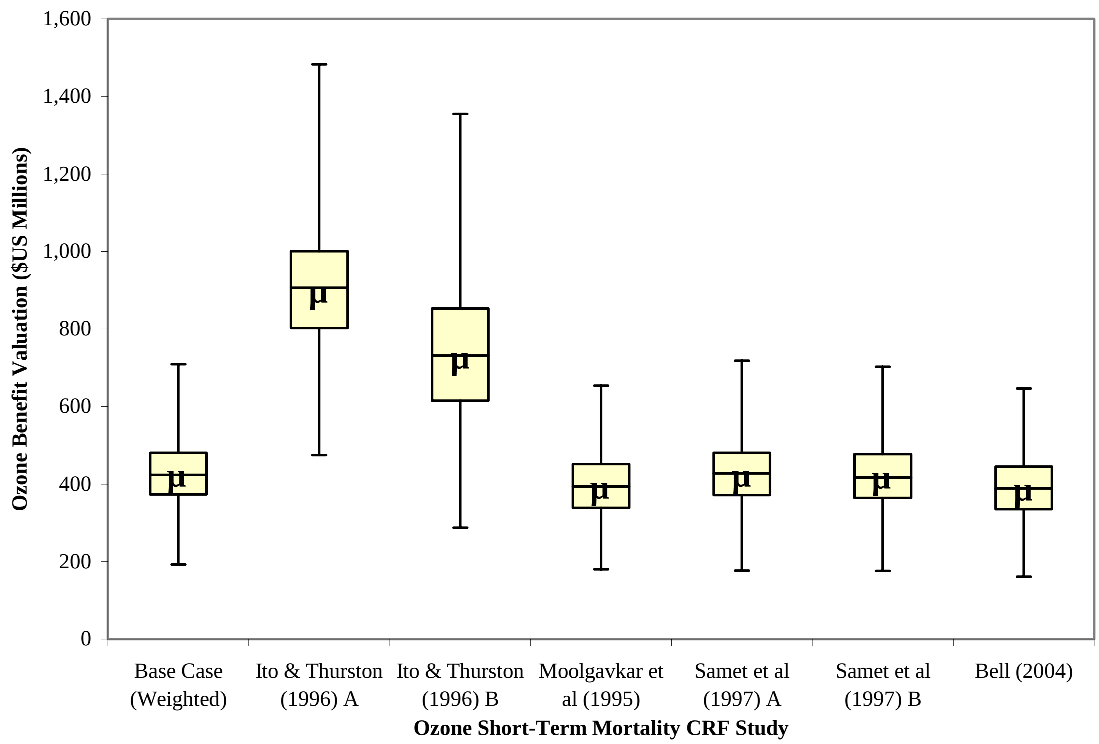


Figure 3-9. Benefits and Costs of Large NO_x Reduction Policy from CAIR Baseline: Alternative VSL Distributions, 2025

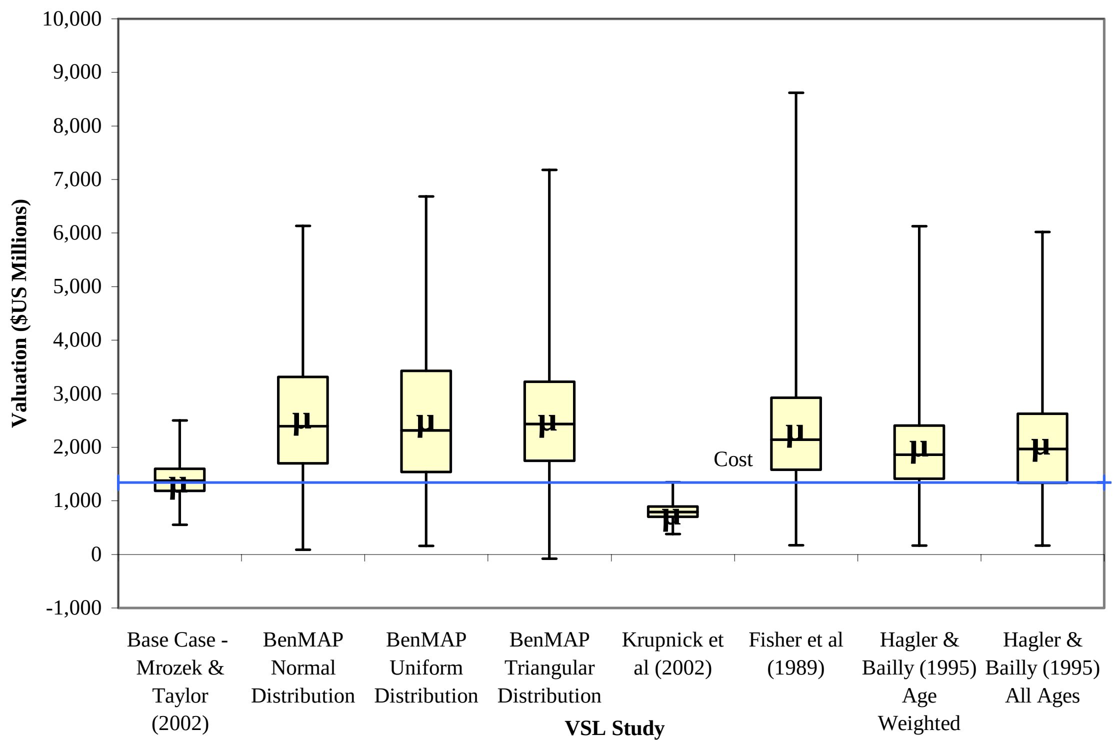


Figure 3-10. Benefits and Costs of Large NO_x Reduction Policy from CAIR Baseline: Equal Weighted VSLs and Mortality Concentration Response Functions (CRFs), 2025

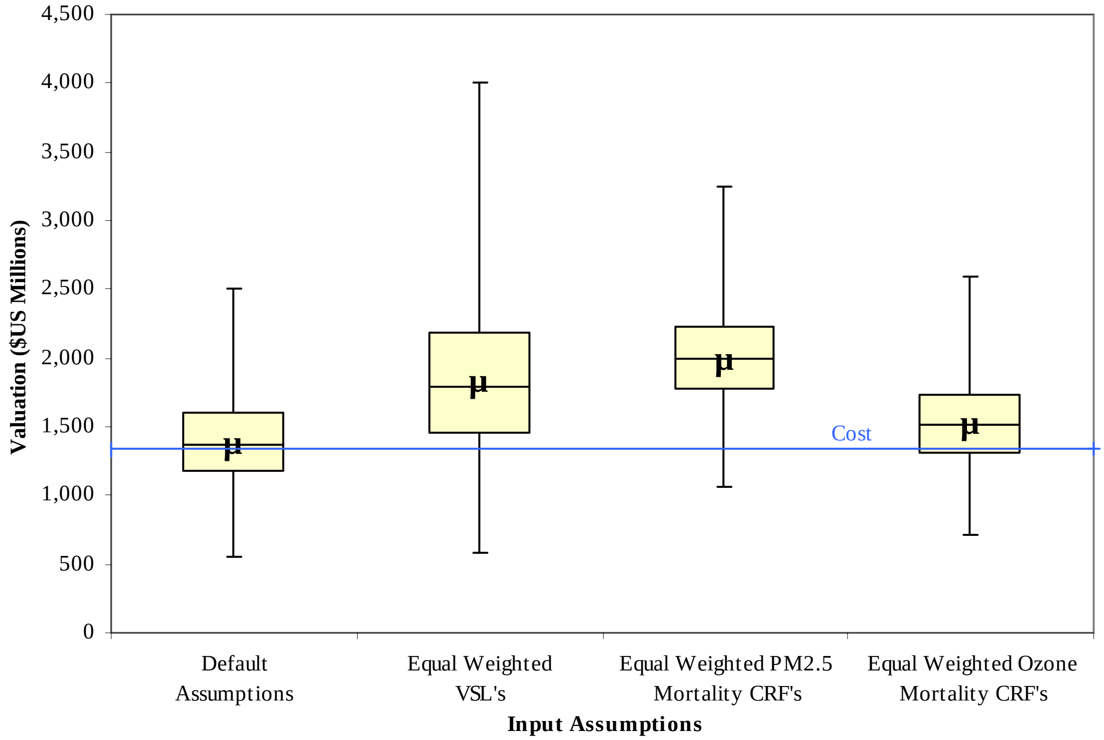


Figure 3-11. Cobweb Plot of Benefits of Large NO_x Reduction Policy from CAIR Baseline: Total Benefit, VSL, PM Mortality, PM Chronic Bronchitis, and Ozone Mortality, 2025 (All 500 Iterations)

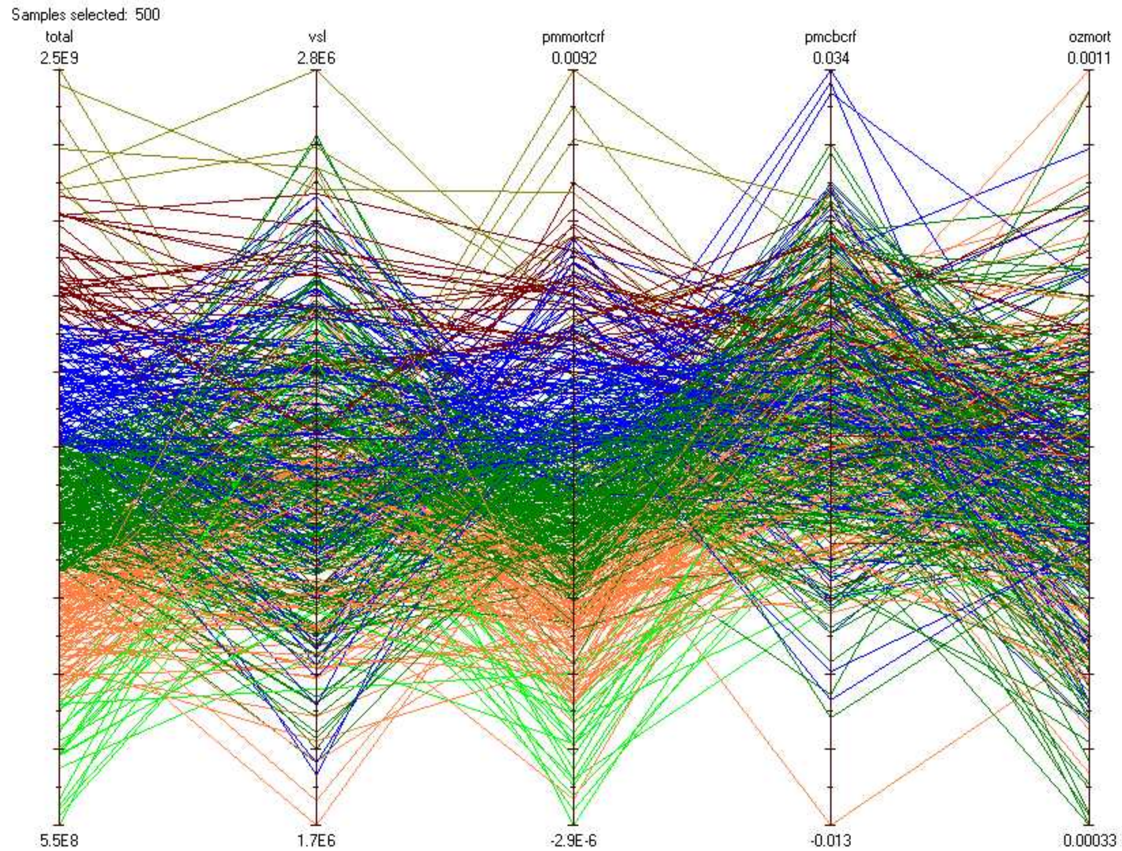
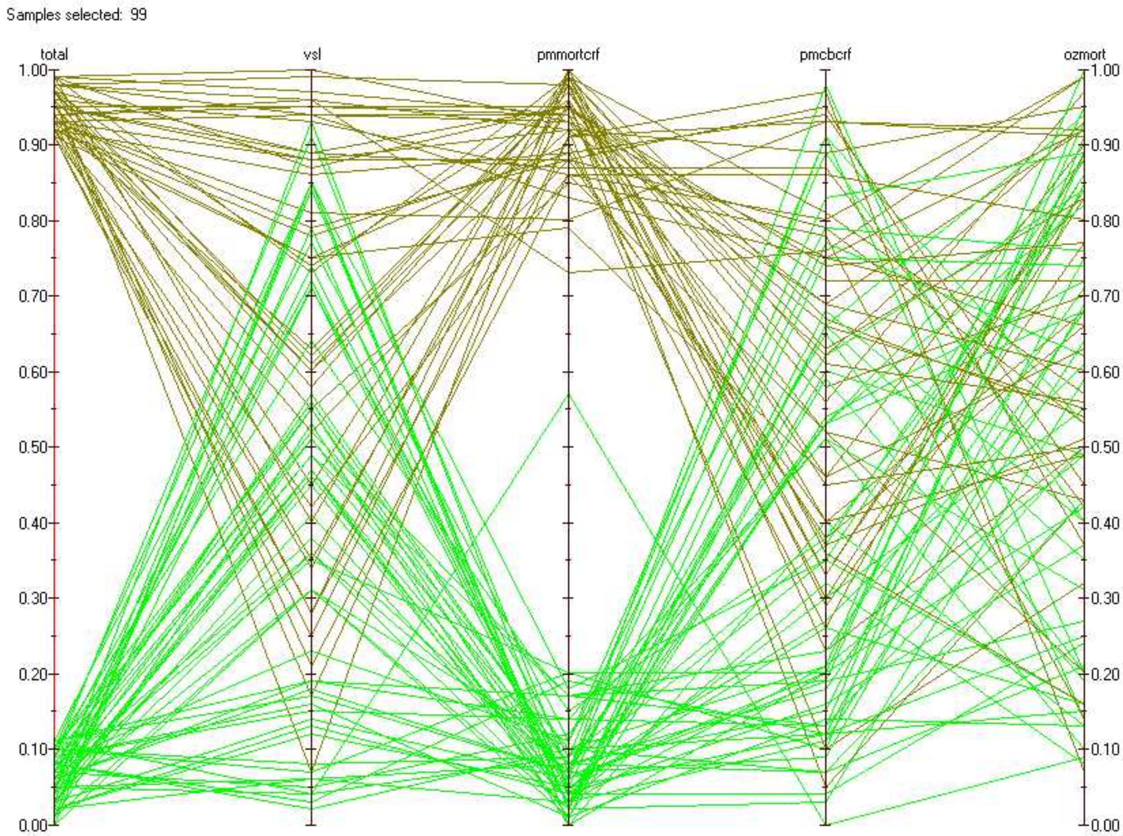


Figure 3-12. Cobweb Plot of Benefits of Large NO_x Reduction Policy from CAIR Baseline: Total Benefit, VSL, PM Mortality, PM Chronic Bronchitis, and Ozone Mortality, 2025 (Top and Bottom 10% of Total Benefits)



Rank Order Correlations: Total Benefits, VSL, and Selected CRF Inputs

	Total Benefits	VSL	PM Mortality CRF	PM Chronic Bronchitis CRF	Ozone Mortality CRF
Total Benefits	1	0.23	0.91	0.16	0.08
VSL	0.23	1	0	0.14	–0.01
PM Mortality CRF	0.91	0	1	–0.02	0.06
PM Chronic Bronchitis CRF	0.16	0.14	–0.02	1	0.08
Ozone Mortality CRF	0.08	–0.01	0.06	0.08	1

Figure 3-13. Mean Net Benefits of Large NO_x Reduction Policy from CAIR Baseline and Change in Mean with Increasing Iterations, 2025

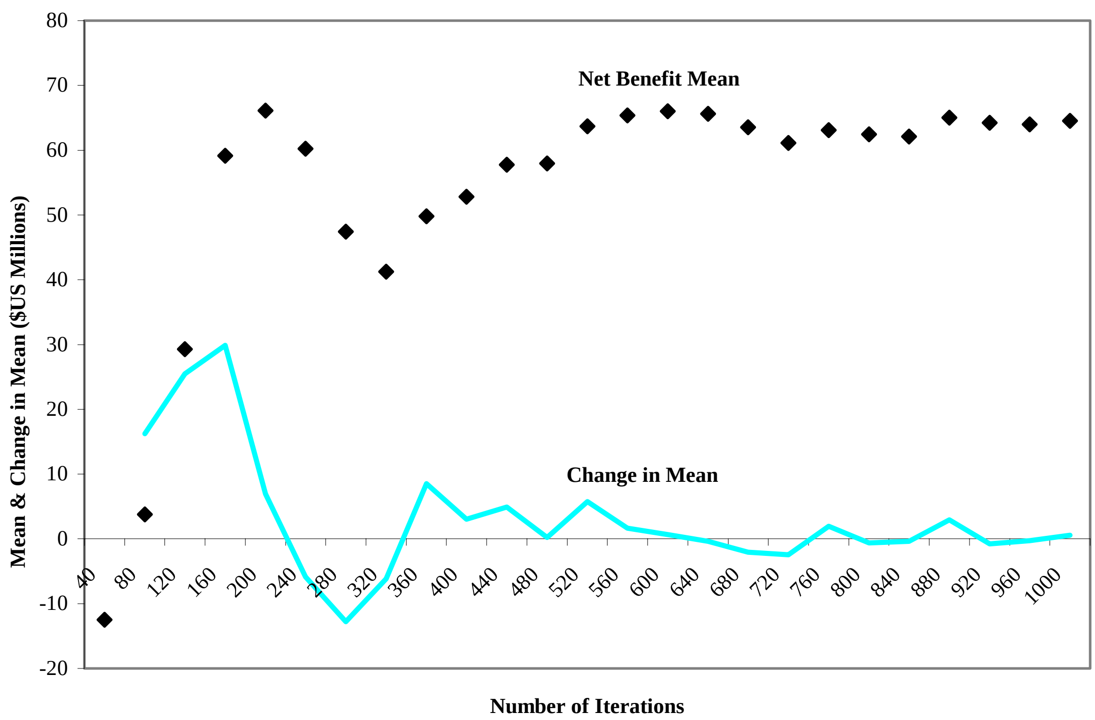


Figure 3-14. Benefits Distribution of Large NO_x Reduction Policy from CAIR Baseline, 2025 (40 Iterations)

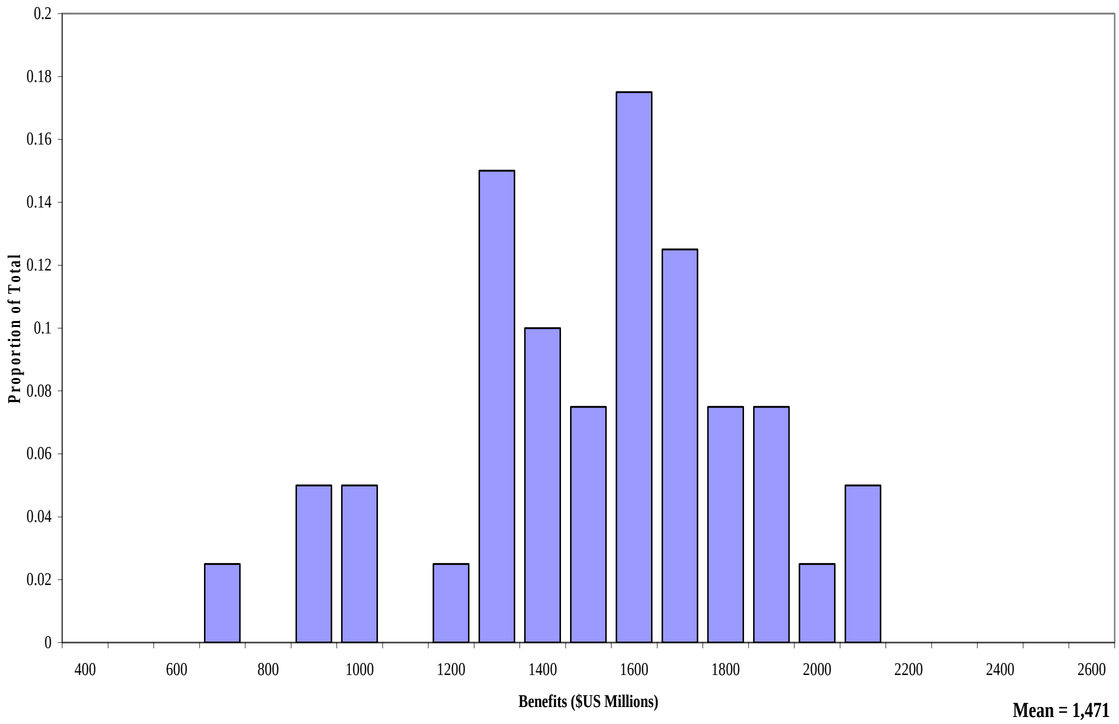
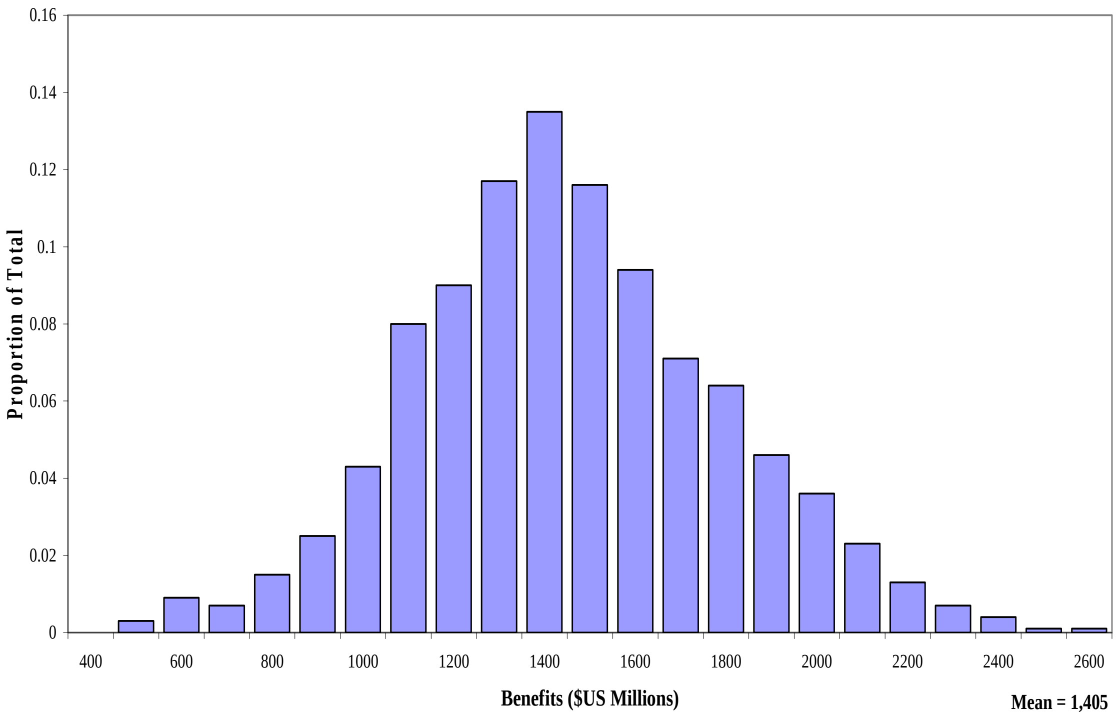


Figure 3-15. Benefits Distribution of Large NO_x Reduction Policy from CAIR Baseline, 2025 (1000 Iterations)



Tables

Table 3-1. Benefits of Large NO_x Reduction Policy from CAIR Baseline: Spatial and Age Distributional Differences between Population Projections, 2025

		Low Population			*High Population*		
	Default Case (Middle Population)	*Location Effect*	*Age Effect*	*Overall Results*	*Location Effect*	*Age Effect*	*Overall Results*
Mean benefits	1,401	1,401	1,425	1,300	1,401	1,375	1,548
%Diff from default case		0.00%	1.68%	–7.20%	0.00%	–1.82%	10.51%
Total population (millions)	347.2	347.2	347.2	316.7	347.2	347.2	391
%Diff from middle population (scale effect)				–8.78%			12.62%

Appendix 3A: The CAIR Baseline

Our baseline is the U.S. Environmental Protection Agency's (EPA's) Clean Air Interstate Rule (CAIR) in combination with the version of EPA's proposed mercury rule that includes mercury trading. Emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) are regulated within a 28-state region, mostly east of the Mississippi, plus the District of Columbia.[1] The State Implementation Plan (SIP) seasonal NO_x cap is included.[2] Regional annual SO_2 allowance distributions are capped at 3.9 million tons beginning in 2010 and 2.7 million tons beginning in 2015. Actual emissions will be higher over the modeling time horizon because of the allowance bank. We follow EPA modeling of the SO_2 CAIR and Title IV within one national trading regime. A single national region is characterized using model results that account for the opportunity to use Title IV allowances within the CAIR region at an offset ratio that changes over time. The actual emission caps that we model are reported in Table 3A-1.

Regional annual NO_x emission distributions are capped at 1.6 million tons beginning in 2010 and 1.3 million tons beginning in 2015. The NO_x caps that we model (Table 3A-1) include an adjustment of about 331,000 tons for units outside the CAIR NO_x region but within the Mid-Continent Area Power Pool (MAPP) and New England electricity regions in the model. National NO_x emissions with the regional cap total 2.55 million tons in 2010 and 2.39 million tons in 2025.

The only difference between the baseline and policy scenarios is that instead of having the CAIR NO_x cap that affects only the states regulated under CAIR, the policy

[1] The 28 states included in the region covered by the proposed version of the CAIR rule are Alabama, Arkansas, Delaware, Florida, Georgia, Illinois, Indiana, Iowa, Kentucky, Kansas, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, New Jersey, New York, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, Texas, Virginia, West Virginia, and Wisconsin.

[2] We find emissions during the summer ozone season within the eastern region increase under the CAIR rule as proposed and the EPA mercury cap when the seasonal NO_x program is terminated, as specified in the draft CAIR rule. Two possible remedies to this increase are tighter annual caps and maintenance of a seasonal cap. The policy scenario we model here is the latter. The policy ensures that emissions of NO_x during the 5-month ozone season do not exceed levels established under current policy to help reduce summer ozone problems. Having two NO_x policies of this sort means that generators located within both the CAIR region and the SIP region must have two permits for every ton of NO_x emitted in the summer season. The dual programs mean that the costs of NO_x controls will be split between two regulatory targets, and the prices of CAIR NO_x allowances are expected to be lower when combined with the SIP Call than when they are not. The final CAIR rule reinstituted a seasonal NO_x program.

scenarios have a more stringent nationwide NO_x cap. Two policy scenarios are modeled. The first (Large) features a nationwide NO_x cap of 2.4 million tons in 2010 and 1.5 million tons in 2015 and thereafter. The second (Moderate) features a nationwide NO_x cap of 2.4 million tons in 2010 and 1.95 million tons in 2015 and thereafter.

Appendix 3B: The Haiku Model

The Haiku model simulates equilibrium in regional electricity markets and interregional electricity trade with an integrated algorithm for choosing the technology for sulfur dioxide (SO_2), nitrogen oxides (NO_x), and mercury emissions control. The model calculates electricity demand, electricity prices, the composition of technologies and fuels used to supply electricity, interregional electricity trading activity, and the emissions of key pollutants.[1] The main data inputs to the Haiku model are listed in Table 3B-1, along with the sources for the associated data.[2]

The model solves for the quantity and price of electricity delivered in 13 regions, for four time periods (superpeak, peak, shoulder, and base load hours) in three seasons (summer, winter, and spring/fall). For each of these 156 market segments, demand is aggregated from three customer classes—residential, industrial, and commercial—each with its own constant elasticity demand function. Estimates of demand elasticities for different customer classes and regions of the country are taken from the economics literature.

The supply side of the model is built using capacity, generation, and heat-rate data for the complete set of commercial electricity plants in the United States from various Energy Information Administration (EIA) datasets. For modeling purposes, these plant-level data are aggregated into 39 representative plants in each region. The capacity for a model plant is determined by aggregating the capacity of the individual constituent plants in a given region that are of the same type as the model plant. However, no region contains every one of these model plants. For example, the New England region does not contain a geothermal plant.

A model plant is defined by the combination of its technology and fuel source (coal, natural gas, oil, hydropower, or nuclear). Steam or gas turbine plants can run on oil or natural gas. Coal is a little different from the other fuels in that it is divided into 14 subcategories on the basis of the region the coal is from and its level of sulfur content

[1] See Paul and Burtraw 2002 for a detailed report on the development and uses of the Haiku model.

[2] The items listed in Table 3B-1 are largely parameters in the model that rely on real-world data or variables derived from real-world data. The Haiku model user also must make assumptions about several inputs, including the discount rate, year in which to base net present value calculations, and expected rate of transmission capacity growth, as well as policy scenario assumptions.

(Table 3B-2). Coal users are categorized by demand regions that have different costs associated with each type of coal, which reflect the varying interregional transport costs. Model plants might switch the type of coal they use to reduce their SO_2 or mercury emissions, which may be more cost-effective than installing new pollution controls. Table 3B-3 gives a list of the various types of model plants.

Investment in new generation capacity and retirement of existing facilities are determined endogenously in a dynamic framework on the basis of capacity-related costs of providing service in the future (known as going forward costs). The model determines investment in and retirement of generation capacity, and new generation capacity is assigned to a model plant representing new capacity of that type. The Haiku model determines the level of new investment in generation capacity and in postcombustion controls as well as the retirement of existing capacity. The model incorporates available information about planned units currently under construction. Generator dispatch in the model is based on the minimization of short-run variable costs of generation. All costs and prices are expressed in 1999 real dollars.

Interregional power trading is identified as the level of trading necessary to equilibrate regional electricity prices (accounting for transmission costs and power losses). These interregional transactions are constrained by the assumed level of available interregional transmission capability as reported by the North American Electric Reliability Council (NERC).

Factor prices, such as the cost of capital and labor, are held constant. Fuel price forecasts are calibrated to match EIA price forecasts from the *Annual Energy Outlook 2005* (EIA 2005a). Fuel market modules for coal and natural gas calculate prices that are responsive to factor demand. Coal is differentiated along several dimensions, including fuel quality and location of supply, and coal and natural gas prices are differentiated by point of delivery. All other fuel prices are specified exogenously.

For control of SO_2, coal-burning model plants are distinguished by the presence or absence of flue gas desulfurization (scrubbers). Unscrubbed coal plants have the option to add a retrofit SO_2 scrubber, and all plants select from a series of coal types that vary by sulfur content and price as a strategy to reduce SO_2 emissions. For control of NO_x, coal-, oil-, and gas-fired steam plants solve for the least costly postcombustion investment from the options of selective catalytic reduction (SCR) and selective noncatalytic reduction (SNCR); coal-fired plants also have the reburn option.

The model accounts for ancillary reductions in mercury associated with other postcombustion controls, including decisions to install retrofit SO_2 scrubbers and SCR

NO_x controls, and includes activated carbon injection (ACI) as another means of reducing mercury emissions. Using ACI only typically has a mercury removal efficiency of 90–95%, and adding on SO_2 wet scrubbers increases this rate to 97%. For bituminous coal, the combination of SCR and SO_2 wet scrubbers yields a removal efficiency of 90%, although this combination is not nearly as effective for sub-bituminous and lignite coal (Table 3B-4). In this analysis we ignore the mercury component of the model.

The Haiku model performs well in comparison with other models including ICF Consulting's Integrated Planning Model (IPM), the standard used for EPA regulatory impact analyses (RIAs) of policies directed at the electric power industry. The model has been compared with other simulation models as part of two series of meetings of Stanford University's Energy Modeling Forum (1998, 2001). In addition, the model is regularly cross-referenced with published studies of EIA and the U.S. Environmental Protection Agency (EPA), and the model performs well under comparable scenarios. The model has been used for several reports and articles that appear in the peer-reviewed literature (Burtraw et al. 2001; Burtraw et al. 2002; Burtraw et al. 2003a, 2003b; Banzhaf et al. 2004) and for analysis on behalf of the EPA and state governments (Palmer et al. 2002, Bharvirkar et al. 2003, Burtraw and Palmer 2003, Palmer et al. 2005).

Appendix 3C: The TAF Model

The output of the Haiku model is emissions of each pollutant by a representative plant within each of 13 North American Electric Reliability Council (NERC) subregions. The emissions are allocated to actual plant locations (latitude and longitude) on the basis of an algorithm that reflects historic utilization and the expected location of new investment. Changes in emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) that result from the policies are aggregated to the state level and fed into the nonproprietary and peer-reviewed Tracking Analysis Framework (TAF) integrated assessment model (Bloyd et al. 1996).[1] TAF integrates pollutant transport and deposition, human health effects, and valuation of these effects at the state level. It is similar to the U.S. Environmental Protection Agency 's (EPA's) Benefits Mapping and Analysis Program (BenMAP) model.

TAF Summary

The TAF model offers a choice of pollution transport modules, which have implications for the pollutants that are included. In the original version of TAF, pollution transport is estimated from seasonal source–receptor (S–R) matrices that are a reduced-form version of the Advanced Source Trajectory Regional Air Pollution (ASTRAP) model. These coefficients, which link SO_2 or NO_x emissions to concentrations of particulate matter of 2.5 microns or less ($PM_{2.5}$; see below) are still used in our base case analysis to link emissions from areas in the "western" United States to other U.S. states (see below).

The base case set of S–R coefficients for the "eastern" United States comes from another model that has been incorporated into TAF: the Urban-to-Regional Multiscale (URM) One Atmosphere Model, known as URM-1ATM. It includes the effects of changes in emissions of NO_x and SO_2 on fine particulate concentrations, as does the

[1] TAF was developed to support the National Acid Precipitation Assessment Program (NAPAP). Each module of TAF was constructed and refined by a group of experts in that field and draws primarily on peer-reviewed literature to construct the integrated model. TAF was subject to an extensive peer review in December 1995, which concluded, "TAF represent[s] a major advancement in our ability to perform integrated assessments" (ORNL 1995). The entire model is available at www.lumina.com\taflist. TAF has been repeatedly updated and is maintained by Resources for the Future (RFF).

original ASTRAP module, as well as the effects of changes in NO_x emissions on atmospheric ozone concentrations.

Health effects are characterized as changes in health status predicted to result from changes in air pollution concentrations. Effects are expressed as the number of days of acute morbidity effects of various types, the number of chronic disease cases, and the number of statistical lives lost. The health module is based on concentration–response functions found in the peer-reviewed literature, including epidemiological articles reviewed in EPA's Criteria Documents that, in turn, appear in key EPA cost–benefit analyses (U.S. EPA 1997, 1999). The health effects modeled are listed in Table 3C-1.

Of these effects, mortality effects are the most important. To characterize these effects, we use a cross-sectional study by Pope et al. (2002) for our default PM mortality concentration–response function. Although this study and others have documented the separate effects of particulate matter less than 10 microns in diameter (PM_{10}), $PM_{2.5}$, and sulfates (a constituent of $PM_{2.5}$) on mortality, none have documented the specific effect of nitrates. Accordingly, we use the separate Pope et al. estimates for the potency of sulfates but assume that nitrates have the potency of the average PM_{10} particle.

TAF assigns monetary values (taken from the environmental economics literature) to the health effects estimates produced by the health effects module. The benefits are totaled to obtain annual health benefits for each year modeled. For the most important aspect, the value of a statistical life, we have used an estimate of $2.324 million (2000$) from a recent meta-analysis by Mrozek and Taylor (2002) of 203 hedonic labor market estimates. This estimate is lower than that used in most previous work and less than half of the $5.5 million estimate used by EPA (U.S. EPA 1997, 1999). The most important reason for this discrepancy is the attribution of wage rate differentials to mortality rate differences in previous studies cited by EPA, whereas Mrozek and Taylor attribute a larger portion of the wage rate differentials to interindustry differences that occur for other reasons.[2] Base case assumptions for TAF are summarized in Table 3C-1.

As with past research, values for chronic morbidity effects (e.g., bronchitis) are transferred from individual studies, often using a conservative cost-of-illness approach. Values for acute effects are predicted from the meta-analysis of Johnson et al. (1997), which synthesized contingent valuation studies of morbidity effects based on their severity according to a health status index and other variables.

S–R Relationships

Two major modeling efforts are used to develop point estimates and uncertainty bounds for linking emissions to concentrations of fine particulates. The more advanced and accepted of the two is the URM model (discussed below), but this simulation model has the disadvantage of being parameterized only for the eastern United States. Thus, to be able to model national S–R relationships, in our base case we supplement this model with S–R relationships for the rest of the United States taken from the ASTRAP regression model. Because ASTRAP provides S–R coefficients for the nation, the coefficients from the eastern United States can be compared with those from the URM to examine the effect of model uncertainty (see below).

ASTRAP has another important feature for our project. It provides "uncertainties" around each of the S–R coefficients, on the basis of climatological variability, measured as statistical confidence intervals from the original regression analyses. URM does not have this feature. However, climatological variability is given expression through a novel scaling procedure that translates S–R information for three episodes into an annual S–R matrix. By varying parameters within this procedure

---

[2] There may be other reasons to suspect that the traditional values are too high. Labor market studies rely on the preferences of prime-age, healthy working males facing immediate and accidental risks of workplace mortality. In contrast, particulate pollution primarily affects seniors and people with impaired health status, and illness may not be apparent until years after initial exposure. This recognition has led to attempts to estimate values for life extensions (Johnson et al. 1998) and future risks (Alberini et al. 2004). New surveys that use contingent valuation to describe mortality risk reductions in a more realistic health context and that are applied to people of different ages and health status, Alberini et al. find that the implied values of statistical lives are far smaller than EPA's estimates, particularly for future risk reductions. However, the effects do not appear to be strongly related to age and, although many conjecture that poor health status would reduce willingness-to-pay, the study finds people in ill health tend to be willing to pay more for mortality risk reductions than people in good health. However, effects of dread and lack of controllability have not yet been factored into these new analyses.

randomly, we generate a set of S–R matrices spanning a reasonable range of climatological variability.

URM-1ATM and SRG Models[3]

This study takes output from URM-1ATM for several air pollution episodes at a detailed geographic scale and uses that information to construct aggregate S–R coefficients for state-level receptors using the Source–Receptor Generator (SRG) model.[4] The episode-specific S–R coefficients are aggregated to annual S–R coefficients using weights developed on the basis of a Classification and Regression Tree (CART) analysis of the episode data.[5] The models that are used to perform these tasks and how they work together are described below.

The URM-1ATM and the Regional Atmospheric Modeling System (RAMS) are used to account for the processes significantly affecting ozone and fine particulate concentrations in the atmosphere, including atmospheric physics, chemical reactions in the atmosphere, cloud and precipitation processes, and wet and dry deposition. RAMS is used to re-create the physics of a historical period of time, providing details and spatial coverage unavailable from observations. URM-1ATM solves the atmospheric diffusion equation (ADE) presented in Equation 3C-1 for the change in concentration, *c*, of pollutant of species *i* with time,

$$\frac{\partial c_i}{\partial t} + \nabla \bullet (\mathbf{u} c_i) = \nabla \bullet (\mathbf{K} \nabla c_i) + f_i + S_i \qquad \text{(3C-1)}$$

where **u** is a velocity field, **K** is the diffusivity tensor, f_i represents the production by chemical reaction of species *i*, and S_i represents sources and sinks of species *i*. As used here, a direct sensitivity capability using the Direct Decoupled Method in Three Dimensions (DDM-3D) is used to calculate the local sensitivities of specified model outputs simultaneously with concentrations (Odman et al. 2002, Russell et al.1988). As shown in Equation 3C-2, the sensitivity S_{ij} of a model output C_i (such as pollutant concentration of species *i*) to specified model inputs or parameters, P_j (e.g., emissions of

[3] Much of this discussion is taken from Shih et al. 2004.

[4] For more information on the URM-1ATM model, see Boylan et al. 2002 and Kumar et al. 1994.

[5] For more information about CART analysis, see Breiman et al. 1984.

NO_x from elevated sources) is calculated as the ratio of the change in output C_i to an incremental change of input or parameter P_j:

$$S_{ij} = \frac{\partial C_i}{\partial P_j} \tag{3C-2}$$

Equations 3C-1 and 3C-2 are solved concurrently and efficiently. The sensitivity in Equation 3C-2 is a local derivative, so a linear assumption is in effect when we extrapolate the result to a nonzero perturbation in emissions. This assumption has been well tested for the pollution concentrations of interest for this study, which include ozone and fine particulates. (A more detailed description of the model is available from Boylan et al. 2002 or Bergin et al. 2004.)

The URM-1ATM model uses a multiscale grid structure encompassing the eastern United States (Figure 3C-1). The finest grids are placed over major source regions (e.g., the Ohio River Valley, where many power plants and large industries are located) and over highly populated regions (e.g., East Coast corridor). This approach allows the evaluation of potential population exposure to pollutants and captures sources related to high population (e.g., automobile exhaust, fast food restaurants). The vertical grid has seven layers that allow different treatments of sources with low- and high-level stacks.

URM-1ATM is applied to three air quality episodes: February 9–13, 1994; May 24–29, 1995; and July 11–19, 1995. These episodes are used to represent winter, spring, and summer weather, respectively. These periods were selected because high-quality and complete data were available and previously modeled and because the data covered large meteorological variation with moderate to high pollution formation. Meteorological information is developed using RAMS as described in Pielke et al. 1992.

Sensitivities—the change in $PM_{2.5}$ concentrations per 1000 tons of NO_x reduction—from the URM-1ATM model are aggregated spatially on the receptor side using the SRG model. The hourly pollutant concentration sensitivity with respect to a uniform 30% reduction in emissions (by states and sources, both elevated and area) and population, for every grid in the entire study domain, are inputs to the SRG Model. The SRG program calculates spatially aggregated (receptor grids) S–R coefficients, population weighted and non-population weighted, for various averaging times (1 hour, 8 hours, and all day) for 22 receptor regions covering a 27-state area.[6] Population-weighted S–Rs are needed for estimating potential health benefits from application of source controls and also give a better proxy for health effects than do area-weighted measures. The area-weighted S–Rs are useful for elucidating the pure spatial and temporal effects of emissions on concentrations.

To use the output from the URM model, which is based on distinct episodes of 6–9 days, in seasonal or annual policy contexts, the episodes must be reweighted to reflect the entire season or year. To reweight the episodes, we follow Deuel and Douglas (1998) in using a CART approach. CART is a nonparametric regression technique that predicts discrete classification (e.g., high–medium–low) levels of a variable of interest (e.g., PM_{10} or ozone levels) by grouping observations on the basis of the similarity of predictive observable (e.g., independent) variables. The CART technique uses a binary decision tree to separate the different values of a classification variable (in this case, $PM_{2.5}$ monitored concentrations). The decision tree consists of a series of binary splits of the independent variables (in this case, meteorological variables), which are chosen to maximize the separation of the dependent variables. Each branch of the resulting tree defines subspaces of the classification variable that are the products of specific combinations of independent variables. These combinations are organized as bins.

Our independent variables include average humidity, precipitation, air pressure, average wind speed, resultant wind speed, temperature, and horizontal sigma (standard deviation of horizontal wind directions). Air quality and meteorological data for this analysis are taken from the Whiteface Mountain Base monitoring station. Other upper air meteorological data was obtained from Radiosonde Data of North America from the National Oceanic and Atmospheric Administration (NOAA). From this data set we used upper air observations from the airport at Albany, New York, as a proxy. We consider only one monitoring station because of budget constraints, but the approach can be easily generalized. Seasonal and annual weights are then based on the proportion of days in each bin for an entire 5-year (1992–1996) period experienced by New York relative to those in our episodes.

[6] Three sets of states and the District of Columbia fully in the model domain are aggregated into multistate receptor regions. Maine, New Hampshire, and Vermont are aggregated into a single region, as are Connecticut, Massachusetts, and Rhode Island and Delaware, Maryland, and the District of Columbia. In addition, 11 states on the western border of the eastern domain are aggregated into a single region.

Consider particulates as an example. First, we group the PM_{10} days into four classes on the basis of observed daily average PM_{10} concentrations (<6 micrograms per cubic meter [μg/m^3], 6–20 μg/m^3, 20–24 μg/m^3, and >24 μg/m^3). CART analysis results include the distribution of bins among the four classes. Use of the CART method provides an approach for segregating days into categories that are representative of certain observed meteorological or air quality conditions. The resulting population of each classification group provides information about the frequency of occurrence of specific meteorological or air quality regimes.

CART repeats the process of splitting for each child node, continuing recursively until further splitting is no longer possible. We then reweight the days in each class by the proportion of days in the season or year, relative to the episodes. For example, if the episodes have fewer PM_{10} days below 6 μg/m^3 relative to the yearly average number of days and more days above 24 μg/m^3, then we would underweight the former and overweight the latter. Then, within each class, we reweight each day by the proportion of days in the same cell of independent variables predicted to cause that class. For example, within the set of cells predicted to cause high-PM_{10} days (>24 μg/m^3), if, relative to the actual number of such days, we find more hot days where the previous day was cool and fewer back-to-back hot days, then we would under-weight the former and over-weight the latter. In this way, the episodes are reweighted to represent the outcomes of interest, and the various types of conditions associated with similar outcomes. We use information on PM_{10} to develop weights for S–R coefficients for fine particulates because only data on PM_{10} were available to us.

To calculate the weight for each of the previously selected episodic day, we use the following formula specifically:

$$1/N\left(\sum_{j=1}^{bk} P_j / \sum_{l_k} P_{lk}\right)\left(n_{lk}/m_{lk}\right)$$

where N is the total number of days included in the dataset; P_j is the number of days in bin j; b_k is the = number of bins with predicted value k of the classification variable; l_k runs over all bins with the classification value k, from which a representative day was chosen; n_j is the number of classified days in bin j; and m_j = number of representative days chosen from bin j.

The selected episodic day in a bin is scaled to the whole bin using the (n_{lk}/m_{lk}). This scale factor is simply the total number of days in the bin over the number of days selected from that bin. Because days are selected in only some of the bins, another scale factor (P_j/P_{lk}) is used to compensate for the bins from which no day is chosen. This double scaling is done for all the days in one class then for all the classes. Finally, the whole scaled factor is normalized by the total number of days N to obtain the weight for each of the previously selected episodic day. The summation of day weight for all days within each episode is used as the annual weight for that episodic S–R matrix. We generated 30 CART decision trees—hence, 30 sets of annual weights—to aggregate the episodic S–R matrices to obtain annual S–R matrices. Table 3C-2 shows 30 sets of annual weights for three episodic S–R matrices.

In this case study, we investigate the sensitivity of the annual weight, for each episodic S–R matrix, with respect to different classification schemes. We randomly generate 30 sets of classification schemes to conduct the CART analysis, which lead to 30 S–R matrices.[7]

We then developed a "potency" index to characterize the average size of the S–R coefficients in each matrix and their impact on benefits. The state-level index takes account of the fraction of emissions reductions in a state and the fraction of the population living in a state. We constructed a weighted average vector from premultiplying the S–R matrix by the state percentage of total national change in emissions (between baseline and policy emissions) and postmultiplying by the state shares of national population. From this calculation, we found that the potency index ranges over 14% and matches the change in mortality benefits from $PM_{2.5}$ reductions (for the states in the URM domain). Because the URM domain is only the eastern United States, capturing the effect of emission changes on receptors in this region but not outside the region, the difference between the highest- and lowest-potency S–R matrices at a national scale varies by only 2%. The S–R matrix with the median potency was used for the base case (Table 3C-3).

ASTRAP Model

The ASTRAP S–R matrices are produced by using regression analysis from data on monitored concentrations and various climatological variables (wind, temperature,

---

[7] In general, the sample size should be determined by (a) the level of confidence required, (b) margin of error tolerated, (c) variability in population studied, and (d) resources available. Sample size 30 is widely used in simulation studies. It is a result involving the Central Limit Theorem.

precipitation) over 11 years. Regression estimates were developed for each of the four seasons. The model captures atmospheric chemistry as NO_x and SO_2 react to form nitrates and sulfates, which are constituents of PM_{10}. It estimates concentrations of these separate constituents of PM_{10}, plus gaseous NO_2 and SO_2. The results were validated against ambient concentration and deposition data by using historical emissions data. Confidence intervals (assuming a normal distribution) around the estimated S–R coefficients were then incorporated into TAF to represent climatological variability. This version of the atmospheric transport module limits benefits to only particulate-related health impacts; however, these impacts account for the vast majority of all benefits according to the major integrated assessment studies of the impacts of electricity generation (Krupnick and Burtraw 1996).

This model has produced results contained in TAF that have three uses in our analysis. The first is to supplement URM S–R relationships for (mostly western) states that lie outside of the domain of URM as part of the base case. The second is to address statistical uncertainty by "turning on" uncertainties in the S–R coefficients for this model (which are based on climatological variability) for the western United States and adding these uncertainties to those for the eastern states from the URM. The third is to address model uncertainty in the emissions to concentrations step by substituting ASTRAP coefficients for URM coefficients within the URM model domain. This is done using both the average coefficients from each model and the distributions from each model.

Appendix 3D: Details of Modeling Uncertainties in Population and Natural Gas Prices

Haiku uses constant elasticity demand functions to forecast demand for electricity by customer class, season, time block, and 13 North American Electric Reliability Council (NERC) subregions of the country. The demand elasticities in these functions come from the literature, and the constant terms are calibrated to match actual historical demand and price data in 1999. The constant terms in the demand function are then grown over time to match trends in demand growth from recent Energy Information Administration (EIA) demand projections.

The baseline electricity demand forecast developed for this project is based on annual electricity demand forecasts from the reference case scenario of the *Annual Energy Outlook* (AEO) for 2005 (EIA 2005a). The AEO 2005 forecasts in turn are based on the U.S. Census Bureau's middle series population projection. For this project we use a combination of state and national population projections from the U.S. Census Bureau.[1]

Alternative Demand Forecasts

For the uncertainty scenario analysis, we developed alternative electricity demand scenarios for all customer classes to be consistent with the alternative population projections obtained from the U.S. Census Bureau. These alternative forecasts are developed under the assumption that differences in regional electricity demand growth across scenarios are roughly proportional to differences in population growth across scenarios. Three sets of population projections (low, middle, and high) are constructed at the state level, and each is aggregated to the NERC subregion level on the basis of population mapping by census track to the NERC subregion that we developed using 2000 Census information. Then ratios of each alternative case to the middle case are calculated at the NERC subregion level and applied to the constant terms in the demand functions to generate alternative demand forecasts and associated complete runs of the Haiku model.

[1] Census projections and methodology are available online at http://www.census.gov/population/www/projections/popproj.html

The construction of the low and high state-level population projections combines three sets of forecasts: national projections from the final projections consistent with the 1990 Census (released in January 2000), interim national projections consistent with 2000 Census (released in March 2004) and interim state projections consistent with 2000 Census (released in April 2005). The national projections released in 2000 include a low, middle, and high projection series at the national level by year of age.[2] The more recent projections do not incorporate uncertainty. However, we take advantage of the most recent estimates by applying the uncertainty from the 2000 data to construct a low and high series around the current state projections.

First, we create a low and high series (by age) around the 2004 projections by multiplying the ratio of the 1999 low and middle series (or high and middle series) by the 2004 middle series projection for a given year for each age. Next, we allocate the high and low national projections to individual states on the basis of state shares of national population by age in the 2005 series. Specifically, we calculate the percentage of population of each age in each state for each year from the 2005 data and apply this to our newly constructed 2004 low and high series. Our result is a set of projections by age at the state level. The implicit assumption in this approach is that there is no state-to-state variation in fertility, mortality, or migration rates between the three national series. Or, put another way, the variation across states in these areas does not change as the total numbers change.

Figure 3D-1 displays our low, middle, and high Census projections for three age groups: population between the ages of 0 and 17 years, 18–64 years, and 65+ years. These projections are at the national level and exclude the population of Alaska and Hawaii; the Tracking Analysis Framework (TAF) model does not consider either of these states. Note that projections for the years 2000–2003 are the same for all three series. The 2004 Census middle projections use independently calculated estimates of past population for these years. Therefore, we do not alter these estimates for the low or high series.

Population projections are available from other sources as well. EIA (2005b) and the U.S. Environmental Protection Agency's (EPA's) Benefits Mapping and Analysis Program (BenMAP) model (Abt Associates 2003) both use unique projections. Each

[2] See Hollman et al. 2000 for a discussion of the U.S. Census Bureau's low, middle, and high methodology.

source uses its own assumptions in calculating future population levels, but baseline estimates all originate in Census 2000 data. Therefore, any long-term differences between series are because of alternative assumptions of the three components of change: fertility, mortality, and migration. Figures 3D-2 and 3D-3 compare Census, EIA, and BenMAP projections. Two graphs are necessary because the BenMAP projections exclude Alaska and Hawaii, whereas the EIA projections include all states; Census data can be adjusted to fit either case.

EIA and BenMAP are very similar to Census middle compared with Census high and low projections. In 2030, expected population for BenMAP is 6.2 million (1.7%) lower than Census middle, whereas Census low is 40.6 million (11.2%) lower than the middle series. EIA is 1.2 million (0.3%) higher than Census middle in 2025[3]; Census high is 44.2 million (12.6%) greater. The proximity of EIA, BenMAP, and Census middle is not surprising because all projections are based on Census data. Because of this similarity, for the rest of our analysis we restrict our attention to Census data. The main advantage to using Census data is that the three projection series (low, middle, and high) contain significant variation and have clear differences in the underlying assumptions. This is what we are most interested in for this project.

Changes to Natural Gas Prices

In Haiku, national natural gas prices are calculated using a linear supply curve whose slope and intercept are derived for each simulated year from AEO 2005 existing and forecasted data. The quantity of natural gas used to compute the price is the sum of the amount consumed in the electricity sector within the model and an estimate of natural gas consumed by other sectors. The national gas price calculated from the supply curve is then marked up for each NERC region on the basis of estimated transportation costs to that region. This regional markup is added to the price derived from the national supply curve.

For this project, the high gas price scenario is developed by multiplying the regional gas prices by 1.3, representing a 30% increase in the price of natural gas, over all regions and time periods. Similarly, the low gas price scenario features regional gas prices decreased by 30%. We also explore a higher gas price scenario of 70%.

[3] Note that EIA data extend to 2025, whereas Census and BenMAP projections go to 2030.

Appendix 3E: An Interactive Graphical Tool for Analyzing Uncertainty

The use of the cobweb plot in Chapter 3 does not do full justice to this technique's power. Thus, in this appendix we develop a simple "toy" model incorporating a threshold in the health effects to better illustrate how the cobweb plot can be used to analyze uncertainty.

Problem

There are two technologies: Clean-n-Costly and Cheap-n-Dirty. An emissions cap must be chosen. Demand produces emissions. As long as emissions are below the cap, the Cheap-n-Dirty technology is deployed. Once the cap is hit, additional Demand is satisfied by Clean-n-Costly. We assume for simplicity that the latter technology produces zero emissions.

The economic benefits are simply equal to demand. There are two types of costs:

- Health costs are equal to all the emissions produced, provided the emissions exceed a dose–response threshold. (Obviously, thresholds apply to concentrations or dose in reality. But our simplification is not material to the example.)
- Abatement costs are equal to the amount of demand that needs be met by Clean-n-Costly. (Full costs from Cheap-n-Dirty are scaled to zero.)

Value is defined as benefits minus costs.

The uncertainties in this problem are of three types:

- Demand is uncertain, its uncertainty is *statistical.*
- Emissions–response threshold (Threshold) is also uncertain; its uncertainty is *model.*
- Cap is a decision variable—the regulator chooses its value. Before choosing, its value is uncertain, and the uncertainty is sometimes termed *volitional.*

We implement this model as follows, purely for purposes of illustration:

- Demand Normal with mean = 100, standard deviation = 20,
- Cap Uniform[30, 130],
- Threshold Uniform[70,100], and
- All distributions are independent.

Analysis

Such a simple model nonetheless poses a complex dependence structure. The interactive tool for visualizing a sample of the joint distribution is called a cobweb plot. Many samples are drawn (in this case, 5000). The variables are arranged as vertical lines. Each sample represents one value of each variable; connecting these values with a jagged line, and one sample thus corresponds to one jagged line. The entire empirical distribution of 5000 samples is shown in Figure 3E-1 as 5000 jagged lines. The leftmost variable (in this case, Demand) is color coded. The highest and lowest sample values are shown above and below, respectively, the corresponding vertical lines.

We recognize the normal distribution for Demand and immediately see that Benefit is equal to Demand and that Clean-n-Costly is equal to abatement cost. Following the red and olive green lines, we see that high demand corresponds to high values of Clean-n-Costly and thus high abatement cost. Other relations are more complex.

Sometimes it is more revealing to plot the percentiles instead of the values themselves (Figure 3E-2). Here we see that Clean-n-Costly has an "atom" at zero—that is, in about 30% of the samples, demand does not exceed the cap, and the amount of Clean-n-Costly deployed is zero.

The real value of cobweb plots lies in the ability to use interactive conditional restrictions. We do this by selecting a subset of samples from the distribution, that is, a subset of the jagged lines. The software supports this clicking on an interval on one of the vertical lines; then all samples are removed except those passing through the selected interval. This operation may be repeated as often as there are samples left to select.

To simplify the visualization, we drop all "intermediate" variables and consider only Value together with Demand, Cap, and Threshold.

Suppose we ask which scenarios lead to high Value. We select the top 2% (say) of the Value variable. Figure 3E-3 shows four conditional cobwebs, corresponding to selecting samples whose Values are90, 70, 0, and in the interval [–45,–85], respectively. Each time we retain approximately 100 scenarios.

Note the nonmonotonic behavior of Cap; it first falls (statistically) as Value drops, and then goes up. One might infer that the best policy would be to set the cap at about 100, but this would be incorrect. Indeed, in Figure 3E-3 we have conditionalized on Value and not on Cap. Bayes' theorem reminds us that the distribution of Cap conditional on Value is NOT the same as the distribution of Value conditional on Cap. When we select the scenarios with Value in the top 2% (top left of Figure 3E-3), we see that the Caps for these scenarios lie close to 100. However, if we select all those scenarios with Cap close to 100 (top right of Figure 3E-4), we do *not* get only high Value scenarios but scenarios with lower value as well.

Taking a decision corresponds to removing volitional uncertainty, and that simply means choosing a cap (i.e., conditionalizing on a volitional variable). Figure 3E-4 shows four successive conditionalizations on Cap corresponding to approximately 125, 95, 70, and 35. We see that for Cap at about 95, the distribution of Value splits: half the conditional samples hit the top of the value distribution, the other half are below zero. This split is presaged in Figure 3E-1, where lines in Value are split into an upper and lower half. As the value of Cap is further lowered, we see that the negative Value scenarios disappear, whereas the positive scenarios drop in value. Thus, for a Cap at about 70, we have no negative Value scenarios and all Values are around 65. Dropping the Cap to 35 yields only Values around 20. The mean and standard deviation of Value for Cap = 95 are 16.7 and 40, respectively. For Cap = 70 they are 68 and 4, respectively. From this we would conclude that the best policy is to set Cap equal to 70.

The top right of Figure 3E-4 leads one to ask, "When the cap is set to 95, what is driving the split? Is the result driven by statistical uncertainty (e.g., demand) or model uncertainty (e.g., Threshold), or neither?" The answer is not apparent from the figure. To answer this we must conditionalize again, selecting first the high, then the low Values from the distribution obtained by selecting first Cap= 95 (Figure 3E-5).

The result, given that we chose Cap = 95, is driven by the model uncertainty (dose–response threshold) and not by statistical uncertainty. When the dose–response threshold is high, value is high, whereas value is low for lower thresholds. We cannot establish a similar relationship between demand (statistical uncertainty) and value.

Comparison with Standard Tools

For purposes of comparison, Table 3E-1 shows some of the standard measures of dependence and sensitivity, most of which are discussed in Chapter 2. The first two columns show the predicted and the base (in this case, input) variables, respectively. The third and fourth columns show the mean and standard deviation of the base variables. The product moment correlation and rank (or percentile) correlation are shown next. Roughly, they show the degree of linear and monotonic dependence between the predicted and base variable. The regression coefficient is the degree to which the predicted mean (Value) covaries with the mean of the base variables. The correlation ratio is always positive and shows the percentage of the variance of the predicted variable that can be explained by the base variable. The partial correlation is similar to the product moment correlation, after removing the effects of the other base variables. (It is interesting to note that the removal of the effects—e.g., of Demand and Cap—strengthens the linear relation between Value and Threshold.) Finally, multiple correlation is the correlation between the predicted variable and the best linear predictor based on all base variables. Relative to the cobweb plots, these measures suffer from being "global." They are averages over the entire set of samples and do not reveal the complex relations in the cobweb plots.

The two-dimensional scatter plots of value against the base variables, with regression lines, are shown in Figure 3E-6. Note that these regression lines are poor predictors of the cloud of points and hence are poor indicators of how Value covaries with each of the base variables. The second scatter plot shows Values against Cap. Here we can see the split in Value as Cap goes from its highest to its lowest value. For cap = 70, we get only the "higher part" of the cloud of points. The cobweb plots have the advantage of showing the other variables as well. We can see (as in Figure 3E-5) that values of Threshold are driving the split in Value for Cap between 80 and 100.

Conclusion

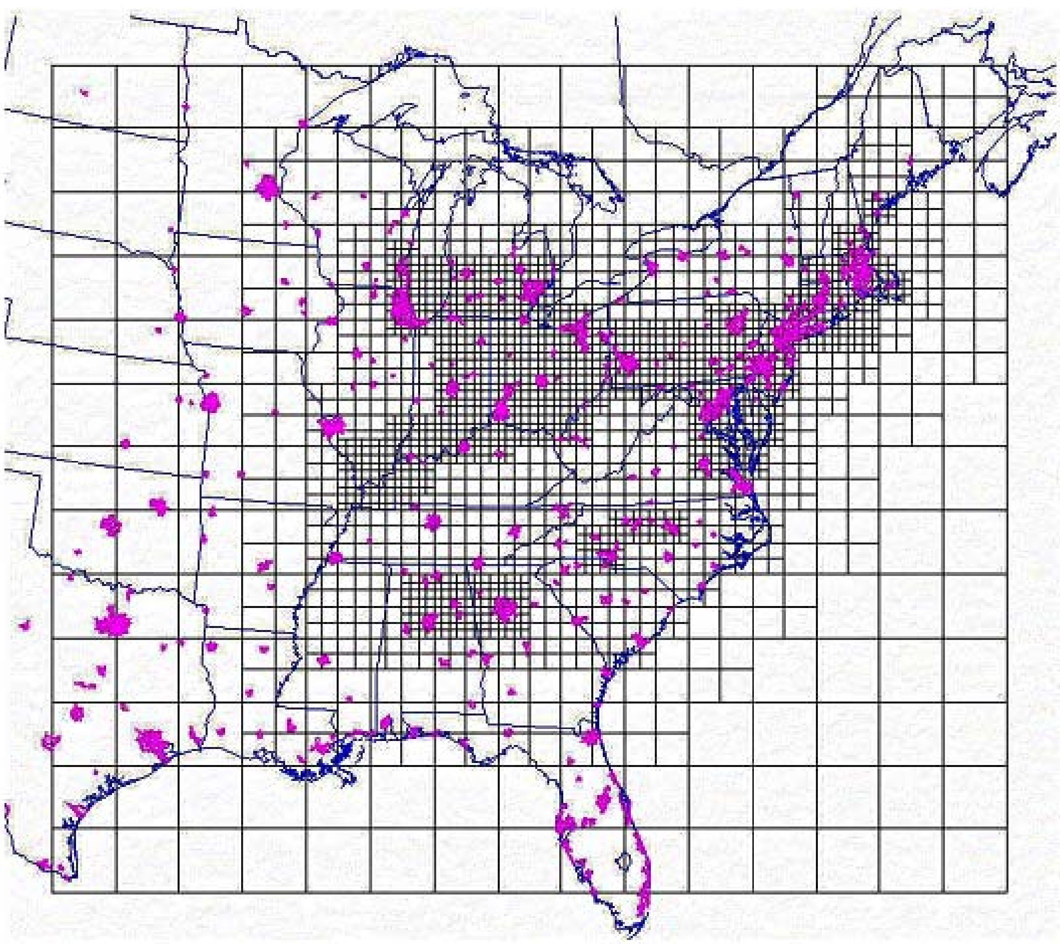
The interactions of the three types of uncertainties in this example show that even very simple problems can lead to very complex interrelationships that analysts—if not decisionmakers—must understand to describe results and ultimately contribute to better decisionmaking.

Appendix 3F: Descriptive Statistics for All Scenario Results

This appendix is available at www.rff.org/makingregulatorychoices.

Appendix Figures

Figure 3C-1. Multiscale Grid Used to Model Changes in Ozone and Particulate Species from Changes in NO_x and SO_2



Note: The finest resolution has horizontal grids of 24 km per side, and the other cells are 48 km, 96 km, and 192 km per side. The shaded areas represent high population densities (urban areas.) Fine-scale cells are placed over areas of high industrial or population densities.

Figure 3D-1. National Population Projections Based on U.S. Census Data: Low, Middle, and High Projections by Age Group, 2000–2030

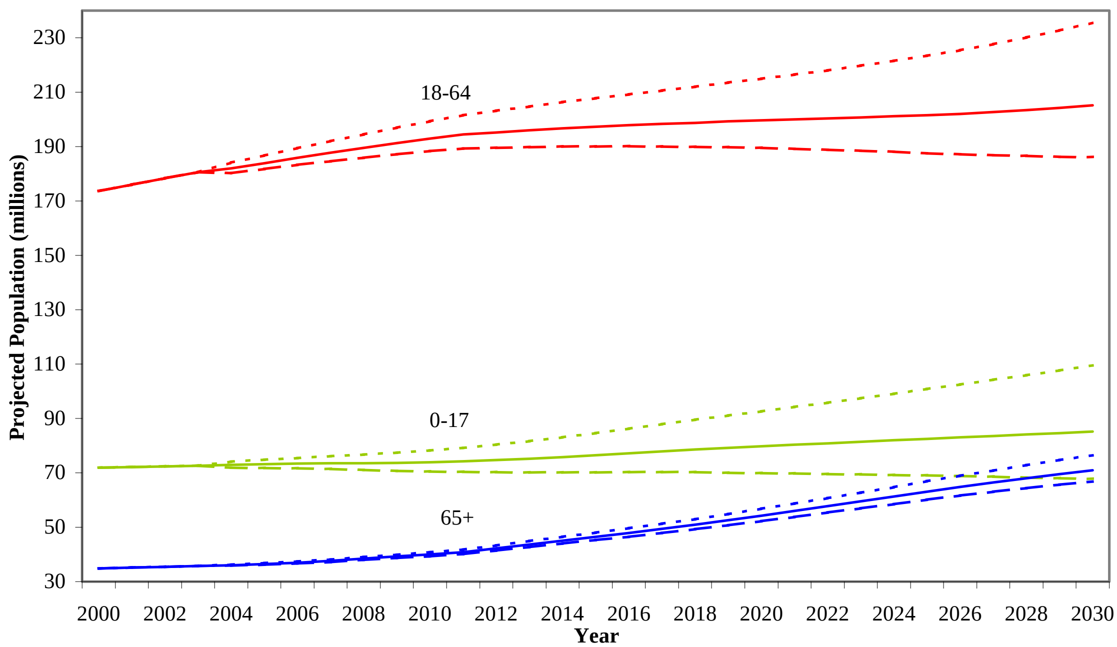


Figure 3D-2. U.S. National Population Projections: Census and BenMAP, 2000–2030 (Excluding AK and HI)

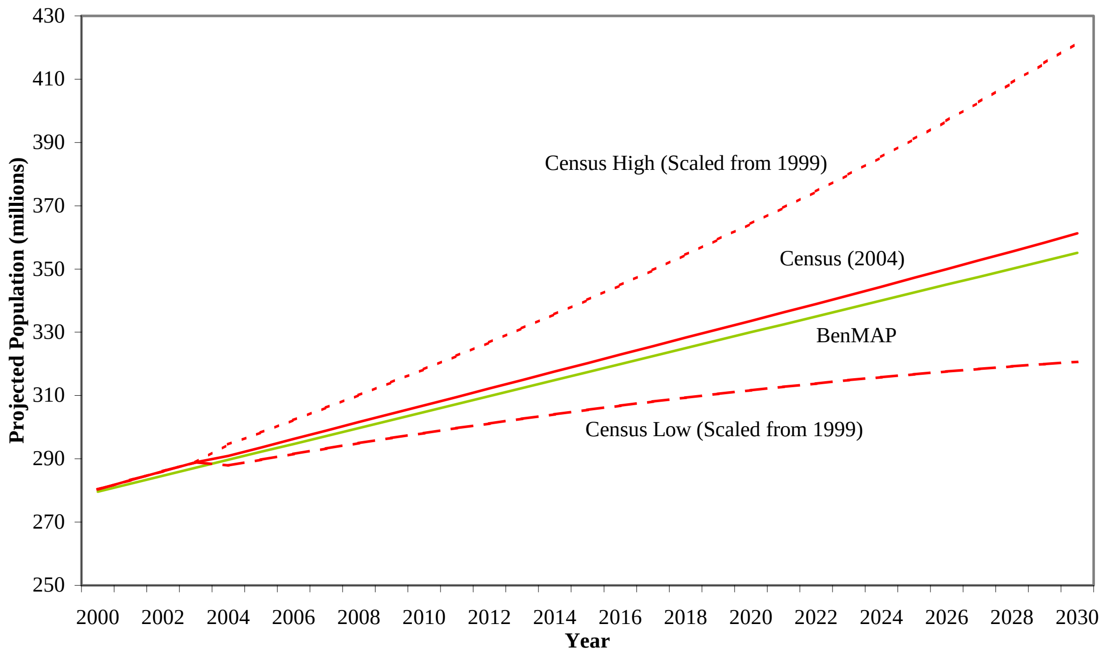


Figure 3D-3. U.S. National Population Projections: Census and EIA, 2000–2025 (Including AK and HI)

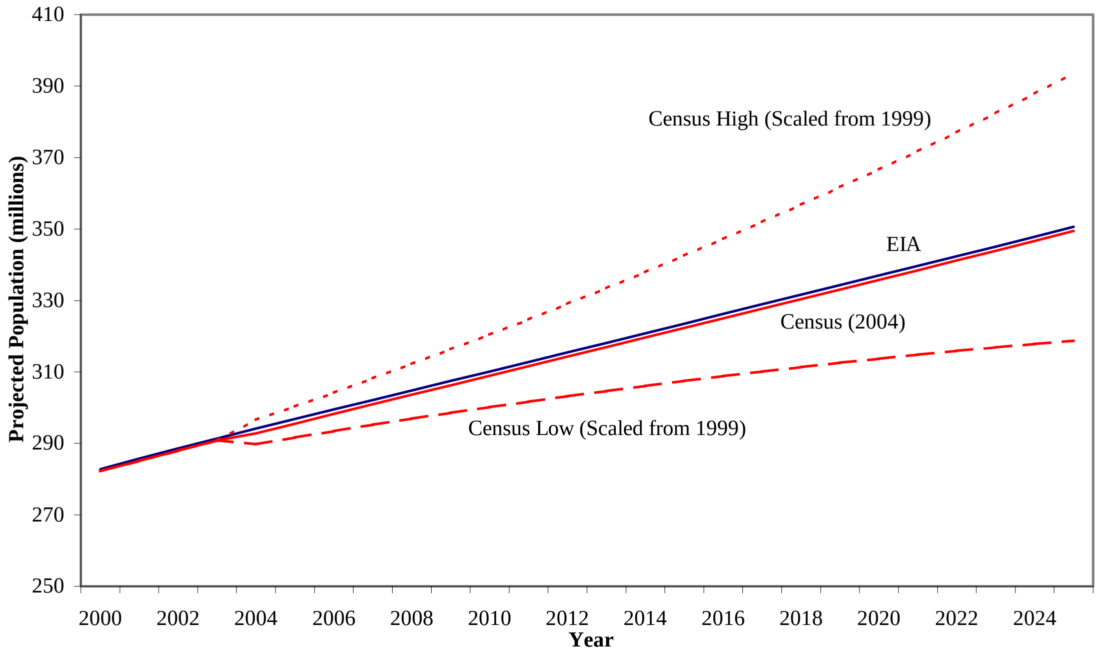


Figure 3E-1. Empirical Distribution (5000 samples)

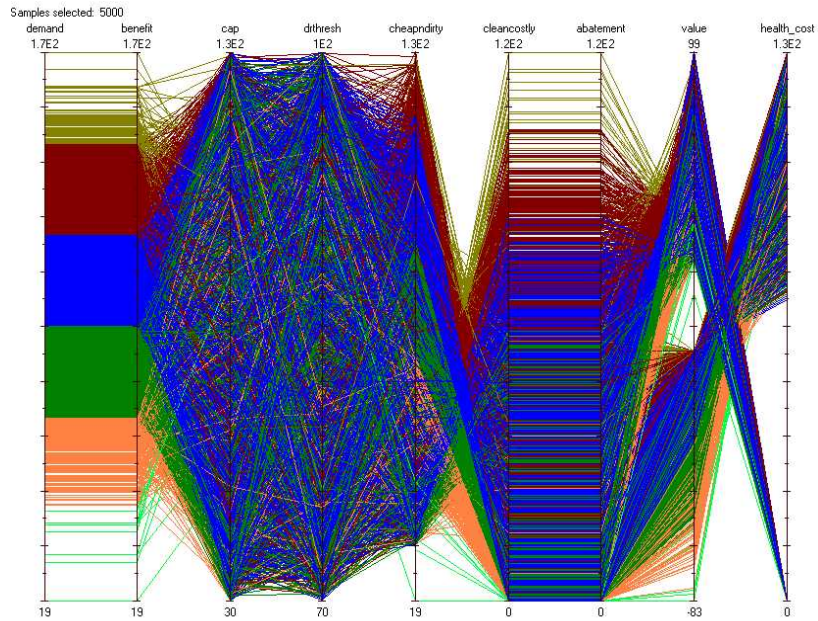


Figure 3E-2. Percentile Cobweb Plot

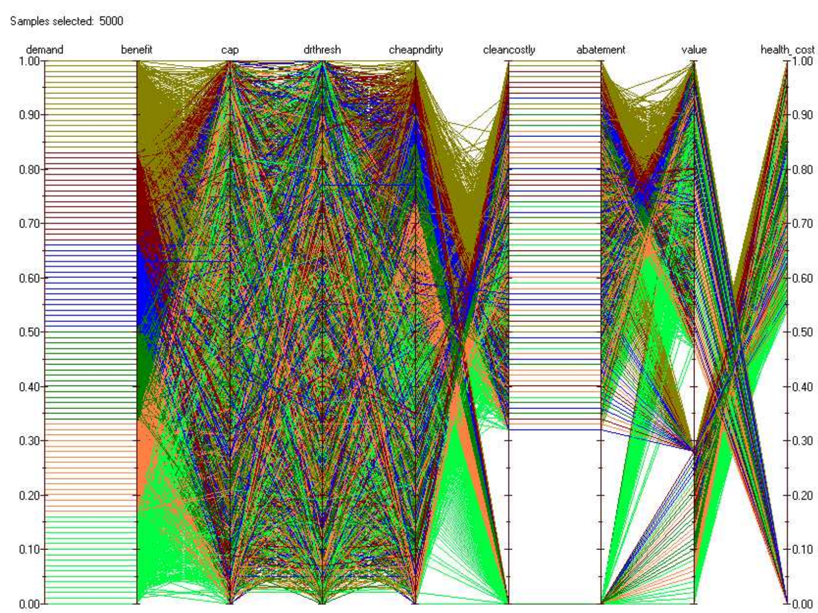


Figure 3E-3. Four Conditional Cobwebs for Value = (90,70,0, [–45,–85])

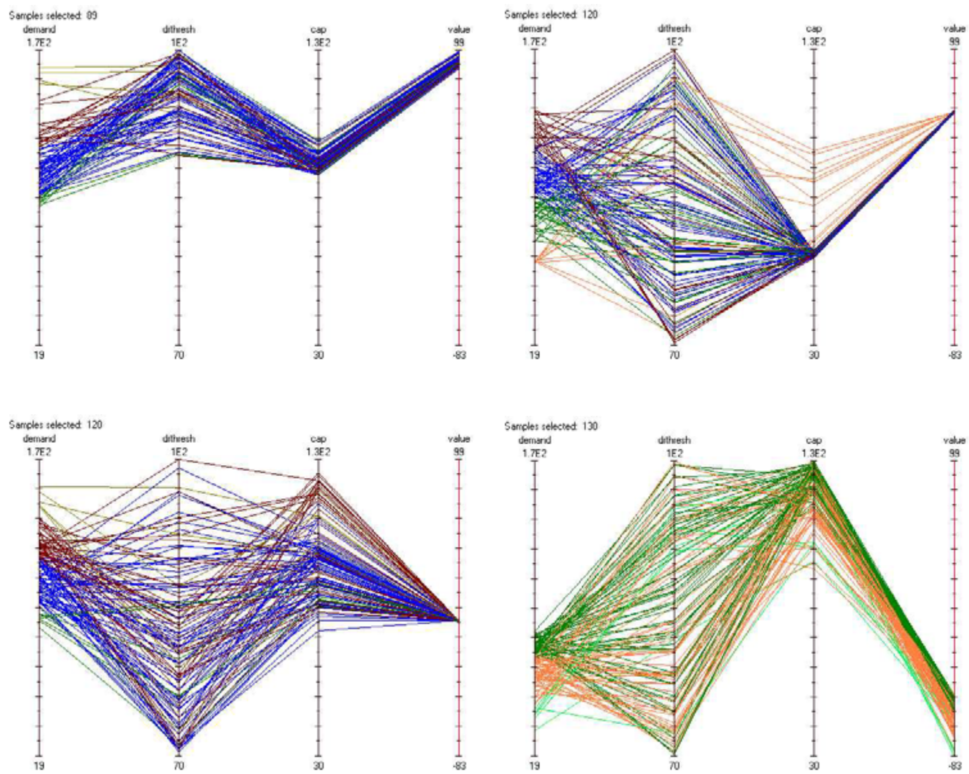


Figure 3E-4. Conditional Cobwebs on Cap = (125, 95, 70, 35)

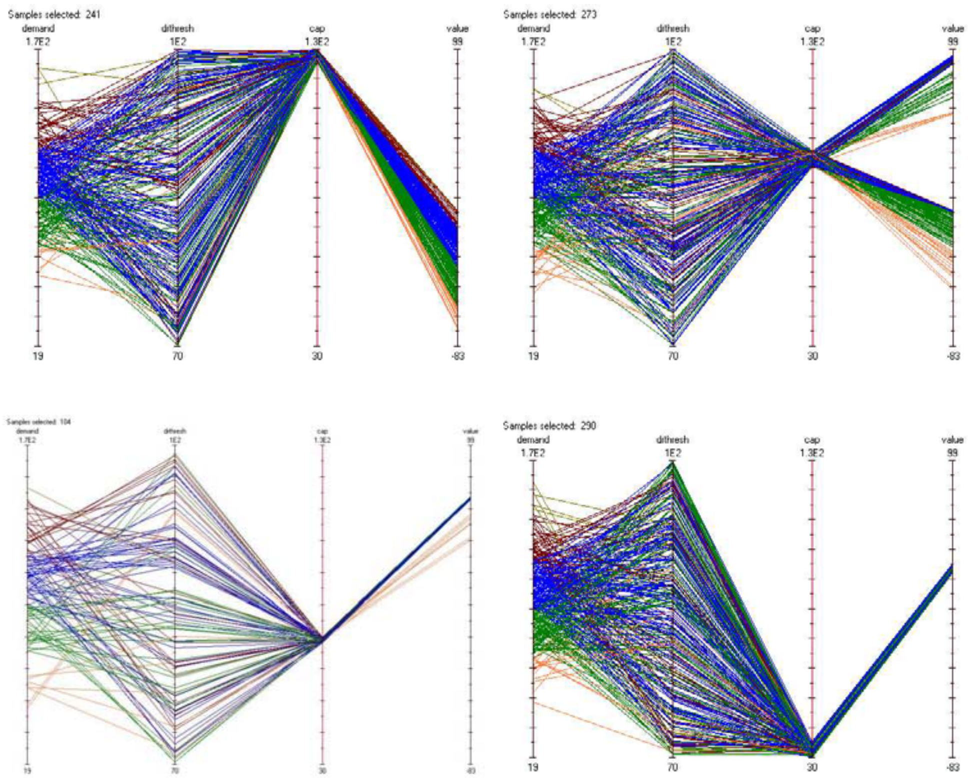


Figure 3E-5. "Double" Conditionalizing on Cap = 95 and Value = High, or Value = Low

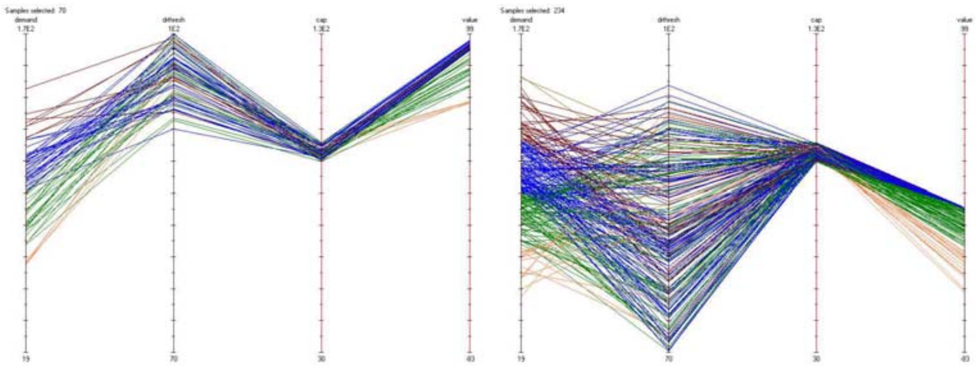
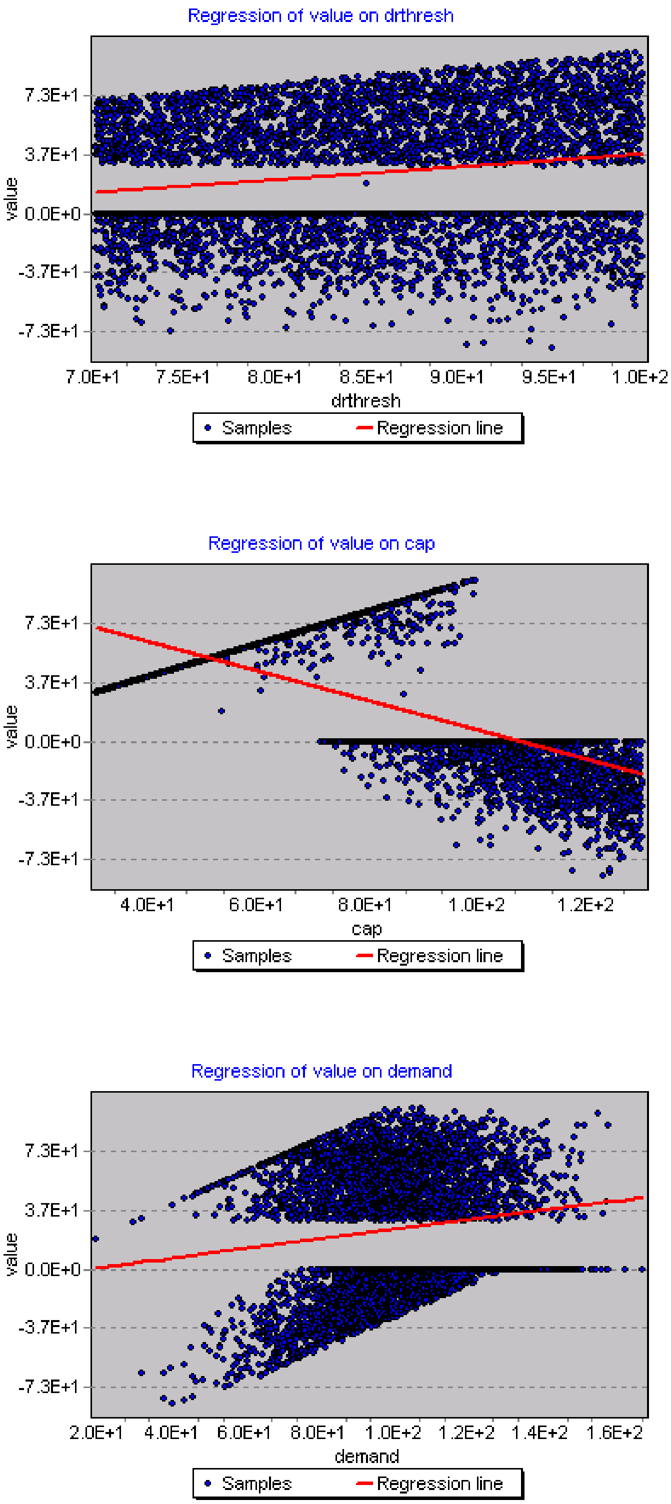


Figure 3E-6. Scatter Plots and Linear Regression Lines



Appendix Tables

Table 3A-1. Annual Emissions under Baseline (CAIR) and Policy as Modeled in Haiku

(tons)	*2010*	*2015*	*2025*
NO_x (million) Baseline	2.550	2.330	2.400
NO_x (million) Policy (Large)	2.400	1.500	1.500
NO_x (million) Policy (Moderate)	2.400	1.950	1.950
SO_2 (million)	6.078	5.001	3.500
Mercury	30.445	27.565	20.700

* NO_x caps include an adjustment of about 331,000 tons for units outside the Clean Air Interstate Rule (CAIR) NO_x region but within the Mid-Continent Area Power Pool (MAPP: MN, IA, NE, SD, ND, and parts of WI and IL) and New England electricity regions in the model

Table 3B-1. Inputs to the Haiku Model

Category	*Variables*	*Source*[a]
Existing generation		
	Capacity	EIA
	Heat rate	EIA
	Fixed and variable operations and maintenance cost	FERC, EIA, EPA
	Existing pollution controls	EPA, RFF
	Planned pollution controls	RFF
	Baseline emission rates	EPA (CEMS/NEEDS)
	Scheduled and unscheduled outage rates	NERC GADS data
New generation facilities		
	Capacity	EIA
	Heat rate	EIA, EPA
	Fixed and variable operating cost	EIA
	Capital cost	EIA
	Outage rates	NERC GADS data
Fuel supply		
	Wellhead supply curve for natural gas	EIA[b]
	Delivery cost for natural gas	
	Minemouth supply curve for coal by region and type of coal	EIA
	Delivery cost for coal	EIA
	Delivered oil price	EIA
Pollution controls		
	SO_2 cost and performance	EPA
	NO_x cost and performance	EPA
	Hg cost and performance	EPA
Transmission		
	Inter-regional transmission capacity	NERC
	Transmission charges	EMF
	Inter and intra regional transmission losses	EMF
Demand		
	Data year demand levels by season and customer class	EIA
	Load duration curve	RFF
	Trends in demand growth by customer class and region	EIA AEO 2004
	Elasticities by customer class	Economics literature

[a] Additional information on data is provided in Paul and Burtraw 2002.

[b] Interpolated on the basis of EIA forecasts.

Notes*:* EIA = Energy Information Administration, FERC = Federal Energy Regulatory Commission, EPA = U.S. Environmental Protection Agency, RFF = Resources for the Future, CEMS = Continuous Emissions Monitoring System, NEEDS = National Electric Energy System Database, NERC = North American Electric Reliability Council, GADS = Generating Availability Data System, EMF = Energy Modeling Forum, AEO= Annual Energy Outlook.

Table 3B-2. Mapping of Coal Supply Categories

Region	*Coal Category*	*2000 Million Short Tons**	*Haiku Coal Supply Mapping*
Northern Appalachia (PA, MD, OH, northern WV)		149.14	
	Medium sulfur (premium)	4.66	NA
	Low sulfur (bituminous)	0.36	NA
	Medium sulfur (bituminous)	72.61	NAMB
	High sulfur (bituminous)	61.41	NAHB
	High sulfur (gob)	10.10	NA
Central Appalachia (southern WV, VA, eastern KY)		258.40	
	Medium sulfur (premium)	47.16	NA
	Low sulfur (bituminous)	65.91	CSALB
	Medium sulfur (bituminous)	145.33	CSAMB
Southern Appalachia (AL, TN)		22.00	
	Low sulfur (premium)	6.82	NA
	Low sulfur (bituminous)	6.03	CSALB
	Medium sulfur (bituminous)	9.15	CSAMB
Eastern Interior (IL, IN, MS, western KY)		88.09	
	Medium sulfur (bituminous)	30.86	EIMB
	High sulfur (bituminous)	56.33	EIHB
	Medium sulfur (lignite)	0.90	NA
Western Interior (IA, MO, KS, OK, AR, TX)		2.42	
	High sulfur (bituminous)	2.42	NA
Gulf (TX, LA, AR)		53.02	
	Medium sulfur (lignite)	36.44	GLML
	High sulfur (lignite)	16.58	GLHL
Dakota (ND, eastern MT)		31.41	
	Medium sulfur (lignite)	31.41	DLML
Powder/Green River (WY, MT)		376.88	
	Low sulfur (bituminous)	1.21	NA
	Low sulfur (sub-bituminous)	345.74	PGLS
	Medium sulfur (sub-bituminous)	29.93	PGMS
Rocky Mountain (CO, UT)		55.80	
	Low sulfur (bituminous)	46.64	SWLB
	Low sulfur (sub-bituminous)	9.16	SWLS

Region	*Coal Category*	*2000 Million Short Tons**	*Haiku Coal Supply Mapping*
Arizona/New Mexico (AZ, NM)		40.43	
	Low sulfur (bituminous)	19.62	SWLB
	Medium sulfur (bituminous)	0.00	NA
	Medium sulfur (sub-bituminous)	20.81	SWMS
Washington/Alaska (WA, AK)		5.91	
	Medium sulfur (sub-bituminous)	5.91	NA

* Source: http://www.eia.doe.gov/oiaf/aeo/supplement/sup_ogc.pdf

Notes: The Haiku Coal Supply Mapping column indicates whether supply for a given region and coal category is modeled individually. Acronyms represent the region and coal category (e.g., NAMB = Northern Appalachian Medium Sulfur Bituminous, CSALB = Central and Southern Appalachian Low-Sulfur Bituminous). NA = region and coal combination not modeled individually.

Table 3B-3. Model Plant Types in Haiku

Existing Plants		*New or Planned Plants*	
	Natural Gas–Fired Combined Cycle		Coal Steam
	Oil Combined Cycle		Conventional Natural Gas–Fired Combined Cycle
	Efficient Natural Gas–Fired Gas Turbine		Natural Gas–Fired Combined Cycle, Combustion Turbine Duct
	Inefficient Natural Gas–Fired Gas Turbine		Conventional Natural Gas–Fired Gas Turbine
	Oil Gas Turbine		Landfill Gas Internal Combustion
	Conventional Hydro		Biomass IGCC
	Hydro Pumped Storage		Wind
	Solar		Advanced Natural Gas–Fired Combined Cycle
	Wind		Advanced Natural Gas–Fired Gas Turbine
	Biomass Steam		Geothermal
	Geothermal		Coal IGCC
	Efficient Natural Gas Steam		
	Inefficient Natural Gas Steam		
	Efficient Nuclear		
	Inefficient Nuclear		
	Oil Steam		
	MSW / Landfill Gas		
	Coal Steam*		

* The model includes several different categories of existing coal steam model plants, which are distinguished by the Energy Information Administration (EIA) coal demand region in which the model plant is located. This distinction brings the total number of model plants from the 29 listed here to 39.

Notes: IGCC = Integrated Gasification Combined Cycle, MSW = Municipal Solid Waste.

Table 3B-4. U.S. EPA Emissions Modification Factors for Mercury

Configuration			*EPA Mercury Removal (%)*		
SO_2 Control	*Particulate Control*	*NO_x Control*	*Bit Coal*	*Sub Bit Coal*	*Lignite Coal*
None	BH	NA	89	73	0
Wet	BH	None	97	73	0
Wet	BH	SCR	90	85	44
Dry	BH	NA	95	25	0
None	CSE	NA	36	3	0
Wet	CSE	None	66	16	44
Wet	CSE	SCR	90	66	44
Dry	CSE	NA	36	35	0
None	HSE/Other	NA	10	6	0
Wet	HSE/Other	None	42	20	0
Wet	HSE/Other	SCR	90	25	0
Dry	HSE/Other	NA	40	15	0

Notes: SO_2 controls: Wet = wet scrubber, Dry = dry scrubber. Particulate controls: BH = baghouse/fabric filter, CSE = cold-side electrostatic precipitator, HSE = hot-side electrostatic precipitator. NO_x controls: SCR = selective catalytic reduction, NA = not applicable; Bit = bituminous coal, Sub = sub-bituminous coal.

Source: http://www.epa.gov/clearskies/techinical.html

Table 3C-1. Base Case Assumptions

Value of a Statistical Life (VSL)		*$Year*
Study	*Weight*	*1999*
Hagler and Bailly (1995) All Ages		
Hagler and Bailly (1995) Age Weighted		
Hagler and Bailly (1995) Under 65		
Hagler and Bailly (1995) Over 65		
Fisher et al. (1989)		
Mrozek and Taylor (2002)	1	
Krupnick et al. (2002)		
BenMAP Normal Distribution		
BenMAP Uniform Distribution		
BenMAP Triangular Distribution		

Population

Census "Middle" Projections

Source Receptor Matrices for PM

Combination of URM and ASTRAP Coefficients - ASTRAP for states not covered by URM Annual NO_x Elevated Matrix #18 (median)

Source Receptor Matrices for Ozone

NO_x Emissions are Grouped by Elevation

$PM_{2.5}$ Options

PM Concentration–Response Functions and Weights		PM Valuation Studies and Weights	
Mortality			
30 and Up		Chosen VSL Study (from above)	
Pope (2002)	1	Mrozek and Taylor (2002)	1
Krewski (2000)			
Dockery (1993)			
Under 1			
Woodruff (1997)	1		

Value of a Statistical Life (VSL)		*$Year*	
Study	*Weight*	*1999*	

Chronic Bronchitis (CB)

Abbey (1995)	1	COI 3% Discount	
		COI 7% Discount	
		WTP Average Severity	1

Nonfatal Heart Attacks (NFHA)

Peters et al. (2001) 18 Up	1	10 Year Med - 5 Year Wage - 3% Discount	1
		10 Year Med - 5 Year Wage - 7% Discount	

Respiratory Hospital Admissions (RHA)

Burnett (1997) None	1	RHA All Ages Valuation	1
Burnett (1997) O_3			
Burnett (1997) NO_2, O_3, SO_2			
Thurston (1994) None			
Thurston (1994) O_3			

Cardiovascular Hospital Admissions (CHA)

Elderly 65 and Up		Elderly 65 and Up	
Moolgavkar (2003) All Cardio	0.979	CHA Valuation 65 Up	1
Ito (2003) Ischemic	0.007		
Ito (2003) Dysrhythmia	0.007		
Ito (2003) Heart Failure	0.007		
18–64		18 and Up	
Moolgavkar (2000) All Cardio	1	CHA Valuation 18-64	1

Asthma Emergency Room Visits (AERV)

Norris et al. (1999) Under 18	1	Smith (1997)	0.5
		Standford (1999)	0.5

Acute Bronchitis in Children (ABiC)

Dockery et al. (1996) 8-12	1	ABiC 1 Day Illness	
		ABiC 6 Day Illness	1
		Dickie and Ulery (2002)	

Value of a Statistical Life (VSL)		*$Year*
Study	*Weight*	*1999*

Upper Respiratory Symptoms in Children (URSiC)

Pope et al. (1991)	1	WTP 1 Day CV WTP 2 Symptoms 1 Day CV Dickie and Ulery (2002)	1

Lower Respiratory Symptoms in Children (LRSiC)

Schwartz and Neas (2000)	1	WTP 1 Day CV WTP 2 Symptoms 1 Day CV Dickie and Ulery (2002)	1

Asthma Exacerbations (AE)

Ostro (2001) Cough	0.3718	WTP 1 Symptom Day Dickie and Ulery (2002)	
Ostro (2001) Wheeze	0.2436	WTP 1 Bad Asthma Day Rowe and Chestnut (1986)	1
Ostro (2001) Short Breath	0.3846	WTP 2x Bad Asthma Day Rowe and Chestnut (1986)	
Vedal (1998) Cough			

Work Loss Days (WLD)

Ostro (1987) 18–64	1	WLD Valuation	1

Minor Restricted Activity Days (MRAD)

Ostro and Rothschild (1989)	1	WTP 1-Day WTP 3 Symptoms Dickie and Ulery (2002)	1

Ozone Choices

Ozone Concentration Response Functions and Weights	Ozone Valuation Studies and Weights

Respiratory Hospital Admissions

Value of a Statistical Life (VSL)		*$Year*	
Study	*Weight*	*1999*	
Elderly 65 and Up		Elderly 65 and Up	
Schwartz All 1995 A		RHA 65 Up Valuation	1
Schwartz All 1995 B	0.5		
Schwartz All 1995 C			
Schwartz All 1995 D	0.5		
Infant <2		Infant <2	
Burnett 2001 A		RHA Under 2 Valuation	1
Burnett 2001 B	1		
Burnett 2001 C			

Asthma Emergency Room Visits

Weisel et al. (1995)	0.5	Smith (1997)	0.5
Cody et al. (1992)	0.5	Standford (1999)	0.5

School Absence Days

Gilliland (2001)	0.08	SAD Valuation study	1
Chen (2000)	0.92		

Minor Restricted Activity Days

Ostro (1989)	1	WTP 1 Day	1
		Dickie and Ulery (2002)	

Short-Term Mortality

Ito and Thurston (1996) A		Chosen VSL Study (from above)	
Ito and Thurston (1996) B	0.0825	Mrozek and Taylor (2002)	1
Moolgavkar et al. (1995)	0.45		
Samet et al. (1997) A			
Samet et al. (1997) B	0.2175		
Bell (2004)	0.25		

Table 3C-2. Annual Weights for February, May, and July Episodes

	February	*May*	*July*
1	0.551	0.221	0.227
2	0.512	0.267	0.221
3	0.486	0.304	0.210
4	0.262	0.477	0.261
5	0.552	0.256	0.192
6	0.494	0.240	0.266
7	0.565	0.227	0.209
8	0.531	0.234	0.236
9	0.477	0.287	0.236
10	0.468	0.270	0.262
11	0.547	0.231	0.222
12	0.400	0.336	0.264
13	0.536	0.254	0.209
14	0.557	0.204	0.239
15	0.552	0.193	0.256
16	0.552	0.179	0.269
17	0.423	0.359	0.218
18	0.526	0.250	0.224
19	0.329	0.440	0.231
20	0.570	0.172	0.258
21	0.418	0.299	0.283
22	0.547	0.219	0.234
23	0.573	0.190	0.238
24	0.515	0.315	0.169
25	0.575	0.170	0.255
26	0.512	0.260	0.227
27	0.554	0.259	0.186
28	0.345	0.431	0.224
29	0.529	0.217	0.254
30	0.438	0.299	0.262

Table 3C-3. S-R Matrix Potency Index Numbers for 30 Annual Elevated NO_x Source Receptor Matrices

Annual Elevated NO_x Matrix	*Index Number*
1	0.01774
2	0.01808
3	0.01833
4*	0.01995
5	0.01786
6	0.01806
7	0.01770
8	0.01788
9	0.01831
10	0.01829
11	0.01779
12	0.01882
13	0.01792
14	0.01766
15	0.01764
16	0.01759
17	0.01880
18*	0.01796
19	0.01951
20	0.01748
21	0.01861
22	0.01775
23	0.01753
24	0.01823
25*	0.01745
26	0.01805
27	0.01786
28	0.01941
29	0.01783
30	0.01852

* Matrix 4 is the largest and Matrix 25 is the smallest, with a 14.29% difference. We use Matrix 18 as the median and default matrix.

Table 3E-1. Standard Sensitivity Measures

Predicted Variable	*Base Variable*	*E[Base Variable]*	*Std[Base Variable]*	*Product Moment Correlation*	*Rank Correlation*	*Regression Coeff.*	*Correlation Ratio*	*Partial Correlation Coeff.*	*Multiple Correlation Coeff.*
Value	demand	99.51366	19.82014	0.149063	0.159445	0.292571	0.027128	0.205414	
Value	cap	80.37920	28.92056	–0.671307	–0.624329	–0.902990	0.579996	–0.692537	0.711889
Value	Thresh	84.71102	8.659721	0.175458	0.173058	0.788205	0.031615	0.252461	

References

Abbey, D.E., B.E. Ostro, F. Petersen, and R.J. Burchette. 1995. Chronic Respiratory Symptoms Associated with Estimated Long-Term Ambient Concentrations of Fine Particulates Less Than 2.5 Microns in Aerodynamic Diameter ($PM_{2.5}$) and Other Air Pollutants. *Journal of Exposure Analysis and Environmental Epidemiology* 5(2): 137–159.

Abt Associates, Inc. 2003. BenMAP Environmental Benefits Mapping and Analysis Program (Version 1.0) and User's Manual. Bethesda, MD: Abt Associates, Inc. Prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards.

Alberini, A., M. Cropper, A. Krupnick, and N. Simon. 2004. Does the Value of Statistical Life Vary with Age and Health Status? Evidence from the U.S. and Canada. *Journal of Environmental Economics and Management* 48(1): 769–792.

Banzhaf, S., D. Burtraw, and K. Palmer. 2004. Efficient Emission Fees in the U.S. Electricity Sector. *Resource and Energy Economics* 26(3): 317–341.

Bell, M.L., A. McDermott, S.L. Zeger, J.M. Samet, and F. Dominici. 2004. Ozone and Short-Term Mortality in 95 US Urban Communities, 1987–2000. *The Journal of the American Medical Association* 292(19): 2372–2378.

Bergin, M.S., J.W. Boylan, J.G. Wilkinson, J.-S. Shih, A.J. Krupnick, and A.G. Russell. 2004. Effects of Multiscale Grid Resolution and Spatial Distribution on 3D Model Performance. Unpublished manuscript.

Bergin, M.S., A.G. Russell, and J.B. Milford. 1998. Effects of Chemical Mechanism Uncertainties on the Reactivity Quantification of Volatile Organic Compounds Using a Three-Dimensional Air Quality Model. *Environmental Science and Technology* 32(5): 694–703.

Bharvirkar, R., D. Burtraw, and A. Krupnick. 2003. *NO_x Emissions Trading and Episodic Air Quality Violations In Maryland. Final Report.* Washington, DC: Resources for the Future. Prepared for the Power Plant Research Program, State of Maryland.

Bloyd, C., J. Camp, G. Conzelmann, J. Formento, J. Molburg, J. Shannon, M. Henrion, R. Sonnenblick, K. Soo Hoo, J. Kalagnanam, S. Siegel, R. Sinha, M. Small, T. Sullivan, R. Marnicio, P. Ryan, R. Turner, D. Austin, D. Burtraw, D. Farrell, T. Green, A. Krupnick, and E. Mansur. 1996. *Tracking and Analysis Framework (TAF) Model Documentation and User's Guide: An Interaction Model for Integrated Assessment of Title IV of the Clean Air Act Amendments.* Argonne, IL: Argonne National Laboratory.

Boylan, J.W., M.T. Odman, J.G. Wilkinson, A.G. Russell, K. Doty, W. Norris, and R. McNider. 2002. Development of a Comprehensive, Multiscale "One Atmosphere" Modeling System: Application to the Southern Appalachian Mountains. *Atmospheric Environment* 36(23): 3721–3734.

Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Burtraw, D., K. Palmer, R. Bharvirkar, and A. Paul. 2001. Cost-Effective Reduction of NO_x Emissions from Electricity Generation. *Journal of Air and Waste Management* 51(10): 1476–1489.

Burtraw, D., K. Palmer, A. Paul, and R. Bharvirkar. 2002. The Effect on Asset Values of the Allocation of Carbon Dioxide Emission Allowances. *The Electricity Journal* 15(5): 51–62.

Burtraw, D., R. Bharvirkar, and M. McGuinness. 2003a. Uncertainty and the Net Benefits of NO_x Emissions Reductions from Electricity Generation. *Land Economics* 79(3): 382–401.

Burtraw, D., A. Krupnick, K. Palmer, A. Paul, M. Toman, and C. Bloyd. 2003b. Ancillary Benefits of Reduced Air Pollution in the United States from Moderate Greenhouse Gas Mitigation Policies in the Electricity Sector. *Journal of Environmental Economics and Management* 45(3): 650–673.

Burtraw, D., and K. Palmer. 2003. *Economic Efficiency and Distributional Consequences of Different Approaches to NO_x and SO2 Allowance Allocation. Final Report.* Washington, DC: Resources for the Future Prepared for the U.S. Environmental Protection Agency. http://www.epa.gov/air/clearskies/pdfs/allocationstudy.pdf. Accessed Feb. 1, 2006.

Deuel, H.P., and S.G. Douglas. 1998. *Episode Selection for the Integrated Analysis of Ozone, Visibility, and Acid Deposition for the Southern Appalachian Mountains.* Technical Report SYSAPP-98/07rl. Prepared for Southern Appalachian Mountains Initiative. San Rafael, CA: ICF Kaiser, Systems Applications International.

Dockery, D.W., C.A. Pope, X.P. Xu, J.D. Spengler, J.H. Ware, M.E. Fay, B.G. Ferris, and F.E. Speizer. 1993. An Association Between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine* 329(24): 1753–1759.

EIA (Energy Information Administration). 2005a. *Annual Energy Outlook 2005 With Projections to 2025.* Washington, DC: U.S. Department of Energy, Office of Integrated Analysis and Forecasting.

———. 2005b. State Population Projections by Census Region, as used in the National Energy Modeling System (NEMS). (Available on request from infoctr@eia.doe.gov.)

Energy Modeling Forum. 1998. *A Competitive Electricity Industry.* Final Report of EMF Working Group 15. Stanford, CA: Stanford University.

———. 2001. *Prices and Emissions in a Restructured Electricity Market.* Final Report of EMF Working Group 17. Stanford, CA: Stanford University.

Fisher, A., L.G. Chestnut, and D.M. Violette. 1989. The Value of Reducing Risks of Death: A Note on New Evidence. *Journal of Policy Analysis and Management* 8(1): 88–100.

Hagler Bailly Consulting, Inc. 1995. *Human Health Benefits Assessment of the Acid Rain Provisions of the 1990 Clean Air Act Amendments.* Final Report. Boulder, CO: Prepared by Hagler Bailly under subcontract to ICF Incorporated (Fairfax, VA) for U.S. EPA, Acid Rain Division.

Haimes, Y.Y. 2004. *Risk Modeling, Assessment, and Management* (2nd ed.). Hoboken, NJ: Wiley.

Hanna, S.R., Z. Lu, H.C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau, and D.A. Hansen. 2001. Uncertainties in Predicted Ozone Concentrations due to Input Uncertainties for the UAM-V Photochemical Grid Model Applied to the July 1995 OTAG Domain. *Atmospheric Environment* 35(5): 891–903.

Hollman, F.W., T.J. Mulder, and J.E. Kallan. 2000. Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100. Working Paper No. 38. Washington, DC: U.S. Census Bureau, Population Division, Population Projections Branch.

Ito, K., and G.D. Thurston. 1996. Daily PM_{10}/Mortality Associations: An Investigation of At-Risk Subpopulations. *Journal of Exposure Analysis and Environmental Epidemiology* 6(1): 79–95.

Johnson, F.R., E.E. Fries, and H.S. Banzhaf. 1997. Valuing Morbidity: An Integration of the Willingness to Pay and Health Status Index Literatures. *Journal of Health Economics* 16(6): 641–665.

Johnson, F.R., W.H. Desvousges, M.C. Ruby, D. Stieb, and P. de Civita. 1998. Eliciting Stated Health Preferences: An Application to Willingness to Pay for Longevity. *Medical Decision Making* 18(2): S57–S67.

Krewski, D., R. Burnett, M. Goldberg, K. Hoover, J. Siemiatycki, M. Jerrett, M. Abrahamowicz, and M. White. 2000. *Reanalysis of the Harvard Six Cities Study and* *the American Cancer Society Study of Particulate Air Pollution and Mortality.* Cambridge, MA: Health Effects Institute.

Krupnick, A., and D. Burtraw. 1996. The Social Costs of Electricity: Do the Numbers Add Up? *Resource and Energy Economics* 18(4): 423–466.

Krupnick, A.J., A. Alberini, M. Cropper, N. Simon, B.J. O'Brien, R. Goeree, and M. Heintzelman. 2002. Age, Health, and the Willingness to Pay for Mortality Risk Reductions: A Contingent Valuation Survey of Ontario Residents. *Journal of Risk and Uncertainty* 24(2): 161–186.

Kumar, N., M.T. Odman, and A.G. Russell. 1994. Multiscale Air Quality Modeling: Application to Southern California. *Journal of Geophysical Research* 99(D3): 5385–5397.

Moolgavkar, S.H., E.G. Luebeck, T.A. Hall, and E.L. Anderson. 1995. Air Pollution and Daily Mortality in Philadelphia. *Epidemiology* 6(5): 476–484.

Mrozek, J.R., and L.O. Taylor. 2002. What Determines the Value of Life? A Meta-Analysis. *Journal of Policy Analysis and Management* 21(2): 253–270.

ORNL (Oak Ridge National Laboratory). 1995. *Peer Review of the Tracking and Analysis Framework (TAF) For Use in the 1996 NAPAP Integrated Assessment.* ORNL/M-4994. Oak Ridge, TN: ORNL.

Odman, M.T., J.W. Boylan, J.G. Wilkinson, and A.G. Russell. 2002. Integrated Modeling for Air Quality Assessment: The Southern Appalachian Mountains Initiative Project. *Journal de Physique IV* 12(PR10): 211–234.

OMB (Office of Management and Budget). 2003. *Circular A-4*. Washington, DC: OMB.

Palmer, K., D. Burtraw, A. Paul, and R. Bharvirkar. 2002. *Electricity Restructuring, Environmental Policy and Emissions: Insights from a Policy Analysis.* Report. Washington, DC: Resources for the Future.

Palmer, K., D. Burtraw, and J.-S. Shih. 2005. Reducing Emissions from the Electricity Sector: The Costs and Benefits Nationwide and in the Empire State. Discussion Paper 05-23. Washington, DC: Resources for the Future. (Also published by the New York State Energy Research and Development Authority, Report 05-02.)

Paul, A., and D. Burtraw. 2002. *The RFF Haiku Electricity Market Model.* Report. Washington, DC: Resources for the Future.

Pielke, R.A., W.R. Cotton, R.L. Walko, C.J. Tremback, W.A. Lyons, L.D. Grasso, M.E. Nicholls, M.D. Moran, D.A. Wesley, T.J. Lee, and J.H. Copeland. 1992. A Comprehensive Meteorological Modeling System: RAMS. *Meteorology and Atmospheric Physics* 49: 69–91.

Pope, C.A., R.T. Burnett, M.J. Thun, E.E. Calle, D. Krewski, K. Ito, and G.D. Thurston. 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to

Fine Particulate Air Pollution. *The Journal of the American Medical Association* 28(9): 1132–1141.

Russell, A.G., K.F. McCue, and G.R. Cass. 1988. Mathematical Modeling of the Formation of Nitrogen-Containing Pollutants. 2. Evaluation of the Effects of Emissions Controls. *Environmental Science and Technology* 22(11): 1336–1347.

Samet, J.M., S.L. Zeger, J.E. Kelsall, J. Xu, and L.S. Kalkstein. 1997. *Air Pollution, Weather, and Mortality in Philadelphia 1973–1988.* Cambridge, MA: Health Effects Institute.

Shannon, J.D. 1981. A Model of Regional Long-Term Average Sulfur Atmospheric Pollution, Surface Removal, and Net Horizontal Flux. *Atmospheric Environment* 15(5): 689–701.

Shannon, J.D. 1985. *User's Guide for the Advanced Statistical Trajectory Regional Air Pollution (ASTRAP) Model.* EPA/600/8-85/016. Washington, DC: U.S. Environmental Protection Agency.

Shih, J.-S., A. Krupnick, M. Bergin, and A. Russell. 2004. Source–Receptor Relationships for Ozone and Fine Particulates in the Eastern United States. Discussion Paper 04-25. Washington, DC: Resources for the Future.

U.S. Census Bureau. 2000. *Annual Projections of the Resident Population by Age, Sex, Race, and Hispanic Origin: Lowest, Middle, Highest Series, and Zero International Migration Series, 1999 to 2100.* Washington, DC: U.S. Census Bureau, Population Division, Population Projections Branch.

———. 2004. *Interim Projections of the U.S. Population by Age, Sex, Race, and Hispanic Origin.* Washington, DC: U.S. Census Bureau, Population Division, Population Projections Branch.

———. 2005. *State Interim Population Projections by Age and Sex: 2004–2030.* Washington, DC: U.S. Census Bureau, Population Division, Population Projections Branch.

U.S. EPA (Environmental Protection Agency). 1997. *The Benefits and Costs of the Clean Air Act: 1970 to 1990.* Washington, DC: U.S. EPA, Office of Air and Radiation/Office of Policy.

———. 1999. *The Benefits and Costs of the Clean Air Act, 1990 to 2010.* Washington, DC: U.S. EPA, Office of Air and Radiation/Office of Policy.

———. 2004. *Final Regulatory Impact Analysis: Control of Emissions from Nonroad Diesel Engines.* Washington, DC: U.S. EPA, Office of Transportation and Air Quality.

Chapter 4: Techniques for Communicating Uncertainty to Agency Decisionmakers

> A big part of my frustration was that scientists would give me a range. And I would ask, "Please just tell me at which point you are safe, and we can do that." But they would give a range, say, from 5 to 25 parts per billion (ppb). And that was often frustrating.
> —Christine Todd Whitman, quoted in *Environmental Science & Technology Online,* April 20, 2005

If the only benefit of uncertainty analysis were improved final point estimates, then the issue of communicating uncertainty to decisionmakers would not loom large. Uncertainty could operate in the background, as part of the modeling process with a final point estimate presented to decisionmakers for evaluation.

However, other reasons for conducting uncertainty analysis crucially depend on the successful communication of the results. Presenting policymakers with a distribution of potential outcomes from various policy options allows them to make judgments about the level of risk they are willing to tolerate regarding their choices. For example, when policymakers are particularly focused on avoiding a specific outcome, a best estimate will be of limited use. The analysis also can provide policymakers with a picture of the state of knowledge underlying the research and prevent a false sense of confidence in the numbers. Additionally, helping decisionmakers understand the key sources of uncertainty can guide research priorities so that the agency can efficiently devote the resources necessary to increase confidence in the analysis.

Communicating uncertainty presents a major challenge, however. In addition to the massive amount of data generated in the course of the analyses are multiple layers of uncertainties, including uncertainty about the underlying science, the appropriate valuation of health improvements, behavioral responses, costs, and future trends. Presenting the results in an intelligible way unavoidably requires significant streamlining and simplification. For example, the National Academies of Science report on the Intergovernmental Panel on Climate Change (IPCC) working group noted that the summary of policymaker emphasized the uncertainties far less than the report itself (cited in Manning 2003). Such streamlining may have implications for the interpretation of the results.

Furthermore, as discussed in Chapter 2, there are myriad types and sources of uncertainty, many of which are cognitively challenging, particularly if the audience lacks technical expertise. Perhaps the major challenge in presenting uncertainty is deciding what to include and what to leave out and then doing so in a manner that is responsive to the needs and capabilities of the audience.

Typically, regulatory analyses have multiple audiences. The core function of a regulatory analysis is to improve government decisionmaking and the development of policy options. The audience in this context is agency policymakers. Although the U.S. Environmental Protection Agency (EPA) has one of the most educated workforces among federal agencies, agency policymakers nonetheless vary considerably in the amount of statistical and other technical training. What policymakers tend to have in common are severe time constraints; therefore, presentations to this audience must be clear and concise.

Another function of the regulatory analysis, embodied in the public release of the regulatory impact analysis (RIA) is to promote public confidence in government decisionmaking through transparency. The audience here is an outside audience, notionally the public but actually technical experts located in academia, think tanks, and interest groups. These outsiders sometimes have as much technical expertise as agency staff and possibly have more time to review the analysis than the agency decisionmakers.

Paradoxically, the presentation of uncertainty in an RIA can be more detailed and complex than is appropriate for the presentation to upper-level agency policymakers. Thus, a challenge for those preparing and communicating the regulatory analysis—as well as for the internal decisionmakers receiving the analysis—is to direct attention to the issues that are likely to be of most interest to outsiders, including potential critics.

In this chapter we consider approaches for effectively communicating to agency decisionmakers. The emphasis here is at a very practical level: the effectiveness of different techniques, rather than what should and should not be included in the presentation. First, we briefly summarize the relevant research on communicating uncertainty. Next, we introduce alternative methods for presenting uncertainty to decisionmakers, review the literature on the effectiveness of the visual presentation of uncertainty, discuss methods for conveying qualitative uncertainty, and suggest some techniques for presenting importance analysis. Finally, we draw a series of conclusions.

Research on the Communication of Uncertainty

Much of the relevant research on presenting uncertainty has been within in the context of risk communication, which overlaps with but is not identical to the problem of communicating uncertainty. A major orientation of risk communication research has been on finding ways of communicating well-defined probabilities so that a lay audience can put risks, particularly very small risks, in appropriate context. Research has shown that lay audiences are not very adept at interpreting probabilities in general and do an extremely poor job of assessing small probabilities (Tversky and Fox 1995). In contrast, decisionmakers in federal agencies are usually facing decisions in which the relevant possible outcomes have much larger probabilities; the issue of trying to interpret and properly contextualize very small likelihoods is not as relevant.

However, a separate strand of the research on risk communication is highly relevant to communicating uncertainty to policymakers. A large literature has developed that examines how the manner in which information is presented can affect its interpretation. In fact, researchers have found many situations that violate *decision invariance,* the principle that "different representations of the same problem should yield the same preference" (Tversky and Kahneman 2000).

This phenomenon has been documented many times over, and a detailed discussion is beyond the scope of this chapter (see Kuhberger 1998 for a review). But to take one example, an option framed in terms of its probability of success (e.g., "The policy has an 80% chance of passing a benefit–cost test") is seen as more attractive than the same option presented in terms of its complementary probability of failure (e.g., "The policy has a 20% chance of failing the test"; Tversky and Kahneman 1981).

Similarly, evidence indicates that interpretations of probabilities can be heavily influenced by the reasons given to support them. A given probability estimate accompanied by positive reasons to justify it is more likely to produce optimism than the same estimate accompanied by negative reasons (Flugstad and Windschutl 2003). For example, in the field of medical risk communication, telling a 37-year old patient that she has a 50% chance of regaining full joint mobility because she is under 40 will make her more optimistic about her outcome than telling her that she has the same probability of success because she is over 35.

Unfortunately, experts are just as susceptible as the general population to the cognitive biases associated with how an issue is framed (Slovic et al.1982). This means that a presentation on uncertainty very possibly could inadvertently provide subtle cues that either increase or decrease the risk aversion of the audience.

The challenges increase when the issue is not simply risk and probabilities but uncertainty about the nature of the probability distribution itself. Such uncertainty is inherent in most regulatory analysis, from model uncertainty to uncertainty about future trends. Here the decisionmaker looking for precision will be disappointed to find that questions about probability estimates and confidence bounds do not have straightforward answers.

This presents a major challenge. Research on risky decisionmaking has shown that people seek to avoid situations with ambiguous probabilities. In a famous paradigm developed by Ellsberg (1961), with a person is given two urns and told that one urn has 50 red balls and 50 black balls, whereas the other urn has black and red balls in an unknown proportion. Numerous studies have generally confirmed that when asked to bet on the outcome of a blind draw from either urn, people prefer the option with precise probabilities. This phenomenon has been called *ambiguity avoidance*. The formal definition of *ambiguity* in this context is "uncertainty about the distribution of probabilities."

In many real-world cases, decisionmakers do not have the option of choosing precise probabilities and are forced to operate under conditions of ambiguity. This raises the question of how they process ambiguous information when making decisions. An intriguing possibility is that when decisionmakers are faced with the sorts of epistemic uncertainties that crop up in regulatory analyses, they will discount the ambiguous information entirely and make decisions that resemble those made by people with no information.

Van Dijk and Zeelenberg (2003) report the results of three experiments that appear to confirm this hypothesis. In one experiment, they asked a control group of respondents if they would be willing to invest in a new Chinese restaurant in a downtown area if they knew that a rival Chinese restaurant would soon be opening in the area. To a second group, they presented the same scenario plus the additional information that a recent survey had shown that 15% of restaurant goers in the downtown area expressed a strong interest in Chinese food. A similar scenario was presented to a third group, except the percentage of restaurant goers in the area expressing a strong interest in Chinese food was raised to 35%. As expected, individuals provided with the 15% number were more willing to invest in the venture than the individuals provide with no extra information, and the individuals told that it was 35% were more likely to invest than the two other groups. However, the researchers told a fourth group that two surveys had been conducted and provided both numbers (15% and 35%). The willingness to invest of this fourth group was less than that of the second and third groups and in fact was not statistically different from that of the control group, which had received no survey results. The two other experiments produced similar results, leading the researchers to conclude that individuals tend to discount ambiguous information, meaning that it does not alter the decisionmaking process.

The reasons for this effect are not clear. One possibility is that choosing under conditions of ambiguity is cognitively demanding. Another possibility is that people may be concerned that their decisions will be subsequently evaluated and have a difficult time providing a rationale for their choices under conditions of ambiguity (Curley et al. 1986).

Discomfort with uncertainty may extend beyond ambiguity and to probabilistic decisionmaking itself. Gneezy et al. (2004) find individuals valuing a lottery less than its worst potential outcome. In one experiment, the willingness to pay for a $50 Barnes and Noble gift certificate is substantially higher than the willingness to pay for a lottery with two potential outcomes: a $100 gift certificate and a $50 gift certificate. They call this phenomenon the uncertainty effect. The fact that the value of the lottery lies outside of the range of possible outcomes violates the principles of most theories of decisionmaking under uncertainty (such as expected utility and prospect theory).

The implications of this research for communicating uncertainty to policymakers is somewhat troubling, because it suggests that an increased emphasis on uncertainty may not produce the desired clarity in the decisionmaking process. Research on the extent to which ambiguity avoidance and the uncertainty effect apply to agency decisionmakers as opposed to the general public would be useful, as would be developing methods to address these challenges.

Approaches to Presenting Quantitative Uncertainty

In this section, we discuss different approaches for presenting the results of quantitative uncertainty analysis. They fall into three general categories: verbal descriptions, numeric presentations, and graphical depictions. We discuss what the research shows about the strengths and weakness of each approach and identify situations in which a particular approach is more likely to be helpful or inappropriate.

Presenting uncertainty may consist of simply showing outcomes from selected scenarios without reference to probabilities. Alternatively, it may involve confidence intervals or distributions for outcomes of interest (e.g., risk reductions, monetized benefits). Often a combination of approaches may be preferred (e.g., providing different scenarios with probabilities or confidence intervals for each, such as probability distributions for net benefits given different assumptions about the value of a statistical life [VSL]).

Verbal Descriptions

In daily life, people frequently use verbal descriptions to convey probabilities. Statements like, "I'm reasonably confident the statement is true" convey a sense of likelihood, albeit in an imprecise manner. Of course the lack of precision implied by using words has its drawbacks. Several empirical studies have attempted to translate verbal probability expressions into numerical equivalents (e.g., what probability is meant by *extremely likely*?). In general, these studies have found that such translation cannot be done in a precise manner, given that interpretations of verbal expressions vary across individuals and contexts (see Budescu and Wallstein 1995 for a review).

The IPCC Working Group I contribution to the Third Assessment Report (WGI-TAR) used a system of verbal descriptors for different findings and projections, with each term linked to a specific range of numerical probabilities. The term *virtually certain* was meant to imply a greater than 99% chance, *very likely* a 90–99% chance, *likely* a 66% to 90% chance, *unlikely* a 10–33% chance, *very unlikely* a 1–10% chance, and *exceptionally unlikely* a less than 1% chance (summarized in Manning 2003).

Patt and Schrag (2003) argue that this system is subject to misinterpretation because individuals evaluate such descriptors as a combination of the probability of the event and the magnitude of its outcome. For example, people tend to interpret the statement "Snow flurries are unlikely" as implying a higher numerical probability estimate than the statement "A hurricane is unlikely." WGI-TAR rated the chances of substantial sea level rise in the twenty-first century as *very unlikely,* which by the above definition means a 1–10% chance, but is arguably open to misinterpretation by ordinary readers as a much smaller numeric probability.

Numeric Presentations

The major numerical presentation format for uncertainty, familiar from most RIAs, is a table with means and confidence intervals (usually the 5th and 95th percentile) presented. These tables have the advantage of presenting information in a concise and easy interpretable format. A normal distribution can be completely described with reference to its mean and standard deviation, so tables are a simple way to provide information on uncertainty with this distributional form. A table included in EPA's recent RIA for the nonroad diesel rule demonstrates how much information can be provided simply through tables (Table 4-1). Another approach is to provide summary statistics that give a picture of the relative uncertainty of variables or predictions (e.g., the commonly used *coefficient of variation*).

There are drawbacks to presenting uncertainty through tables or summary statistics. The foremost issue is that complex tables and summary statistics for many variables may not hold the audience's attention. Additionally, numeric probability estimates, although more precise than verbal descriptions, are not immune to misinterpretation or bias because of the framing impacts and other biases described above (Flugstad and Windschitl 2003).

Graphical Displays

Graphical displays offer several advantages for communicating uncertainty (see Lipkus and Hollands 1999 for a discussion). First, graphics can reveal data patterns that are difficult to convey through other means. Second, for certain problems (e.g., comparing risks), graphs facilitate the processing of information better than numbers alone. Finally, graphs are more compelling than words and numbers and thus do a better job of holding an audience's attention.

The simplest graphical presentation of uncertainty is scenarios that illustrate different outcomes under alternative assumptions, such as a high, medium, and low value for an input. For example, Figure 4-1 is an IPCC graphic representing numerous outcomes over time under different assumptions but conveys no information about their relative plausibility. When information is insufficient to allow for the presentation of uncertainty in probabilistic terms, presenting scenarios may be the only viable option. Even when outcomes can be probabilized, scenario presentations may be sufficient if the outcomes under all plausible assumptions are relatively homogenous (e.g., if, under the range of all reasonable assumptions, net benefits are assumed to be positive).

The obvious drawback with scenario approaches is that they provide little insight on the relative probability of occurrence. For displaying uncertainty in probabilistic terms, the three most common graphical techniques are box-and-whisker plots, probability density functions (PDFs), and cumulative density functions (CDFs).

Box-and-whisker plots (like the one in Figure 4-2) are well suited for displaying summary statistics such as means, medians, ranges, and fractiles. Research has shown that they are effective in presenting this summary information to audiences, mainly because the information is labeled directly on the graph (Ibrekk and Morgan 1987). Although box-and-whisker plots provide no information about the shape of the distribution, they may be in many cases sufficient for the needs of the policymaker (e.g., if the issue is simply whether the confidence interval around a net benefit estimate excludes negative net benefits).

In some cases, the policymaker may seek information that goes beyond summary statistics to the actual shape of the distribution. In such cases, PDFs and CDFs are preferred, each with its relative strengths and weaknesses. PDFs (e.g., Figure 4-3) enable easy identification of the relative probabilities of different values as well as the mode of the distribution. They can be particularly helpful for highlighting mulitimodal distributions (e.g., net benefits gravitate around very low and very high numbers). One disadvantage of PDFs is that they can be can be somewhat “noisy” and do not allow for easy interpretation of important elements of the distribution, such as means and fractiles. CDFs (e.g., Figure 4-4) are far less noisy than PDFs and are particularly helpful when what is of interest is the fractiles of the distribution (e.g., the probability that net benefits will be above a certain level). A downside is that, as with PDFs, audiences have a difficult time extracting summary information (e.g., means) from a CDF plot.

Because of the complementary strengths and weaknesses of PDFs and CDFs, Ibrekk and Morgan (1987) recommend using a CDF and a PDF together, with the mean clearly labeled on both as a solid point.

Although less commonly used, pie charts like Figure 4-5 are useful for emphasizing proportions and may be appropriate when a policymaker is interested in the probability of a high-consequence event. For example, researchers at the Massachusetts Institute of Technology’s (MIT’s) Joint Program on the Science and Policy of Global Change have developed a roulette wheel to highlight the estimated probability of relatively large climate changes under different policies.

A more complicated presentation is required when uncertainty exists in two dimensions (e.g., uncertainty about both the dose–response functions and the VSLs). One approach is to present the probability distributions in a series of separate graphs, but it is probably more efficient and informative to superimpose distribution on the same graph, as long as the number of distributions is not too large. When multiple dimensions are involved (e.g., different benefit estimates under different VSL assumptions for different scenarios), the two approaches can be combined with a series of charts with overlaid distributions, as in Figure 4-6.

Alternatively, the uncertainty can be represented by bars around a line graph. For example, in Figure 4-7, the x-axis represents the assumed value of the statistical life, the y-axis represents net benefits, and the error bars provide uncertainty information. Similarly, in Figure 4-8, the x-axis represents the year of interest and the y-axis annual mean health benefits. Uncertainty also can be presented in three dimensions through probability surfaces, as seen in Figure 4-9, where the elevation represents the probability.

A less commonly used approach is the triple scatterplot shown in Figure 4-10 (suggested by Anscombe [1973] and discussed by Cleveland and McGill [1984]). The centers of the circles represent the x and y coordinates and form an ordinary Cartesian graph. The area of the circles is proportional to the third variable z (in this case, the probability).

Research on the Effectiveness of Visual Presentations of Uncertainty

Two major areas of research relevant to the graphical presentation of uncertainty are (a) the performance of different graphical formats in how well they allow the accurate extraction of quantitative information and (b) the impact different graphical formats have on decisionmaking and whether different graphical presentations of the same information induce different responses. Of particular interest is whether different types of presentations make an audience more or less risk averse.

The first area allows for objective evaluation. One can test how accurately the audience extracts quantitative information from the visual display. Cleveland and McGill (1984) rank 10 elementary perceptual tasks in terms of the accuracy of quantitative extraction: positions along a common scale (e.g., line graphs, bar charts), positions on common nonaligned scales (e.g., scatterplots), and angles (e.g., pie charts) perform best, whereas areas, volume, shading, and color saturation perform poorly. Cleveland and McGill also recommend against using curve difference charts (Figure 4-11), because people have a hard time accurately judging the vertical length between the two curves. In Figure 4-11, the panels in the left group map to the corresponding panels on the right group. Although the different pairs of curves look very similar, in fact, the graphs represent markedly varying net differences. Thus, superimposing cost and benefit curves may tend to mislead the audience about the shape of the net benefit curve.

Other research has shown that individuals exhibit biases in estimating physical magnitudes and particularly tend to underestimate large areas and volumes (Stevens and Gallanter 1957). Lipkus and Hollands (1999) recommend against using volume and area charts for presenting uncertainty because of these perceptual biases.

The issue of how different modes of presentation affect judgments is not testable in the same way as the question efficacy of communicating quantitative information. There is no one correct degree of risk aversion, and thus it is impossible to rank presentation performance on the basis of an objective standard. Siebenmorgan et al. (2000) compare how the use of bar graphs and PDFs affect investors' interpretation of asset riskiness. Providing historical information on asset returns in the form of a density function instead of bar graphs led to greater estimates of asset volatility and risk. The density representations made respondents more conscious of the extremes. The authors conclude, "Given that nominally equivalent presentation formats lead to different impressions of asset risks, which translate into differences in investment behavior, and given that no gold standard exists to indicate a correct level of perceived risk, policymakers need to realize that decisions about the appropriate content and format of financial risk communication cannot be made in an objective or value free fashion" (Siebenmorgan et al. 2000, p. 17).

Similarly, in an examination of how well visual displays of risk communicated low-probability events, Stone et al. (1997) find that adding graphics to the numeric presentations increases participants willingness to pay for risk reductions; in other words, it increases their level of risk aversion.

The effectiveness of graphical presentation also can be measured through audience evaluation (e.g., how they judge the clarity and utility of the format). The most relevant research on this topic for our purposes is a Thompson and Bloom (2000) study based on interviews and focus groups with EPA risk managers in the early 1990s. Part of the exercises involved presenting different graphical displays of risk. The researchers found a strong preference for graphics that were not too busy (e.g., many policymakers found a tornado graph fairly confusing). The risk managers rated the PDF format most favorably, although some said they would have preferred a CDF (not included in the focus group presentation). Thompson and Bloom see the results as consistent with Ibrekk and Morgan's (1987) recommendation to present CDFs and PDFs jointly with the mean clearly labeled.

Presenting Qualitative Uncertainty

The discussion has focused on presenting quantitative uncertainty around the outcomes of the analysis. Another important context for presenting uncertainty is characterizing the state of knowledge and the degree of confidence around key parameters and assumptions. In its most basic form, this can involve qualitative descriptions about the degree of knowledge about various inputs into the analysis. In other cases, there may be enough information to provide numerical or graphical presentations (e.g., a probability distribution around a well-sampled variable). In cases where data are sparse or it is impossible to generalize from past results (e.g., forecasting technological progress), expert elicitation techniques may be used to generate probability distributions.

Despite the growing sophistication in the state of the art of quantitative uncertainty analysis, qualitative evaluation remains a fundamental aspect of communicating uncertainty, particularly presenting the assumptions behind the analysis and the state of the knowledge underpinning the analysis. Verbal presentation is particularly well suited for this task (e.g., giving policymakers an overview of the types of uncertainties around input variables). It can take the form of simple tables with the descriptions of various uncertainties around each relevant variable, or it may include an assessment of the hypothesized directionality of the effect on estimates (Table 4-2).

More formalized approaches can be used as well. The Numeral Unit Spread Assessment Pedigree (NUSAP) system first proposed by Funtowicz and Ravetz (1990) is designed for multidimensional uncertainty assessment and aims to capture both qualitative and quantitative dimensions of uncertainty. One component of NUSAP, *pedigree analysis,* is a systematic, multiple-criteria approach to evaluating the strength of the knowledge base, by assigning scores of 0–4 to the variables on different criteria, such as the strength of the proxy, empirical support, theoretical basis, the rigor of the method, and the extent of validation. The scores are then either averaged or normalized by the maximum possible score to yield a score (van der Sluijs et al. 2004; also see Lee et al. 1995 for an application).

Although not commonly used, symbols can be an effective technique for communicating qualitative uncertainty. For example, the *Consumer Reports*-style graphical presentation in Table 4-3 characterizes the levels the state of knowledge of variables in an economic analysis

Importance Analysis

Finally, we present some approaches for conveying the relative importance of uncertainties to policymakers. Communicating the linkages between uncertainty in input variables and uncertainty in outcomes is a key goal of this activity. With an understanding of these linkages, policymakers can gain insights into the importance of particular parameters or assumptions. This knowledge, in turn, can be useful in making judgments about policy options. It can also help identify and prioritize targets for further research.

Perhaps the most familiar technique is the tornado graph, simply a stacked bar chart (Figure 4-12). The bars can represent correlation coefficients for the input variables with the output variables or the effect on the output from changing the input variable by some amount (e.g., 1 standard deviation). The bars are stacked in descending order from the variable with the highest correlation or impact. Tornado graphs are useful for showing both the magnitude of the relationships and the directionality.

A scatterplot technique can be used within the NUSAP system to identify uncertain variables that have a particularly strong impact on the results. Variables are plotted on a Cartesian plane, with the x-axis representing the pedigree score described above (from strong to weak) and the y-axis representing the sensitivity of the results to changes in the variable. Variables appearing in the top right section are of most interest, because they significantly affect the results and are characterized by a high degree of uncertainty. In Figure 4-13, which is based on a study of volatile organic compounds (VOCs) emissions from paint in the Netherlands, the variable of interest is the assumed VOC percentage on imported paint.

More detailed information can be displayed through the use of pairwise scatter plots, which visually present correlations between input variables and outcomes as well as the complete picture of correlations between input variables and outcomes as well as among input variables. In Figure 4-14, the variable IGN_HEAD is the outcome of interest, and other variables are inputs into the model. Low values of the variable BAT result in very constrained outcomes for the variable IGN_HEAD (Cooke and Van Noortwijk 1999).

Conclusion

Much of the research on communicating uncertainty has been in the context of the risk communication paradigm, which in general involves much smaller probabilities than those that agency decisionmakers face. The focus on risk communication has also meant an orientation toward finding ways of communicating to lay audiences. To the extent that attention has focused on communicating risk and uncertainty to agency decisionmakers, the emphasis has been on the content (what to include) rather than the form (how to present it).

A notable exception is the Thompson and Bloom (2000) study, based on interviews with EPA risk managers in the early 1990s. This research focused primarily on risk communication, and much has transpired since it was completed. Many issues have grown in importance since then, such as the role of physical effects and cost-effectiveness analysis versus monetary benefits and cost–benefit analysis, distinguishing between statistical and model uncertainty, addressing nonquantifiable variables, and the role of expert judgment. Thus, additional research on the relative effectiveness of techniques for communicating these complex issues to specialists is warranted.

Communicating uncertainty involves more than just choosing the right mode of presentation. One must be aware of subtle cues in the presentation that may bias interpretations. In fact, the possibility that different methods of presentation may induce different decisions is troubling. Under conditions of uncertainty, the rational decision depends on the decisionmaker's degree of risk aversion. But there is no "right" level of risk aversion and therefore no objective criteria for ranking presentation methods along this dimension.

Finally, much of the work on communicating uncertainty has focused on the communicators and finding ways to improve presentation methods. Given the importance of correctly appreciating the implications of uncertainty, perhaps more work should be done on the audience side. Agencies should consider providing training decisionmakers on how to think about the complex issues associated with uncertainty in regulatory analyses.

Figures

Figure 4-1.

Source: IPCC 2000, p. 7

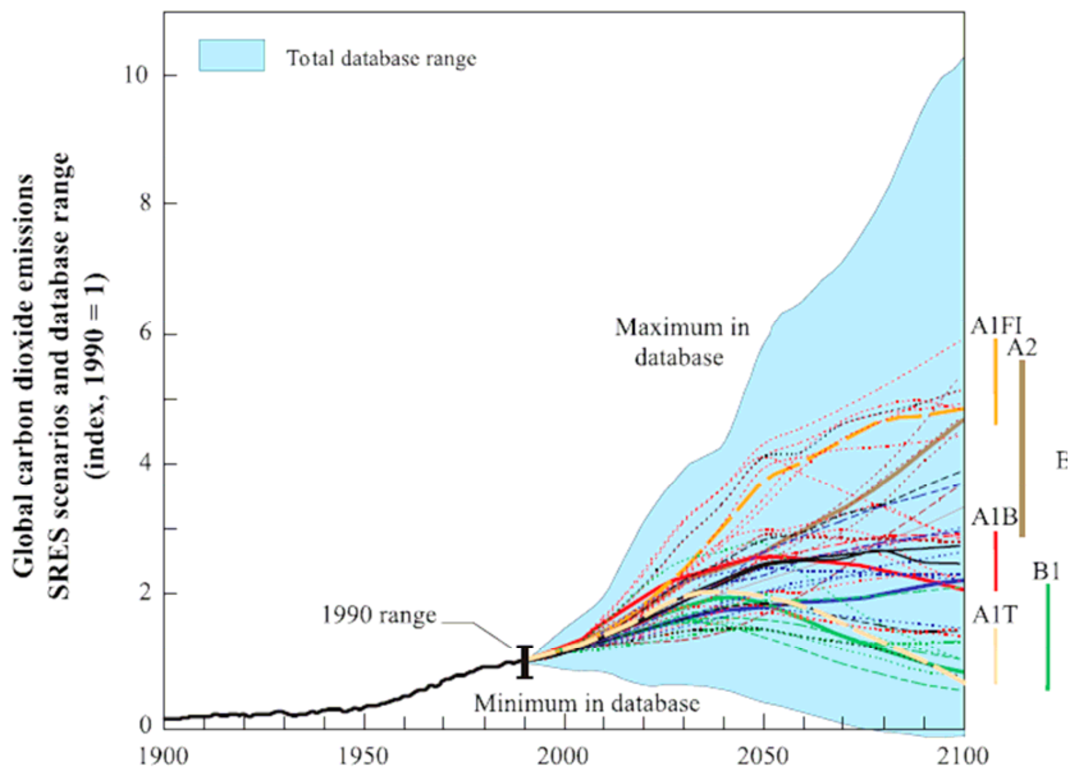


Figure 4-2.

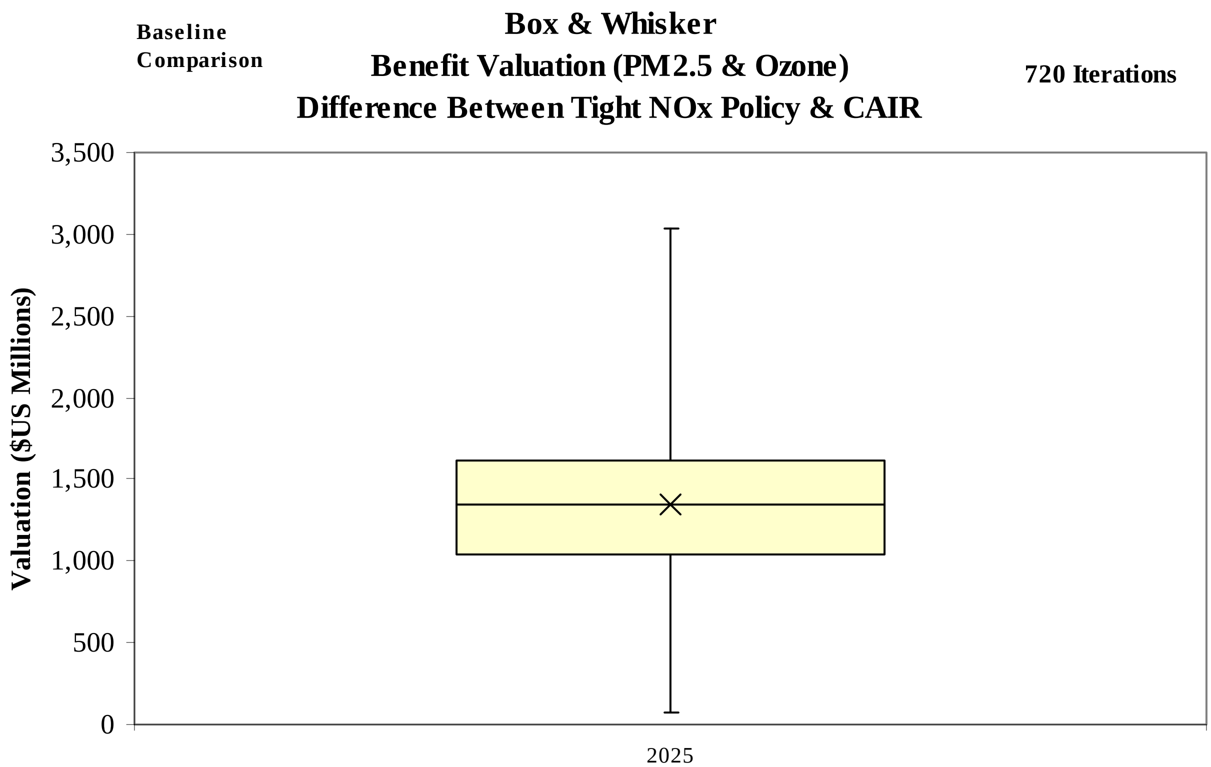

Baseline
Comparison
Box & Whisker
Benefit Valuation (PM2.5 & Ozone)
Difference Between Tight NOx Policy & CAIR
720 Iterations
Valuation ($US Millions)
3,500
3,000
2,500
2,000
1,500
1,000
500
0
2025

Figure 4-3.

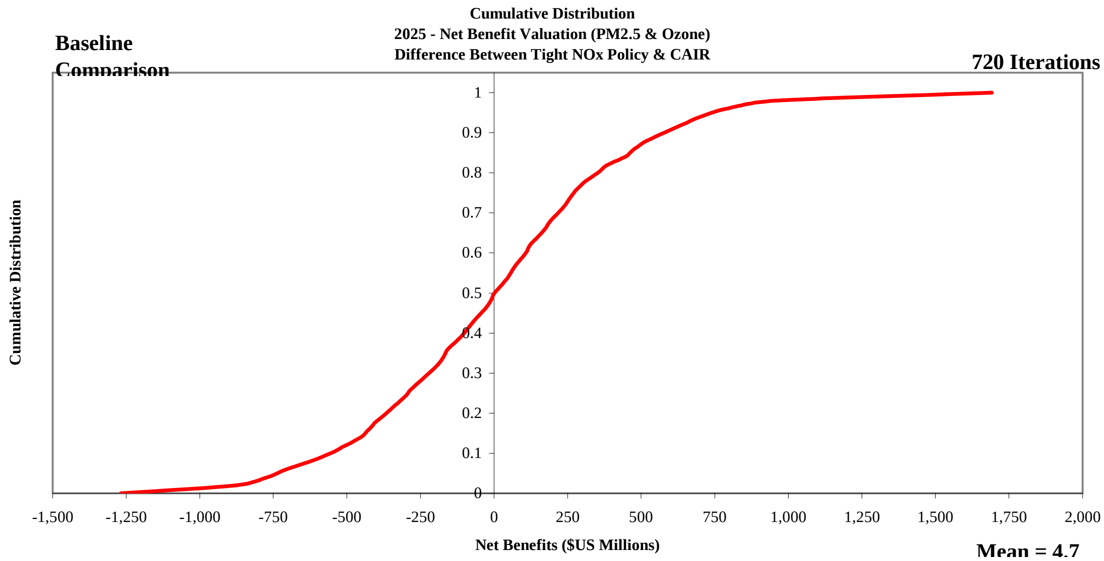


Figure 4-4.

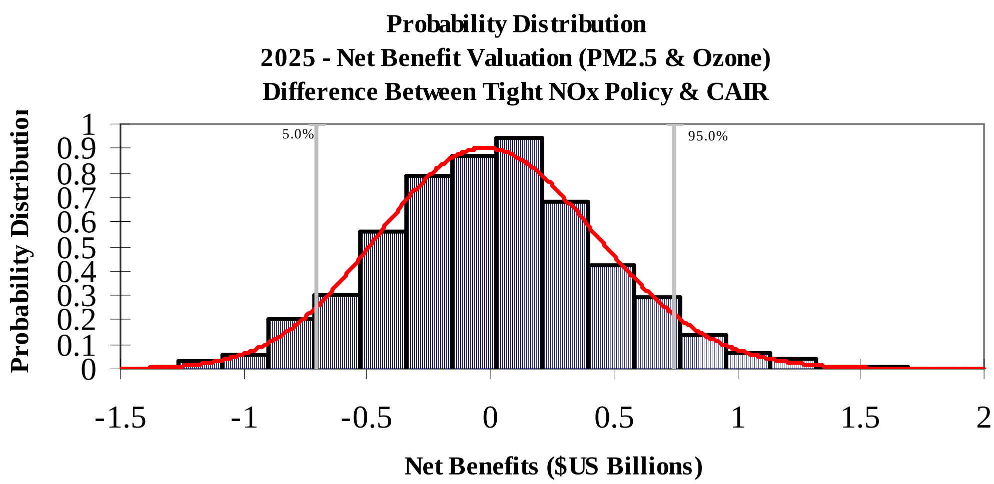


Figure 4-5.

Source: "The Greenhouse Gamble," http://web.mit.edu/globalchange/www/wheel.degC.html

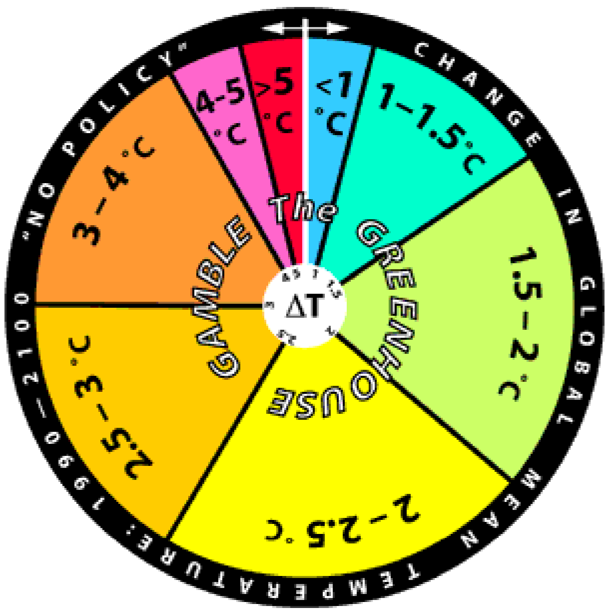


Figure 4-6.

Source: Morgan and Henrion 1990, p. 247

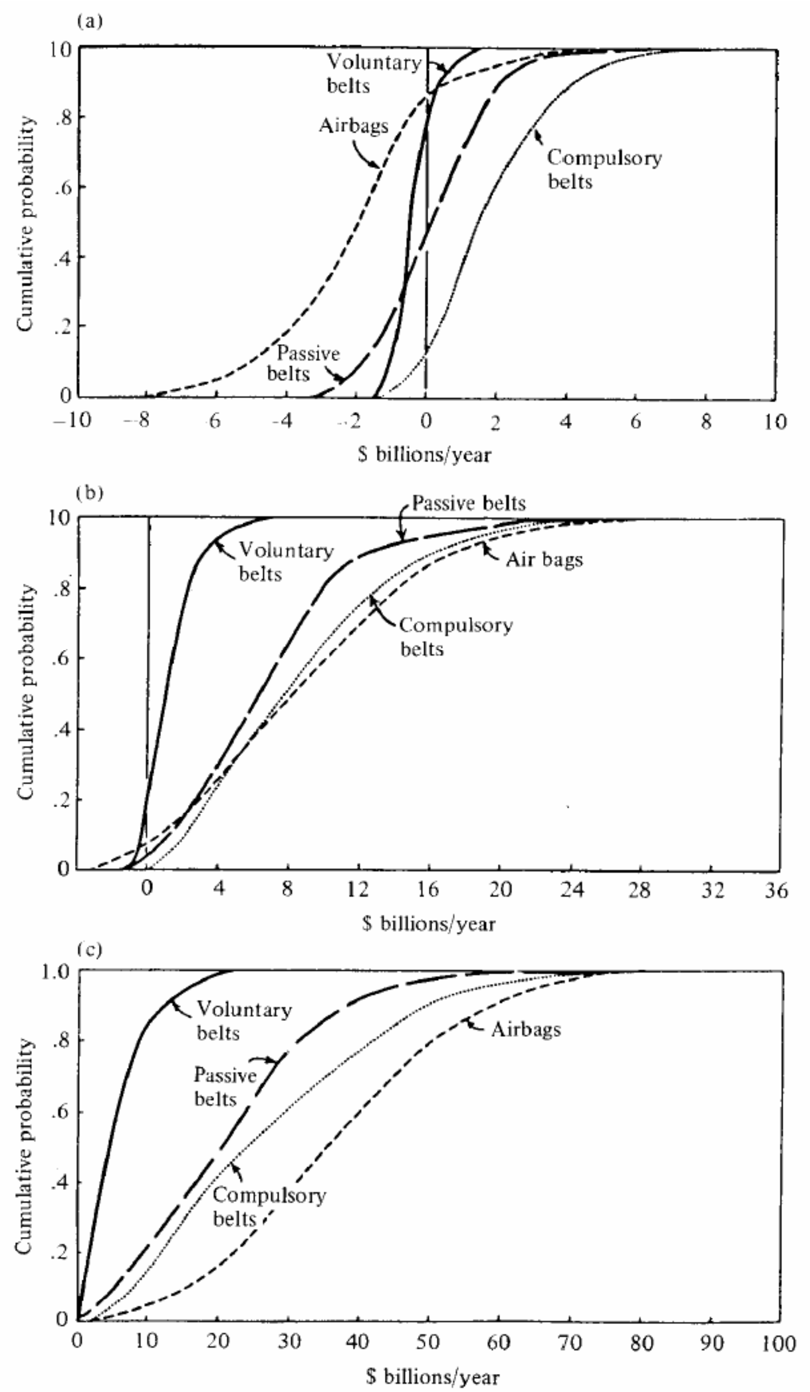


Figure 4-7.

Source: Morgan and Henrion 1990, p. 248

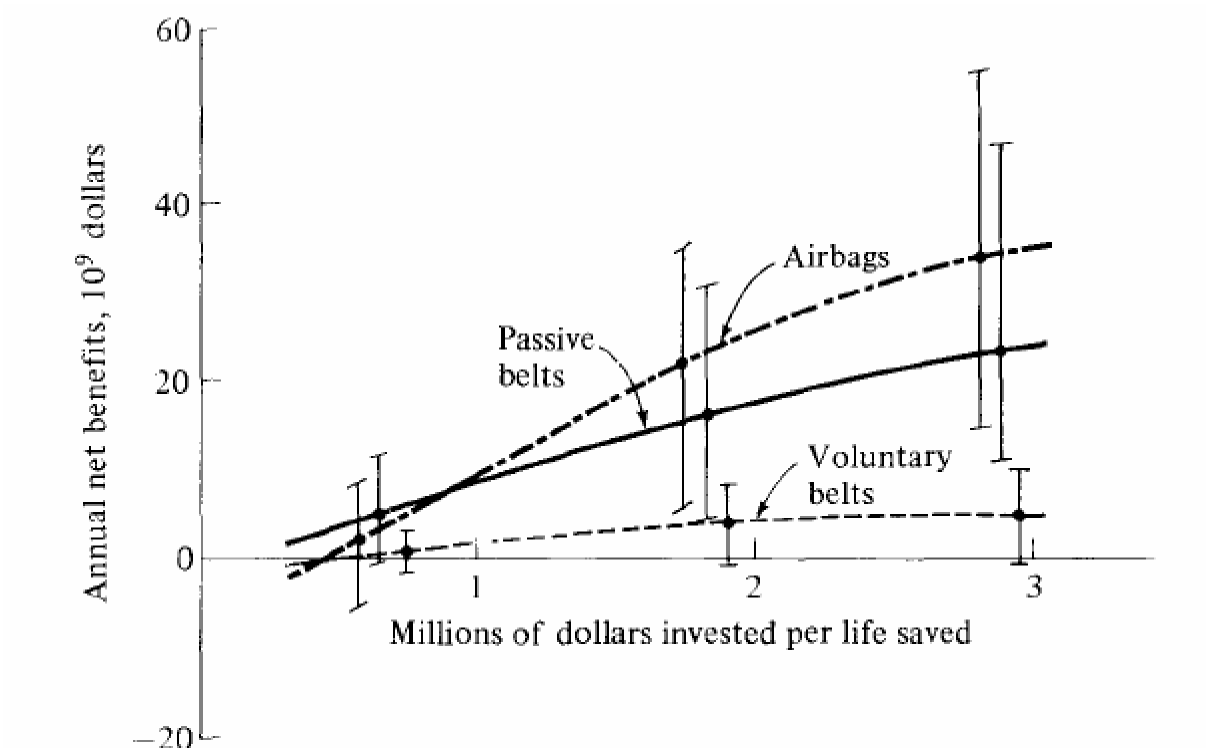
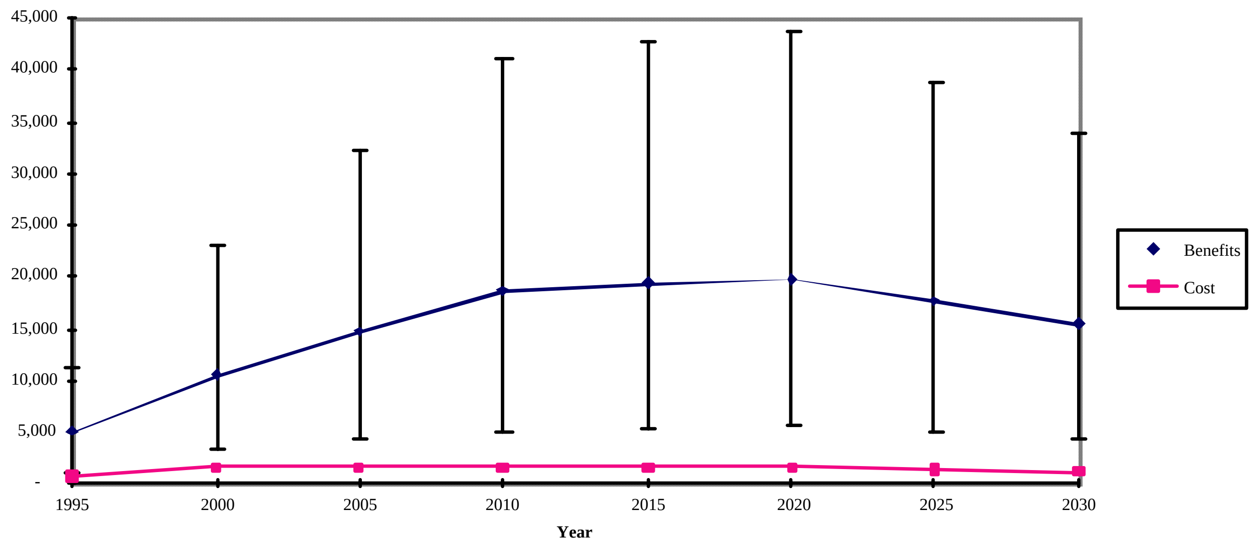


Figure 4-8.

Source: Burtraw et al. 1998



Annual mean total health benefits with 90% confidence intervals compared with expected annualized costs

Figure 4-9.

Source: Allen and Stainforth 2002

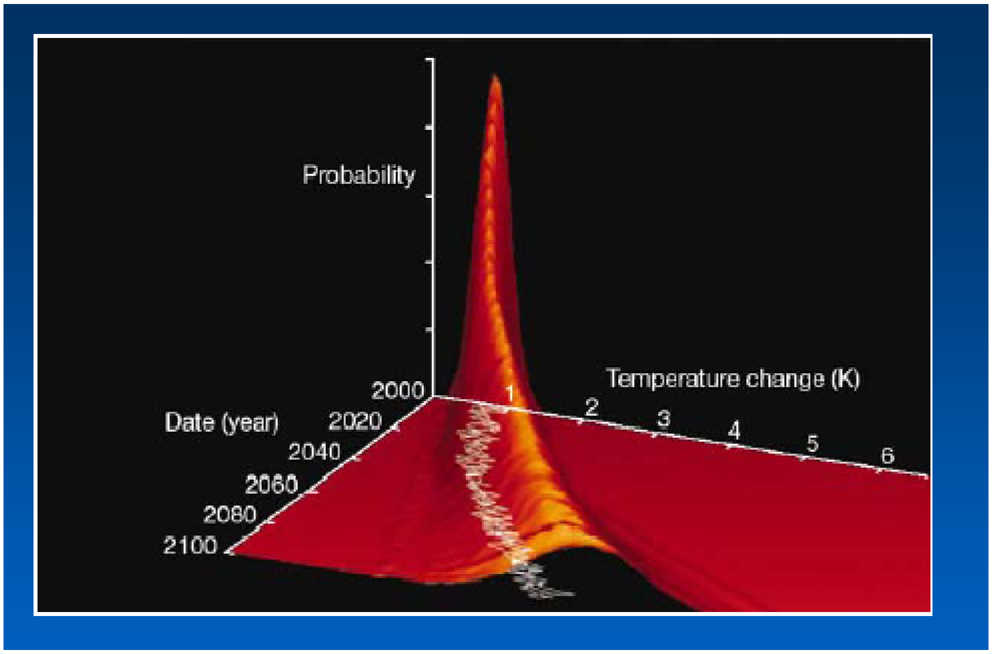


Figure 4-10.

Source: Cleveland and McGill 1986

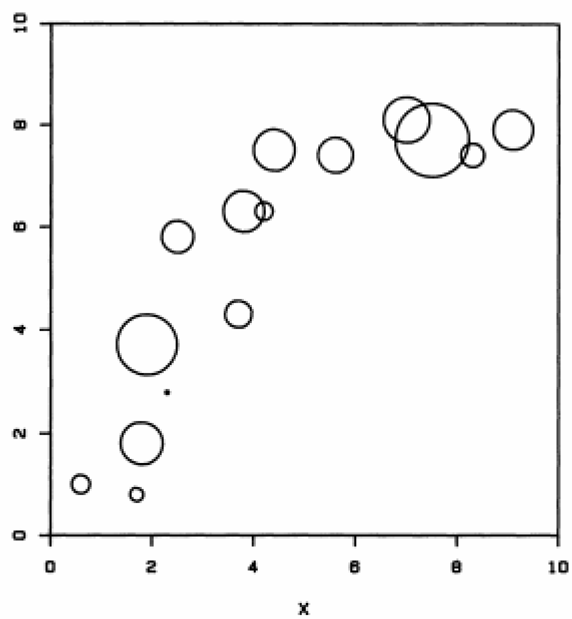


Figure 9. Triple scatterplot.

Figure 4-11.

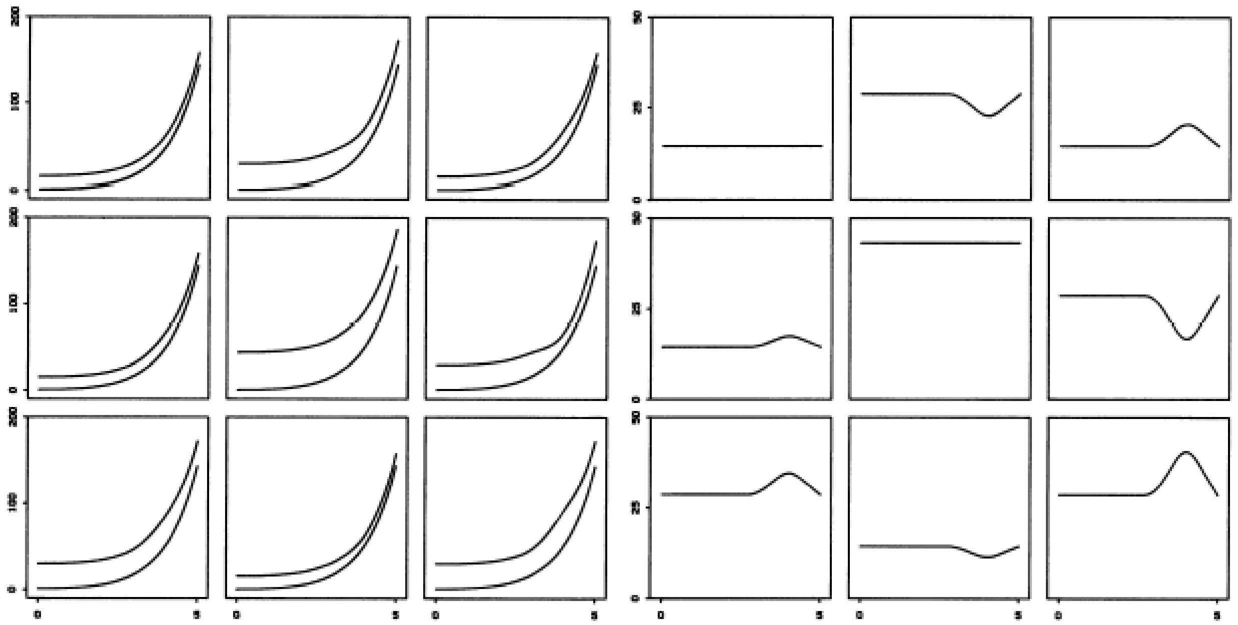


Figure 26. Curve-difference chart.

Figure 27. Curve differences.

Figure 4-12.

Source: Roozenberg and Nicholson 2003

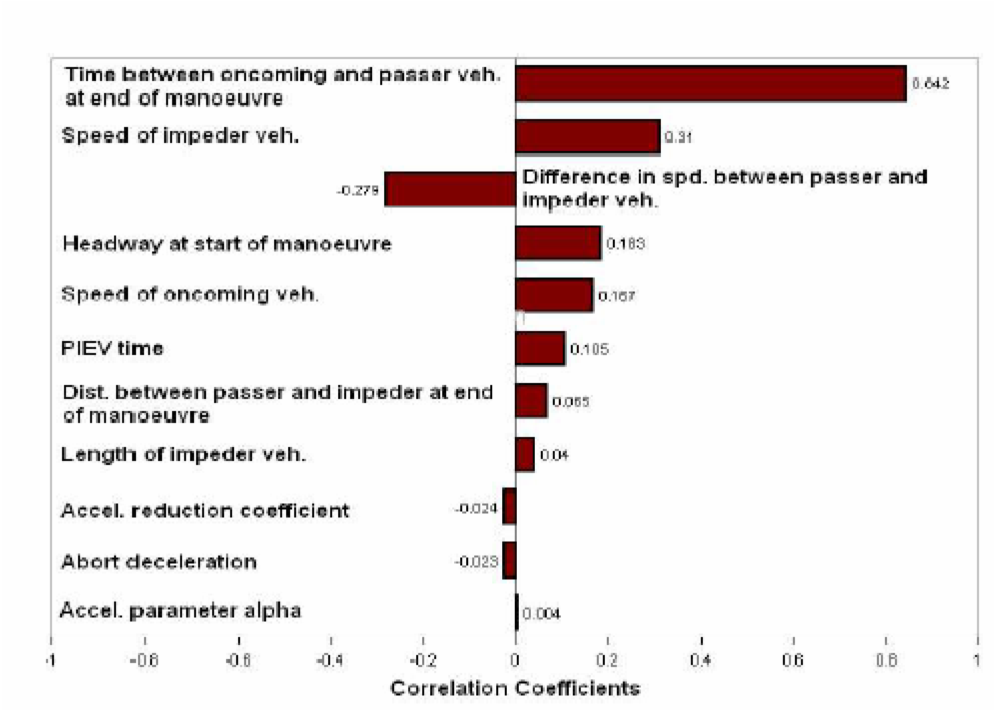


Figure 4-13.

Source: van der Sluijs et al. 2003

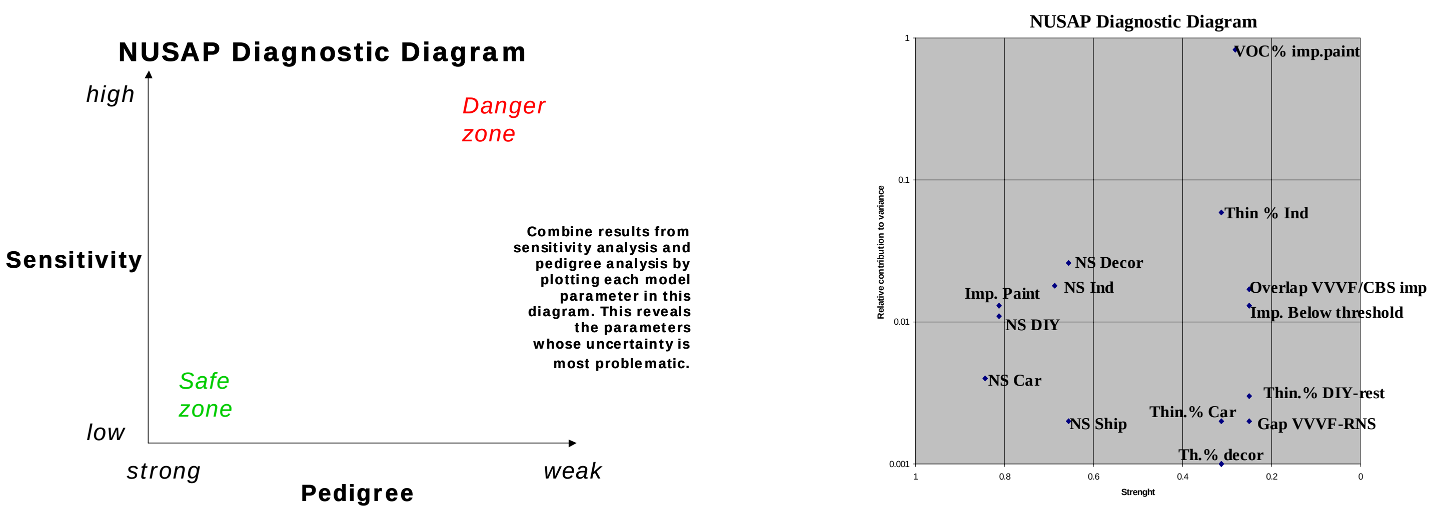
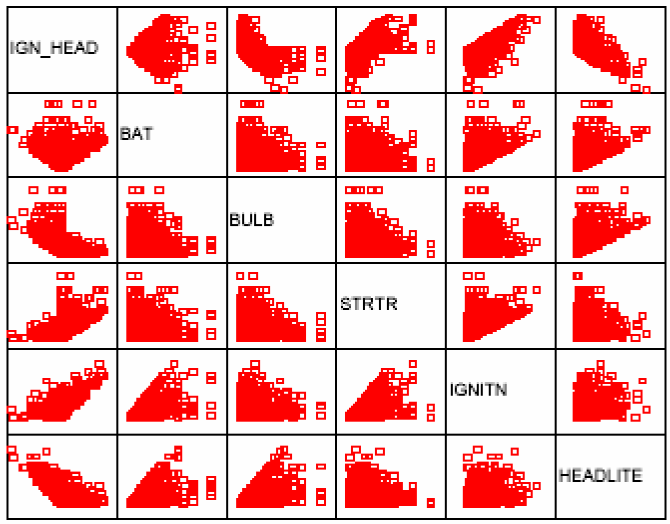


Figure 4-14.

Source: Cooke and Van Noortwijk 1999



Tables

Table 4-1.

Source: U.S. EPA 2004, p. 9-212 (PDF p. 1252)

Table 9B-2.
Distribution of Value of Annual Human Health and Welfare Benefits in 2030 for the Modeled Preliminary Control Option of the Non-Road Diesel Rule[A]

Endpoint	Monetary Benefits[B, C] (Millions 2000$, Adjusted for Income Growth)		
	5th Percentile	Mean	95th Percentile
Premature mortality[D]			
Long-term exposure, (adults, >30yrs)	$20,000	$89,000	$180,000
Long-term exposure (child <1yr)	$40	$180	$350
Chronic bronchitis (adults, 26 and over)	$200	$2,800	$9,400
Non-fatal myocardial infarctions (adults, 18 and over)	$300	$1,400	$3,300
Hospital Admissions from Respiratory Causes[E]	$17	$36	$54
Hospital Admissions from Cardiovascular Causes[F]	$59	$96	$130
Emergency Room Visits for Asthma (children, <18)	$1.3	$2.2	$3.4
Acute bronchitis (children, 8-12)	($0.2)	$5.9	$15
Lower respiratory symptoms (children, 7-14)	$1.1	$2.9	$5.4
Upper respiratory symptoms (asthmatic children, 9-11)	$0.9	$3.7	$7.7
Work loss days (adults, 18-65)	$140	$160	$180
Asthma exacerbations (asthmatic children, 6-18)	$0.2	$11	$29
Minor restricted activity days (adults, age 18-65)	$200	$340	$500
Recreational visibility (86 Class I Areas)	$1,700	$1,700	$1,700
Unquantified Benefits	B	B	B
Monetized Total[G]	$23,000+B	$96,000+B	$200,000+B

[A] The benefit estimates provided in this table are based on the modeled air quality data for the preliminary control option used in the Non-Road Diesel proposal analysis and do not reflect the predicted emission reductions of the final rule's stringency levels. In the primary estimate in Chapter 9, the modeled benefits were scaled to the level necessary to reflect the predicted emission reductions of the final rule. The estimates provided in this table have not been scaled to the rule's stringency level, as the scaling methodology adds a new element of uncertainty that cannot be appropriately characterized here. These estimates should not be compared with the primary estimate provided in the chapter, but could be compared to results presented in Apprendix 9A.
[B] Monetary benefits are rounded to two significant digits.
[C] Monetary benefits are adjusted to account for growth in real GDP per capita between 1990 and 2030.
[D] The valuation of mortality assumes the 5 year distributed lag structure described earlier. Impacts of alternative lag structures are provided in a sensitivity analysis in Appendix 9C. Results reflect the use of 3% and 7% discount rates consistent with EPA and OMB's guidelines for preparing economic analyses (US EPA, 2000c, OMB Circular A-4).
[E] Respiratory hospital admissions for PM includes admissions for COPD, pneumonia, and asthma.
[F] Cardiovascular hospital admissions for PM includes total cardiovascular and subcategories for ischemic heart disease, dysrhythmias, and heart failure.
[G] B represents the monetary value of the unmonetized health and welfare benefits. A detailed listing of unquantified PM, ozone, CO, and NMHC related health effects is provided in Table 9-1.

Table 4-2.

Source: U.S. EPA 2004, p. 9-37 (PDF p. 1076)

Table 9-8
Primary Sources of Uncertainty in the Benefit Analysis

1. Uncertainties Associated With Health Impact Functions
– The value of the ozone or PM effect estimate in each health impact function. – Application of a single effect estimate to pollutant changes and populations in all locations. – Similarity of future year effect estimates to current effect estimates. – Correct functional form of each impact function. – Extrapolation of effect estimates beyond the range of ozone or PM concentrations observed in the study. – Application of effect estimates only to those subpopulations matching the original study population.
2. Uncertainties Associated With Ozone and PM Concentrations
– Responsiveness of the models to changes in precursor emissions resulting from the control policy. – Projections of future levels of precursor emissions, especially ammonia and crustal materials. – Model chemistry for the formation of ambient nitrate concentrations. – Lack of ozone monitors in rural areas requires extrapolation of observed ozone data from urban to rural areas. – Use of separate air quality models for ozone and PM does not allow for a fully integrated analysis of pollutants and their interactions. – Full ozone season air quality distributions are extrapolated from a limited number of simulation days. – Comparison of model predictions of particulate nitrate with observed rural monitored nitrate levels indicates that REMSAD overpredicts nitrate in some parts of the Eastern US and underpredicts nitrate in parts of the Western US.
3. Uncertainties Associated with PM Premature mortality Risk
– No scientific literature supporting a direct biological mechanism for observed epidemiological evidence. – Direct causal agents within the complex mixture of PM have not been identified. – The extent to which adverse health effects are associated with low level exposures that occur many times in the year versus peak exposures. – The extent to which effects reported in the long-term exposure studies are associated with historically higher levels of PM rather than the levels occurring during the period of study. – Reliability of the limited ambient $PM_{2.5}$ monitoring data in reflecting actual $PM_{2.5}$ exposures.
4. Uncertainties Associated With Possible Lagged Effects
– The portion of the PM-related long-term exposure mortality effects associated with changes in annual PM levels would occur in a single year is uncertain as well as the portion that might occur in subsequent years.
5. Uncertainties Associated With Baseline Incidence Rates
– Some baseline incidence rates are not location-specific (e.g., those taken from studies) and may therefore not accurately represent the actual location-specific rates. – Current baseline incidence rates may not approximate well baseline incidence rates in 2030. – Projected population and demographics may not represent well future-year population and demographics.
6. Uncertainties Associated With Economic Valuation
– Unit dollar values associated with health and welfare endpoints are only estimates of mean WTP and therefore have uncertainty surrounding them. – Mean WTP (in constant dollars) for each type of risk reduction may differ from current estimates due to differences in income or other factors. – Future markets for agricultural products are uncertain.
7. Uncertainties Associated With Aggregation of Monetized Benefits
– Health and welfare benefits estimates are limited to the available effect estimates. Thus, unquantified or unmonetized benefits are not included.

Table 4-3.

Source: Burtraw et al. 1998

Categories ● high ⭘ high-mid ◒ mid ⊙ low-mid ○ low	1. Link Between Science and Economics: Are benefit endpoints well established? Does science provide infomation needed for economic analysis?	2. Economic Methods: Are economic methods adequately developed?	3. Data Availability: Is data available from science and from economics for an assessment of benefits?	4. Expected Benefit: Are expected benefits large?	5. Value of Additional Information: With the goal of improving benefit estimates, what is the relative short-term return on investment?
Health: Mortality	⭘	⭘	⭘	●	●
Health: Morbidity	⭘	⭘	⭘	⭘	⭘
Visibility	⭘	◒	⊙	⭘	⭘
Materials and Cultural Resources	⊙	◒	○	⭘	⭘
Nonuse Values: Ecosystem Health	⊙	⊙	⊙	●	◒
Aquatics: Recreation	⭘	●	⊙	⊙	◒
Forests: Recreation	⊙	⭘	○	⊙	◒
Agriculture and Commercial Forestry	⭘	●	◒	◒	⊙
Radiative Forcing	⊙	○	○	⊙	○

References

Allen, M.R., and D.A. Stainforth. 2002. Towards Objective Probabilistic Climate Forecasting. *Nature* 419: 228.

Anscombe, J.F. 1973. Graphs in Statistical Analysis. *The American Statistician* 27: 17–21.

Budescu, D., and T. Wallsten. 1995. Processing Linguistic Probabilities: General Principles and Empirical Evidence. *The Psychology of Learning and Motivation* 32: 275–318.

Burtraw, D., A. Krupnick, et al. 1998. Costs and Benefits of Reducing Air Pollutants Related to Acid Rain. *Contemporary Economic Policy* 16: 379–400.

Cleveland, W., and R. McGill. 1984. Graphical Perception: Theory, Experimentation, an Application to the Development of Graphical Methods. *Journal of the American Statistical Association* 79(387): 531–554.

Cooke, R.M., and J.M. van Noortwijk. 2000. Graphical Methods. In *Sensitivity Analysis,* edited by A. Saltelli, K. Chan, and E.M. Scott. New York, NY: Wiley.

Curley, S. P., J. F. Yates, et al. 1986. Psychological Sources of Ambiguity Avoidance. *Organizational Behavior and Human Decision Processes* 38: 250–256.

Ellsberg, D. 1961. Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics* 75(4): 643–669.

Flugstad, A., and P. Windschitl. 2003. The Influence of Reasons in Interpretations of Probability Forecasts. *Journal of Behavioral Decisionmaking* 16: 107–126.

Funtowicz, S.O., and J.R. Ravetz. 1990. *Uncertainty and Quality in Science for Policy*. Dordrecht, the Netherlands: Kluwer Academic.

Gneezy, U., J. List, et al. 2004. *The Uncertainty Effect: When a Risky Prospect is Valued than its Worst Possible Outcome.* Chicago, IL: University of Chicago, Center for Decision Research.

Ibrekk, H., and M.G. Morgan. 1987. Graphical Communication of Uncertain Quantities to Non-Technical People. *Risk Analysis* 7(4): 519–529.

IPCC (Intergovernmental Panel on Climate Change). 2000, Summary for Policymakers. In *Special Report on Emissions Scenarios.* http://www.grida.no/climate/ipcc/emission/001.htm (accessed Feb. 16, 2006).

Kuhberger, A. 1998. The Influence of Framing on Risky Decisions: A Meta-Analysis. *Organizational Behavior and Human Decision Processes* 75: 23–55.

Lee, R., D. Burtraw, and A. Krupnick. 1995.*The External Costs and Benefits of Fuel Cycles. A Study by the U.S. Department of Energy and the Commission of the European Communities.* Prepared by Oak Ridge National Laboratory and Resources for the Future. New York, NY: McGraw-Hill Utility Data Institute, December.

Lipkus, I., and J. G. Hollands. 1999. The Visual Communication of Risk. *Journal of the National Cancer Institute* 25: 149–163.

Manning, M. 2003. The Difficulty of Communicating Uncertainty: An Editorial Comment. *Climatic Change* 61: 9–16.

Morgan, M.G., and M. Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York, NY, Cambridge University Press.

Patt, A., and D. Schrag. 2003. Using Specific Language to Describe Risk and Uncertainty. *Climatic Change* 61: 17–30.

Roozenberg, A., and A. Nicholson. 2003. *Required Passing Sight Distance for Rural Roads: A Risk Analysis.* Wellington, New Zealand: Institution of Professional Engineers New Zealand. http://www.ipenz.org.nz/ipenztg/ipenztg_cd/cd/2003_pdf/10_Roozenburg_Nicholson.pdf (accessed Feb. 1, 2006).

Siebenmorgen, N., E. Weber, et al. 2000. *Communicating Asset Risk: How the Format of Historic Volatility Information Affects Risk Perception.* Mannheim, Germany: University of Mannheim.

Slovic, P., B. Fischoff, et al. 1982. Facts versus Fears: Understanding Perceived Risk. In *Judgment Under Uncertainty: Heuristics and Biases,* edited by D. Kahneman, P. Slovic, and A. Tversky. New York, NY: Cambridge University Press, 463–492.

Stevens, S. S., and E. Galanter. 1957. Ratio Scales and Category Scales for a Dozen Perceptual Continua. *Journal of Experimental Psychology* 54: 377–411.

Stone, E., J. F. Yates, et al. 1997. Effects of Numerical and Graphical Displays on Professed Risk Taking Behavior. *Journal of Experimental Psychology: Applied* 3(4): 243–256.

Thompson, K., and D. Bloom. 2000. Communication of Risk Assessment Information to Risk Managers. *Journal of Risk Research* 3(4): 333–352.

Tversky, A,. and C. Fox. 1995. Weighing Risk and Uncertainty. *Psychological Review* 102: 269–283.

Tversky, A., and D. Kahneman. 1981. The Framing of Risky Decisions and the Psychology of Choice. *Science* 211: 453–458.

———. 2000. Rational Choice and the Framing of Decisions. In *Choices, Values and Frames,* edited by D. Kahneman and A. Tversky. New York, NY: Cambridge University Press, 209–223.

U.S. EPA (Environmental Protection Agency). 2004. *Final Regulatory Impact Analysis: Control of Emissions from Nonroad Diesel Engines.* EPA420-R-04-007. Washington, DC: U.S. EPA.

van der Sluijs, J., P. Kloprogge, J. Risbey, and J. Ravetz. 2003. *Towards a Synthesis of Qualitative and Quantitative Uncertainty Assessment: Applications of the Numeral, Unit, Spread, Assessment, Pedigree (NUSAP) System.* Presented at the International

Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling, Aug. 19–21, Rockville, MD.
van der Sluijs, J., M. Craye, et al. 2004. Combining Quantitative and Qualitative Measures of Uncertainty in Model Based Foresight Studies: The NUSAP System. Presented at the UPEM Conference, Copenhagen, Denmark.
van Dijk, E., and M. Zeelenberg. 2003. The Discounting of Ambiguous Information in Economic Decision Making. *Journal of Behavioral Decisionmaking* 16: 341–352.

Chapter 5: Presentation of Uncertainty Information to High-Level EPA Decisionmakers

The most sophisticated uncertainty analysis will go to waste if it is not effectively presented to the individuals making the policy choice. As discussed in Chapter 4, this means taking into account their background and familiarity with the material as well as the advantages and disadvantages of different presentation techniques. Unfortunately, there remains much to learn about how best to present risk and uncertainty information to high-level policymakers.

To shed light on this question, we interviewed seven former high-level decisionmakers at the U.S. Environmental Protection Agency (EPA). The interviews were structured around a mock briefing, which included several different graphical presentations of uncertainty. The respondents were asked about their reaction to the different approaches, how the graphics influenced their decisionmaking, and their general thoughts on the treatment of uncertainty in regulatory analyses.

All interview subjects had served in senior positions at EPA at the level of assistant administrator or deputy administrator during the period 1989–2004. The decision was made to restrict the respondents to individuals at the highest level of the decisionmaking process because these individuals were less likely to have detailed knowledge of the techniques of uncertainty analysis and because their interpretation of uncertainty could have potentially large ramifications.

The interviews were designed to elicit feedback on several aspects of decisionmaking under uncertainty. The first was how well different graphical displays perform in conveying information about uncertainty to decisionmakers. On one level, this was a question of how easily the decisionmakers could process the information presented in the visuals after a brief introductory explanation. On a more subjective level was the question of whether the policymakers believed that the visuals aided their decisionmaking process.

A second area of inquiry was how uncertainty analysis (and technical analysis in general) is used by decisionmakers. The kind of technical analysis presented in Chapter 3 is but one of the important pieces of information that agency decisionmakers typically consider. In a real-world situation, other factors also may play important roles (e.g., the state of the economy, the political climate, and historical experience with regulating the pollutant). In the course of our presentation, decisionmakers were given an opportunity to indicate the other types of information they would rely on when making a decision.

Third was the question of how policymakers make decisions under uncertainty. The policy choice was structured around two options with similar mean net benefits but different variances. The decision between the two policies therefore involved choosing between spending larger amounts of money for potentially high net benefits or spending less money and avoiding potentially high net costs. This part of the discussion provided insight into the sorts of frameworks and heuristics policymakers that rely on in making such choices.

Finally, the policymakers for their perspectives on how uncertainty analysis should fit into the institutional decisionmaking structure of the agency. Issues raised included how much of this type of analysis is necessary for top-level decisionmakers and how process refinements can enable the analysis to positively affect regulatory decisionmaking.

The presentation was based on an analysis quite similar to that presented in Chapter 2, although it was modified to make the policy choice a closer call. We asked each interview subject for a decision on a hypothetical proposed tightening of the Clean Air Interstate Rule (CAIR). They were presented with three options: (a) doing nothing, (b) an intermediate option of reducing the nitrogen oxide (NO_x) cap by an additional 20% below baseline in 2020, and (c) a more stringent option of reducing the NO_x cap 40% from baseline in 2020. We asked the decisionmakers to imagine that they were making the decision in the 2012–2015 time frame to reduce complications regarding discounting, future regulatory developments, and the like. The interviewees were told that their names would be listed (as a group) in the final report to EPA but that none of the responses would be attributed to individuals.[1] Although a formal script was not used in the actual interviews, a stylized version of the information presented and the questions posed is contained in Appendix 5A (available on request from Alan Krupnick). Summaries of the discussions are presented in Appendix 5B.

---

[1] The following individuals were interviewed: Don Clay, Terry Davies, Linda Fischer, David Gardiner, Lynn Goldman, Hank Habicht, and Tracy Meehan.

The presentation began with a brief overview of the goals of the project and recent developments regarding the incorporation of uncertainty in regulatory analysis. The decisionmakers were then given information on the role of NO_x as a ozone and particulate matter precursor and the health consequences of the resulting pollution. They also received a brief description of the CAIR rule along with a summary of EPA's estimated benefits of the rule. They were then given a description of the three proposed options.

Graphical Material

The decisionmakers were presented with a series of slides containing either tables or figures and asked for their reaction to them. After each slide, they were asked if the material provided them with enough information to decide on which option to choose and if so what their decisions were. The presentation included seven slides in total: two tables and five figures.

Physical Effects

The first table (Table 5-1) shows the impacts of the two proposed policies in terms of physical health impacts and costs in 2025. As expected, the tighter option is more expensive but averted more mortality and morbidity. However, there is fairly wide uncertainty around the health estimates, with the confidence interval for mortality extending between 122 and 810 premature deaths for the stringent option and between 65 and 443 premature deaths for the intermediate option. Because of modeling limitations, no uncertainties around costs are presented.

Table 5-1 generally was well-received by all respondents, but several said they would have liked to see cost-effectiveness calculations, such as the cost per life saved or the cost per ton reduced. In fact, several decisionmakers made back-of-the-envelope calculations for these numbers. Several respondents were troubled by the lack of uncertainty about costs and said it was their impression that costs were frequently overstated. Many respondents said the confidence intervals raised the question of what factors were driving the uncertainty and requested more information on this subject. Some of the interviewees were surprised that both benefits and costs were fairly linear; a doubling of required emissions reductions resulted in an approximate doubling of both costs and benefits. (It was then explained that this was an artifact of the modeling analysis. For other pollutants and other control options, nonlinear outcomes might result.)

Monetized Benefits

The second table (Table 5-2) presents the results from the cost–benefit analysis, showing total benefits, costs, and net benefits in 2025. Although both options had different total benefits and costs, they were virtually identical with respect to the best estimate for net benefits: $10 million for the intermediate option and $11 million for the stringent option. However, the two policies had very different ranges for net benefits, with the more stringent option ranging from –$831 million to $854 million and the intermediate option ranging from –$455 million to $474 million.

Reaction to this table was more mixed. Most respondents said that it was helpful, but only in conjunction with the first table (Table 5-1). Two said they were skeptical of monetization and thought that it didn't provide much guidance on close calls like this decision. One respondent said that monetization is useful for making an "apples-to-apples" comparison but that it is hard to convey the concept to the general public; and for purposes of selling the policy, the information in Table 5-1 was more useful. Two interviewees were surprised that the net benefits were so similar for the two policies. They noted that in their time at EPA, they often felt presentations were structured around a prechosen option and one or more straw-men options, making the process almost preordained. All subjects said that in a close call like this decision, other factors such as political considerations would loom large.

Pie Chart

The first figure we showed to the policymakers was a simple pie chart displaying the probabilities that the policies would produce positive net benefits (Figure 5-1). The analysis shows that both policies had about the same probability of achieving net benefits (about 53%), so the charts were quite similar. Some of the interviewees found this chart helpful and said it gave them more confidence in their decision. Others found it too crude and said it did not provide much more information than was already contained in the tables. Some who said it was not helpful also said they could imagine cases in which it might be helpful (e.g., if the two policies differed greatly in the probability of achieving net benefits).

Box-and-Whisker Plot, PDF, and CDF

The decisionmakers were then shown three of the most common graphical formats for presenting uncertainty: a box-and-whisker plot, a probability density function (PDF), and a cumulative density function (CDF).

In general, the policymakers were not positive about the box-and-whisker diagram (Figure 5-2); several said it was repetitious and did not provide much additional information beyond the tables. Some felt that the presentation focused their attention on the middle of the distribution and away from the extremes, although one noted that the plots showed that the policies could have large negative net benefits, and opponents might seize on that number. Three interview subjects said they would have rather seen total benefits presented in the place of net benefits, because the latter bundled too much information into a single number.

The interviewees responded much more favorably to the PDF (Figure 5-3), saying it was more intuitive and persuasive. One respondent said the PDF was an easier graphic to frame a discussion around than the others. Interestingly, a majority of the respondents said the PDF made them gravitate toward the intermediate option. This is consistent with findings by Siebenmorgen et al. (2000) that PDFs make people more sensitive to extreme values and consequently more risk averse.

Only one interview subject was familiar with CDFs, so Figure 5-4 seemed to require the most detailed explanation. Most of the decisionmakers did not find the CDF very helpful, even after explanation. In fact, the sole individual who had experience with CDFs would not be inclined to present it to most decisionmakers because it was hard to explain. One decisionmaker, however, said the CDF was a more neutral presentation than the PDF because, unlike the PDF, it did not push toward the intermediate option.

Sources of Uncertainty

The final figure was a graph showing the relative contributions of several key variables to the uncertainty associated with the estimate of net benefits (Figure 5-5). Four of the respondents found it very helpful; in fact, in several cases, the graphic was moved up in the presentation in response to specific questions about the factors most responsible for the uncertainty. One of the interviewees said it was helpful because it identified different types of uncertainty: “If the uncertainty is coming from population projections, I can do something about that—choose which one I think is most reasonable. If it's coming from the science, I can't do much about that."

The two individuals who said the graph was not useful said that the mortality concentration–response function and the value of the statistical life were the biggest sources of uncertainty and that they had assumed this going into the presentation.

Also, one of the respondents who rated the graph as helpful said the wide band around mortality raised questions about the quality of the studies used in the analysis.

The Policymakers' Decisions

All interview subjects said that, in practice, their decisions would require much more information than was included in this presentation. They questioned distributional issues, the political balance of power, and the current state of the economy. With this caveat, all respondents were willing to choose an option. One favored doing nothing, and one favored the stringent option. Three favored the intermediate option. One narrowed the choice to either the intermediate or the stringent option and said the final choice would depend on political factors, such as the relative strength of EPA vs. the Office of Management and Budget (OMB) in the administration. The final decisionmaker would prepare both the stringent and intermediate for review with a recommendation to go with the stringent option.

The decisionmaker who chose to do nothing cited the large uncertainties associated with both policies and the fact that the estimated net benefits for both were close to zero. The decisionmaker who chose the stringent option was willing to accept the possibility of large negative net benefits for the potential of achieving very large health gains. Reasons given for supporting the intermediate option included the tighter error bounds, the option to do more later, and the fact that counting the omitted ecological benefits would push the net benefits into positive territory.

Views on Uncertainty in the Process

After the presentation of the information drawn from the case study, we queried the interviewees about their overall views on the use of uncertainty information in decisionmaking. We asked them whether they felt the information aided their decisionmaking and about their degree of comfort in delegating potentially significant technical decisions about the analysis to staff.

All respondents said they found the presentation of uncertainty helpful. Many identified information on the sources of uncertainty as an important input to decisionmaking. The presentation of uncertainty was also useful for giving decisionmakers insight into how confident they should be in their decision, and some said it would be very useful in those cases where the ranges of net benefits were all positive or all negative. Additionally, some decisionmakers cited the need to see everything that potential critics might see so they are better prepared to defend the policy.

On the issue of delegation, interviewees said the complexity of the issues made delegation unavoidable. As one said, "If you weren't comfortable with it, you would either quit or go insane. That's why you have a staff." Many favored building safeguards into the process, for example, having different offices of EPA analyze the regulation separately or at least review the underlying details of the analyses. Another suggestion was to have outsiders like the Science Advisory Board (SAB) review and approve the uncertainty techniques used. One respondent said that senior decisionmakers often don't know what they have delegated: "They don't know what they don't know. … I often find out more and more about what I didn't know when I made a particular decision."

Miscellaneous Observations

In addition to the responses that came out in the structured interviews, respondents also made general observations on uncertainty and environmental decisionmaking.

- *The need for context.* Every decisionmaker said providing context was crucial to helping inform the decisionmaking process. The kinds of context that they wanted included information on the magnitude of the problem; how much had been done previously to address this problem; how the costs of the proposed regulations compared with those of other regulations; distributional patterns of costs and benefits, both geographically and across demographic groups; and political factors, such as where different interest groups stood on the proposal(s).

- *A preference for cost-effectiveness numbers.* Nearly all decisionmakers seemed more comfortable thinking in terms of cost-effectiveness (e.g., dollars per life saved or dollars per ton reduced) rather than in cost–benefit terms. Although none were hostile to the concept of monetization, many felt that presenting information in terms of net benefits collapsed too much information into a single number.

- *Different styles of decisionmaking.* Several interviewees noted that there were likely to be important differences in the responses to the formal presentation of uncertainty analyses based on the decisionmakers' backgrounds. Those with a technical background were more comfortable with the kind of presentation that was we used in this project. However, several respondents observed that many EPA decisionmakers have legal training and that lawyers tend to be more comfortable with an argument-based approach that lays out the pros and cons of the different options. One decisionmaker said that lawyers liked to debate the proposed policy and that "I can remember being moved by how strongly someone made their case."
- *The importance of political considerations.* Many decisionmakers framed their discussions around how the analysis would help or hinder their ability to sell the policy to the White House or to outsiders. One noted that uncertainty analysis had a downside in that it armed critics like OMB, who would be tempted to seize on lower-bound numbers. One decisionmaker said they might be tempted to select the tight option even though they favored the intermediate option, because they would rather negotiate with OMB starting from a more stringent standard. This policymaker likened developing a regulation to running a gauntlet and lamented the adversarial nature of the process. One policymaker cautioned against placing too much emphasis on politics first: "Get the facts first before you bring on the politics. … If you do the politics first, you will get confused. In the end, you want to know how far you had to move from the most analytically defensible options."

Conclusions

Performance of Graphical Presentations

All respondents found the tables informative and were able to interpret them without difficulty. Several made rough cost-effectiveness calculations of their own on the basis of the information in the tables. They found the PDFs most familiar of the graphical displays, and they appeared most comfortable using them to make policy judgments. Almost all respondents said the PDFs made them more inclined to choose the intermediate option, even if they ended up favoring the stringent or do nothing options. This raises questions of whether PDFs might create a bias toward tighter spread options.

The policymakers had more difficulty understanding the CDF, and many required a detailed explanation of how to interpret the graph. In addition, most said it was not helpful to their decisionmaking. The respondents easily grasped the pie charts (showing the likelihood of benefits exceeding costs for each option), but many said they did not provide helpful information. One reason for this may be the nature of the policy choice in this exercise, which resulted in two virtually identical pie charts; pie charts may still be useful for emphasizing differences when they exist. The box-and-whisker plots were also easily understood, even by those who were unfamiliar with them. However, most said these graphs did not provide much more information than was already presented in the tables. All policymakers understood the graph showing the sources of uncertainty and most found it very helpful.

Given the results from this admittedly small sample, tables and PDFs appear to be best suited for communicating to high-level decisionmakers. One caveat is that only two options were presented in this exercise. As the number of options or scenarios increases, PDFs tend to look fairly busy, and comparative box-and-whisker plots might be a cleaner way to convey the same information. In addition, because PDFs tend to move respondents toward options with less uncertainty, we should provide additional explanation on the implications of different choices (e.g., forgoing the opportunity for higher risk reductions in exchange for avoiding the possibility of higher net costs).

Uncertainty Analysis in Context

One clear message from this exercise is the importance of presenting technical analysis in context. The respondents were quite forthcoming about the kinds of additional material they would need to make an informed decision. As one put it, "The answer is not in these graphs."

The policymakers were particularly interested in the regulatory history of the pollutant and a summary of how the current proposal compared with past efforts, particularly in terms of cost-per-ton and cost-per-health-improvement. They also were interested in more information on the size of the current problem. Other types of contextual information requested included how the current proposals compared with other regulatory options, the degree of political opposition to the initiatives, a description of nonmonetized benefits, and impacts on vulnerable groups.

Decisionmaking Rules of Thumb

The mock policy options themselves represented something of a close call. They had best estimates of around $10 million in net benefits each but very different spreads. The more aggressive policy offered the potential for very high net benefits if NO_x emissions turn out to have a strong impact on mortality. Conversely, the policy also has the possibility of high net costs if the actual health impacts are small and the increased expenditure produces little in the way of additional health benefits.

Policy choices were not unanimous. In general, decisionmakers strongly favored the intermediate option, and only one of the policymakers favored doing nothing. The reduced possibility of making a major error was the main reason given for the attractiveness of the intermediate option. But beyond this, it is hard to generalize about the policymakers' decisions.

Incorporating Uncertainty into the Decisionmaking Process

Several respondents said that in order for uncertainty analysis and technical analysis in general to be most credible, an internal system of checks and balances should be created. It could be accomplished by having multiple offices within EPA look at the analysis or by establishing an external peer review process. Additionally, some argued for a presentation format that took the form of arguments pro and con for different regulatory options. One policymaker said that it might be better to have input from OMB at an early stage of the process, so that potential points of controversy could be identified earlier.

Figures

Figure 5-1. Probability that Policies Produce Net Benefits in 2025: Comparison of Stringent and Intermediate NO_x Caps

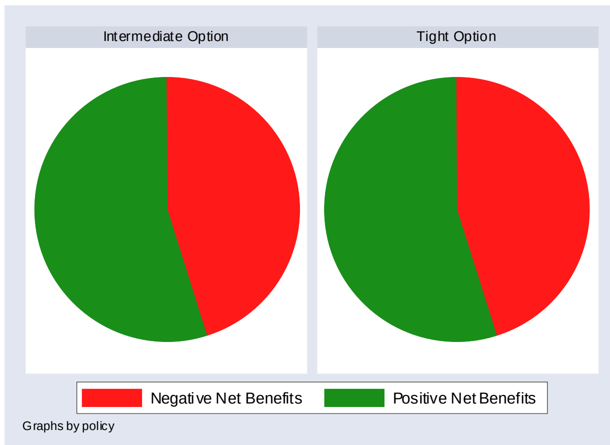


Figure 5-2.

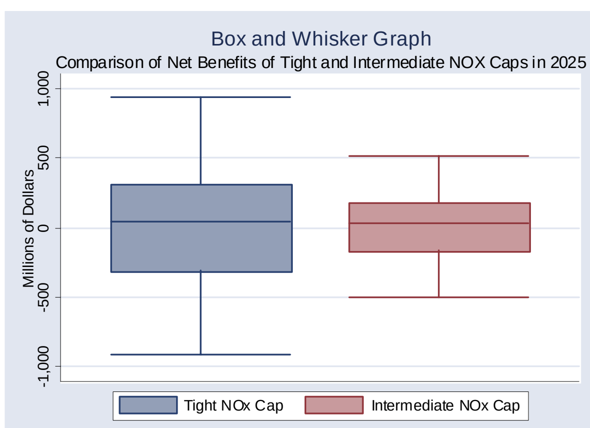


Figure 5-3.

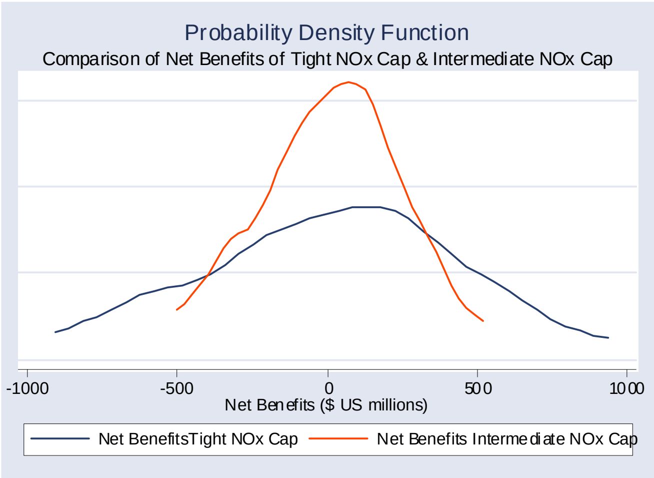


Figure 5-4.

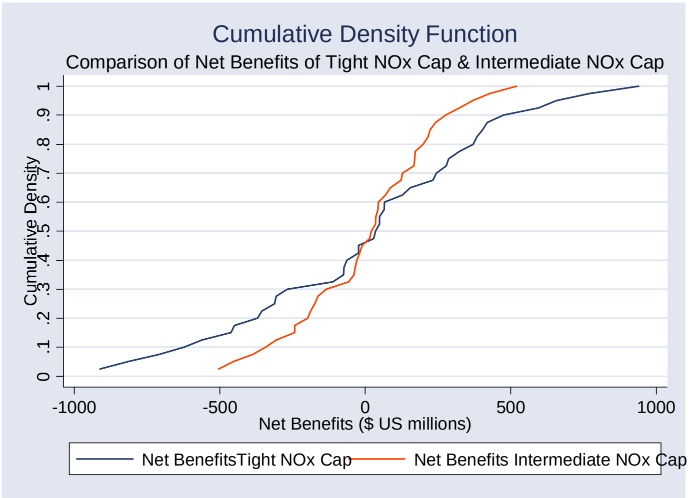
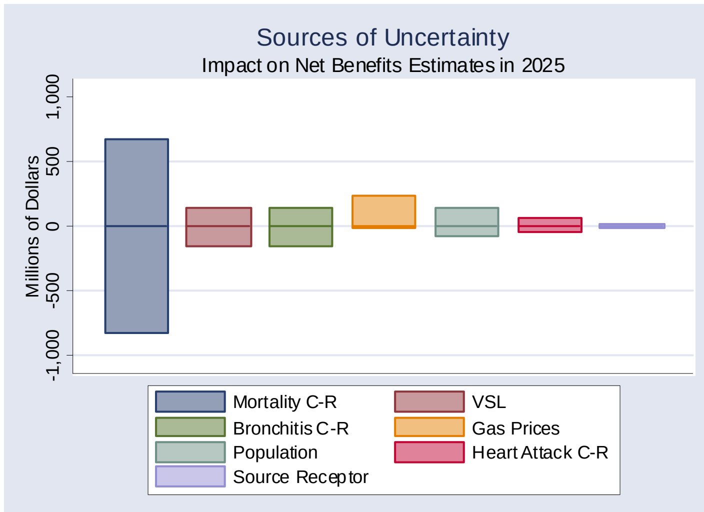


Figure 5-5.


Sources of Uncertainty
Impact on Net Benefits Estimates in 2025
Millions of Dollars
1,000
500
0
-500
-1,000
Mortality C-R
VSL
Bronchitis C-R
Gas Prices
Population
Heart Attack C-R
Source Receptor

Table

Table 5-1.

	Comparison of Tight NOx Cap and Intermediate NOx Cap Policies					
	Averted Physical Impacts in 2025					
	Tight NOx Cap			Intermediate NOx Cap		
	Mean	95% CI lower bound	95% CI upper bound	Mean	95% CI lower bound	95% CI upper bound
Mortality	466	122	810	254	65	443
Cardiovascular Hospital Admissions Admissions/Year	409	47	771	230	27	434
Non-Fatal Heart Attacks Cases/Year	995	338	1652	543	187	900
Respiratory Hospital Admissions Admissions/Year	2611	1550	3672	841	512	1169
Cardiovascular Hospital Admissions Admissions/Year	338	204	471	197	123	272
Asthma Emergency Room Visits/Year	598	358	838	265	150	380
Cost (millions $)	1340			710		

Table 5-2.

	Comparison of Tight NOx Cap and Intermediate NOx Cap Policies					
	Net Benefits in 2025					
	Tight NOx Cap			Intermediate NOx Cap		
	Mean	95% CI lower bound	95% CI upper bound	Mean	95% CI lower bound	95% CI upper bound
Total Benefits ($ US millions)	1351	509	2194	720	255	1184
Costs ($ US millions)	1340			710		
Net Benefits ($ US millions)	11	-831	854	10	-455	474

Appendix 5A: Stylized Version of the Script Used for Discussions with Senior EPA Officials

This appendix is available at www.rff.org/makingregulatorychoices.

Appendix 5B: Summaries of Individual Interviews

This appendix is available at www.rff.org/makingregulatorychoices

Chapter 6: Conclusions and Recommendations

In this final chapter, we present a series of conclusions from this project as well as several recommendations and ideas for further research. Overall, the U.S. Environmental Protection Agency (EPA) and other regulatory agencies face many challenges in responding to the Office of Management and Budget's (OMB's) *Circular A-4*. Arguably, EPA's early and continuing work in developing guidelines for Monte Carlo analysis and in incorporating formal quantitative assessments into many major studies and regulatory impact analyses (RIAs) puts EPA in a strong position—likely ahead of most other agencies—in responding to *Circular A-4*.

At the same time (as detailed in this report), there are substantial opportunities to improve the analytic rigor of the assessments and, simultaneously, to improve communication of the findings of the assessments to decisionmakers in the agency and to outsiders, including the general public. Some promising opportunities and the challenges they represent are summarized in this chapter.

Conclusions

Typologies for Characterizing Uncertainty

The literature is full of overlapping and competing typologies for classifying uncertainties. We were able to classify the uncertainties we quantified as well as identify (and classify) some uncertainties we did not quantify. We judged this process to be a useful part of our case study. Our interview with Karen Palmer, Darius Gaskins Senior Fellow in the Quality of the Environment Division at Resources for the Future, about the Haiku model led us to realize that, in the process of working through the uncertainty analysis, we identified some uncertainties that we otherwise would have missed. In our judgment, this exercise would be productive at any stage of uncertainty analysis, whether at the beginning (from a model design standpoint) or later (when reviewing and interpreting results).

Techniques for Representing Uncertainty

Many techniques are available for portraying and analyzing uncertainty at what we call the analyst level—that is, a technical stage before data reduction and well before results are brought to decisionmakers. As analysts ourselves, we found that the box-and-whisker plot was the most useful of the simple tools for portraying and thinking about the consequences of quantified uncertainties in our case study.

We also investigated some advanced tools, most notably, the cobweb plot. Because the relationships in our case study are generally linear and continuous, the benefits of this technique are not highlighted to the fullest extent. Thus, we created a simple model with a threshold in the concentration–response (C–R) function to give the reader a better example of the power of this approach (see Appendix 3E in Chapter 3).

Representing Uncertainties in the Case Study: Successes and Challenges

First, it is not difficult to model additional uncertainties beyond those traditionally modeled in RIAs, such as C–R and valuation uncertainty. For instance, we have shown how population uncertainties can be modeled using standard Bureau of the Census population series, and how introducing uncertainties in future natural gas prices is relatively easy. However, in both these cases, we introduced uncertainty in nonparametric ways (i.e., through a sensitivity analysis, rather than through a distributional analysis). Thus, we have not yet demonstrated an approach to statistical modeling of uncertainty on the cost side.

At the same time, introducing such uncertainties on the cost side necessitates running the baseline and policy scenarios again. Thus, it is generally not possible to predict how net benefits will change for alternative values of the uncertain parameters. We see this in the result that both lower and higher natural gas price forecasts resulted in higher net benefits than the base-case forecast.

Second, on the benefit side, we had success in introducing uncertainties in source–receptor (S–R) coefficients. Those introduced through the Advanced Source Trajectory Regional Air Pollution (ASTRAP) reduced-form model significantly affected benefits. At the same time, the uncertainties we introduced through our Urban-to-Regional Multiscale (URM) uncertainty analysis had little effect on benefits. However, scaling our derived S–R matrix for specific episodes to be representative of annual S–R relationships using the Classification and Regression Tree (CART) statistical technique was a novel approach to accounting for weather variability.

Third, our means of portraying the information generated by this complex case study may offer some useful guidance to EPA. The box-and-whisker plots were very convenient and concise means of summarizing large quantities of information, making it possible to place multiple plots on the same graph, such as for Figure 3-7, which succinctly shows the model uncertainty in the particulate matter–mortality C–R function component of the benefits analysis. Related are Figures 3-5, 3-11, and 3-12, which attribute uncertainties in aggregate benefits to the uncertainties in each component of the base-case benefits analysis. Together, the figures allow the identification of those components in which further research would make the greatest contribution to narrowing the overall uncertainty in benefits. Additionally, we found the cobweb plots useful because they enable users to visualize complex relationships that may not be apparent in other graphical or numerical representations.

Our fourth finding will come as no surprise to analysts, The knottiest issue in implementing the uncertainty provisions of *Circular A-4* is determining how to systematically address model uncertainties while merging such work with a treatment of statistical uncertainties to reveal uncertainties in the overall costs and benefits. After such assessments are performed, the challenge is to present them in an understandable and efficient way—a more challenging task than presenting statistical or model uncertainties alone.

Key Findings from the Uncertainty Communication Literature

Although a great deal of research has been conducted on the communication of uncertainty and risk, very little attention has been focused on the means of communicating the results of such analyses to policymakers. Instead, the orientation has been toward understanding how to present uncertainty to lay audiences and help them put low-probability risks in appropriate context. The issue of communicating uncertainties associated with climate change to policymakers garners increasing attention, with a focus on high-consequence outcomes. But the issue of communicating uncertainty in a typical regulatory decisionmaking process remains largely unexplored.

Psychological research on decisionmaking under uncertainty has uncovered numerous instances in which decisions are influenced simply by the manner in which a problem is presented. Because decisionmakers (and even experts) are just as susceptible to these cognitive biases as the general population, the data analyst's choice of presentation format could influence a policymaker's decision. Furthermore, some evidence suggests that as the emphasis on uncertainties increases, so does the probability that decisionmakers will lose confidence in the overall analysis.

Research on the effectiveness of different graphical techniques has demonstrated that box-and-whisker plots, probability density functions, and cumulative density functions perform relatively well in allowing the audience to accurately extract quantitative information. Area and volume presentations can be misleading and cause viewers to underestimate large magnitudes and thus should be avoided. Ways to convey the importance of uncertain variables is one emerging area of interest. Beyond standard tornado graphs, novel approaches such as radar graphs, cobweb plots, and pairwise scatterplots offer ways to present large amounts of information in an economical manner, although these approaches might be too complex for a nontechnical audience. (See Chapter 3 and Appendix 3E for examples of cobweb plots.)

Communicating Uncertainties to Decisionmakers

We conducted in-depth interviews with seven former EPA assistant or deputy administrators in which we presented the basic results of the case study using alternative metrics and graphics, then solicited their opinions about the presentations. From their responses, a number of observations can be made.

First, the interviewees are rather heterogeneous in backgrounds and in interest in and familiarity with uncertainty assessments. This heterogeneity no doubt led to differences in the ease with which they interpreted alternative metrics and graphics portraying the results of our case study. Therefore, we find it difficult to generalize about the techniques used and challenges encountered in communicating these types of results.

Nevertheless, even with the limited sample, interviewees were most comfortable with the use of probability density functions (PDFs) and simple tabular formats rather than the complex graphics more commonly used by analysts (and favored by us)—namely, box-and-whisker plots, cumulative density functions, and circle charts. We conjecture that the box-and-whisker plot would be increasingly useful and the PDFs less so as the number of variables considered increased.

Beyond PDFs and simple tabular formats, the former decisionmakers also favored other graphs. For example, they were particularly interested in the graphic displaying the relative importance of the various factors considered in the uncertainty analysis. On several occasions, they specifically asked about relative importance even before the graphic was presented.

The interviewees also were interested in identifying any factors for which uncertainty might be an important issue but that had been excluded from formal uncertainty analysis. This list included factors that could, in principle, be included in uncertainty analysis (e.g., uncertainties in the efficiency of NO_x removal technology) as well as those that would always lay outside such analyses (e.g., institutional and political factors).

Conclusions from the NO_x Case Study

The experience of performing a case study yielded its own set of conclusions that are suggestive about the usefulness of uncertainty analysis, quite apart from technical issues. Our decisionmaking process to choose air pollution policy for the case study reminded us that air pollution policy is probably the best-case situation at EPA or other agencies for studying uncertainty because this area has well-developed integrated assessment models and a history of examining both statistical and modeling uncertainties. Other areas of regulatory activity are less favorable in terms of data availability and modeling capability. At the same time, the *Circular A-4* requirement should stimulate EPA in its data and modeling efforts.

Recommendations

On the basis of the conclusions noted above and our overall experience with this project, we offer the following recommendations.

1. Hold a workshop on introducing uncertainties into air quality modeling.

We have presented two examples of how S–R uncertainties can be modeled, but neither of these examples aligns perfectly with EPA modeling priorities in this area. Our first example was with the ASTRAP reduced-form model in which the S–R coefficients “came with” distributions around the coefficients. This approach is consistent with Monte Carlo simulation of uncertainties, but such models are not the favored models at EPA.

Our second approach used a detailed simulation model (an approach favored by EPA) to derive a series of S–R matrices for each day of each episode, then used a CART statistical analysis to scale the results for a year of weather and, in so doing, to introduce variability about weather. The usefulness of this approach remains to be seen.

The ASTRAP approach introduces uncertainty in the underlying meteorological data, but clearly neither approach introduces uncertainty in the other underlying parameters of the simulation models—an approach we judge to align well with EPA modeling priorities. For instance, rate constants—parameters governing the transformation of emissions into concentrations of pollutants in the air—are uncertain. So distributions of such parameters could be sampled in Monte Carlo analysis, at least in theory.[1] In practice, limitations on computational resources might preclude such an approach, because it already takes as much time to compute the consequences of an episode for air quality as it does the episode itself.

Because of the computation and perhaps conceptual challenges of introducing uncertainty in these S–R relationships, a workshop with top air quality modelers and uncertainty analysts could advance the state of the art.

2. Hold a workshop on introducing population demographics into modeling of costs and benefits.

Our work incorporates only the most basic variation in projections of future population, but such variation can have important impacts on both costs and benefits. More sophisticated analysis could account for differences in migration and fertility by region of the county. Correlations in regional population growth, both positive and negative, exist between regions.

EPA uses population data developed by the Bureau of the Census and other organizations. The agency could consider convening a conference to facilitate model development that would better and more consistently accommodate the subtle population demographic uncertainties on both benefits and costs.

[1] In practice, some limited air quality modeling analyses have accounted for rate constant uncertainty (See Bergin et al. 1998 and Hanna et al. 2001 in the Chapter 3 References.)

3. EPA should consider adding certain types of uncertainties to its RIAs for air pollution policies.

EPA should consider adding population uncertainties to its RIAs. We are not familiar with all the models that EPA uses to make cost estimates for RIAs. However, EPA's counterpart of the Haiku model—the Integrated Planning Model (IPM)—is very complex, and its creators should be asked to estimate the ease with which population projections could be manipulated.

EPA also should consider adding natural gas price uncertainties and similar types of uncertainties to its RIAs. In previous studies, the EPA has asked for the IPM authors to conduct scenario analyses using different natural gas prices. If the cost of doing such analysis with IPM is not prohibitive, it should be common practice around variables that could contribute the most to uncertainty in costs or profiles of emissions.

4. EPA should expand its institutional capacity for addressing and communicating uncertainty.

The findings of interviews we conducted with the former EPA decisionmakers, plus additional conversations with EPA senior staff, indicate that information about uncertainty assessments—including both how the analysis was conducted and how results were interpreted—is quite limited. Accordingly, we recommend that a formal effort be initiated to inform and educate senior staff at various levels about uncertainty assessments. This effort could take several forms, but one possible approach is the development of simplified case study materials that could be used in seminars and related events. A similar approach was successfully used in the agency in the mid-1980s to inform senior staff (and others) about then-emerging work on risk assessment.

5. EPA should enhance its research support for issues raised in this report to better address the challenges of* Circular A-4*.

EPA research on this topic (beyond our project) is already under way. As discussed in Chapter 2, EPA's Council for Regulatory Environmental Models (CREM) has initiated several projects with implications for the treatment of uncertainty in RIAs, such as a draft guidance document on environmental models, an online Models Knowledge Base, and a series of regional seminars.

CREM also has teamed with the National Academies of Science and participated in symposia with the Woodrow Wilson International Center to address the role of uncertainty in environmental modeling and decisionmaking at EPA. These efforts show an increased attention to including uncertainty in the development of new models and the analysis of results.

Although focus has centered on defining the desirable qualities of models (e.g., resembles real-world interactions, addresses different types of uncertainty, provides transparent documentation and analysis, reconciles alternate model predictions, permits assessment of the quality or accuracy of model results), more research is needed to meet the challenge of putting these ideas into practice and ensuring that results are presented in the most useful and compelling way. Our research contributes to this goal.

Further Research

Extensions to Our Interview Protocols

Given the small interviewee sample for this project, a sensible next step would be to solicit the interest in and preferences for alternative presentation formats among a broader group of agency decisionmakers. Possible expansions could involve additional interviews with individuals at the level of assistant or deputy administrator or the inclusion of additional levels of EPA management (e.g., office directors or division directors). Another possible avenue would be to conduct interviews with consumers of uncertainty analysis from outside the agency, for example, members of nongovernmental organizations.

In addition, our interviews highlighted some issues that were not part of the core presentations. For instance, many decisionmakers inquired about the distributional impacts of the policies. This information was not contained in the mock briefing, but technical analysis of distributional impacts could be integrated into future presentations. Given the decisionmakers' interest in equity, it might be useful to explore the effectiveness of different presentation options for this topic.

The decisionmakers also expressed a great deal of interest in the sources of uncertainty. Because of time constraints, we presented only one type of graphic on this subject. Showing decisionmakers more approaches might yield useful insights for communicating uncertainty within the agency.

Addressing Model Uncertainties

As noted earlier, probably the hardest challenge for uncertainty analysis is addressing model uncertainty. Decisionmakers can rightly ask whether the analyst's chosen "models" are representative of models in the literature and, more pointedly, whether the analysis will pass peer review on this criterion. The ultimate issues are how weights will be assigned to alternative models of a given phenomenon in the literature and by whom.

Decisionmakers may want to perform this task, but this possibility seems unlikely. If the analyst is asked to do it, then there are several alternatives. In our case study, we swapped different "models" for C–R functions and valuation functions in and out of our integrated analysis and examined the effect this had on the distribution of net benefits. Alternatives include "offline" approaches, such as a meta-analysis that already integrates across alternative models and could incorporate statistical uncertainties as well.

Alternatively, the analyst may not want or feel qualified to make these judgment calls; in such a case, an appropriate response might be to mount an expert elicitation. In our judgment, too little guidance is available for choosing among these alternative options, let alone to decide how to perform an expert elicitation. Therefore, this area is ripe for research.

Further Work on Our Case Study

Our case study was fairly complex. Even so, the basic structure could be modified to learn more about the best methods to use. First, we could examine additional policies; we examined only two policies for NO_x reductions since the 2005 Clean Air Interstate Rule (CAIR; discussed extensively in Chapter 3). The advantage is that we could investigate whether the initial counterintuitive result (i.e., with comparable gas prices, costs of the policy relative to the appropriate baseline fall with both higher and lower natural gas prices) is a local or a global result.

Another potentially interesting task would be to investigate the effect on net benefits of more Benefits Mapping and Analysis Program (BenMAP) functions. So far, our investigations have been only illustrative, rather than comprehensive. A third option is to bring uncertainties from the Regional Modeling System for Aerosols and Deposition (REMSAD) model, which drives BenMAP, into our Monte Carlo integrated assessment model and, in particular, to explore how to make this model stochastic and compare its results to those of the other air quality models. A fourth option is to explore ways to simultaneously capture both statistical and model uncertainty graphically.

Model Development Incorporating Uncertainty in Engineering–Economic Models

Another option for future research arises from the complexity and size of the Haiku model, which precluded its running in Monte Carlo simulation mode, even if we were to somehow obtain distributions of our uncertain input variables. We are developing an algorithm that augments the current convergence method in Haiku with goal-seeking functions in an effort to make formal Monte Carlo analysis feasible, but this capability is not yet available.

A major technical advance would be to incorporate formal treatments of uncertainty within a highly parameterized model and to have the uncertain variables measured consistently on both the cost and benefit sides of the analysis.

Further Research on Uncertainty Analyses and Metrics

One issue that we have not examined is choosing the appropriate metrics to display the results of an analysis such as ours. This issue has several layers of complexity. The first is on the cost side. Our analysis focused on costs as an economist would measure them (i.e., as changes in consumer plus producer surplus changes). Alternatively, the results could have been presented in terms of compliance costs—certainly an easier concept to grasp, but possibly misleading. Indeed, perhaps decisionmakers prefer some more politically driven measure, such as whether electricity stays below 10 cents/kWh. Unfortunately, consideration of such measures was beyond the scope of our project but is essential for appropriately communicating the results of RIAs to decisionmakers.

A second issue is on the benefit side. Many effects of NO_x reductions are not listed, because they cannot be quantified or monetized or because the analyst judges them to be insignificant. How such nonquantifiable effects should be displayed is another key issue outside the scope of this project (alternatives including the Numeral Unit Spread Assessment Pedigree (NUSAP) system are discussed in Chapter 4).

A third issue is which aggregate measures are most appropriate for communication. Cost-effectiveness analysis would have the analyst calculate and communicate costs per key unit of benefit (e.g., mortality avoided, emissions, ambient concentrations, or physical health effects) without monetization of benefits. Or, following the recent report from the Institute of Medicine, the measure could aggregate over multiple physical (health) endpoints, like quality adjusted life years.

A fourth issue is the treatment of time. Our case study focused entirely on the end year for the analysis (2025) rather than the time path of net benefits for each scenario. This simplification may not be necessary for uncertainty communication.

Research on Legal and Other Implications of Using Uncertainty Analysis

EPA should examine the legal implications of conducting formal uncertainty analyses. For example, are such assessments likely to affect the agency's legal defense of its regulations?

A corollary issue concerns the possible reaction of stakeholders or the press to the uncertainty analyses. Would the existence of such analyses support or undermine agency decisions? Some form of survey research might yield valuable insights on this latter issue.

Value of Information Analyses

When used as a tool in an interview format, the uncertainty typology drawn from the literature (Chapter 2) was useful in eliciting information and identifying uncertainties. A screening tool developed along these lines may help EPA to identify sources of uncertainty in a systematic fashion.

However, a comprehensive accounting of sources of uncertainty does not imply that each source deserves equal attention. Given that resources are limited, EPA would be well served to identify priorities for both analysis of uncertainty in specific applications and research initiatives that could make the treatment of uncertainty more rigorous.

We suggest that EPA undertake a new research initiative to develop a value of information framework to assess the potential contribution of uncertainty in each stage of a model against the uncertainty that may characterize the final outcome.

Glossary

ΔLOG	difference in log-odds ratios
ACI	activated carbon injection
ACS	American Cancer Society
ADE	atmospheric diffusion equation
AEO	U.S. Energy Information Administration's *Annual Energy Outlook*
ANOVA	analysis of variance
ASTRAP model	Advanced Source Trajectory Regional Air Pollution model
BBNs	Bayesian belief networks
BEA	break-even analysis
BenMAP	Benefits Mapping and Analysis Program
CAA	Clean Air Act
CAIR	Clean Air Interstate Rule
CAMR	Clean Air Mercury Rule
CART	Classification and Regression Tree
CASAC	Clean Air Science Advisory Committee
CC	correlation coefficient
CDF	cumulative distribution function
C–R	concentration–response
CREM	Council for Regulatory Environmental Models
CSA	conditional sensitivity analysis
CV	contingent valuation
DDM-3D	Direct Decoupled Method in Three Dimensions
DSA	differential sensitivity analysis
EIA	U.S. Energy Information Administration
EPA	U.S. Environmental Protection Agency
FAST	Fourier Amplitude Sensitivity Test
FORM	first-order reliability methods
GPS	Global Positioning System
HTBR	hierarchical tree-based regression
IEEE	Institute of Electrical and Electronics Engineers

IPCC	Intergovernmental Panel on Climate Change
IPM	ICF Consulting's Integrated Planning Model
LHS	Latin Hypercube Sampling
MAPP	Mid-Continent Area Power Pool
MC	Monte Carlo
MIT	Massachusetts Institute of Technology
NAPAP	National Acid Precipitation Assessment Program
NAS	National Academies of Science
NERC	North American Electric Reliability Council
NOAA	National Oceanic and Atmospheric Administration
NO_x	nitrogen oxides
NRC	National Research Council
NRSA	nominal range sensitivity analysis
NUSAP system	Numeral Unit Spread Assessment Pedigree system
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PCC	partial correlation coefficient
PDF	probability density function
$PM_{2.5}$	particulate matter less than 2.5 microns in diameter
PRCC	partial rank correlation coefficient
PSA	probabilistic sensitivity analysis
QMC	quasi-Monte Carlo
R^2	coefficient of determination
RAMS	Regional Atmospheric Modeling System
RC	regression coefficient
RCC	rank correlation coefficient
REMSAD	Regional Modeling System for Aerosols and Deposition
RFF	Resources for the Future
RIAs	regulatory impact analyses
RRC	rank regression coefficient
SAB	Science Advisory Board
SCR	selective catalytic reduction
SIP	State Implementation Plan
SNCR	selective noncatalytic reduction
SO_2	sulfur dioxide

SORM	second-order reliability methods
S–R	source–receptor
SRC	standardized regression coefficient
SRG model	Source–Receptor Generator model
SRRC	standardized rank regression coefficient
TAF model	Tracking Analysis Framework model
URM-1ATM	Urban-to-Regional Multiscale One Atmosphere Model
VCE	variance of the conditional expectation of prediction
VOCs	volatile organic compounds
VOI	value of information
VSL	value of a statistical life

INDEPENDENT. BALANCED. OBJECTIVE.

RESOURCES

FOR THE FUTURE

1616 P STREET, NW • WASHINGTON, DC 20036-1400
TELEPHONE: (202) 328-5000 • FAX: (202) 939-3460
WWW.RFF.ORG