IEc

Expanded Expert Judgment Assessment of the Concentration-Response Relationship Between $PM_{2.5}$ Exposure and Mortality

Final Report | September 21, 2006

prepared for:

Office of Air Quality Planning and Standards

U.S. Environmental Protection Agency

Research Triangle Park, NC 27711

prepared by:

Industrial Economics, Incorporated

2067 Massachusetts Avenue

Cambridge, MA 02140

INDUSTRIAL ECONOMICS, INCORPORATED

EXECUTIVE SUMMARY

In 2002, the National Research Council (NRC) report, *Estimating the Public Health Benefits of Proposed Air Pollution Regulations*, presented the results of an NRC review of U.S. Environmental Protection Agency's (EPA's) benefits assessment methods for evaluating proposed regulations of air pollutants. The NRC committee approved of EPA's benefits analysis framework but provided a number of recommendations for improving EPA's characterization of uncertainty in benefit analysis. The NRC report recommended that probability distributions for key sources of uncertainty be developed using available empirical data or through formal elicitation of expert judgments in situations where scientific data are limited or conflicting.

The effect of changes in ambient fine particulate matter ($PM_{2.5}$) levels on mortality constitutes a key component of the EPA's approach for assessing potential health benefits associated with air quality regulations targeting emissions of $PM_{2.5}$ and its precursors. Avoided premature deaths constitute, on a dollars basis, between 85 and 95 percent of the monetized benefits reported in EPA's retrospective and prospective Section 812A benefit-cost analyses of the Clean Air Act (U.S. EPA, 1997 and 1999) and in Regulatory Impact Analyses (RIAs) for rules such as the Heavy Duty Diesel Engine/Fuel Rule (U.S. EPA, 2000) and the Non-road Diesel Engine Rule (EPA, 2004). Because it is such a large component of benefits, obtaining a good characterization of uncertainties regarding the mortality effects of changes in $PM_{2.5}$ exposure could capture the largest portion of uncertainty characterization of an entire benefit analysis (aside from unquantified or unmeasurable benefit endpoints).

In response to the 2002 NRC report, EPA has taken steps to incorporate the formal elicitation of expert judgments into uncertainty analyses for the benefits of air pollution rules affecting $PM_{2.5}$. Since 2003, Industrial Economics (IEc) has worked with a team of EPA and OMB analysts (collectively, the Project Team)[1] to conduct two expert judgment studies of the mortality impacts of $PM_{2.5}$ for EPA's Office of Air and Radiation (OAR). The first was a pilot study of five experts conducted in 2003 and 2004 aimed at exploring and refining the application of expert elicitation methods in the context of air pollution policy, and the second, which began in late 2004, was a full-scale study of 12 experts that built on the experience gained from the pilot and incorporated numerous refinements. This report documents the full-scale study of expert judgments concerning the impact of a

[1] The term "Project Team" refers to IEc staff and subcontractors, including Dr. Katy Walker, a specialist in expert judgment elicitations, and Patrick Kinney, an expert in the health effects of $PM_{2.5}$, plus a group of analysts from several EPA offices (Office of Air and Radiation, Office of Research and Development, and Office of Policy, Economics, and Innovation) and from the Office of Management and Budget.

one μg/m^3 change in ambient, annual average $PM_{2.5}$ on annual, adult, all-cause mortality in the U.S.

Expert elicitation uses carefully structured interviews to elicit from each expert his best estimate of the true value for an outcome or variable of interest as well as his uncertainty about the true value. This uncertainty, expressed as a subjective probabilistic distribution of values, reflects each expert's interpretation of theory and empirical evidence from relevant disciplines and ultimately his beliefs about what is known and not known about the subject of the study.

Exhibit ES-1 provides an overview of the steps of the expert elicitation process followed for this study, which includes the following elements: development of an elicitation protocol, selection of experts, development of a briefing book, conducting elicitation interviews, the use of expert workshops prior to and following individual elicitation of judgments, as well as the expert judgments themselves.

The expanded $PM_{2.5}$-mortality elicitation involved personal interviews with 12 health experts who have conducted research on the relationship between $PM_{2.5}$ exposures and mortality. The elicitation interview consisted of a protocol of carefully structured questions about the nature and magnitude of the relationship between changes in annual average $PM_{2.5}$ and annual, adult, all-cause mortality in the U.S. The Project Team developed the interview protocol between October 2004 and January 2006. Development of the protocol was informed by an April 2005 symposium held by the project team where numerous health scientists and analysts provided feedback, by detailed pre-testing with independent EPA scientists in November 2005, and by discussion with the participating experts at a Pre-elicitation Workshop in January 2006.

The 12 experts participating in the study were selected through a two-part peer-nomination process and included eight experts in epidemiology, three in toxicology, and one in medicine. The peer nomination process was designed to obtain a balanced set of views and serves to minimize the influence of the analysts and sponsors on expert selection. The experts selected for participation are shown in Exhibit ES-2.

EXHIBIT ES-1: OVERVIEW OF THE EXPERT JUDGMENT PROCESS

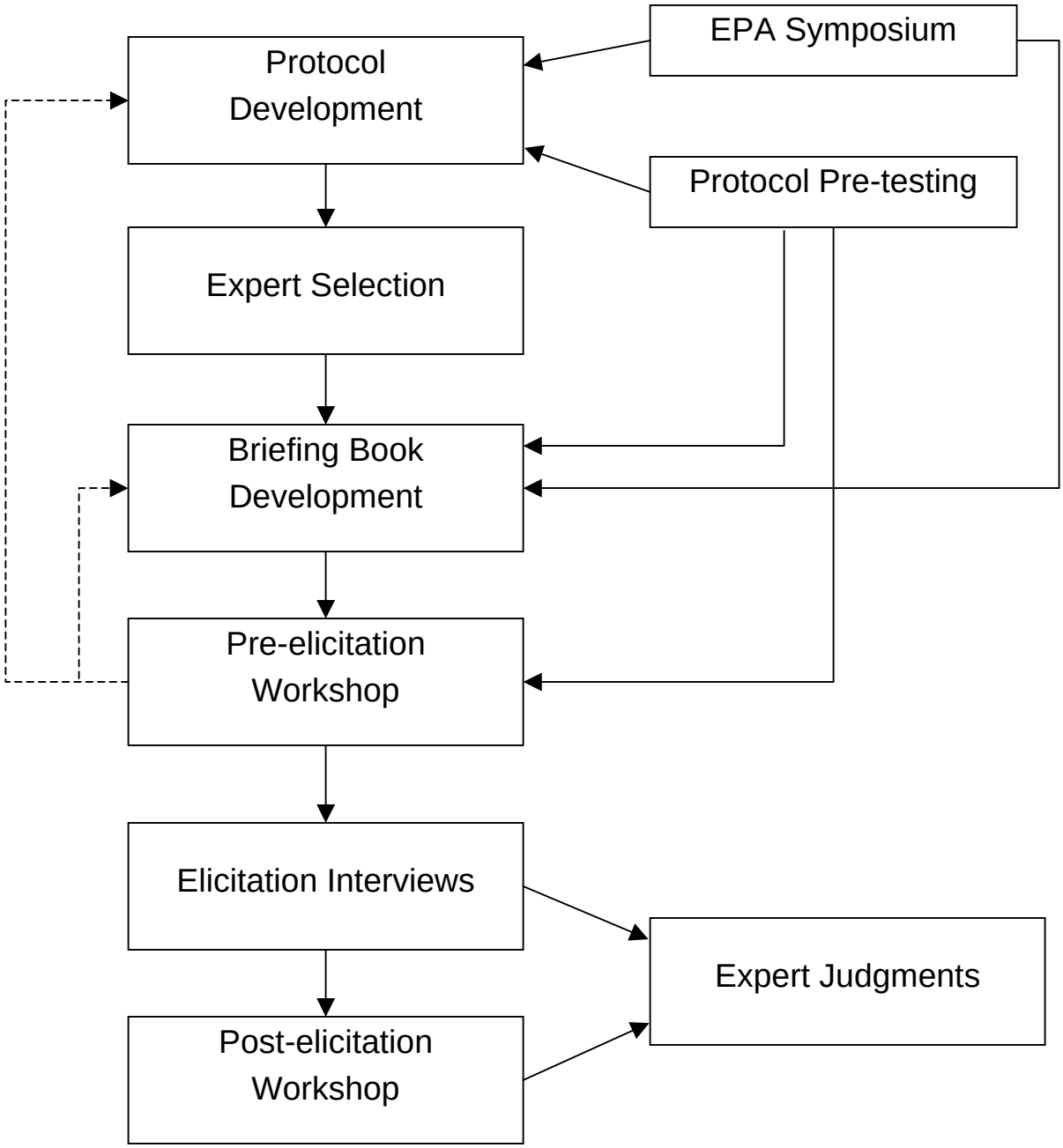


EXHIBIT ES-2: FINAL EXPERT LIST

NAME	AFFILIATION
Dockery, Doug W.	Harvard School of Public Health
Ito, Kazuhiko	New York University School of Medicine
Krewski, Daniel	University of Ottawa
Künzli, Nino	University of Southern California Keck School of Medicine (currently at Institut Municipal d'Investigació Mèdica (IMIM) - Center for Research in Environmental Epidemiology, Barcelona, SPAIN)
Lippmann, Morton	New York University School of Medicine
Mauderly, Joe	Lovelace Respiratory Research Institute
Ostro, Bart D.	California Office of Environmental Health Hazard Assessment
Pope, C. Arden III	Brigham Young University
Schlesinger, Richard	Pace University
Schwartz, Joel	Harvard School of Public Health
Thurston, George D.	New York University School of Medicine
Utell, Mark	University of Rochester School of Medicine and Dentistry

In January 2006, the Project Team held a Pre-elicitation Workshop with the selected experts. The workshop educated the experts as to the objectives of the study and the expert judgment process. It also fostered discussion among the participants about the key evidence available to answer questions in the protocol and provided an opportunity for experts to provide feedback on the protocol prior to beginning the interviews.

The elicitation interviews were conducted between January and April 2006. Each expert was provided a briefing book of reference materials and a copy of the elicitation protocol prior to the interviews. Each interview lasted approximately 8 hours and covered both qualitative and quantitative questions. The qualitative questions probed experts' beliefs concerning key evidence and critical sources of uncertainty and were intended to make the conceptual basis for their quantitative judgments explicit. These questions covered topics such as potential biological mechanisms linking $PM_{2.5}$ exposures with mortality; key scientific evidence on the magnitude of the PM-mortality relationship; sources of potential error or bias in epidemiological results; the likelihood of a causal relationship between $PM_{2.5}$ and mortality, and the shape of the concentration-response (C-R) function. The main quantitative question asked each expert to provide a probabilistic distribution for the average expected decrease in U.S. annual, adult, all-cause mortality associated with a 1 $\mu g/m^3$ decrease in annual average $PM_{2.5}$ levels. In addressing this question, the experts first specified a functional form for the $PM_{2.5}$ mortality C-R function and then developed an uncertainty distribution for the slope of that function (the mortality impact per unit change in annual average $PM_{2.5}$), taking into account the evidence and judgments discussed during the qualitative questions.

When answering the main quantitative question, each expert was instructed to consider that the total mortality effect of a 1 $\mu g/m^3$ decrease in ambient annual average $PM_{2.5}$ may reflect reductions in both short-term peak and long-term average exposures to $PM_{2.5}$. Each expert was asked to aggregate the effects of both types of changes in his answers.

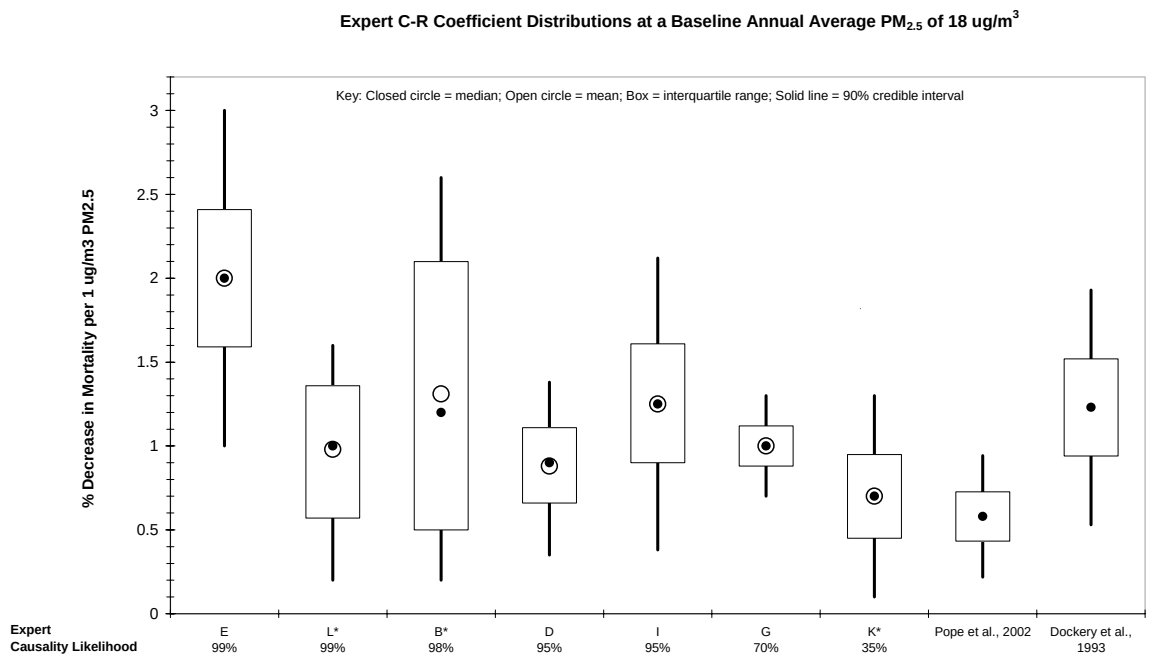
Each expert was given the option to integrate their judgments about the likelihood of a causal relationship and/or threshold in the C-R function into his distribution or to provide a distribution "conditional on" one or both of these factors. The interviewers asked each expert to characterize his distribution by assigning values to fixed percentiles (5^{th}, 25^{th}, 50^{th}, 75^{th}, 95^{th}). To assist experts in the elicitation process, the interviewers provided real-time feedback during the interviews in the form of graphs and example calculations, using spreadsheet tools and Internet teleconferencing. During the interviews, experts were able to view their responses plotted onto a distribution using a software interface. They then adjusted their estimates until the distribution represented the views they expressed during the day-long interview.

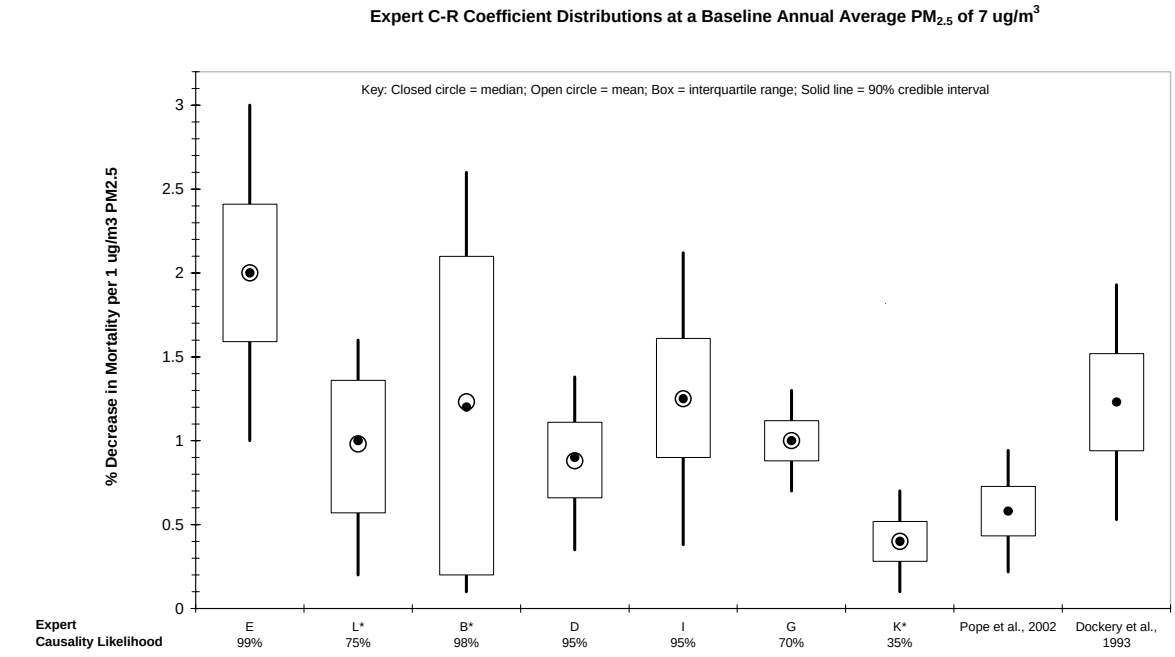
Following the interviews, the Project Team reconvened the experts for a Post-elicitation Workshop in June 2006. At this workshop, the Project Team anonymously shared the results of all experts with the group, and gave experts the opportunity to raise issues that may have emerged during the interviews for further discussion with the group. The workshop was not intended to promote consensus, and the Project Team made no effort to encourage experts to change their responses. Experts were, however, allowed to modify their responses privately if they chose to do so based on insights gained during workshop discussions, provided they included the detailed rationale for their changes. Four experts made adjustments to their judgments following the workshop; one of them made changes to the percentiles of his distribution. (A sensitivity analysis conducted on the results showed minimal impact of these changes when compared to the results prior to this workshop.)

In Exhibits ES-3 and ES-4 we display the responses of the experts to the main quantitative elicitation question. The distributions provided by each expert, identified by the letters A through L, are depicted as box and whisker plots with the solid circle symbol showing the median (50th percentile); the open circle showing the mean;[2] the box defining the interquartile range (bounded by the 25th and 75th percentiles); and the ends of the "whiskers" defining each expert's 5th and 95th percentiles. Each expert's stated best estimate of the likelihood of a causal relationship between $PM_{2.5}$ and mortality is shown on the x-axis and the experts are arrayed in order of decreasing certainty of causality. Exhibit ES-3 displays the distributions for the experts who chose to provide a distribution conditional on the existence of a causal relationship between $PM_{2.5}$ and mortality; Exhibit ES-4 shows the distributions for the group who chose to integrate their judgments about the likelihood of causality directly into their distribution. Each figure displays the expert distributions for two different $PM_{2.5}$ levels, 18 $\mu g/m^3$and 7 $\mu g/m^3$, to observe the implications of four experts' (B, F, K, and L) assumptions about non-linearities in the C-R function and about differing degrees of uncertainty in the slope of the function across specific ranges of PM. Also, as a point of reference for the results, we include box plots of two epidemiology studies used in EPA benefit analyses – Pope et al., 2002 and Dockery et al., 1993.

[2] The mean of the expert's distribution was not directly elicited. IEc generated this value using Crystal Ball™ statistical modeling software.

EXHIBIT ES-3: GROUP 1: UNCERTAINTY DISTRIBUTIONS FOR $PM_{2.5}$- MORTALITY C-R COEFFICIENTS; CONDITIONAL ON THE EXISTENCE OF A CAUSAL RELATIONSHIP





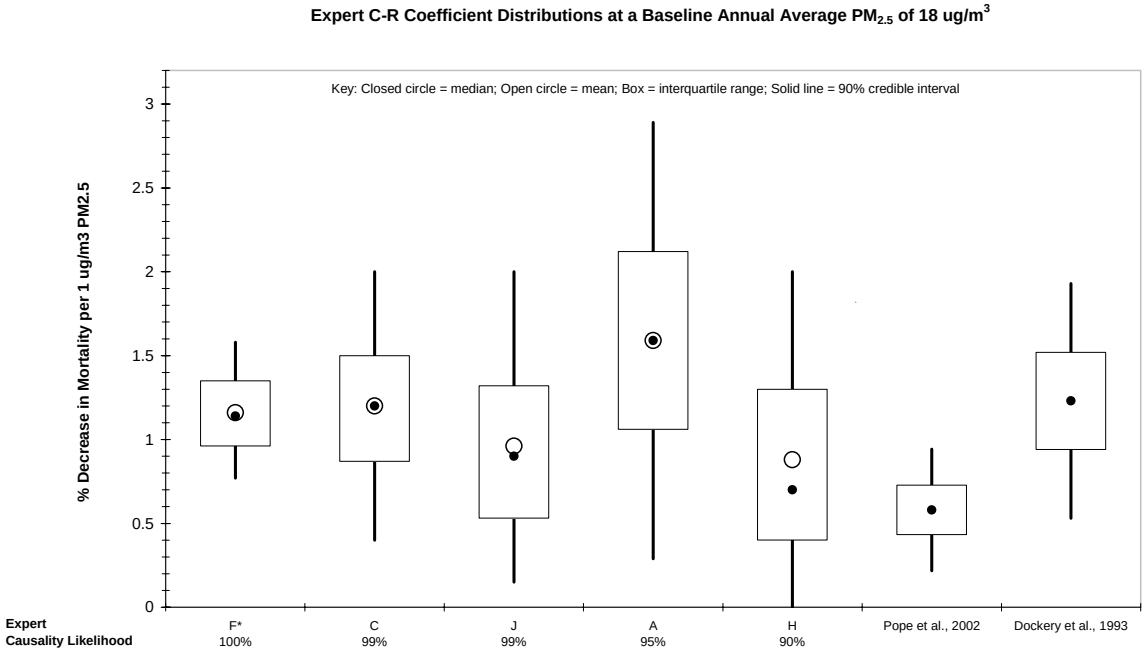
Note: Box plots represent distributions as provided by the experts to the elicitation team. When asked, experts in this group preferred to give conditional distributions and keep their probabilistic judgment about the likelihood of a causal or non-causal relationship separate.

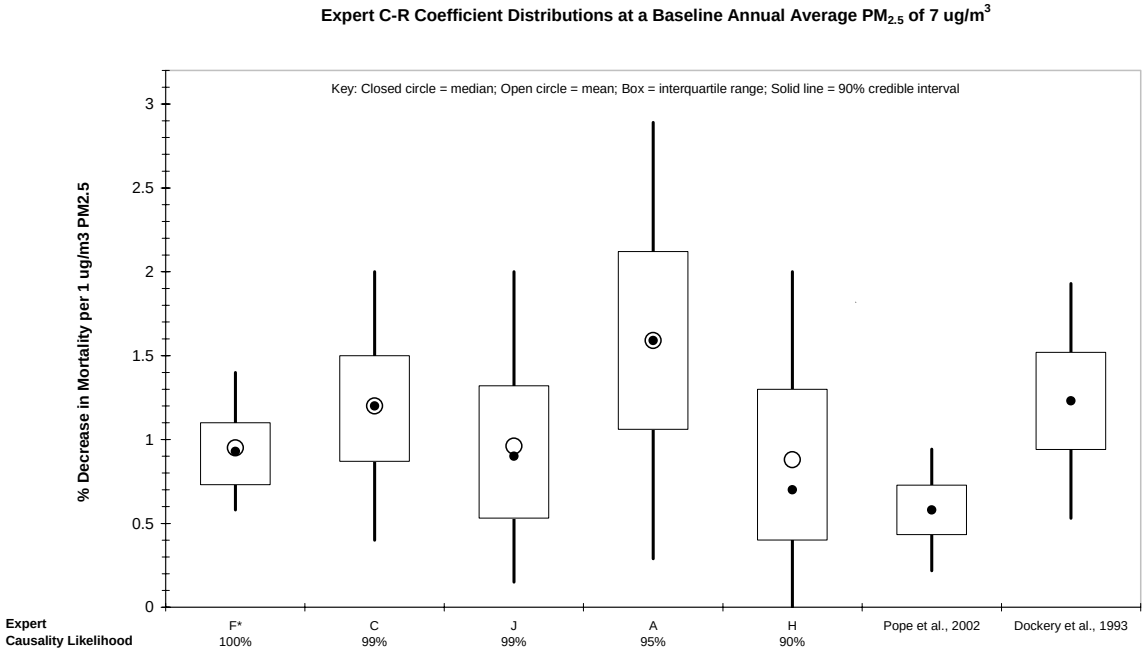
*Experts' C-R coefficient distribution changes between 7 and 18 $\mu g/m^3$.

Expert K specified a threshold (not shown).

Expert L provided two different likelihoods of causality for his C-R coefficient distributions at 7 and 18 $\mu g/m^3$, although his distribution appears in the same location in both graphs.

EXHIBIT ES-4: GROUP 2: EXPERT UNCERTAINTY DISTRIBUTIONS FOR THE $PM_{2.5}$-MORTALITY COEFFICIENT INCORPORATING THE EXPERT'S LIKELIHOOD OF A CAUSAL RELATIONSHIP





Note: Box plots represent distributions as provided by the experts to the elicitation team. When asked, experts in this group preferred to give distributions that incorporate their likelihood that the $PM_{2.5}$ mortality association may be non-causal

*Experts' C-R coefficient distribution changes between 7 and 18 μg/m^3.

Among the experts who provided distributions that were conditional on the existence of a causal relationship (Exhibit ES-3), median estimates ranged from a 0.7 to 1.6 percent decrease in annual, adult, all-cause mortality per 1 $\mu g/m^3$ decrease in annual average $PM_{2.5}$. For five of the seven experts in this group the effect of ultimately integrating their judgments about causality in a benefits assessment will be small, given the high likelihoods of causality they expressed. Median benefits estimated using the distributions of the two experts in this group who expressed greater doubt about a causal relationship will show more significant declines. Among the experts who directly incorporated their views on the likelihood of a causal relationship into their distributions (Exhibit ES-4), the median estimates ranged from a 0.7 to 1.6 percent decrease in annual, adult, all-cause mortality per 1 $\mu g/m^3$ decrease in annual average $PM_{2.5}$.

Because of the different approaches taken by the experts to characterizing their views about uncertainty in the $PM_{2.5}$ mortality relationship, it is difficult to compare all the distributions. Direct comparison can only be done when the distributions are applied to the same scenario in a benefits analysis. However, certain observations and conclusions can be drawn from these plots and from the experts' responses to the qualitative questions:

- Experts in this study tended to be confident that $PM_{2.5}$ exposure can cause premature death. Ten of twelve experts believed that the likelihood of a causal relationship was 90 percent or higher. The remaining two experts gave causal probabilities of 35 and 70 percent. Recent research in both epidemiology (e.g., Jerrett et al., 2005, Laden et al., 2006) and toxicology (e.g., Sun et al., 2005) significantly contributed to experts' confidence.
- Only one of 12 experts explicitly incorporated a threshold into his C-R function.[3] The rest believed there was a lack of empirical and/or theoretical support for a population threshold. However, three other experts gave differing effect estimate distributions above and below some cut-off concentration. The adjustments these experts made to median estimates and/or uncertainty at lower $PM_{2.5}$ concentrations were modest.
- Experts relied upon a core set of cohort epidemiology studies to derive their quantitative estimates, mainly those associated with the ACS and Six Cities cohorts. The Six Cities results tended to be weighted more highly by experts in this study than in the pilot study. The greater emphasis on Six Cities appeared to result from corroborating evidence in the recent Six Cities follow-up (Laden et al., 2006) and from concerns about potential exposure misclassification issues and/or effect modification in the ACS cohort (see below).

[3] Expert K indicated that he was 50 percent sure that a threshold existed. If there were a threshold, he thought that there was an 80 percent chance that it would be less than or equal to 5 $\mu g/m^3$, and a 20 percent chance that it would fall between 5 and 10 $\mu g/m^3$.

- Although the quantitative question asked experts to consider mortality changes due both to short-term and long-term $PM_{2.5}$ exposures, all experts based their median effect estimates on effects due to long-term exposures. Short-term exposure effects were sometimes used to derive lower-bound effect estimates.
- Confounding of epidemiological results tended to be a minor concern for most experts. Only one of 12 experts expressed substantial concern about confounding as a source of error in the key literature on $PM_{2.5}$ and premature mortality.
- Experts' concerns regarding potential negative bias in the ACS main study results due to effect modification (see Pope and Dockery, 2006) and/or exposure misclassification (Jerrett et al., 2005; Willis et al., 2003; and Mallick et al., 2002) led many experts to adjust the published results upwards when considering the percentiles of their distribution.
- A sensitivity analysis conducted using a simplified benefits analysis demonstrated that no individual expert's distribution of effect estimates had more than a plus or minus 8 percent impact on an overall, pooled distribution of effects. The influence of individual experts appeared symmetrically distributed.

Compared to the pilot study, experts in this study were in general more confident in a causal relationship, less likely to incorporate thresholds, and reported higher mortality effect estimates. The differences in results compared with the pilot appear to reflect the influence of new research on the interpretation of the key epidemiological studies that were the focus of both elicitation studies, more than the influence of changes to the structure of the protocol. They may also reflect differences in the composition of the expert panel. The experts' distributions in the current study display a similar diversity in the degree of uncertainty expressed as in the pilot study. The variation in the experts' responses reflects a number of factors, including differences in their views about the degree of uncertainty inherent in key epidemiological results from long-term cohort studies, the evidential support for a causal relationship, and, for one expert, the shape of the C-R function. In almost all cases, however, the spread of the uncertainty distributions elicited from the experts exceeded the statistical uncertainty bounds reported by the most influential epidemiologic studies, suggesting that the expert elicitation process was successful in developing more comprehensive estimates of uncertainty for the $PM_{2.5}$ mortality relationship.

TABLE OF CONTENTS

APPENDICES

EXHIBITS

EXHIBITS (CONTINUED)

CHAPTER 1 | INTRODUCTION

The effect of changes in ambient fine particulate matter ($PM_{2.5}$) levels on mortality constitutes a key component of the U.S. Environmental Protection Agency's (EPA's) approach for assessing potential health benefits associated with air quality regulations targeting emissions of $PM_{2.5}$ and its precursors. Avoided premature deaths constitute, on a dollars basis, between 85 and 95 percent of the benefits reported in EPA's retrospective and prospective Section 812A benefit-cost analyses of the Clean Air Act (U.S. EPA, 1997 and 1999) and in Regulatory Impact Analyses (RIAs) for rules such as the Heavy Duty Diesel Engine/Fuel Rule (U.S. EPA, 2000) and the Non-road Diesel Engine Rule (EPA, 2004). Because it is such a large component of benefits, uncertainties regarding the mortality effects of changes in $PM_{2.5}$ exposure could have a significant impact on the range of plausible benefit values reported for air pollution rules and on the interpretation of the results of benefit analyses.

In 2002, the National Research Council (NRC) report, *Estimating the Public Health Benefits of Proposed Air Pollution Regulations*, presented the results of an NRC review of EPA's benefits assessment methods for evaluating proposed regulations of air pollutants. The NRC committee approved of EPA's benefits analysis framework but provided a number of recommendations for improving the implementation of that framework, including recommendations for improving uncertainty analysis. The committee recommended that EPA take steps to improve its characterization of uncertainties associated with key components of its health benefits analyses, such as PM-related mortality effects. The NRC report recommended that probability distributions for key sources of uncertainty be developed using available data or, where scientific data are limited or conflicting through formal elicitation of expert judgments.

In response to the 2002 NRC report, EPA has taken steps to incorporate the formal elicitation of expert judgments into uncertainty analyses for the benefits of air pollution rules affecting $PM_{2.5}$. Since 2003, Industrial Economics (IEc) has conducted for EPA's Office of Air and Radiation (OAR) two expert judgment studies of the mortality impacts of $PM_{2.5}$; the first was a pilot study of five experts conducted in 2003 and 2004 aimed at exploring and refining the application of expert elicitation methods in the context of air pollution policy, and the second, which began in late 2004, was a full-scale study of 12 experts that built on the experience gained from the pilot and incorporated numerous refinements. This report documents the full-scale study of the formal elicitation of expert judgments concerning the mortality impacts of a 1 $\mu g/m^3$ change in ambient, annual average $PM_{2.5}$.

In the remainder of this introductory chapter, we discuss the purpose and scope of the full-scale study, discuss the relationship of this study to the pilot, and provide a road map to the rest of the report.

1.1 PURPOSE AND SCOPE

The purpose of this project is to provide a more complete characterization, both qualitative and quantitative, of the uncertainties associated with the relationship between reductions in ambient $PM_{2.5}$ (measured as total gravimetric mass) and mortality. The study was designed to produce results that EPA can apply in its benefit models when preparing future regulatory analyses.

The full-scale study involved personal interviews with 12 health experts who have conducted research on the relationship between $PM_{2.5}$ exposures and mortality. These experts were selected through a peer-nomination process and included experts in epidemiology, toxicology, and medicine. The elicitation interview consisted of a protocol of carefully structured questions about the nature and magnitude of the $PM_{2.5}$-mortality relationship, developed by IEc in consultation with a Project Team of EPA and Office of Management and Budget (OMB) staff scientists (hereafter "Project Team"). The protocol included both qualitative and quantitative questions. The questions requiring qualitative responses probed experts' beliefs concerning key evidence and critical sources of uncertainty and were intended to make the conceptual basis for their quantitative judgments explicit.[4] The main quantitative question sought experts' probabilistic judgments about the average expected decrease in U.S. adult, annual all-cause mortality associated with a 1 $\mu g/m^3$ decrease in annual average $PM_{2.5}$ levels. Experts were instructed to consider that the total mortality effect of a 1 $\mu g/m^3$ decrease in ambient annual average $PM_{2.5}$ may reflect reductions in both short-term peak and long-term average exposures to $PM_{2.5}$. Experts were asked to characterize the distribution of possible values for the mortality effect (hereafter, "uncertainty distribution") by assigning values to fixed percentiles (5^{th}, 25^{th}, 50^{th}, 75^{th}, 95^{th}). They were also asked for the minimum and maximum values of the distribution.

The elicitation focused on the concentration-response (C-R) function relating $PM_{2.5}$ mass changes with mortality, and was not intended to characterize the uncertainty surrounding the role of specific PM components (e.g., diesel particulates) or sources (e.g., power plants) in the $PM_{2.5}$-mortality relationship. Also this study did not ask experts to characterize the time sequence of any mortality reductions following a 1 $\mu g/m^3$ decrease in ambient annual average $PM_{2.5}$.

[4] These questions covered topics such as potential biological mechanisms linking $PM_{2.5}$ exposures with mortality; key scientific evidence on the magnitude of the PM/mortality relationship; sources of potential error or bias in epidemiological results; the likelihood of a causal relationship between $PM_{2.5}$ and mortality, and the shape of the C-R function.

1.2 RELATIONSHIP TO THE PILOT STUDY

The current study is not an extension of the pilot. Rather, it is a separate study that features a number of refinements intended to improve the elicitation process. It also reflects more recent research on $PM_{2.5}$-related mortality that has become public since the pilot study. To provide some perspective on the current study, we provide below an overview of the pilot study and its results and then discuss the key refinements incorporated into the full-scale study.

1.2.1 OVERVIEW OF THE PILOT STUDY

In 2003 and 2004, the Project Team conducted a pilot-scale elicitation study with five experts to explore the effectiveness of expert judgment techniques for characterizing uncertainty and to explore the use of the expert judgment results in the context of economic benefits analysis (IEc, 2004) (hereafter, "pilot study"). In particular, the pilot study was designed to provide feedback on the efficacy of the protocol developed and the analytic challenges, as well as to provide insight regarding potential implications of the results on the degree of uncertainty surrounding the C-R function for $PM_{2.5}$ mortality.

Like the full-scale study, the pilot study consisted of individual interviews featuring a series of qualitative and quantitative questions about the nature of the $PM_{2.5}$-mortality relationship. One key difference was that the quantitative questions in the pilot protocol asked experts to provide separate uncertainty distributions for changes in mortality due to long-term exposure (a 1 $\mu g/m^3$ decrease in annual average $PM_{2.5}$ levels) and changes in mortality due to short-term exposure (a 10 $\mu g/m^3$ reduction in daily average $PM_{2.5}$ levels). A detailed report on the pilot study (IEc, 2004); the results of a peer review of the pilot conducted in the summer of 2004; and descriptions of EPA's use of the pilot results as part of its uncertainty analysis for assessing benefits of the Non-road Diesel Engine Rule (EPA, 2004), the Clean Air Interstate Rule (EPA, 2005a), and the Guidelines for Best Available Retrofit Technology (BART) (EPA, 2005b) can be found at the following website: http://www.epa.gov/ttn/ecas/benefits.html.

For changes in annual average $PM_{2.5}$ levels (which are the focus of the present study), experts in the pilot study estimated median values ranging from zero to a 0.7 percent reduction in mortality associated with a 1 $\mu g/m^3$ decrease in annual average $PM_{2.5}$ levels in the U.S. Four of the five experts provided median values of 0.5 percent or less, and the combined mean effect across all experts was 0.33 percent. Most experts were influenced most heavily by results of the American Cancer Society (ACS) cohort, including the HEI reanalysis (Krewski et al., 2000) and the Pope et al. follow-up (2002). Experts placed less weight on the results of the Six Cities cohort (Dockery et al., 1993), despite citing numerous strengths of that study during the interviews. Experts varied in their level of confidence that the relationship between mortality and long-term $PM_{2.5}$ changes was causal, with three experts providing probabilities of a causal relationship in the 40 to 50 percent range and two providing probabilities in the 80 to 90 percent range. Two of the five experts incorporated thresholds into their distributions (IEc, 2004).

The peer review of the pilot study was generally positive but raised three main issues. First, reviewers recommended that we provide opportunities for communication between

experts before and/or after the elicitation interviews, which was not possible in the pilot due to time constraints. Second, they encouraged us to take steps to minimize the potential that experts anchor too closely to individual studies (i.e., anchoring and adjustment bias), which can lead to overconfidence in judgments.[5] Third, several reviewers expressed reservations about the reporting of combined (or pooled) expert judgments in the pilot report.[6]

1.2.2 IMPROVEMENTS IN THE FULL-SCALE STUDY

The full-scale study features numerous improvements intended to address both comments from the peer review of the pilot and issues identified by the Project Team while conducting the pilot. Changes from the pilot study include the following:

- **Shorter protocol.** This study features a shorter protocol focused on total changes in adult, all-cause mortality associated with a 1 μg/m3 decrease in annual average $PM_{2.5}$ in the U.S. The protocol for the pilot study was too long and difficult to complete in an 8-hour day, because it required elicitation of two separate distributions, one for short-term and one for long-term impacts of $PM_{2.5}$ exposures. As a result, the interviews often felt rushed, which may have had an impact on the quality of the judgments we received. Based on our experience in the pilot study and advice received from health experts who reviewed early drafts of the protocol, the full-scale study takes a more integrated approach that covers issues related to short- and long-term exposure, but elicits a single C-R function that integrates experts' judgments about the impacts of changes in peak and long term exposure associated with a 1 μg/m3 decrease in annual average $PM_{2.5}$ in the U.S.
- **Additional expert review and feedback on protocol.** Protocol development benefited from additional review and feedback from the scientific community. The Project Team conducted an EPA Symposium in April 2005 with more than 30 internationally-recognized experts knowledgeable about PM health effects. The purpose of the EPA Symposium was to present EPA's plans for conducting the full-scale elicitation and to get feedback on the structure and content of the protocol.
- **Larger expert panel.** This study interviewed a larger panel of 12 experts selected from an unrestricted pool of potential experts via a peer nomination process. The panel includes expertise in epidemiology, toxicology, and medicine. Expert selection in the pilot used 5 experts selected from a restricted pool derived from the

[5] "Anchoring and adjustment" refers to when an expert begins his estimates with, or "anchors" on, a particular value then develops confidence intervals by adjusting that value to account for various factors that influence his judgment. Some expert judgment research has shown that use of this approach may lead to overconfidence, and thus poor accuracy, because respondents often fail to adjust confidence intervals adequately for what they do not know.

[6] One reviewer did not feel that it was appropriate to combine the individual distributions into a single estimate. Another preferred that expert results be applied in benefits analysis individually and then pooled, rather than being combined as part of the expert judgment study. Some reviewers commented that the combination of judgments using averaging across experts might generate results with which none of the experts would agree. While not endorsing combination of responses, some of the experts indicated that if results must be combined, they would be most comfortable with a process that uses equal weighting.

membership of two NAS committees. The small sample size of the pilot study was a concern noted by the EPA Science Advisory Board (SAB) Health Effects Subcommittee, which reviewed early plans for the pilot study (EPA SAB, 2004).

- **Improved briefing book materials.** The study featured improved briefing book materials for experts, including a CD containing over a hundred relevant studies, plus background information pages with data on air quality in the U.S., population demographics, health status, summaries of published effect estimates, and data on other factors (air conditioning use, housing stock, PM composition) that may affect experts' judgments. While the pilot study relied more heavily on compendia such as EPA's criteria document for PM for briefing book materials, we supplemented these materials in the full-scale study to provide clear and easy access to relevant studies and key data useful for developing quantitative estimates specific to the current U.S. population.
- **Pre-elicitation Workshop.** In response to peer-review comments, the full-scale study included a Pre-elicitation Workshop held in January 2006 to better prepare experts in advance of the interviews. The workshop provided training on the history of subjective judgment, and opportunity for review and comment on the elicitation protocol, and encouraged experts to share and critique data and analyses they believed were relevant to the questions in the protocol.
- **Better accounting of experts' views on relevant scientific issues.** The new protocol employed a more systematic approach to cataloguing and assessing expert views related to confounding, effect modification, exposure issues, and other potential sources of error or bias in published mortality effect estimates. During the pilot study, the Elicitation Team at times found it challenging to make sure that each expert systematically addressed in his quantitative estimates each of the major issues he may have raised in the conditioning phase of the protocol. Following a method used in a protocol designed for a study of climate change (Morgan et al., 2001), we modified the protocol for the full-scale study to give each expert a set of cards on which we asked him to write down key factors he wished to discuss. The cards were then used to organize and rank the factors, and to ensure that all factors are discussed and evaluated.
- **Greater flexibility.** The new protocol was designed to allow experts greater flexibility in specifying the shape of the C-R function and in eliciting the values for their uncertainty distributions. Our experience from the pilot elicitation indicated that the pilot protocol was not sufficiently explicit about the options available for characterizing the C-R function. We revised the protocol to facilitate the elicitation of judgments for experts who wished to specify a non-linear C-R function and/or incorporate uncertainty about thresholds. Also, at the EPA Symposium, several experts indicated that they would prefer to be able to explore specifying parametric distributions to assist them in estimating their intermediate percentiles, such as the 25^{th} and 75^{th}. The real-time feedback system, described below, provided alternative approaches to characterizing the uncertainty distributions.

- **More extensive real-time feedback.** The elicitation of judgments is a complex and iterative process. The pilot study lacked mechanisms to provide experts with real-time data and graphs that would allow them to better visualize, evaluate, and refine their initial judgments. The interviews for the expanded study included a new real-time feedback system that used spreadsheet models and Crystal Ball™ probabilistic modeling software to provide experts with graphs and data during the elicitation. The system, provided via Internet teleconferencing allowed for experts to visualize their distributions, to assess the effect of judgments about causality and threshold, to compare their results against published mortality effect estimates, and to estimate and compare the change in deaths associated with specific $PM_{2.5}$ reductions against U.S. mortality data for major causes of death.[7]
- **Post-elicitation Workshop.** The current study also included a workshop held following completion of the interviews (hereafter, "Post-elicitation Workshop") that allowed for a final discussion of themes that emerge in expert responses, differences in interpretation of key studies used to support responses, and any areas of confusion that arose during the interviews. The goal of this workshop was not to promote consensus, but again to ensure that all experts had access to the same information, could have outstanding questions addressed, and were able to provide feedback to EPA on the study. Experts had the opportunity to adjust their estimates following the workshop. As in the case of the Pre-elicitation Workshop, the Post-elicitation Workshop addressed a frequent concern expressed by peer reviewers of the pilot study that there should be more opportunity for communication between the experts.
- **Reporting of individual expert distributions only.** In response to concerns raised by peer reviewers related to the potential pitfalls of generating and presenting a single combined estimate as well as other feedback received during the planning for the pilot and full-scale studies, we have chosen to present only the individual expert distributions in this report, preserving the diversity of opinion across experts on this topic. However, we have conducted a sensitivity analysis using data pooled across experts from a simplified mortality benefits analysis (see Chapter 4 for further details).

1.3 ORGANIZATION OF THIS DOCUMENT

The remainder of this document is organized into four chapters. Chapter 2 describes IEc's analytical approach to conducting the expert judgment study, including the design of an elicitation protocol, testing of the protocol, selection of experts, expert workshops, and the interview process. Chapter 3 presents the results of the assessment, summarizing expert responses to both quantitative and qualitative questions. Chapter 4 provides discussion of both the quantitative findings of the study and the elicitation process itself. Chapter 5 includes the overall conclusions of the study.

[7] The elicitation team used either WebEx™ or Go To Meeting™ Internet conferencing software during the interviews.

CHAPTER 2 | ANALYTICAL APPROACH

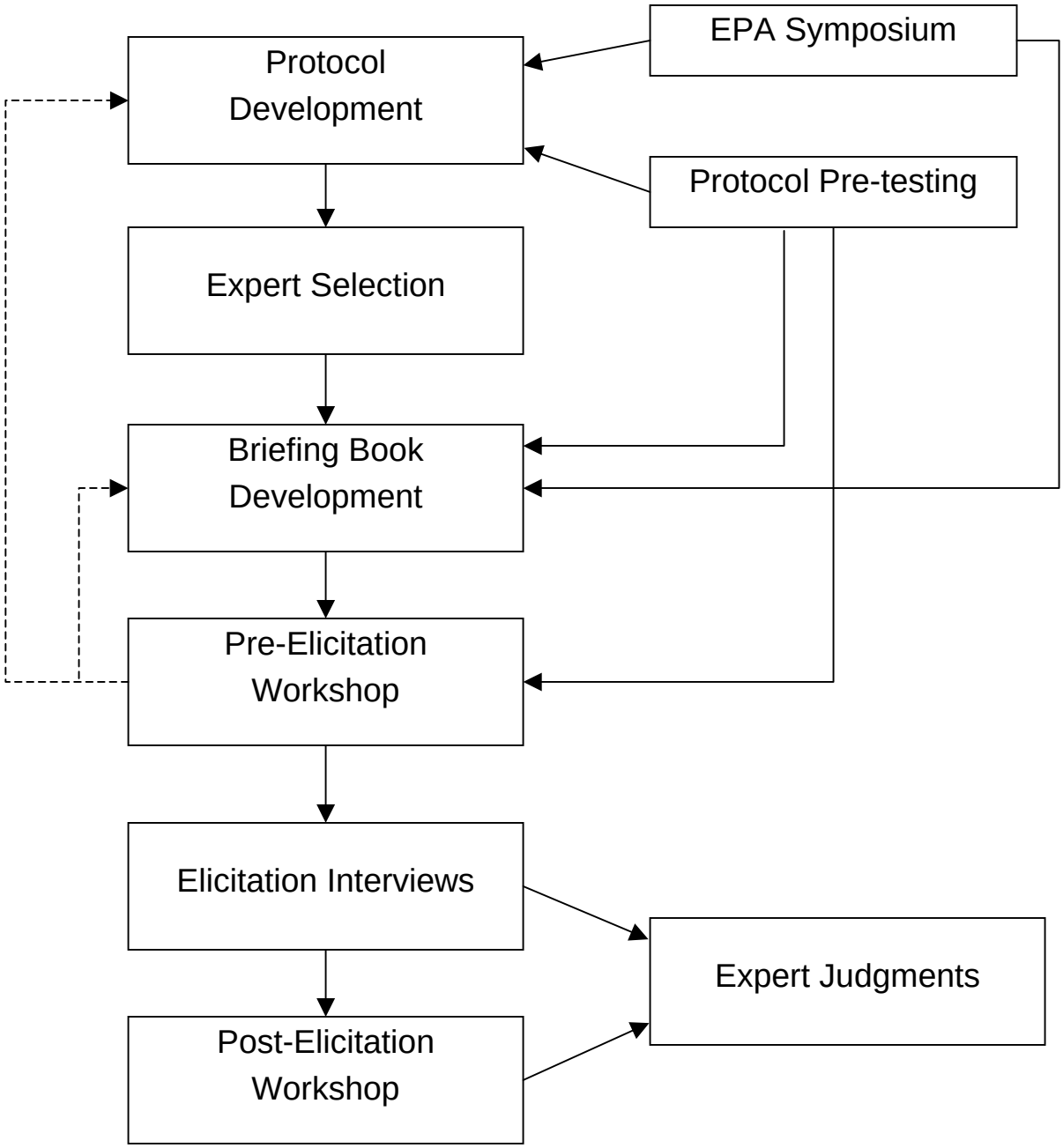
While researchers have been studying the process of eliciting and interpreting expert judgments for several decades, no single accepted, standardized method has emerged for this type of assessment. However, numerous studies conducted over the last two decades (e.g., Morgan et al., 1984 & 2001; Evans et al., 1994; Walker et al., 2001) have often shared common elements such as use of criteria for the selection of experts, the use of a detailed written protocol, the preparation of briefing materials, the elicitation of judgments in individual interviews, and where feasible, the convening of experts prior to and following the elicitation.

The motivation behind the development of the elements of formal elicitation studies is ultimately to help experts provide judgments that are informative, unbiased and well-calibrated. Specifically, they are intended to help avoid some of the well-documented heuristics and biases (Kahneman et al., 1982) that lead to poorly calibrated judgments. These heuristics include the tendency to rely on data or studies most cognitively available or recent (the availability heuristic), and the tendency to adhere too closely to published or initial estimates and failing to make adequate adjustments for key uncertainties (anchoring and adjustment heuristic). These and other heuristics can contribute to expert over-confidence (i.e., the tendency to express more certainty about a quantitative estimate than is warranted).

IEc has relied on the experience of earlier studies to inform the design of the overall process for this study. Exhibit 2-1 outlines the expert judgment process of this study. The main elements of the process included development of an elicitation protocol, selection of experts, development of a briefing book, the elicitation interviews, and expert workshops prior to and following individual elicitation of judgments. The diagram also indicates the influences of certain elements in the process on the development of others. For example, developing the protocol was an iterative process that was influenced not only by Project Team discussion, but also by the EPA Symposium and pre-testing of the protocol with two independent EPA scientists with expertise in the study of $PM_{2.5}$-mortality. In addition, discussions during the Pre-elicitation Workshop influenced elements of the protocol as well as the briefing book. Finally, the expert judgments were elicited during the personal interviews with members of the project team, but experts were allowed to adjust their responses following the Post-elicitation Workshop.

The following sections discuss each of the main elements of the expert judgment process described above. We also provide discussion of the decisions reached by the project team regarding the evaluation and presentation of experts' judgments.

EXHIBIT 2-1: OVERVIEW OF THE EXPERT JUDGMENT PROCESS



2.1 PROTOCOL DEVELOPMENT

In a formal elicitation process, the interview protocol serves several purposes. It ensures that experts are answering the same, clearly specified question;[8] makes explicit the critical assumptions and rationale underlying the expert's judgments; encourages experts to think critically about potential uncertainties; and helps the expert avoid some of the common pitfalls and biases when providing subjective judgments about uncertainty (Otway and von Winterfeldt, 1992). The Project Team developed a protocol over a period of 16 months that included a clear statement of the questions to be answered, including the critical assumptions underlying the main quantitative question, and provided a process for identifying and raising key issues, evidence, and interpretations upon which the expert could base his quantitative judgments.

The protocol evolved through a series of steps beginning with collaborative discussions with the Project Team aimed at scoping and structuring the problem that would be posed to the experts. A draft protocol was next reviewed and discussed by health experts at an expert elicitation symposium sponsored by EPA.[9] Following revisions to the protocol based on recommendations from the EPA symposium, the protocol underwent pre-testing with two independent EPA experts on $PM_{2.5}$ mortality. The Project Team finalized the protocol in January 2006, following the Pre-elicitation Workshop with the selected expert panel (discussed below).

The following sections discuss the development of the protocol, in particular the process of structuring the problem, the influence of the EPA Symposium, and the protocol pre-testing. It also describes the final structure of the protocol. A copy of the protocol used in the interviews is included as Appendix A.

2.1.1 STRUCTURING THE PROBLEM

The expert judgment literature discusses two broad approaches to elicitation of judgments; an aggregated and a disaggregated approach. As the term implies, an aggregated approach asks the expert to estimate the quantity of interest directly; for example, the numbers of newspapers sold in the U.S. in a particular year. In a disaggregated approach, the expert (or group of experts) would be asked to construct a model for estimating the quantity of interest and would be asked directly about the inputs to that model (e.g., population in each state, percentage of the population that reads newspapers). The theory behind the disaggregated approach is that it is easier for experts to answer questions about the intermediate quantities than about the total quantity. Relatively few studies have been done to compare the two approaches, however. Morgan and Henrion (1990) note that studies that have attempted to document the superiority of the disaggregated approach have shown mixed results.

---

[8] The "clairvoyance" test is often used to determine if a question is well-specified. The clairvoyance test asks whether an all-knowing individual would be able to answer the question without asking for additional information (Hora, 1992).

9 IEc also received valuable early input from Dr. Andrew Wilson of the Harvard Center for Risk Analysis, a researcher in expert judgment elicitation.

The Project Team carefully considered the relative advantages and disadvantages of the two approaches. A major advantage of the disaggregated approach is a more structured and transparent characterization of the key inputs and sources of uncertainty in the final quantity of interest. However, this method does require additional time and resources to develop not only a set of input variables, but also a model structure (or in some cases, multiple models) specifying the mathematical relationship between the variables on which the experts can agree prior to the individual elicitations. An aggregated approach can be easier to implement, both because it does not require extensive consultation with the expert panel to reach consensus on a model structure and because can reduce the length of the elicitation protocol. However, these benefits must be weighed against concerns about transparency of expert responses and the difficulties faced by experts who will be expected to mentally integrate the various problem components to provide their judgments.

The Project Team opted to pursue a largely aggregated approach, with some disaggregated elements. The protocol elicits a single distribution of the mortality impact of a unit change in annual average $PM_{2.5.}$ This effect distribution integrates the mortality impacts of the changes in both short-term (e.g., peak 24-hour) and longer-term $PM_{2.5}$ exposures that may contribute to a 1 $\mu g/m^3$ reduction in annual average ambient $PM_{2.5}$. The Project Team selected this approach for several reasons. First, our experience with the pilot study showed that eliciting separate distributions for the mortality impacts of short-term and long-term changes in exposure was very time consuming and difficult to complete in the 8-hour timeframe of the interview. Second, the team believed the large panel of experts recruited for this study made it unlikely that we could reach agreement among the group on a quantitative model for disaggregating the PM mortality effect estimate. Attendees of the EPA Symposium (see below) also supported this approach.

In an effort to ameliorate concerns about the aggregated approach we developed a comprehensive and detailed set of conditioning questions and also incorporated some quantitative disaggregated elements. The team designed the conditioning questions to promote transparency and provide a thorough understanding of the foundation of beliefs underlying each expert's quantitative distribution, including the ranking of key elements of uncertainty and careful documentation of key evidence in support of judgments. In addition, we gave experts the option to disaggregate two potentially challenging elements of an aggregated model: the likelihood that a causal relationship does or does not exist linking $PM_{2.5}$ exposures with mortality, and the likelihood and location of a threshold for PM mortality effects.

2.1.2 EPA SYMPOSIUM

The protocol development was also influenced by an EPA Symposium on the proposed expert elicitation project in April of 2005. The purpose of the EPA Symposium was to discuss the role of expert elicitation in EPA's efforts to improve the characterization of uncertainty in regulatory benefits analysis, particularly in the relationship between long-term $PM_{2.5}$ exposures and mortality. Attendees of the first day of the symposium included representatives of EPA and OMB, a group of invited scientists with expertise in the health effects of $PM_{2.5}$, and members of the public.

During the first half of the EPA Symposium, which was open to the public, members of the Project Team provided an overview of the use of expert elicitation in uncertainty analysis, including a presentation of the results of the pilot study and their use in an RIA as well as a discussion of plans for the full-scale expert elicitation study on $PM_{2.5}$ and mortality. This portion of the EPA Symposium concluded with an opportunity for audience members to comment on the design and content of the elicitation.

The second half of the EPA Symposium consisted of a closed meeting with Project Team members and the invited scientists to carefully review the draft elicitation protocol for the full-scale study. A series of moderated discussions were held to help assess whether the draft protocol was focused on the right questions, that it was comprehensive (i.e., addressing key elements of uncertainty) and that the questions and associated assumptions were clearly specified.

Following the symposium the Project Team and additional representatives from EPA and OMB met for half a day to discuss possible modifications to the draft elicitation protocol in response to suggestions made during the EPA Symposium. Comments by members of the public, the invited experts, OMB, and EPA were considered when revising the elicitation protocol and other study design aspects. Examples of specific decisions informed by the EPA Symposium included the following:

- The questions should focus on $PM_{2.5}$ (mass concentration); data on $PM_{2.5}$ components are too limited to support meaningful judgments;
- The questions should explicitly encourage experts to think broadly about evidence and theories from all relevant disciplines, and should be clear to both epidemiologists and non-epidemiologists;
- The protocol should provide experts with the option to either include judgments about causality either directly into their effect distribution or separately;
- The main quantitative question should focus on all-cause mortality as the outcome, rather than eliciting separate C-R functions for specific causes of death;
- The protocol should employ an aggregated approach to eliciting total mortality effects from long-term and short-term $PM_{2.5}$ exposures, rather than a disaggregated approach in which judgments about the effects of short-term and long-term exposures are elicited separately;
- The Project Team should include PM values in the study ranging from background up to the upper end of concentrations found in epidemiologic studies (4-30 $\mu g/m^3$);
- The protocol should specify that housing stock and levels of susceptibility in the U.S. population will remain constant for purposes of the study and co-pollutant concentration is unknown and remains a source of uncertainty;
- The briefing book should include data on the current U.S. population, such as health status and educational attainment;
- The interview should feature real-time feedback tools to allow experts to visualize their distributions as well as the impact of their views on causality and threshold.

In addition, the tools would provide experts with a "back of the envelope" calculation to show how their judgments translate into number of deaths as a "reality check."

2.1.3 PROTOCOL PRE-TESTING

Pre-testing of the protocol using experts not on the expert panel is critical to developing a well-functioning protocol. It enables the elicitation team to test the clarity of the questions and enables the team to practice and refine the process for administering the protocol. IEc conducted two full-length pre-tests of the protocol with two EPA experts in PM-related mortality: Dr. Tony Huang of EPA's National Health and Environmental Effects Research Laboratory (NHEERL), an expert in PM toxicology in humans, and Dr. Lucas Neas of EPA's NHEERL, an expert in PM epidemiology. The pre-tests were conducted in November of 2005. The pre-test subjects were not paid for their participation.

The pre-tests provided valuable input to the protocol development. Changes made to the protocol in response to the pre-tests included clarifications to the question about population thresholds and to the main quantitative question, the addition of a section where experts could discuss any important topics not previously covered (e.g., publication bias), and the inclusion of questions asking the experts to describe the strengths and limitations of the studies used to support their answers on mechanisms, causality, and threshold.

The pre-tests were also useful for identifying topics that would be helpful to discuss with experts at the workshop held before the Pre-elicitation Workshop. For example, it was evident that some confusion remained over the distinction between the causality and threshold question. Therefore, this was included as a topic discussion.

2.1.4 PROTOCOL DESIGN

The structure of the protocol is outlined in Exhibit 2-2. The protocol was divided into three parts: the preview and assumptions, conditioning questions regarding issues and evidence, and the quantification of a C-R function. (For experts who did not attend the Pre-elicitation Workshop, we also spent some time at the beginning of the interview discussing the objectives of the elicitation project and reviewing the expert elicitation process.)

The ultimate goal of the protocol was to get the experts to answer the following question:

> **What is your estimate of the true percent change in annual, all-cause mortality in the adult U.S. population resulting from a permanent 1 μg/m^3 reduction in annual average ambient $PM_{2.5}$ across the U.S.? In formulating your answer, please consider mortality effects of both reductions in long-term and short-term exposures.**

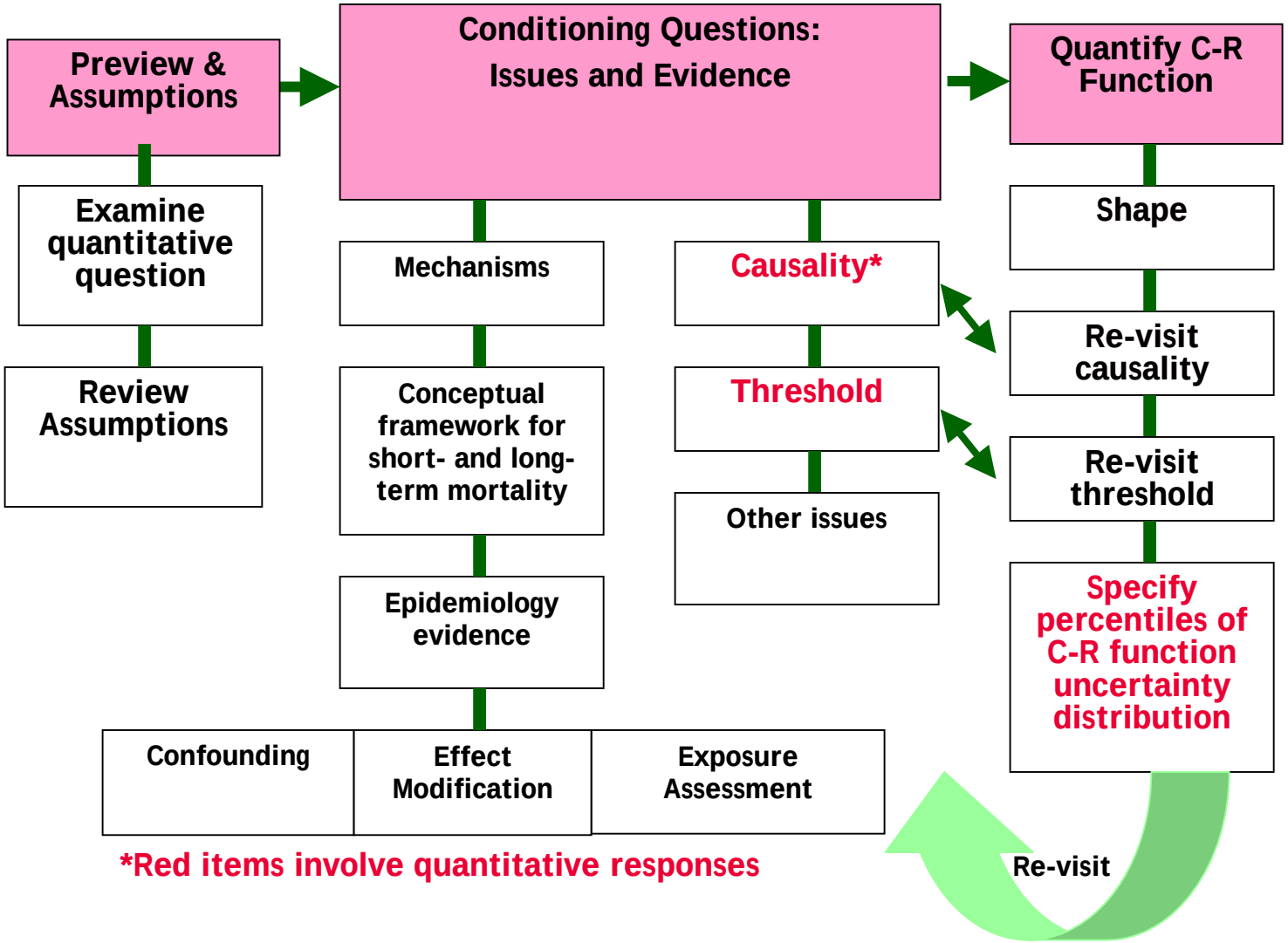
The protocol is divided into the following three parts:

1) Preview and Assumptions. In this section the elicitation team previewed the key quantitative question concerning the mortality impact of changes in annual average $PM_{2.5}$ concentrations in the U.S. In addition to emphasizing the ultimate goal of the elicitation, this section gave each expert the opportunity to examine carefully the underlying assumptions about factors such as historical and baseline exposures, regulatory implementation that might result in the hypothetical change in $PM_{2.5}$ concentrations, and characteristics of the U.S. population over time. After the preview, the expert had the opportunity to ask clarifying questions and raise any concerns regarding the specification of the main question and the assumptions.

2) Conditioning Questions. This section consisted of largely qualitative questions about factors to consider when characterizing the relationship between $PM_{2.5}$ exposure and premature mortality. These questions covered the following categories:

- Scientific evidence supporting physiological mechanisms linking $PM_{2.5}$ exposures to mortality;
- Key causes of death associated with $PM_{2.5}$ exposures;
- Conceptual framework for mortality effects of short-term and long-term $PM_{2.5}$ exposures;
- Role of study design in capturing effects of changes in annual average $PM_{2.5}$ exposures;
- Epidemiologic evidence for the impact of exposures to $PM_{2.5}$ on mortality;
- Evidence and impacts of confounding;
- Evidence and impacts of effect modification;
- Evidence and impacts of exposure misclassification;
- Likelihood of a causal relationship between $PM_{2.5}$ exposures and mortality;
- Potential thresholds in the C-R function; and
- Other influential factors (e.g., selection bias, statistical methodology, publication bias)

EXHIBIT 2-2: STRUCTURE OF THE ELICITATION PROTOCOL



These questions encouraged the experts to thoroughly and comprehensively consider key concepts, theories, and scientific evidence concerning the nature and potential magnitude of the $PM_{2.5}$/mortality C-R relationship. The expert's responses provided a conceptual foundation to support his quantitative judgments in the final section of the protocol. In answering these questions, experts were encouraged to rely on evidence from multiple disciplines and to consider impacts of both short-term and long-term $PM_{2.5}$ exposure, to the extent they believe both contribute to the mortality impacts of changes in annual average $PM_{2.5}$.

Unlike the pilot study, this protocol did not include questions about the relative toxicity of different PM components and sources. While the issue of differential toxicity remains a topic of interest to EPA, we removed these questions because of concerns expressed by the experts in the pilot study and the experts attending the EPA Symposium that available research is insufficient to support meaningful judgments on this issue.

3) Elicitation of Quantitative Judgments. The final section of the protocol presented the key quantitative question to be elicited:

> **What is your estimate of the true percent change in annual, all-cause mortality in the adult U.S. population resulting from a permanent 1 μg/m^3 reduction in annual average ambient $PM_{2.5}$ across the U.S.? In formulating your answer, please consider mortality effects of both reductions in long-term and short-term exposures. To characterize your uncertainty in the C-R relationship, please provide the 5th, 25th, 50th, 75th, and 95th percentiles of your estimate.**

In this section, the expert specified a functional form for the $PM_{2.5}$ mortality C-R function and then developed an uncertainty distribution for the slope of that function (the mortality impact per unit change in annual average $PM_{2.5}$). Experts could specify a functional form with a single slope that applies across the entire range of $PM_{2.5}$ annual average values specified in this study (4-30 μg/m^3) or could specify a more complex function by segmenting the study range of $PM_{2.5}$ and specifying different slopes for different segments.[10] The expert was also asked in this section to consider whether and how to factor his quantitative responses to the questions

[10] The $PM_{2.5}$ study range was developed through review of EPA monitoring data for the U.S., consultation with the EPA Project Team, and discussions during the EPA Symposium.

about the likelihood of a causal relationship and the potential for a population threshold into his distribution.

As shown in Exhibit 2-2, the elicitation was an iterative process. During the elicitation of the probabilistic distribution of $PM_{2.5}$ mortality effect estimates in the last part of the protocol, the elicitation team revisited responses to conditioning questions to ensure the expert's distribution was consistent with the judgments provided in the second section of the protocol regarding the various sources of uncertainty and their potential impacts.

2.2 EXPERT SELECTION

A well-designed expert selection process for an expert elicitation study should ensure that the final panel of experts has the appropriate expertise to address the questions posed to them and represents, as a group, a reasonably balanced range of respected scientific opinion on the issues being addressed. Previous studies (e.g., Hawkins and Graham, 1990) have also identified several additional criteria: the process should be explicit and reproducible, be reasonably cost-effective and straightforward to execute, and should minimize the level of control of the researcher conducting the elicitation.

We conducted expert selection in two parts, both of which relied on a peer nomination process and chose 12 experts for participation in this study. The first phase of the expert selection process was designed to select nine experts. The initial decision to include nine experts was based on several factors, including 1) a literature search that found most of the elicitation studies conducted to date (60 percent) use panels of six to eight experts, and 90 percent use panels of 11 or fewer experts (Walker, 2004); 2) it was deemed that nine experts would provide a balanced set of views on this topic; 3) the pilot study was criticized for the small panel size of five experts (EPA SAB, 2004); 4) government agencies are required to undergo an Information Collection Request process for the Paperwork Reduction Act if information is collected from more than nine individuals; and 5) resource and time requirements increase with each additional expert. The first phase of the selection process included the following steps:

1. **Identifying Experts to Provide Peer Nominations.** IEc conducted a literature search and publication count to identify a group of experts in the field who have published articles in peer-reviewed journals on the mortality and/or morbidity effects of PM exposure. We performed a search of pertinent literature published in the past 60 years (1945-2005) using the search engine Web of Science.[11] We used the results of the literature search to rank the authors based on the total number of first, second and last authorships combined. We selected the 32 highest-ranking authors to serve as nominators. The 32 nominators obtained from the publication count represented a range of expertise and perspective on the elicitation topic.

[11] The search included the following search terms: ("air pollution" or "particulate matter" or "fine particles" or "$PM_{2.5}$" or "PM_{10}") and ("death" or "mortality" or "survival" or "morbidity") and ("epidemiology" or "human" or "time-series" or "cohort" or "clinical"). IEc acknowledges support from Andrew Wilson of the Harvard Center for Risk Analysis in conducting the publication count.

2. **Obtaining Peer Nominations.** To encourage nomination of a broad array of experts on this topic, we divided the pool of potential nominators into four groups of eight. Potential nominators were assigned randomly to a group. We asked nominators to provide 10 names of individuals that met general criteria common to all groups. We also asked that the individuals nominated meet a set of criteria that was specific to each group (See text boxes for the criteria.) For instance, nominators in Group 1 received a list of criteria that included all of those listed in the "General Selection Criteria" as well as the two criteria listed in the "Group Specific Criteria" under "Group 1." Exhibit 2-3 shows the group assignments. We received 25 sets of nominations overall (seven from Groups 1 and 2; five from Group 3; and six from Group 4; nominators who provided lists are shown in bolded italics in Exhibit 2-3).[12]

GENERAL SELECTION CRITERIA

1. Ideal experts should possess the educational background and/or experience to both display a thorough understanding of results from the epidemiological literature addressing the relationship between chronic exposure to $PM_{2.5}$ and mortality, and to evaluate these results in the context of other evidence pertinent to the $PM_{2.5}$/mortality issue, such as relevant toxicological and physiological literature.

2. Experts may include primary scientific researchers as well as prominent individuals from scientific panels, institutions, journal editorial boards, and other such groups who, through their educational background and experience, are in a position to carefully interpret the key evidence regarding $PM_{2.5}$ exposures and mortality.

3. The overall set of experts nominated should be a balanced group that reflects the full range of respected scientific opinions concerning the strength of the evidence linking premature mortality with ambient $PM_{2.5}$ concentrations.

4. The nominees should all be based in either the U.S. or Canada.

[12] While this study did not allow nominators to self-nominate, eight experts in Exhibit 2-3 were nominated by others in the list and were ultimately selected as part of our expert panel. Given that the process of identifying nominators produced a list of frequently published researchers in the area of PM mortality, we believe it is reasonable that there would be significant overlap between the nominators and the selected experts.

GROUP SPECIFIC CRITERIA

We would like you to nominate experts that you feel:

Group 1:

- Are the most knowledgeable about the relationship between long-term $PM_{2.5}$ exposures and mortality; and/or
- Have studied in-depth the uncertainties and methodological limitations of existing cohort studies on $PM_{2.5}$ and mortality.

Group 2:

- Have made the most significant contributions to our understanding of the potential underlying biological mechanisms of the $PM_{2.5}$/mortality relationship; and/or
- Have made the most significant contributions to our understanding of the likelihood of a causal relationship between $PM_{2.5}$ and mortality.

Group 3:

- Display significant experience analyzing the relationship between $PM_{2.5}$ and mortality through participation in expert committees and workshops, and/or publication of review articles; and/or
- Display significant experience analyzing and applying the PM/mortality literature within a risk assessment and/or policy context.

Group 4:

- Are conducting innovative, cutting-edge research investigating the relationship between $PM_{2.5}$ and mortality; and/or
- Have made the most significant contribution to our understanding of the relationship between health effects and $PM_{2.5}$ exposures.

3. **Ranking and Selecting Experts.** We ranked experts based on the number of peer nominations they received within each specific group. We also compiled the groups into a combined ranked list. We then selected the top two nominees from each of the four groups, plus the most highly nominated expert from the remaining combined list, for a total of nine experts.[13]

4. **Replacing Experts.** If an invited expert was unwilling or unable to participate in the assessment, IEc replaced the expert with the next most highly nominated candidate in that expert's group, provided they were nominated by at least half of the respondents in that group. Otherwise, IEc replaced the expert with the next most highly nominated expert overall.

To aid the experts in the nomination process, we described the overall objectives of the expert elicitation project and provided them with an alphabetical list of first and last authors with at least two authorships as identified through the literature search, who are based either in the U.S. or Canada. To increase response rates, we followed-up the mailings with phone calls to the nominators.

[13] When faced with a group of two or more experts with the same number of nominations, we randomly selected from the tied set of experts.

EXHIBIT 2-3: NOMINATORS DERIVED FROM PUBLICATION COUNT

NOMINATOR	AFFILIATION
GROUP 1	
Burnett, Richard T.	Health Canada
Ito, Kazuhiko	New York University School of Medicine
Katsouyanni, Klea	University of Athens Medical School
Saldiva, Paulo Hilario Nascimento	University of Sao Paulo
Zanobetti, Antonella	Harvard School of Public Health
Devlin, Robert B.	U.S. EPA
Dockery, Douglas W.	Harvard School of Public Health
Pope, C. Arden	Brigham Young University
GROUP 2	
Brunekreef, Bert	Utrecht University
Diaz, Julio	Central University of Public Health, Madrid
Ghio, Andew J.	U.S. EPA
Samet, Jonathan M.	Johns Hopkins University
Sunyer, Jordi	Institut Municipal d'Investigacio Medica (IMIM)
Costa, Daniel L.	U.S. EPA
Ostro, Bart D.	California Office of Environmental Health Hazard Assessment
Peters, Annette	GSF-National Research Center for Environment and Health
GROUP 3	
Braga, Alfesio Luis Ferreira	University of Santo Amaro
Hoek, Gerard	Utrecht University
Jerrett, Michael	University of Southern California
Künzli, Nino	University of Southern California
Lee, Jong-Tae	Yonsei University
Spengler, John D.	Harvard School of Public Health
Anderson, H. Ross	St. George's Hospital Medical School
Schwartz, Joel	Harvard School of Public Health
GROUP 4	
Ballester, Ferran (Diez)	Valencian School of Studies for Health
Brauer, Michael	University of British Columbia
Dominici, Francesca	Johns Hopkins University
Goldberg, Mark S.	McGill University
Krewski, Daniel	University of Ottawa
Levy, Jonathan I.	Harvard School of Public Health
Lipfert, Frederick W.	Independent Consultant
Thurston, George D.	New York University School of Medicine
Note: Names in bolded italics indicate those who responded with a list of nominations.	

While this process featured a good acceptance rate and yielded nine experts, the panel exhibited less diversity in expertise than we anticipated in design, with most experts being epidemiologists. In an effort to increase representation of the biological, medical, and toxicological disciplines, we conducted a second phase of selections. EPA sought additional nominations of experts in these fields based on nominations provided by the Health Effects Institute (HEI). The general criteria for nominations were the same as for the first part of the selection process (Holmstead, 2005). HEI provided EPA with an unranked list of 10 nominees, plus two alternates (O'Keefe, 2005).14 IEc used a random ordering process to determine the order in which these experts would be contacted, with a goal of inviting three additional experts from this list, for a total of 12 experts overall. If an expert declined, either the specified alternate for that expert or the next expert on the list was contacted. The acceptance rate for this second step was not as high as the first; IEc recruited three additional experts to participate.15 Exhibit 2-4 presents the final list of the 12 participating experts.

EXHIBIT 2-4: FINAL EXPERT LIST

NAME	AFFILIATION
Dockery, Doug W.	Harvard School of Public Health
Ito, Kazuhiko	New York University School of Medicine
Krewski, Daniel	University of Ottawa
Künzli, Nino*	University of Southern California Keck School of Medicine (currently at Institut Municipal d'Investigació Mèdica (IMIM) - Center for Research in Environmental Epidemiology, Barcelona, SPAIN)
Lippmann, Morton	New York University School of Medicine
Mauderly, Joe	Lovelace Respiratory Research Institute
Ostro, Bart D.	California Office of Environmental Health Hazard Assessment
Pope, C. Arden III	Brigham Young University
Schlesinger, Richard	Pace University
Schwartz, Joel	Harvard School of Public Health
Thurston, George D.	New York University School of Medicine
Utell, Mark	University of Rochester School of Medicine and Dentistry

*Dr. Künzli was based in the U.S. at the time of expert selection, and subsequently began a sabbatical in Barcelona midway through the project.

The following experts were contacted, but declined to participate: Dr. Richard Burnett; Dr. Jonathan Samet; Dr. Michael Brauer; Dr. Carol Henry; Dr. Judith Graham; Dr. John Balmes; Dr. Gunter Oberdorster; and Dr. James Crapo. The most frequently cited reason for declining was scheduling concerns given the level of commitment asked of the experts.

[14] The alternates were provided for specific nominees, based on similarity of background and expertise.

[15] Three experts from the first group declined to participate, whereas five experts from the HEI list declined.

2.3 BRIEFING BOOK

All experts were sent a "briefing book" binder at least two weeks in advance of his interview (IEc, 2006). The purpose of the briefing book was to provide experts with a baseline set of materials to assist them in preparing for their elicitation interview; however, experts were free to consider other materials not included in the briefing book. The briefing book contained the following materials:

- The elicitation interview protocol;
- A CD containing over 150 relevant papers and compendia, searchable both alphabetically and by topic area;[16]
- A set of background information pages with recent U.S. data on air quality, health status, population demographics and other topics that may factor into the experts' probabilistic judgments; and
- Background materials, including a document describing factors to consider when providing probability judgments in order to avoid potential sources of bias, and an excerpt from 2002 National Research Council report on estimating public health benefits of proposed air rules.

Experts were given the opportunity to comment on the protocol and other materials either at the Pre-elicitation Workshop (discussed below) or privately, and where necessary, materials were updated and re-sent to all experts in response to the comments received prior to commencement of the first interview. Examples of the briefing book materials are included in Appendix B.

2.4 PRE-ELICITATION WORKSHOP

The expert panel was invited to participate in a day-long workshop in January of 2006. The workshop was designed to introduce the project, provide background information to the experts on expert judgment and the elicitation process, and to foster discussion about the key evidence available to answer the questions posed by the study.

The first half of the workshop consisted of presentations by the Project Team focused on educating the experts about the history of expert judgment elicitation, the objectives of this study, the design of the protocol, and the overall structure of the project. The experts also participated in a calibration exercise designed to familiarize them with the process of providing subjective probabilistic judgments. Another important goal of the morning session was to obtain feedback on the protocol.

The second half featured structured discussion sessions on three topics related to the $PM_{2.5}$-mortality issue (causality, shape of the C-R function, and evidence for quantitative estimates of the mortality effect of $PM_{2.5}$ exposure). In each discussions session, one

[16] For some of the topic areas, publications on the CD were restricted to those published since 2000 due to a large volume of studies in that area. The binder also included a list of additional PM health studies that had been compiled separately by EPA's National Center for Environmental Assessment for another purpose. Experts were able to request copies of these or any other papers they felt were relevant to the topic.

expert presented evidence related to the topic being discussed, and two other experts served as discussants. The purpose of these sessions was to both introduce important evidence for making judgments about each topic and to provide an opportunity for a critical discussion among all the expert participants about the strengths and limitations of that evidence. The Project Team was explicit that the goal of these discussions was not to encourage a consensus within the group, but to allow each expert to develop his own informed judgments.

Although every effort was made to find a date for the workshop that was amenable to all of the experts, eight of the 12 experts were able to participate in person for the full workshop; one participated by telephone for part of the day; and three were unable to participate. A workshop summary and copies of presentation slides from the workshop were sent to all 12 experts. Experts who did not attend the workshop were encouraged to review these materials carefully. In addition, papers cited at the workshop that were not already included in the briefing book were sent to all 12 experts in advance of their interviews, as was the finalized protocol, reflecting minor revisions requested by the experts at the workshop.[17]

2.5 ELICITATION INTERVIEWS

As in the pilot study, the Project Team chose to elicit the judgments of each expert individually during a personal interview. This choice reflects a preference based on a review of the expert judgment literature, and the goals of the project. Numerous approaches for obtaining subjective judgments from groups of experts with differing opinions have been proposed and demonstrated in the expert judgment literature (Morgan and Henrion, 1990; Cooke, 1991). Approaches vary widely in the degree of interaction between experts during or after the process and in whether the process is intended to achieve group consensus. Unfortunately, little research has been done that rigorously examines the relative ability of the different methods to achieve well-calibrated results (Morgan and Henrion, 1990). Most of the research comparing methodologies has involved almanac-type questions (e.g., "What is the height of Mount Everest?") and has shown little difference in the quality of the results obtained by the various approaches. The research does suggest that interactions between experts can increase rather than decrease the problem of overconfidence (and thus, poor calibration) (Morgan and Henrion, 1990). Whether these findings can be extrapolated to studies like this one that feature complex questions posed to relevant experts requires further research.

Prior to the interviews, IEc provided each expert a copy of the protocol and the briefing book materials, including the pre-elicitation workshop summary. The material was provided to allow each expert to familiarize himself with the questions to be asked and the resources available in the briefing book.

[17] IEc clarified that experts should assume for the purpose of this exercise that the regulatory action will achieve proportional reductions in all $PM_{2.5}$ components, and added differential impacts of $PM_{2.5}$ sources and/or $PM_{2.5}$ components to the examples of factors experts may address in the Other Issues question.

The elicitations were conducted between January 31 and April 28, 2006. The Elicitation Team consisted of two interviewers, one experienced in the elicitation of expert judgments, Dr. Katherine Walker, and one with expertise in PM exposure assessment, Dr. Patrick Kinney of Columbia University. Dr. Kinney was selected for his expertise in the subject matter of the elicitation and for his ability to objectively evaluate the responses of the various experts. Mr. Henry Roman and/or Ms. Tyra Gettleman of IEc also participated in the latter half of each interview, using internet-based conferencing software to provide experts with real-time feedback on their quantitative responses, as discussed below. Eleven of the 12 interviews were conducted at the expert's offices; one was conducted at Dr. Kinney's office.[18]

Most of the elicitations were conducted over the course of a single 8-hour day. Due to scheduling constraints, one of the elicitations was conducted over two days, the first half being completed the afternoon of the first day and the second half the next morning. Typically, covering the introductory material took about an hour, while the remaining time was split approximately evenly between answering the conditioning questions, and answering the quantitative questions.

For each question in the protocol, experts were asked to think systematically about the relevant evidence and to consider any sources of uncertainty, error, or bias that might influence their interpretation of that evidence. The elicitation team requested that experts cite specific studies or other evidence supporting their judgment on a particular issue, and also prompted experts to consider specific evidence that would support an opposing or alternative position. This approach gave experts the opportunity to more fully evaluate the robustness of their conclusions, allowing them to make adjustments to their judgments in light of the full range of evidence or explain why they found the alternative evidence unpersuasive.

For the conditioning questions related to the influence of confounding, effect modification, and exposure misclassification on published epidemiological studies, the elicitation team asked the expert to write each factor he wished to discuss on a card. The team then asked the expert to physically group and rank the cards in response to questions about the direction and size of the impact of these factors on the results from each of the epidemiological studies the expert cited as most relevant.

For the quantitative questions, each expert was first asked to specify his assumptions about the overall shape of the $PM_{2.5}$-mortality C-R function for the range of $PM_{2.5}$ concentration changes specified. This step included questions about the functional form for the C-R relationship (e.g., linear, log linear, piece-wise linear, curvilinear and whether he planned to incorporate a threshold).

Another key decision each expert faced at this juncture was whether he wished to incorporate his quantitative judgments about the likelihood of a causal relationship

[18] One expert's schedule could not accommodate an interview at his office during the timeframe of this study. The expert was able to travel to New York during this time period, however, and he agreed to conduct the interview at Dr. Kinney's office at Columbia University instead.

directly into his characterization of uncertainty or whether he wished to have it be incorporated at a later stage. Experts who chose the latter approach were asked to acknowledge that they planned to develop their uncertainty distributions about the $PM_{2.5}$-mortality relationship conditional on the assumption that a causal relationship exists.

As part of the process of eliciting quantitative values, we asked each expert to identify the most important factors that might contribute to bias in published, peer-reviewed estimates of the percent increase in mortality associated with PM exposures and to quantify the effect of such biases on the overall uncertainty in those estimates. Each expert was then instructed to consider whether his quantitative responses needed to be adjusted to account for those potential biases.

Each expert was then asked to estimate the percent reduction in all-cause mortality associated with a 1 $\mu g/m^3$ decline in $PM_{2.5}$ associated with several fixed percentiles; minimum, 5^{th}, 25^{th}, 50^{th}, 75^{th}, 95^{th}, and maximum of the distribution intended to describe his uncertainty in the magnitude of the "true" but unknown relationship (hereafter, his "uncertainty distribution) In an effort to minimize use of the "anchoring and adjustment" heuristic, the protocol was designed to begin with the theoretical basis for bounding the estimates (min, max, 5^{th} and 95^{th} percentiles), although some experts found this approach counter-intuitive preferred to begin with the median.

Each expert was also given the opportunity to use spreadsheet tools developed by the Elicitation Team that provide real-time feedback regarding the expert's responses. The Elicitation Team used two spreadsheets that were shared via an internet conferencing software. The first spreadsheet ("Spreadsheet Tool 1") was used to record the expert's estimated values for each of the fixed percentiles, his likelihood of a causal relationship, and his quantitative views on the existence of a threshold. The expert could either specify all of the percentile values of the distribution directly, or could specify a parametric distribution to assist with estimating some percentiles and/or to visualize the final distribution. Experts who chose the parametric approach could specify two or more parameters (e.g., 5^{th} and 95^{th} percentile or mean and standard deviation) of a parametric distribution (e.g., normal, Weibull) that described their judgments. The Elicitation Team used Spreadsheet Tool 1, along with Crystal Ball™ probabilistic software (CB), to build the distribution and then showed the expert a probability density function (PDF) and/or a cumulative density function (CDF) representation of his distribution.

If the expert specified that his mortality effect distribution was conditional on the existence of a causal relationship between $PM_{2.5}$ exposures and mortality, the protocol originally called for the Elicitation Team to incorporate the expert's estimate of the causal likelihood into the distribution probabilistically.[19] The Elicitation Team then displayed a

[19] The elicitation team multiplied the expert's uncertainty distribution by a Yes/No distribution representing the expert's likelihood of a causal relationship to illustrate how his view of causality would impact his distribution when applied in a benefits analysis. For example, if an expert specified a causal likelihood of 95 percent, his distribution would be multiplied by one for 95 percent of the time and by zero for five percent of the time. This approach assumes that the expert's causality distribution and conditional mortality effect distribution are independent.

PDF and CDF of his distribution incorporating the causal likelihood for the expert to review.[20]

As in the pilot study, when experts used approaches to characterize the concentration response function as non-linear, or incorporated thresholds, and/or developed distributions conditional on a causal relationship, the elicitation team thought it was important for experts to be able to visualize the integrated impact of these various assumptions on an overall uncertainty distribution for the percent reduction in mortality. The elicitation team accomplished this step using Spreadsheet Tool 1 and CB. Crystal Ball was used to sample from the expert's uncertainty distribution and his distribution of possible threshold levels, in relationship to the 2002 distribution of population-weighted $PM_{2.5}$ annual average concentrations in the U.S., obtained from EPA's benefits analysis model, BenMAP.[21] A similar approach was also used to illustrate the effect of specifying different slopes of the C-R function for different ranges of baseline annual average $PM_{2.5}$.

The second spreadsheet ("Spreadsheet Tool 2") was intended to show the expert alternate displays of his distribution as well as some of the implications of his distribution for estimates of mortality in the U.S. The first display in Spreadsheet Tool 2 showed each expert a graph plotting his uncertainty distribution as well as the analogous distributions from selected epidemiologic studies. Alternatively an expert could view his uncertainty distribution in the form of juxtaposed box plots of mortality estimates from the epidemiologic studies on which the expert relied for his quantitative estimates.[22] Finally, an expert could view two "back of the envelope" calculations based on his distribution.[23] The first was an estimate of the number of annual deaths that might be avoided if the annual average $PM_{2.5}$ concentration in the U.S. were reduced from 12 to 11 $\mu g/m^3$. The second was an estimate of the number of annual deaths that might be avoided if the annual average $PM_{2.5}$ concentration in the U.S. were reduced from 12 to 4 $\mu g/m^3$.[24] The expert was then shown a bar graph plotting the estimated annual deaths attributable to this drop in air pollution compared with annual deaths attributed to other major risk factors

---

[20] As discussed later under the discussion of the post-elicitation workshop, the Project Team agreed that this approach was not suitable approach for combining these two elements of the experts' judgments, and subsequently has presented them separately. The result of this change is that the uncertainty distributions of those experts who incorporated causal likelihood estimates are not directly comparable to those who did not and they are therefore presented separately.

[21] The Example Applied distribution was created as follows: On each iteration of CB, a value for baseline PM2.5 was selected from the BenMAP distribution, a value for the effect threshold was selected from the expert's threshold distribution, and a mortality effect estimate was selected from the expert's distribution incorporating causality. For each iteration, if the PM2.5 value equaled or exceeded the threshold, then the value from the distribution incorporating causality was selected for the example applied distribution; otherwise zero was selected. The process was repeated 10,000 times to generate the Example Applied distribution.

[22] All box plots consisted of the median, interquartile range, and 90% confidence interval.

[23] In order to perform the "back of the envelope" calculations of avoided deaths, we used the following damage model $D = P \times M \times (\exp(\beta \times \Delta PM) - 1)$, where: D = Number of Annual Deaths Avoided; P = U.S. Population (data taken from the Census Bureau website (www.census.gov)); M = Background mortality rate in the U.S. (deaths/100,000 population) (data taken from the Centers for Disease Control and Prevention (www.cdc.gov)); β = expert's C-R coefficient (percent change in mortality per 1 $\mu g/m^3$ change in $PM_{2.5}$ divided by 100); and ΔPM = change in annual average $PM_{2.5}$ ($\mu g/m^3$).

[24] In this example, 4 $\mu g/m^3$ is assumed to be the background level.

(e.g., smoking) and to the major causes of death in the U.S. The calculations were intended to provide a "reality check" for the expert. The two calculations were performed for the expert's 5th, 50th, and 95th percentiles of his distribution, and were also completed for other values at the request of the expert.

The Elicitation Team took extensive notes during the interviews. In some cases, experts may have written or sketched responses to certain questions. Following each interview, IEc provided the expert with a summary of his qualitative and quantitative judgments for review, adjustment and/or confirmation of his responses.

To maintain confidentiality, each expert was assigned a randomized letter between A and J with which his judgments would be associated in this report. We provided confidentiality to allow experts the freedom to express candid, independent opinions even if they should differ from those he has expressed publicly or from those of his employer.

2.6 POST-ELICITATION WORKSHOP

The project team held a final workshop with the experts following the completion of the elicitation interviews. The objectives of this workshop were to:

- Anonymously share the results of all experts with the group;
- Highlight areas where expert opinion varied for possible additional discussion;
- Clarify points of confusion that may have emerged during the interviews;
- Discuss data not available to all experts at outset of interviews;
- Give experts the chance experts to raise issues with which they had struggled during the interview for a broader discussion; and
- Encourage critical review by the experts of their judgments.

The Post-elicitation Workshop was not intended to force or otherwise promote a consensus. If points of agreement on issues occurred naturally, we documented them. No efforts were made to encourage experts to change their responses. All expert responses were presented using randomly assigned letters. Individual expert summaries prepared following the elicitations were not provided to the rest of the expert group.

The Post-elicitation Workshop was held in June of 2006. Eleven of the 12 experts participated in at least part of the workshop, and ten were present for three quarters of the discussions. The 4-hour workshop consisted of an overview of the qualitative and quantitative results from the 12 interviews, highlighting both areas of general agreement and areas where expert opinion varied. The overview was followed by a series of discussion sessions focused on topics identified by the elicitation team that exhibited significant variation in opinion across experts and that were viewed as having a significant impact on the results. The purpose of these discussions was not to promote consensus, but rather to clarify points of confusion or differences in interpretation of specific questions. The workshop included an open discussion session where experts could raise topics for discussion or clarification, as well as a session where experts could

provide feedback on the elicitation process. A discussion of the results of this workshop can be found in Chapter 3.

Following the Post-elicitation Workshop, a meeting summary and other follow-up materials requested at the Workshop were sent to all 12 experts. All experts were provided an opportunity to revise their judgments privately following the workshop, on the basis of insights gained from the Post-elicitation Workshop discussions, and were given a form to complete if they opted to make changes. Experts were allowed to modify both their quantitative probabilistic distributions as well as other quantitative and qualitative responses given during the interview. Experts were asked to clearly describe the changes requested and provide a detailed rationale suitable for inclusion in this report. Where necessary, the Elicitation team followed up with the experts by phone to clarify information submitted on the modification form.

2.7 EVALUATING EXPERTS' JUDGMENTS

To understand how experts' performance is assessed, both in the broader expert judgment field and for this project, it is important to understand that each expert's probabilistic judgment is a reflection of his or her own state of knowledge. (How well or with what degree of certainty does he think he can predict the quantity of interest?) One measure of his success is therefore related to how well he knows the limits of his knowledge. In the expert judgment field, this measure is known as calibration. An expert is well-calibrated if, for example, when asked to give his 90 percent confidence intervals for 100 predictions, his intervals contain the true value 90 percent of the time. A second important measure of the quality of an expert's judgment is "informativeness," an attribute reflecting the breadth of his confidence intervals. Two experts, one giving very broad intervals and the other very narrow, can both be well calibrated, but the latter is more informative. Finally, good judgments should be unbiased. That is, even if an expert is well-calibrated using the definition above, his median should fall close to the expected "true" median for the predicted value.

Of particular concern for analysis and decision-making is that research has suggested that experts tend to be overconfident, and therefore poorly calibrated (see Morgan and Henrion, 1990 for a review). In essence, they think they know more about the subject than they do, and therefore they express greater certainty in their predictions than is warranted. Their confidence intervals tend to be overly narrow, causing them to "miss' the true value entirely, or they may provide biased estimates. In the calibration example discussed above, an overconfident expert's 90 percent confidence intervals would contain the true value less than 90 percent of the time. Such inaccuracy in a key input to an analysis could have significant effects on the outcome.

At the same time, we do know that experts can perform well making predictions in their own areas of expertise. Studies have also shown that experts who receive regular feedback on their judgments (e.g., weather forecasters, see Murphy and Winkler, (1992); and physicians, see Winkler and Poses, (1993)). Walker et al. (2003) demonstrated that

exposure experts asked to predict benzene concentrations in EPA Region V were relatively well-calibrated.

Techniques do exist to objectively evaluate experts' performance. The "gold standard" for judging their performance requires that the "truth," for example the true $PM_{2.5}$ C-R function, become known. For studies like this one, the gold standard is clearly beyond reach due to data gaps. Other investigators have used additional sets of questions, for which the truth can subsequently be known, to assess the calibration of experts (Cooke, 1991). This approach requires careful selection of a set of questions that are likely to be valid predictors of calibration on the questions that are the focus of the assessment. Neither of these calibration alternatives was available for this study. Concern about feasibility of devising calibration questions that would be equally applicable to experts with the variability in technical background represented by the expert panel led the Project Team to decide not to incorporate a calibration component into this study.

In the absence of these calibration measures, we developed the protocol, followed elicitation procedures, held workshops, and provided extensive briefing materials with the objective of helping the experts avoid some of the common biases and errors of judgment (also referred to as heuristics) that can lead to poor calibration. Both during and subsequent to the interviews, we have evaluated the results of this study considering whether: 1) the judgments were statistically coherent; 2) the judgments were reasonably consistent with the rationales given by the experts; and 3) whether the process did help experts to avoid some of the common pitfalls associated with giving subjective judgments. The results of this evaluation are described in Chapter 4.

2.8 APPROACHES TO PRESENTING RESULTS

When faced with differing judgments across experts, analysts must give careful thought to whether and how to combine these judgments into a single value or distribution. Many investigators (e.g., Hawkins and Graham, 1990; Winkler and Wallsten, 1995; and Morgan et al., 1984) have preferred to keep expert judgments separate in order to preserve the diversity of opinion on the issues of interest. In such situations, the range of values expressed by the experts can help decision-makers by serving as inputs to sensitivity analyses of analytical models and thereby bounding possible outcomes. Individual judgments can also illustrate dichotomies of opinion arising from different disciplinary perspectives or from the rational selection of alternative theoretical models or data sets (Morgan and Henrion, 1990).

For this analysis, IEc has presented only the individual quantitative distributions of the mortality effect estimate elicited from the 12 experts interviewed. This approach differs from the Pilot Study, where we presented both individual expert distributions and a single, combined distribution using equal weights for each of the five experts. The Project Team discussed this issue extensively during the planning phase of this project and considered feedback on this issue from several sources. The decision not to combine expert responses was ultimately based on consideration of peer review comments on the pilot study, advice from EPA's Science Advisory Board (SAB), discussions from EPA's

April 2005 symposium on the full-scale PM Expert Elicitation study, and the composition and nature of the expert panel.

The external peer review of the pilot study yielded extensive comments on the issue of combining of expert responses. The peer reviewers disagreed about whether experts' judgments should be combined as part of the expert elicitation study and none identified a generally agreed-upon method of combining expert responses. One reviewer felt strongly that expert judgments should not be combined, cautioning that the combined distribution could produce a result that none of the experts would endorse, and another recommended that combination should only occur outside of the expert elicitation study (e.g., through the pooling of benefit estimates generated by running each expert's distribution individually through a benefits model.) No consensus approach to this issue emerged from the external peer review.

The Health Effects Subcommittee (HES) of EPA's SAB Advisory Council on Clean Air Compliance Analysis provided EPA with a limited review of the pilot study methodology as part of its review of plans for EPA's second prospective Analysis of the Costs and Benefits of the Clean Air Act (EPA-SAB-COUNCIL-ADV-04-002, March 2004). In its recommendations, the committee strongly emphasized the presentation of individual judgments:

> "…the HES advises EPA to present the entire collection of individual judgments; to carefully examine the collection of individual judgments noting the extent of agreement or disagreement; to thoughtfully assess the reasons for any disagreement; and to consider formal combinations of judgments only after such deliberation and with full awareness of the context ..."

At the April 2005 EPA Symposium, the issue of combining results was discussed among the invited health experts, some of who cited concerns over the equal weighting of results in the pilot study. Several attendees suggested the use of weights derived from the content and quality of responses.[25] Ideally, these weighting systems would address problems of uneven calibration (i.e., accuracy) and informativeness (i.e., precision) across experts, as well as potential motivational biases (Cooke, 1991).[26] However, implementation can be problematic. Weights determined based on the analyst's judgment or the judgment of the other experts are likely to have a significant subjective component and may themselves be subject to motivational biases; more objective methods that rely on an expert's answers to calibration questions with known answers are appealing, but not universally endorsed, in part because of the difficulties of identifying appropriate

[25] The literature includes several approaches that have been used to assign weights to individual experts based on content and expertise. Weights can be assigned based on the analyst's opinion of the relative expertise of each expert; on a quantitative assessment of the calibration and informativeness (i.e., precision) of each expert based on their responses to a set of calibration questions (as described in Cooke, 1991); or on weights assigned by each expert, either to him or herself or to the other experts on the panel (see Evans et al., 1994 for an example of this approach).

[26] "Motivational bias" refers to the willful distortion of an expert's true judgments. The origins of this bias can vary, but could include, for example, a reluctance to contradict views expressed by one's employer or a deliberate attempt to skew the outcome of the study for political gain.

questions for measuring expertise in the subject matter. The Project Team ultimately did not identify a preferable weighting scheme for combining responses.

Finally, the Project Team considered the selection process for the expert panel and its implications for combining judgments. A key objective of the selection process was to ensure representation of the full range of respected scientific opinion. However, the resulting panel is not a statistical sample; it does not reflect the relative weights of the various opinions expressed by the experts. Thus, a combination approach using equal weighting, for example, might overweight some opinions and underweight others. Lacking data on the prevalence of opinion in the expert universe, the team felt it would be inappropriate to present a mathematically combined distribution based on this sample.

The Project Team also considered behavioral approaches to combining expert responses, which require experts to interact in an effort to reach a consensus opinion. As in the pilot study, we chose not to consider a behavioral approach for several reasons. First, because there is disagreement regarding certain aspects of the likely nature of the PM-mortality relationship, we viewed the potential for achieving consensus to be very limited. Second, because the experts were selected to reflect the range of respected scientific opinions on this issue, individual elicitations enabled us to examine and preserve the variability in the experts' responses. Finally, the use of individual elicitations also avoids the dominance of the group opinion by any one individual and "attempts to preserve the unique perspective of each expert" (Wolff et al., 1990).

CHAPTER 3 | RESULTS

This chapter of the report presents the experts' responses to the questions posed in the elicitation protocol. We divide the results into two sections. We begin by summarizing expert discussions in response to the conditioning questions, which covered topics addressing the key evidence for or against a $PM_{2.5}$-mortality relationship as well as specific characteristics of such a relationship. The conditioning section generally follows the order in which the topics were raised in the elicitation interview. We then summarize the experts' responses to Part 4 of the protocol, including, their judgments concerning the shape of the concentration-response (C-R) function and their quantitative estimates of the percent change in annual, all-cause mortality in the adult U.S. population resulting from a permanent for a 1 $\mu g/m^3$ reduction in annual average $PM_{2.5}$ concentration across the U.S.

In the sections that follow, we identify the key rationales, key sources of data, and major uncertainties behind the experts' quantitative estimates of uncertainty, highlighting the important commonalities and differences among the experts' opinions. Where feasible, we also present tables summarizing the responses of each of the experts on a particular issue. (To preserve confidentiality of responses, we refer to experts using the letters A through L, which were randomly assigned to them.) Detailed summaries of each expert's responses to the protocol questions can be found in the Technical Support Document for this study (IEc, under development).

3.1 RESPONSES TO CONDITIONING QUESTIONS

The first half of each interview was devoted to addressing a broad set of "conditioning questions." The goal of these questions was to help each expert bring to mind and critique the scientific evidence they thought to be relevant to answering the quantitative questions about the $PM_{2.5}$-mortality relationship. Experts were encouraged to consider both relevant theory as well as empirical evidence from a wide range of scientific disciplines (e.g., epidemiology, clinical medicine, toxicology, exposure assessment). Because experts (as well as lay people) tend to bring to mind the evidence with which they are most familiar or to which they have been most recently exposed, the role of the elicitors in this discussion is to bring forward evidence that an expert may not have considered. While the protocol introduced the conditioning questions in a particular sequence, experts were offered the opportunity to respond in the order that made the most sense to them.

3.1.1 MAJOR CAUSES OF AND MECHANISMS FOR $PM_{2.5}$-RELATED MORTALITY

The quantitative characterization of $PM_{2.5}$-related mortality focused on total mortality. Consequently, the first set of questions in the protocol was important for understanding what the experts' thought the principal "drivers" of total mortality might be, such as the causes of death, the biological mechanisms, and the relative importance of long-term exposures versus short-term exposures in contributing to total mortality. These discussions also provided an important foundation for the experts' later evaluations of the strength of the causal relationship and their articulation of the potential existence of thresholds in the C-R function.

Experts were first asked to discuss whether they thought the mechanisms for PM related mortality differed between short-term exposures and long-term exposures. Experts A, B, C, F, G, H, K, and L chose to discuss the effects separately and the remaining four experts discussed the effects together. Despite their distinguishing between the effects of long and short-term exposures, it became clear that these eight experts saw some overlap across the two temporal domains.[27]

Regardless of experts' decisions to discuss the mechanisms and causes of death from short-term and long-term exposures separately or together, the types of effects often fell into similar categories. The main causes of death resulting from short-term exposures to $PM_{2.5}$ discussed by the experts were related to acute cardiac or respiratory events. Some experts also mentioned ischemic stroke as another plausible outcome. The three main contributors to mortality from long-term exposure discussed were cardiovascular mortality, chronic respiratory disease, and lung cancer.

Although the summaries below discuss some of the general patterns observed and the literature cited, it is important to recognize that there was often substantial variation in the depth and detail provided by individual experts. Some provided more general statements about kinds of evidence; others provided either detailed diagrams of mechanistic pathways or cited individual papers in support of particular viewpoints.

Cardiovascular disease

The mechanism that most experts thought was most plausible for explaining acute cardiac effects resulting from short-term particle exposures involved an oxidative stress response to particles deposited in the lung or translocated systemically, with release of reactive oxygen species leading to acute inflammation, cytokine release, compromise of the cardiac endothelium, and a fatal cardiac event. Acute responses could involve changes in blood coagulability/viscosity (Seaton et al., 1999; Peters et al., 2001; Riediker et al.,

[27] In particular, most of the experts cited the potential role of the "oxidative stress" pathways in which exposures to particles lead to inflammation and the release of cytokines, fibrinogen and other factors that may either directly impact the lung or be transported to and affect the heart or brain. Experts thought that this pathway represented a link between mechanisms for effects from short-term and long-term exposures. Pope et al. (2004a) and Ghio et al. (2000 & 2004) were often cited for the basic oxidative stress hypothesis. Additional scientific research supporting this hypothesis most frequently included work by Godleski (2000); Wellenius et al. (2003); Costa and Kodavanti (2003), and Sun et al. (2005). Expert E also cited work by Gurgueira (2002), Evelson and Gonzalez-Flecha (2000), and Rhoden (2004) in support of the role of reactive oxygen species and their impact on inflammation in the endothelium (cells lining the inner walls of blood vessels).

2004), and decreased plaque stability leading to the potential for increased risk of myocardial infarction (Zeka et al., 2005; Wellenius et al., 2003; Dominici et al., 2006). Experts A, E, G, and L also discussed ischemic stroke as potential causes of death, since it too could be triggered by this mechanism. Several experts mentioned "endothelial dysfunction" as another contributor to cardiovascular mortality from short-term exposures. Expert E cited a study by O'Neill et al. (2005) on brachial artery reactivity; Expert K mentioned Brook et al. (2002); Expert L mentioned Yamawaki et al.'s (2006) work on endothelial dysfunction and carbon black, though more in connection with a discussion of long-term exposures.

A second short-term cardiac mechanism discussed by several experts was related to particle-induced changes in the autonomic nervous system. Most experts discussed evidence showing that particle exposures may be associated with disturbances in the autonomic nervous system leading to changes in heart rate variability and other alterations. Here experts cited research on changes in heart rate variability (Pope et al., 1999; Gold et al., 2000; Creason et al., 2001; Devlin et al., 2003; Schwartz et al., 2005) and the defibrillator studies (e.g., Peters et al., 2000; Dockery et al., 2005). Two experts (E, L) thought that it would ultimately be found that this mechanism is related to the oxidative stress pathway, citing Schwartz et al. (2005).

Discussions of the cardiovascular effects of long-term exposures to fine particles were dominated by the hypothesis that chronic oxidative stress and inflammation in the endothelium contributing to accelerated formation of atherosclerotic plaque. The Sun et al. (2005) work was particularly influential in these discussions. This study reported progression of atherosclerosis in a strain of mice genetically pre-disposed to atherosclerosis (ApoE-/-) chronically exposed to fine particles. Many of the experts noted that this study provided a critical empirical link between studies showing the release of reactive oxygen species, thickening of the lining of carotid arteries (e.g., Kunzli et al., 2005), endothelial dysfunction, and epidemiologic studies showing increased cardiovascular mortality in populations exposed to fine particles. Expert L noted that works by Suwa et al. (2002) and Goto et al. (2004) were also supportive in this regard.

Respiratory Mortality

The second cause of death most experts discussed in connection with exposure to fine particles was respiratory-related diseases. Several experts expressed the view that the evidence was weaker for respiratory than for cardiac effects, due in part to diminished statistical power to detect respiratory related deaths (e.g., deaths from respiratory disease are less prevalent and are more likely to be miscoded).

Several experts expressed a role for short-term exposures to particles in contributing to respiratory mortality. These experts generally indicated that short-term PM exposures contribute to respiratory mortality largely through exacerbation of existing conditions (e.g., Chronic Obstructive Pulmonary Disease (COPD), asthma, allergies, other underlying pulmonary disease, or heart disease). Experts D and F specifically discussed alteration in immune responses; F cited studies by Gilmour et al. (2002) and Zelikoff et al. (2003) showing increased mortality rates in animals exposed to both streptococcus and

particles compared with animals exposed to streptococcus alone and Plopper and Fanucchi (2000) showing altered immune defenses in exposed animals.

Experts E, F, G, H, I, J, K, and L chose to discuss a potential role for long-term exposures to particles in contributing to increased respiratory mortality. They generally focused on mechanistic and epidemiological evidence for changes in lung function and the implications for mortality. Expert E discussed particle related development of COPD, positing a role for reactive oxygen species in promoting inflammation, mucous hyper-secretion, and structural damage to the lung. He cited Saldiva et al.'s (2002) concentrated air particles (CAPs) study showing increased capillary wall thickness in the lungs of animals exposed to particles compared to unexposed animals and epidemiological evidence from the Adventists Health and Smog (AHSMOG) studies (Abbey et al., 1991 & 1999; McDonnell et al., 2000; Chen et al., 2005), Schwartz (1993), and the Children's Health Study (Avol et al., 2001; Gauderman et al., 2004). In support of the mechanism for lung damage, Experts F and H cited other CAPs studies in animals by Ghio et al. (2000) showing increased inflammation as indicated by blood fibrinogen. Experts E, G, K, and L all brought up the Children's Health Study in California (Avol et al., 2001; Gauderman et al., 2004) showing diminished lung function growth in high-pollution communities, and decreases in lung function in children who moved from lower to higher pollution areas as evidence of the possible contribution of chronic particle exposure to changes in lung health and ultimately to respiratory mortality. Expert K was not convinced that these changes were particle specific, noting that NO_2 appeared to play a role. Other experts suggested that NO_2 might have been a surrogate for traffic-related particulate matter.

Views on the possible contribution of particle exposure to the development of asthma were limited. Two experts cited growing evidence that early exposures to particles might contribute to the development of asthma; Expert K cited Diaz-Sanchez et al. (1999) as some evidence for a role of particles in the development of asthma. Expert L did not think the evidence on particles and asthma development provided a clear story.

Cancer

Several experts discussed lung cancer in connection with long-term exposures to $PM_{2.5}$. Lung cancer arose in discussions of short-term exposures only to the extent that cancer patients might be a susceptible subgroup for short-term exposure effects.

The most common view expressed was that a connection between long-term PM exposure and lung cancer was scientifically plausible. Many expressed the basic argument that particulates contain mutagenic and/or carcinogenic chemicals that can act as initiators and/or promoters and that lung inflammation and the resulting cell turnover can act as a promoter. Several experts raised smoking as an analogy.

However, most noted that the epidemiologic data remained limited for lung cancer. Several experts cited Pope et al. (2002) as the strongest epidemiologic evidence of a link between particulate exposure and lung cancer and thought that findings from several other studies, though not statistically significant, were generally supportive. The strength of opinion about cancer ranged from Expert B, who thought there was a greater weight of

evidence supporting a link between lung cancer and particle exposure than for cardiovascular mortality, to Expert K who argued that the link was not strong and likely a function of poor control for confounding by smoking (noting the unreliability of personal smoking histories).

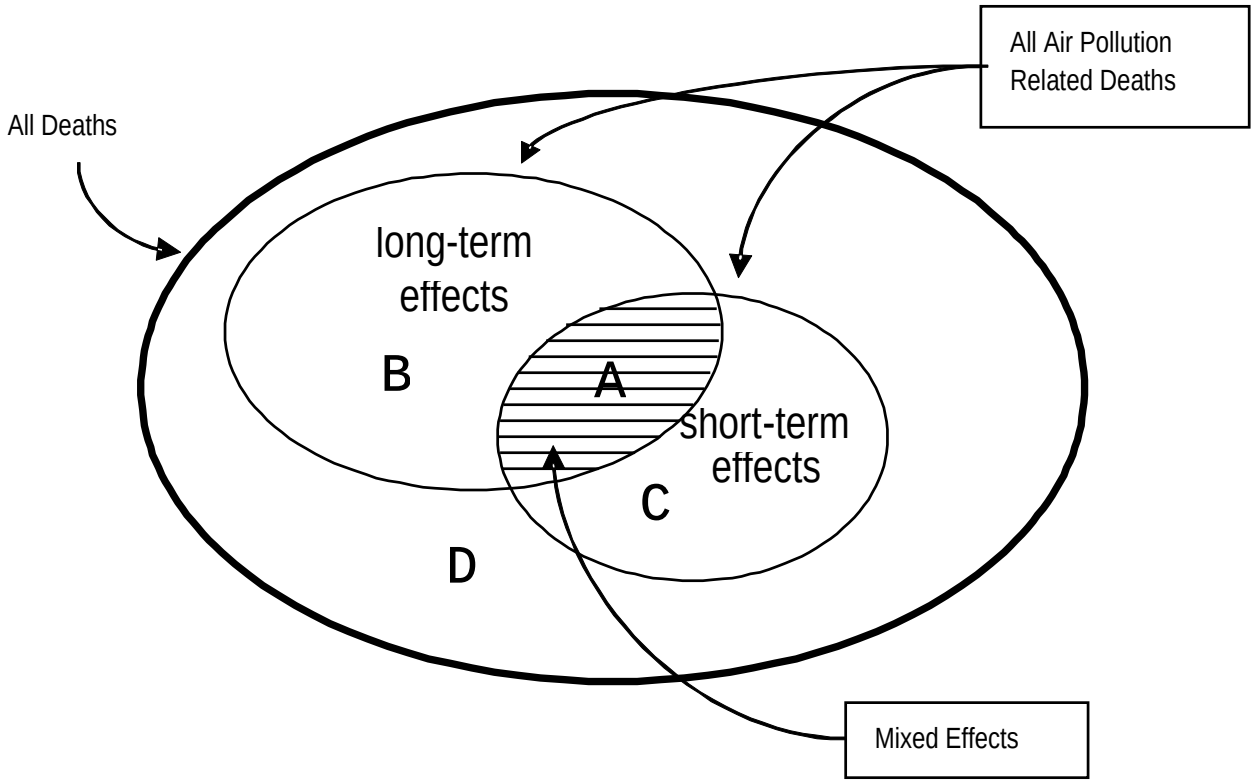
3.1.2 RELATIVE IMPORTANCE OF MORTALITY IMPACTS FROM SHORT-TERM AND LONG-TERM EXPOSURE CHANGES

The goal of this elicitation was ultimately to obtain a quantitative estimate for the <u>total</u> annual mortality effect that may result from a reduction in annual average $PM_{2.5}$ including both changes in short-term (e.g., 24-hour) and long-term exposures to $PM_{2.5}$. Developing an aggregated mortality estimate that reflects reductions in both short-term peak and long-term average exposures to $PM_{2.5}$ requires both assessing the relative contribution of short-term and long-term impacts to the overall change in mortality and thinking about the potential for overlap in the signals reported in different types of epidemiological studies (e.g., how much of the short-term impacts are captured by long-term cohort studies).[28] To assist experts in approaching this problem, we discussed with each of them the conceptual framework first presented in Kunzli et al. (2001). This framework uses Venn diagrams to describe the relationship between the deaths attributable to long-term exposures and those attributable to short-term exposures to fine particles (see Exhibit 3-1). Experts were encouraged to review and critically evaluate this diagram and discuss their own views about the relative contributions of different types of mortality impacts, developing alternative frameworks if necessary.

Nearly all experts found the structure described by Kunzli et al. (2001) a useful starting point for discussions though most discussed possible modifications to this basic conceptual framework. While the experts were not asked to specifically quantify the relative contributions of the different types of mortality, many discussed their views of the relative sizes of the different sets of mortality impacts. Exhibit 3-2 presents a summary of the experts' views on this issue based on responses to this and other sections of the protocol.

[28] For example, cohort studies focus primarily on analyzing the impact of long-term exposures to $PM_{2.5}$ but may also capture some of the impact of short-term variations in exposure during the cohort follow-up period. Time-series studies analyze the impacts of daily or short-term variations in PM concentrations and can characterize the cumulative impact of exposure over a few days, but not over a longer period of time. Those who rely exclusively on cohort studies may not account for all of the mortality impacts of short-term exposures, and therefore, may underestimate total mortality impacts. Those who rely on a sum of effects estimated in both cohort and time-series studies may overestimate mortality impacts.

EXHIBIT 3-1: CONCEPTUAL FRAMEWORK FOR SHORT-TERM AND LONG-TERM MORTALITY EFFECTS



Circle sizes do not reflect relative effects. (Adapted from Kunzli et al., 2001).

Category of Cases	Impact of Air Pollution	
	Underlying frailty due to air pollution	Occurrence of death (event) triggered by air pollution
A	Yes	Yes
B	Yes	No
C	No	Yes
D	No	No

A: Air pollution increases both the risk of underlying diseases leading to frailty and the short-term risk of death among the frail. For example, patients with chronic bronchitis that has been enhanced by long-term air pollution exposure may be hospitalized with an acute air pollution-related exacerbation of their illness leading to death shortly afterward.

B: Air pollution increases the risk of chronic diseases leading to frailty but is unrelated to timing of death. For example, a person's suffering from chronic bronchitis may be enhanced by long-term ambient air pollution exposure but the person may die due to acute pneumonia acquired during a clean air period.

C: Air pollution is unrelated to risk of chronic disease but short-term exposure increases mortality among persons who are frail. For example, a person with diabetes mellitus may be susceptible to heart attacks due to long-standing coronary disease; in such a case, an air pollution episode may trigger the fatal infarction leading to death.

D: Neither underlying chronic disease nor the event of death is related to exposure to air pollution.

EXHIBIT 3-2: RELATIVE IMPORTANCE OF SHORT-TERM AND LONG-TERM IMPACTS

RESPONSE	EXPERTS	COMMENTS
Emphasized longer-term impacts	A, B, C, E, F, G, H, I, J, L	Emphasized categories A, B, and/or A+B. Category C viewed as small contributor to overall mortality.
Uncertain	D	Thought he had insufficient information to discuss relative sizes of sets and their overlap, but relied on cohort studies for quantification
Emphasized short-term impacts	K	Emphasized the importance of peak exposure episodes in exacerbation of underlying disease process.

All but two experts placed much more weight on long-term exposure changes as drivers of changes in mortality. This conclusion was based on epidemiological evidence, supported by toxicological and clinical evidence for plausible mechanisms. These experts emphasized the potential for cumulative chronic damage over time from PM exposure leading to increased frailty and corresponding risk of death (Category B), with or without the involvement of short-term exposures as acute triggers of mortality. All experts in this group thought that the short-term mortality impacts not included in relative risk estimates from the cohort studies represented a very small percentage of the total mortality impact.[29] Though not explicitly addressed by the protocol, some experts did discuss their views on latency for mortality related to long-term $PM_{2.5}$ exposures. Most experts believed that the cardiovascular and respiratory mortality effects observed in the cohort studies are primarily due to exposures over the previous five to 10 years. Experts mentioned intervention studies, as well as the Six Cities follow-up analysis of Laden et al. (2006), as empirical support for latency periods measured in years rather than decades. Experts noted that we still lack empirical data on this important question.

One expert (K) expressed greater confidence that the mortality impacts reflect changes in short-term peak exposures. He found the body of evidence for short-term impacts more complete, compelling, and coherent than that for long-term impacts. Expert K expressed the belief that the mortality impact largely represents acute exacerbation of underlying disease processes. He also noted that his interpretation of this framework would

[29] In support of this position, Experts H and J cited studies that cast doubt on the hypothesis that time-series results represented "harvesting" (i.e., the displacement of mortality among the frail by a very short time period). H cited an analysis by Zeger et al. (1999) that simulated the potential effect of harvesting on a dataset and could not demonstrate the anticipated harvesting effects upon analysis of the data. In addition, he indicated that papers by Joel Schwartz (2000) found increasing coefficient size with longer time window, which "is the opposite of … harvesting." Expert J cited the work by Zeger and Schwartz as well as a study of the harvesting issue by Dominici et al. (2003). He thought that the evidence suggests, "that daily time-series studies utilizing only short-term time, day to day variability, are observing more than just the phenomena of short-term harvesting or mortality displacement. These results suggest the daily time-series studies capture only a small amount of the overall health effect of long-term-related exposure to particulate air pollution."

significantly shrink the size of the sets of air pollution-related deaths (circles B and C) relative to All Deaths (circle D).

One expert (Expert D) thought the categories of death postulated in the Kunzli diagram made sense conceptually, but thought he lacked sufficient data to inform a decision regarding the relative sizes and overlap of the different categories. He did indicate that he was comfortable that the estimates from cohort studies of long-term exposures are capturing the percent change in mortality associated with changes on in annual average $PM_{2.5}$ levels.

3.1.3 KEY EPIDEMIOLOGICAL EVIDENCE

Discussions of key epidemiological evidence supporting or refuting a relationship between $PM_{2.5}$ and mortality permeated the interviews. As part of the conditioning questions, experts first discussed the role of epidemiologic study design in characterizing the total impacts of $PM_{2.5}$ on mortality and then were asked to describe the characteristics of an "ideal" epidemiologic study for addressing this issue. The discussion then turned to the strengths and limitations of the existing epidemiological evidence, and how it compares with their ideal study. The experts revisited this topic during the quantitative questions while deciding the relative emphasis they would place on each study in developing their quantitative estimates. Experts were not required to provide quantitative weights, but were asked to indicate when or how they used specific studies in developing quantitative estimates. This section focuses on the theoretical discussion and the initial survey and evaluation of studies the experts found relevant.

The Role of Epidemiological Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

The experts were first asked to describe the types of epidemiologic study designs that they thought were most useful for estimating the change in total annual mortality related to a permanent reduction in ambient $PM_{2.5}$ concentration. The protocol then asked the experts to explain the extent to which the study designs that they chose capture effects from short-term and long-term exposures.

All of the experts thought that cohort studies would capture a large portion of the annual deaths from a drop in annual average ambient $PM_{2.5}$ exposures. Experts thought that these studies captured long-term effects as well as some short-term effects. Most of the experts thought that cohort studies missed deaths caused by very short-term exposure-responses. However, most thought that this small proportion of deaths missed by cohort studies was very small in relation to those captured by cohort studies. For example, Expert F estimated that roughly 97% of the mortality effects reported by cohort studies were due to long-term exposures with the remaining 3% attributable to cumulative short-term exposure effects, whereas Expert L thought that the short-term effects missed by the cohort studies might be around 0.5% per 10 $\mu g/m^3$ (i.e., similar to the effect estimates found in the NMMAPS study (Samet et al., 2000a & b)), which is equivalent to about five one-hundredths of a percent for a change of 1 $\mu g/m^3$. Expert E thought that cohort studies looking at changes in air pollution over time were the most directly relevant for measuring the mortality effects of changes in annual average $PM_{2.5}$ concentrations. He

thought that the Six Cities follow-up (Laden et al., 2006) is the only such study that has assessed the effect of long-term annual average PM changes over time on mortality.

All of the experts also thought time-series studies with various lag lengths could be useful in capturing true short-term effects (e.g., lags up to one week) and potentially some intermediate length effects (e.g., lags up to 1-2 months), depending on the length of the lag. Expert H thought that since the majority of short-term effects were already captured by the cohort studies, that time-series studies were more informative in establishing plausibility than in quantifying mortality effects. Ultimately, as discussed below, most experts did not rely on time-series studies when creating their quantitative distributions, except in some cases to support a lower-bound effect estimate.

All of the experts discussed intervention studies to some extent when answering this question. Some thought of them as supporting evidence, and not directly useful for quantifying the C-R function. Others thought these studies quantified some intermediate-length effects (e.g., between time-series and cohort) and some short-term effects. Expert G thought they were useful for determining which PM components might be causing the mortality. Expert A indicated that intervention studies would be most informative for the quantitative question because he thought most of the mortality effects occur in weeks to months. However he noted that the re-analysis of the Six Cities data (Laden et al., 2006), which looks over the last 10 years, might capture some long-term effects as well.

Experts H and J thought that cross-sectional or ecologic studies would capture the same effects as cohort studies (long-term and some short-term), but Expert H expressed concern about the lack of control for individual level confounders in these studies. Expert F thought that case-crossover studies would capture the same effects as time-series studies (short-term). Expert L indicated that case-control studies were useful for examining long-term effects for specific outcomes, such as lung cancer or COPD.

Ideal Epidemiologic Study

Experts were asked to describe an ideal epidemiologic study for answering the specific quantitative question posed by the protocol. The intent of this question was to provide experts with a "gold standard" against which to evaluate existing epidemiologic studies. The following are characteristics of an ideal epidemiologic study that were mentioned by several experts:

- Geographically representative of the entire U.S. (e.g., monitoring sites across the country);
- Collection of information on individual risk factors and residential information both at the beginning and throughout the follow-up period;
- Large sample size that is representative of the general U.S. population;
- Collection of genetic information from cohort members to identify and assess potential effect modifiers;
- Monitoring of individual exposures (e.g., with a personal monitor);

- Collection of data on levels of several co-pollutants (not only those that are monitored for compliance purposes);
- Accurate characterization of outcome (i.e., cause of death);
- Follow-up for a long period of time, up to a lifetime; and
- Prospective study design.

Experts E and J thought it would be useful to include cities that have differing changes in air pollution over time. Expert G said he would conduct a cohort and time-series study in the same population. He would then conduct studies in three cities, each of which has a different level of PM (although of a similar mix). Expert K expressed a preference for having an intervention in exposure. Expert L thought that having spatially resolved exposure data at the neighborhood level would be useful.

Epidemiologic Evidence for the Impact of Exposures to $PM_{2.5}$ on Mortality

Experts were asked to discuss which specific existing epidemiologic studies they think are informative for addressing the quantitative question and to describe their strengths and limitations.

Exhibit 3-3 indicates the epidemiologic studies mentioned by each expert in response to the conditioning question. All of the experts cited the following long-term, cohort-based studies as major evidence in support of a positive relationship between ambient annual average $PM_{2.5}$ concentrations and mortality:

- The Six Cities Cohort Studies (including the original study (Dockery et al., 1993), the reanalysis (Krewski et al., 2000a & b) and the follow-up (Laden et al., 2006)); and
- The ACS Studies (including the original study (Pope et al., 1995), the reanalysis (Krewski et al., 2000a & b), the follow-up (Pope et al., 2002) and analysis of follow-up data with a focus on cardiovascular outcomes (Pope et al., 2004))

The experts were uniform in their statements that the original Six Cities and ACS studies were well-conducted and that their results proved robust upon extensive reanalysis by Krewski et al. (2000a & b). One key strength of the Six Cities study often noted was that it was designed specifically for the purpose of evaluating the relationships between air pollution and health. The Six Cities study was prospective in nature, included recruitment of representative samples of subjects in each community, had reasonable control for possible confounders and effect modifiers, and set up purposeful air monitoring at sites chosen to characterize cohort exposures. In addition, some experts thought the choice of six cities with a wide range of PM levels was a strength. Frequently cited limitations of the Six Cities study included the small sample size, limited number of cities, and concerns about the representativeness of the six cities for the U.S. as a whole (since important regions of the U.S., such as the Southwest, Midwest and California, were not represented).

EXHIBIT 3-3: EPIDEMIOLOGIC STUDIES DISCUSSED BY EXPERTS WHILE ANSWERING CONDITIONING QUESTIONS

	Six Cities (Dockery et al., 1993; Krewski et al., 2000a & b; Laden et al., 2006)	ACS (Pope et al., 1995, 2002 & 2004; Krewski et al., 2000a & b)	ACS LA Reanalysis (Jerrett et al., 2003 & 2005	Utah Valley (Pope et al., 1989, 1991, 1996; Ghio et al., 2004)	Netherlands Cohort Study (Hoek et al., 2002)	AHSMOG (Abbey et al., 1991 & 1999; McDonnell et al., 2000; Chen et al., 2005)	Veteran's (Lipfert et al., 2000, 2003 & 2006)	Dublin Study (Clancy et al., 2002)	Elderly Californians Study (Enstrom et al., 2005)	Hong Kong Study (Hedley et al., 2002)	APHEA[1]	NMMAPS (Samet et. al., 2000a & b)	Willis et al., 2002	Canadian Time-Series Studies (Burnett et al., 2000 & 2003)	Finkelstein et al., 2004	MESA Cohort[2]	Mallick et al., 2002	Filluel et al., 2005	Woodruff et al., 1997	Women's Health Initiative[2]
Expert A	✓	✓	✓	✓	✓			✓			✓	✓				✓			
Expert B	✓	✓	✓			✓	✓				✓	✓	✓	✓			✓			
Expert C	✓	✓	✓	✓				✓		✓										
Expert D	✓	✓	✓	✓			✓		✓											
Expert E	✓	✓	✓		✓								✓							
Expert F	✓	✓	✓																	
Expert G	✓	✓		✓						✓										
Expert H	✓	✓				✓	✓													
Expert I	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓								✓	
Expert J	✓	✓	✓		✓	✓	✓		✓						✓					✓	✓
Expert K	✓	✓	✓					✓												
Expert L	✓	✓	✓		✓	✓			✓											
Total:	12	12	10	5	5	5	5	4	4	3	3	2	2	1	1	1	1	1	1	1

[1] The Air Pollution and Health - A European Approach (APHEA) includes a large group of studies. For full list of papers, please consult http://airnet.iras.uu.nl/products/reports_and_annexes/APHEA/APHEA_publications.pdf.

[2] Study not yet published at the time of the interview.

Many of the experts praised the ACS study for its large sample size, the large number of cities, broad geographic scope, and collection of individual risk factors, including those pertaining to cancer and cardiovascular disease. The limitation of the ACS study mentioned by most experts was the method of recruitment for the study, which resulted in a group with higher income, more education, and a greater proportion of whites than is representative of the general U.S. population. Also, several experts thought that the ACS exposure assessment was more problematic than the Six Cities study; the ACS study had to rely on whatever monitors were available to the study. This led to a single monitor representing exposure for an entire metropolitan area (containing several counties), whereas the Six Cities study often had exposures assigned at the county level.

Ten out of the twelve experts discussed the ACS Los Angeles (ACS LA) reanalysis by Jerrett et al. (2005). Experts explained that this study examined a subset of ACS cohort members in LA, assigning them to exposures based on a model estimating $PM_{2.5}$ in the same zip code as their residence. The effect estimates were two- to three-fold higher than the original ACS study. Experts thought that the better spatial resolution in this study was its main strength, and many believed that the higher effect estimates reflected reduced exposure errors. However, several experts were concerned that it only included one city with a specific mix of PM that may differ from the rest of the U.S. Experts F and H thought that only including one city may lead to spatial autocorrelation and residual confounding and Expert F added that there could be within-MSA mobility impacts on exposure. Two experts discussed other analyses examining the effect of improving exposure characterization. Expert B discussed a paper by Mallick et al. (2002) that performed a hypothetical analysis that attempted to correct for exposure misclassification in the Six Cities data. This study also found two- to three-fold higher effect estimates than the original study. Experts B and E mentioned a paper by Willis et al. (2003) that only included those individuals in the ACS cohort who lived in the same county as the exposure monitor. There was a doubling of the sulfate coefficient for all-cause mortality in the restricted cohort compared to including all individuals in the metropolitan-area. Although this coefficient was for sulfates, he thought that it provided evidence that better spatial resolution reduces exposure misclassification and leads to increased effect estimates.

Three other cohorts that were either discussed by experts or brought up during the interviews by the elicitors included the Veterans' Association (VA) Cohort Studies (Lipfert et al., 2000, 2003 & 2006), AHSMOG study of Seventh Day Adventists in southern California (Abbey et al., 1991 & 1999; McDonnell et al., 2000; Chen et al., 2005), and a study of elderly Californian's (Enstrom et al., 2005). Generally, experts did not weight these studies as highly as the cohort studies described above and only Expert J used these studies to inform his uncertainty in the C-R coefficient. Many of the experts were not familiar with the methodological details of these studies. Of the experts that did discuss the studies, many thought that they were generally supportive of the PM-mortality relationship and could be used qualitatively, but not quantitatively. Some of the main limitations of the AHSMOG study mentioned by experts included small sample size, too many subgroup analyses, and poor exposure characterization. Some experts thought the analytical approach used in the VA study was unclear. Experts thought all three studies

included populations that were not representative of the general U.S. population and also were concerned that they were not published in, and thus did not receive peer review by, epidemiologic journals.[30]

Eight experts discussed intervention or exposure change studies during this section of the protocol. The studies most often cited by experts included the "Dublin Study" (Clancy et al., 2002), the "Hong Kong Study" (Hedley et al., 2002), the "Utah Valley Studies" (Pope et al., 1989, 1991 & 1996; Ghio et al., 2004) and the Six Cities "Change" analysis (Laden et al., 2006). Six of these experts thought that these studies provided supporting evidence for plausibility, rather than informing the quantitative question. The exception is the Laden et al., 2006 "change" estimate, which was relied on by a two experts (see Exhibit 3-14). The most commonly cited strengths of these studies include large changes in exposure and avoidance of potential confounders associated with differences between populations. Expert E also thought that the Laden et al., 2006 was the only existing study that directly addressed the question posed by the protocol. Commonly discussed limitations included potential confounding by time-varying covariates, the fact that not all of these studies used $PM_{2.5}$ as an exposure measure, and the fact that they do not capture all of the long-term effects, as discussed above.

3.1.4 CONFOUNDING

Confounding is an issue that often is raised when interpreting quantitative results of epidemiologic studies. In the present context, a potential confounder is a variable that is both an independent risk factor for premature mortality and also is correlated with $PM_{2.5}$ concentrations. Only if both conditions are met can a variable act as a confounder. If not adequately controlled in the design and/or analysis of an epidemiology study, confounders could lead to biased results (either upwards or downwards) in the effect estimate for $PM_{2.5}$. Because of the potential importance of this issue in the present context, discussion of potential confounders was a key component of the elicitation interview.

Questions on confounding were directed at understanding each expert's views on the influence, if any, of confounding in the specific studies he cited as most relevant to his quantitative judgments concerning the form, magnitude, and uncertainty in the C-R function for mortality related to ambient $PM_{2.5}$ exposure. Experts began by listing what they believed to be potential confounders of the relationship between exposure to $PM_{2.5}$ and mortality in relevant studies of long-term and/or short-term exposures. Examples from the epidemiologic literature were provided as a starting point for discussion. Experts then discussed the theoretical rationale (e.g., biological or toxicological mechanism) or empirical evidence (e.g., clinical, epidemiological, animal, or exposure studies) for the impact of each potential confounder on the $PM_{2.5}$-mortality effect.

[30] The VA study assessed male Veterans with high blood pressure, originally recruited to assess the efficiency of anti-hypertension drugs, and the AHSMOG study involves only Seventh Day Adventists, a relatively small pool likely to be representative of a healthier population. The Enstrom et al. paper only included elderly Californians.

Experts discussed their views on the extent to which each potential confounder had been adequately controlled for in the relevant epidemiologic studies. For confounders that were considered to have not been adequately controlled, experts discussed the direction and magnitude of the possible resulting bias in the $PM_{2.5}$ effect estimate. We asked experts to try to classify the magnitude of bias on a three-level scale: 1 for minimal bias; 2 for medium bias; and 3 for major bias. The quantitative interpretation of this scale was left to the expert to define.

Experts mentioned a number of potential confounders in the interviews, but most of them believed that most or all of the key potential confounders had been adequately controlled for in the key studies they cited (i.e., the ACS and Six Cities studies, and their re-analyses and extended analyses: Pope et al., 1995 & 2002; Krewski et al., 2000a & b; Jerrett et al., 2005; Dockery et al., 1993; Laden et al., 2006).[31] For this reason, few made adjustments to their quantitative judgments to take account of bias due to confounding (see below). Three of 12 (F, G, D) made some mention of confounding (by co-pollutants and/or occupational exposures) in deriving their median estimate for the mortality effect of $PM_{2.5}$.

Contextual socio-economic status (SES) variables and co-pollutants were the two categories of potential confounders that were discussed most extensively and for which moderate concern regarding bias was sometimes noted.

- **Contextual SES.** The ACS and Six Cities cohort studies both assessed SES based on educational attainment, which is only one of many possible measures of economic status. The Jerrett et al. (2005) re-analysis of the Los Angeles ACS cohort explored the influence of an extended set of Census-based SES variables on the mortality- $PM_{2.5}$ relationship. In general, adding more of these "contextual" variables tended to diminish but not eliminate the $PM_{2.5}$ effect estimate. Some Experts viewed this as evidence that contextual SES variables are confounders that, if not taken into account, lead to positive bias in the $PM_{2.5}$ effect estimate. Others, such as Expert L, pointed out that some of the contextual SES variables might be surrogates for exposures to $PM_{2.5}$, in which case including them would lead to a negative bias in the $PM_{2.5}$ effect estimate. Ultimately however, no experts adjusted their median effect estimates for this factor.
- **Co-pollutants.** Co-pollutants were also frequently discussed. Most experts considered the SO_2 effect observed in the ACS study as implausible since SO_2 levels were quite low and most SO_2 deposits in the upper respiratory system. Several experts were of the opinion that SO_2 can be viewed as a surrogate for PM in the ACS study, since sulfate particles are derived from SO_2. Expert D made a case for confounding by volatile and semi-volatile organic compounds freshly emitted from motor vehicles. He noted that these compounds have been shown in laboratory studies to have relevant health effects, and also correlate with patterns

[31] Potential confounders cited included smoking, age, socio-economic position (at both the individual and contextual levels), ethnicity, occupation, co-pollutants, diet, exercise/obesity, weather, pre-existing health status, time-varying trends in health-care, indoor exposures, and differential migration.

of $PM_{2.5}$ concentrations in the ambient air, making them plausible confounders. Expert D adjusted his median effect estimate downwards slightly to account for the positive bias that may exist in the cohort studies. Expert K also noted concern about un-regulated co-pollutants, but made no adjustment for this factor. Expert G believed that the Six Cities study was biased upwards due to confounding by co-pollutants (NO2, ozone, SO_2, semi-volatile organics), and took account of this in deriving his median effect estimate. In the context of later discussions of exposure issues, Expert J mentioned the difficulty of directly addressing the quantitative question that was the focus of the elicitation (i.e., what is the mortality effect of a change in $PM_{2.5}$ concentration, holding all co-pollutants constant). Since $PM_{2.5}$ contrasts (spatial as well as temporal) occur in concert with contrasts in co-pollutants (due to overlapping sources), it is difficult to quantify the marginal impact of $PM_{2.5}$. If one wants the marginal effect of $PM_{2.5}$ alone, which is the stated focus of the elicitation, the PM effect estimates in key epidemiology studies could be moderate over-estimates of the true PM effect for the U.S. population since they incorporate some unknown quantity of co-pollutant effects. This issue was also discussed by other experts, as well as the related issue of differential toxicity for different particle components.

Among the other potential confounders, occupational exposures and smoking were frequently mentioned. Only Expert F adjusted his median effect estimate downwards slightly to account for possible bias due to occupational exposures. Smoking was mentioned by most experts, while noting that both the ACS and Six Cities studies controlled well for this factor using individual-level smoking histories. Only Expert K thought that smoking had not been adequately controlled for in the cohort studies. Expert K rated smoking as a potential level 1-2 positive bias, but ultimately did not incorporate this into his quantitative judgments.

3.1.5 EFFECT MODIFICATION

Effect modification is a term used by epidemiologists to refer to the phenomenon in which the magnitude of an estimated health effect (e.g., percent change in mortality for a 1 $\mu g/m^3$ change in exposure to $PM_{2.5}$) differs for different groups of people defined on the basis of some variable such as age, gender, race, economic status etc. The variable used to stratify the population is referred to as the effect modifier. Whereas uncontrolled confounding may result in biased and invalid effect estimates, effect modification does not threaten the validity of a study. It can, however, influence the generalizability of an effect estimate reported by an individual study to the general population. This can happen if the distributions of the effect modifier differ in the general population as compared with the study population. Depending on the direction and magnitude of this difference, the effect reported by the study may underestimate or overestimate the effect in the general population. Because of the potential importance of this issue in the present context, discussion of effect modification was a key component of the elicitation interview.

Discussions were directed at understanding experts' views on the influence, if any, of effect modification in the specific studies upon which they planned to base their

quantitative judgments concerning the form, magnitude, and uncertainty in the C-R function for mortality related to ambient $PM_{2.5}$ exposure. Experts began by listing what they believed to be potential effect modifiers in the relevant studies of long-term and/or short-term exposures. Examples from the epidemiologic literature were provided as a starting point for discussion. Experts then discussed the theoretical rationale (e.g., biological or toxicological mechanism) or empirical evidence (e.g., clinical, epidemiological, animal, or exposure studies) for the impact of each effect modifier on the $PM_{2.5}$-mortality effect. Experts discussed their views on the influence each effect modifier may have had on the relative risks reported in the relevant epidemiologic studies, and whether these reported risks were likely to be underestimates or overestimates of the average effect for the full U.S. adult population. Experts discussed the direction and magnitude of the possible under- or overestimation. We again asked Experts to try to classify the magnitude of the under- and overestimation on a three-level scale: 1 for minimal; 2 for medium; 3 for major. The quantitative interpretation of this scale was left to the Expert to define.

Experts mentioned a broad array of factors that could serve as effect modifiers of long- and/or short-term mortality effects of $PM_{2.5}$, including educational attainment and related SES factors, co-pollutants such as SO_2, housing characteristics like air conditioner use, race, smoking, and pre-existing conditions including diabetes, obesity, systemic inflammation, and genetic predisposition. Of these, educational attainment, co-pollutants, and race were each discussed by three or more experts.

Several experts noted that educational attainment was reported by Krewski et al. (2000a & b) to be an effect modifier in both the ACS and Six Cities studies, with mortality effects of $PM_{2.5}$ greatest in the sub-populations with less than a high school education and lowest or non-existent in the sub-population with more than a high school education. Because of the recruitment strategy employed in the ACS study, that population under-represented persons with less than high school education (11 percent) compared to statistics for persons 25 years or older in the nation as a whole (30 percent in 1980). In the Six Cities study population, 28 percent had less than a high school education, similar to the national average in 1980. These statistics suggested to many experts that the effect estimates reported by the ACS study were likely to underestimate effects for the adult U.S. population. Several experts noted that education per-se is unlikely to be responsible for the observed effect modification, but rather that education may represent a constellation of factors including housing, income, access to medical care, exercise, diet, underlying health status. Effect modification by educational attainment was discussed at some length during the pre-elicitation workshop, where Dr. Pope presented numerical results from a then-unpublished analysis that re-calculated ACS RR's based on re-weighting on educational attainment prevalence reported in the Six Cities study. These calculations resulted in a 30-50 percent increase in the RR compared to that originally reported in the ACS study. Several experts referred to this analysis during the elicitation. Six experts (A, C, E, G, I, J) chose to take account of this in deriving their median effect estimate in the quantification, either by inflating the reported ACS effects by 30-50 percent, or by placing weight on the Six Cities study, which, unlike the ACS study, did not under-represent persons with less than a high school education as compared to the

U.S. adult population. Dr. Pope's analysis has been published recently as part of a critical review of PM health effects (Pope and Dockery, 2006).

Among the other potential effect modifiers discussed, three experts (C, E, J) discussed race, mainly on a theoretical basis. Both the ACS and Six Cities studies under-represented non-whites in their populations, but experts noted that there are few empirical data demonstrating effect modification by race to-date. Some noted that it's difficult to separate race from various measures of SES. In the end, none made any quantitative adjustments for race.

Co-pollutants were also raised as theoretical effect modifiers, with little empirical evidence, by three experts (F, H, K). Higher SO_2 levels in the northeastern U.S. was noted as a possible explanation for high reported $PM_{2.5}$ effects in the Six Cities study. Expert F referred to this issue in deriving his quantitative judgments. Housing factors like air conditioner use were mentioned by two experts (E, K), mainly from the perspective of modifying indoor exposure to PM of outdoor origin.

EXHIBIT 3-4: SUMMARY OF EXPERTS' VIEWS ON EFFECT MODIFIERS

EFFECT MODIFIER	EXPERT RESPONSES											
	A	B	C	D	E	F	G	H	I	J	K	L
Education	⊗		⊗		⊗	x	⊗	x	⊗	⊗	x	x
Co-pollutant						⊗[1]		x			x	
Housing					x						x	
Race			x		x					x		
Smoking						x						
Underlying Susceptibilities (e.g., diabetes, obesity, systemic inflammation, COPD)					x							
Genetic Polymorphisms					x							
Air Conditioning					x							

Key:
X = Discussed
⊗ = Discussed and adjusted median to account for factor

[1]Specifically for SO_2

3.1.6 EXPOSURE ISSUES

Exposure assessment is both an essential and particularly challenging component of air pollution epidemiology. The challenge for exposure assessment is to measure health-relevant contrasts in exposures with as much accuracy and precision as possible. Exposure contrasts in currently-available epidemiology studies of particulate matter are based on either concentration variations occurring over time within a single location or variations in average concentrations across locations. In either case, population-based

studies typically rely on one or more central-site ambient monitors in a city to estimate air pollution exposure contrasts. This raises questions about the relationship between temporal/spatial *concentration* contrasts measured at the central site and temporal/spatial *exposure* contrasts experienced by the population at risk. Uncertainties in this relationship may arise due to variations in ambient concentrations over space within a city, incomplete penetration of ambient pollution into homes and workplaces, and patterns of population activity. Indoor sources may also contribute significantly to individual $PM_{2.5}$ exposures. In addition to uncertainties related to central site vs. individual exposures, there may be uncertainties regarding the timing of relevant exposures. In a cohort study of mortality, what is the relevant period of exposure that leads to increased mortality risk? Furthermore, what uncertainties are introduced if cohort members migrate to or from a location with different long-term $PM_{2.5}$ concentrations? Finally, there are uncertainties in both whether there are differences in toxicities of different $PM_{2.5}$ component species (e.g., sulfates, nitrates, elemental and organic carbon, metals) and the extent to which such potential differential toxicity may be responsible for differences in the relationship between health and PM exposure observed in different studies. Such differential toxicity would have implications for comparing studies conducted in different cities as well as interpreting analyses that use city-to-city differences in fine PM mass to characterize variations in exposure.

The elicitation interview probed experts' views on the influence, if any, of these exposure issues on their judgments concerning the form, magnitude and uncertainty in the C-R functions for mortality related to ambient $PM_{2.5}$ exposure, with particular reference to the key studies upon which they planned to base their quantitative judgments. Experts began by listing and defining what they believed to be the most influential exposure issues, and then discussed the theoretical rationale (e.g., biological or toxicological mechanism) or empirical (e.g., clinical, epidemiological, animal, or exposure studies) evidence for the impact of each potential exposure issue on the $PM_{2.5}$-mortality effect. Experts were then asked to characterize the direction and magnitude of any under- or overestimate in reported mortality relative risks (RRs) from specific studies that may have resulted from the exposure issue. The magnitude of under- or overestimation of RRs was reported on a three level scale, with one indicating minimal, two indicating medium, and three indicating major effect. The quantitative interpretation of this scale was left to the expert to define.

Uncertainties in population exposures assessed using central-site monitoring was raised by all experts as an important issue, and in many cases as a major issue (level 3), and nine experts took this issue into account when deriving their median effect estimate of the mortality effects of a 1 $\mu g/m^3$ change in $PM_{2.5}$. Most considered this to be a more important uncertainty for the ACS main study, where central sites were used to represent exposure over entire metropolitan areas, and many thought that this issue caused underestimation of the effects of $PM_{2.5}$ on mortality. The reason citied for this underestimation was the well-known effect of exposure measurement error ("misclassification") in biasing epidemiological effect estimates towards the null. Experts noted that the higher RRs reported in Six Cities might relate in part to finer spatial scale of monitoring there vs. ACS. Also the Jerrett LA ACS analysis, which

featured a finer spatial resolution of exposure assessment than the original ACS study and reported results more comparable to Six Cities, was often cited in support of this idea. However, some experts thought that the larger effects reported by Jerrett might also be due to a different mix of particle components in LA vs. the U.S. as a whole, perhaps related to motor vehicle pollution, or might reflect some bias due to spatial confounding by SES. Two experts also referred to the Willis et al. (2003) reanalysis of sulfate data from the ACS cohort, using only monitors located in the county of residence of cohort members. The RR based on the county-level subset was approximately double that in the full cohort where the unit of observation had been the metropolitan area. Also noted was the theoretical analysis of Mallick et al. (2002). One expert pointed out that there is likely to be an optimal spatial resolution that is fine enough to avoid spatial error but large enough to capture the spatial activity range of community residents. Several experts noted that a recent unpublished analysis by Jerrett and colleagues for NYC gave contrasting results to those observed in the LA analysis. These data were also discussed at some length at the Post-elicitation Workshop. However, given the lack of detailed information on this unpublished work, the NYC results were not relied upon by experts for the purposes of the elicitation.

Nine of the experts discussed whether the relevant time course of historical exposures was well-captured in the published cohort studies. A mismatch could introduce additional random exposure misclassification or could theoretically result in upward bias in effect estimates. Theoretically, as noted by Expert J, "if the time scale [for long-term effects] is on the order of decades and we use more recent $PM_{2.5}$ measures where the rank ordering is the same, but the [concentrations] are lower, then we are overestimating the effects." In other words, a more recent, more-narrow range of measured exposure differences across cities would serve as a surrogate for a larger, health-relevant, but unmeasured range of exposure differences that existed in the past. After considering the evidence and theory, however, experts decided not to take this into account in their quantification, noting that the bias would be significant only if latency periods were very long (i.e., multiple decades), for which there is little epidemiologic evidence.

A few experts thought that migration of cohort members out of the community of initial residence could lead to additional uncertainties, leading to underestimate of RR. This effect was thought to be greater for the ACS study, which lacked follow-up data on residence after initial enrollment. Differential migration by SES might lead to different degrees of misclassification of exposure, and might explain some of the effect modification noted in ACS study. No experts took this issue into account in quantifying their median effect estimate.

Expert J mentioned the difficulty of directly addressing the quantitative question that was the focus of the elicitation (i.e., what is the mortality effect of a change in $PM_{2.5}$ concentration, holding all co-pollutants constant). Since $PM_{2.5}$ contrasts (spatial as well as temporal) occur in concert with contrasts in co-pollutants (due to overlapping sources), it is difficult to quantify the marginal impact of $PM_{2.5}$. If one wants the marginal effect of $PM_{2.5}$ alone, which is the stated focus of the elicitation, the PM effect estimates in key epidemiology studies could be moderate over-estimates of the true PM effect for the U.S.

population since they incorporate some unknown quantity of co-pollutant effects. This issue was also discussed by other experts, as well as the related issue of differential toxicity for different particle components, but in the end, experts concluded that there was insufficient empirical data to guide any quantification of these influences. It should be noted that this issue is really one of confounding by co-pollutants. Other issues raised but left un-quantified by experts included the use of estimated rather than measured $PM_{2.5}$ data in the Six Cities follow-up analysis (Laden et al., 2006), and the influence of air conditioning use on penetration of ambient $PM_{2.5}$ indoors. Experts thought that these issues might introduce some biases, but were not able to quantify them.

EXHIBIT 3-5 SUMMARY OF EXPERTS' VIEWS ON EXPOSURE ISSUES

EFFECT MODIFIER	EXPERT RESPONSES											
	A	B	C	D	E	F	G	H	I	J	K	L
Spatial resolution	⊗	⊗	⊗	x	⊗	⊗	x	⊗	⊗	⊗	x	⊗
Migration						x				⊗		
Co-pollutants										⊗		
Temporal changes in exposure		x	x		x	x	x	x	x	x		x
Estimated $PM_{2.5}$	x			x	x							
$PM_{2.5}$ composition			x	x					x			x
Air conditioning use			x									

Key:
X = Discussed
⊗ = Discussed and adjusted median to account for factor

3.1.7 LIKELIHOOD OF A CAUSAL RELATIONSHIP BETWEEN LONG- AND SHORT-TERM $PM_{2.5}$ EXPOSURES AND MORTALITY

One of the more influential discussions held with each expert involved his assessment of the likelihood of a causal relationship between annual average exposures to $PM_{2.5}$ and total annual all-cause mortality (from both short-term and long-term exposures). Experts were asked to estimate this likelihood after completing most of the other conditioning questions, which were designed to evaluate the scientific evidence on causes of and mechanisms for $PM_{2.5}$-related mortality, epidemiological evidence, and factors relating to the quality of the key epidemiological studies (e.g., confounding, effect modification, exposure issues). The question began by having each expert talk generally about the types and/or strength of scientific evidence he would like to have in order to believe that an exposure/response relationship is actually causal. Experts were not required to follow a particular rubric or set of causal criteria. In addition to providing a framework from which to view individual experts' assessments of causality, this component of the question also provided some insight into some of the differences between experts' approaches to weighing evidence.

Each expert was then asked to estimate quantitatively the probability that the relationships between short-term exposures and mortality, between long-term exposures and mortality, or both were causal. They were asked for a "best estimate" as well as for a range of probabilities to give a sense of the level of uncertainty in their estimate.

Although this question had been discussed both at the EPA expert elicitation symposium and at the pre-elicitation workshop, the estimation of the causal likelihood and its use in the characterization of uncertainty was difficult and controversial in practice. Questions about the clarity of the question on causal likelihood and its use in the development of the final quantitative uncertainty distributions were a major focus of discussions at the post-elicitation workshop.

Some experts questioned whether the causal likelihood should be linked to different concentrations, that the question should be "causal at what concentration?" A related question was whether they might think differently about the likelihood of a change in mortality for a small incremental change in concentration (e.g., 1 $\mu g/m^3$) versus a large incremental change (e.g., 25 $\mu g/m^3$). In our view, these types of questions could be addressed in discussions of thresholds or in the mathematical expression of the concentration response relationship. We anticipated an answer to, "What is the likelihood that there is a causal relationship between reductions in annual average exposures to $PM_{2.5}$ (including reductions in both short-and/or long-term exposures) and changes in mortality at $PM_{2.5}$ levels currently experienced in the U.S. (e.g., annual averages of 4-30 $\mu g/m^3$)?"

Asking the experts to develop probabilities reflecting some joint judgment about the role of short-term and long-term exposures also added to difficulties in answering the question and interpreting responses. As indicated in the summaries in the exhibit, though many experts thought long-term exposures were dominant determinants of mortality, several thought repeated short-term exposures were important. Separating their roles was difficult.

Three experts B, D, and K discussed an additional interpretation of the question. Expert D made a distinction between the likelihood that there is a causal relationship between $PM_{2.5}$ (on a mass concentration basis) and the likelihood that particles are "causing all the effects ascribed to them." He thought the former was highly likely but that the latter was very low, but appropriately based his response on the first question. Expert B echoed this view: "It's not to say that $PM_{2.5}$ is the only air pollutant that may have a causative role in mortality, but certainly if you ask does $PM_{2.5}$ contribute to mortality, my answer is almost certainly yes." Expert K appeared to base his likelihood of a causal relationship at least in part on his assessment of whether particles *per se* were responsible for the mortality effects observed in studies.

Next, experts sometimes found it difficult to make the conceptual distinction between a statistical confidence interval and the uncertainty assessment for this project. As noted by Expert D, "on statistical grounds, 95% confidence is sufficient grounds for 'conviction" (i.e., for rejecting the null hypothesis of no relationship). However, the development of a subjective probability distribution allows for judgments to be made over the whole

interval from 0 to 1 on the basis of an aggregate assessment of the strength of evidence. Some experts found it to be almost a conceptual conflict to say that they were highly certain of a causal relationship (often 95%) but then to have to reflect that they were implicitly putting a 5% probability of no causal relationship. In fact, we are talking about two entities -- statistical confidence intervals on a particular finding and a subjective confidence interval or "credible" interval as it is sometime referred to in Bayesian statistics.

Basis for Causal Likelihood

Although not specifically mentioned by name, many of the experts drew on one or more of the basic Bradford Hill causal criteria (Hill et al., 1965). The following are conditions that the experts frequently mentioned as helpful for assessing whether an association is causal:

- Consistency (across epidemiological study designs);
- Biological Plausibility;
- Coherence (see associations across range of health outcomes);
- Temporal Relationship (between exposure and outcome);
- Strength (of the association);
- Analogy (to other exposures);
- Specificity (of response); and
- Statistically significant findings (robust to concerns about confounding, effect modification).

Not all experts focused on the same attributes, nor relied on the same studies when citing support for particular attributes.

Short term vs. Long term Exposures

Prior to developing their quantitative assessment of the likelihood of a causal relationship between total annual all-cause mortality and annual average ambient $PM_{2.5}$ exposures, experts were asked whether they would like to make a distinction between the likelihood of a causal relationship for short-term and long-term exposures in preparation for their overall assessment. Five experts (B, F, G, I, J, K) initially indicated they would, though only one gave separate causal likelihood for short-term exposures (Expert F placed 90% confidence in the causal relationship between short term exposures and mortality versus 100% for long term). Ultimately, most of the experts relied on assessment of the strength of the literature relating to the effects of long-term exposures. In their discussions of the conceptual framework for mortality effects, most expert indicated that long-term effects predominate (See Section 3.1.2)

Likelihood of a Causal Relationship

Exhibit 3-6 summarizes the probabilities given by each of the experts to describe their assessment of the likelihood of a causal relationship between exposures to $PM_{2.5}$ and total

mortality. The first column shows each expert's "best estimate", the estimate on which they would place their greatest probability weight, while the second column gives the range of values they thought were plausible. The final column provides qualitative summary statements, if given, to describe the expert's qualitative sense of the strength of the relationship. As part of this question, however, we had them consider the range of possible values as they evaluated alternative ways of weighing or interpreting the existing data. For example, they were asked to think about what factors or constellation of factors made it least plausible that there was a causal relationship and to contrast that with what factors made the relationship more plausible. This approach was sometimes helpful as an intermediate step.

As indicated in Exhibit 3-6, 10 of the 12 of the experts thought there was strong scientific support for a causal relationship with best estimates ranging from 90 to 100%. These experts most often cited the following kinds of arguments: the consistency in the pattern of epidemiological results across multiple study designs; the robustness of the Six Cities and ACS epidemiological results under the scrutiny of Health Effects Institute (HEI) re-analyses (Krewski et al., 2000a & b); the importance of intervention studies as "experiments" establishing the temporal relationship between decline of exposures and in mortality; the support from toxicological studies in animals (in particular, but not restricted to, the recent Sun et al. studies (2005)) and from *in vitro* studies for a plausible biological explanation. Only one expressed absolute certainty that there was no alternative explanation. Most expressed an unwillingness to eliminate all doubt, some indicating that "nothing is for certain."

The two experts who expressed stronger reservations about the plausibility of the causal relationships focused on similar concerns with the available scientific data although they ultimately selected different likelihoods. In our discussions, both essentially argued that that statistical associations from epidemiological studies, in particular the cohort studies, were not sufficient to establish a causal link. Expert K in addition raised at least two concerns that undermined his confidence in epidemiological studies; he thought that there was still a potential for confounding by smoking in cohort studies relying on questionnaire data and he thought that the absence of personal exposure data in both cohort and more recent defibrillator studies made it difficult to link mortality definitively to particular exposures. Both thought the Sun et al. (2005) study was an important new contribution but questioned whether the mouse model or exposures were truly relevant to humans and chronic ambient exposures to $PM_{2.5}$.

EXHIBIT 3-6: LIKELIHOOD OF A CAUSAL RELATIONSHIP BETWEEN $PM_{2.5}$ AND ALL-CAUSE MORTALITY

EXPERT	"BEST ESTIMATE"	RANGE	QUALITATIVE STATEMENTS
F	100	100	--
J	99	80-99	"[P]retty likely … the literature is pretty compelling … that there is a causal relationship both for short-term exposure and long-term."
C	99	99	"[V]ery, very likely"
L	99	90-100	"The evidence in the short-term … is so strong that … [it] is actually enough to say that it is a cause of death … I think it's extremely hard to argue that this is only an acute effect story and nothing else."
E	99	80-99	"I'm pretty convinced, … but nothing is for certain"
B	98	90-99	"[E]xtremely likely"
A	95	70-95.5	--
I	95	80-100	--
D	95	90-100	"[E]xtremely high likelihood for both"
H	90	80-95	--
G	70	60-80	The state of the science "strongly suggests a causal relationship."
K	35	5-50	"[I] don't have at this point great confidence that … bringing down the levels currently experienced in the U.S. are going to change mortality."

-- Expert did not give a summary statement about likelihood of a causal relationship.

3.1.8 THRESHOLDS

The protocol asked experts for their judgments regarding whether a threshold exists in the $PM_{2.5}$ mortality C-R function. The protocol focused on assessing expert judgments regarding theory and evidential support for a population threshold (i.e., the concentration below which no member of the study population would experience an increased risk of death).[32] If an expert wished to incorporate a threshold in his characterization of the concentration-response relationship, the team then asked the expert to specify the threshold $PM_{2.5}$ concentration probabilistically, incorporating his uncertainty about the true threshold level.

From a theoretical and conceptual standpoint, all experts generally believed that individuals exhibit thresholds for PM-related mortality. However, 11 of them discounted the idea of a population threshold in the C-R function on a theoretical and/or empirical basis. Seven of these experts noted that theoretically one would be unlikely to observe a population threshold due to the variation in susceptibility at any given time in the study population resulting from combinations of genetic, environmental, and socioeconomic factors.[33] All 11 thought that there was insufficient empirical support for a population threshold in the C-R function. In addition, two experts (E and L) cited analyses of the ACS cohort data in Pope et al. (2002) and another (J) cited Krewski et al. (2000a & b) as supportive of a linear relationship in the study range.

Seven of the experts favored epidemiological studies as ideally the best means of addressing the population threshold issue, because they are best able to evaluate the full range of susceptible individuals at environmentally relevant exposure levels. However, those who favored epidemiologic studies generally acknowledged that definitive studies addressing thresholds would be difficult or impossible to conduct, because they would need to include a very large and diverse population with wide variation in exposure and a long follow-up period. Furthermore, two experts (B and I) cited studies documenting difficulties in detecting a threshold using epidemiological studies (Cakmak et al. 1999, and Brauer et al., 2002, respectively). The experts generally thought that clinical and toxicological studies are best suited for researching mechanisms and for addressing thresholds in very narrowly defined groups. One expert, B, thought that a better understanding of the detailed biological mechanism is critical to addressing the question of a threshold.

One expert, K, believed it was possible to make a conceptual argument for a population threshold. He drew an analogy with smoking, indicating that among heavy smokers,

---

[32] As part of this section, the elicitation team reviewed the key assumptions about the study population for this elicitation. We assume the following about the U.S. population: 1) the population is 25 years of age or older, 2) the distribution of susceptible individuals across population reflects current patterns of susceptibility, and 3) this pattern will remain the same. The impact of projected changes in the age distribution on mortality, however, is incorporated directly into EPA's benefits model.

[33] Ten of the experts addressed this issue in an integrated fashion with respect to impacts of short-term and long-term exposures, or focused on long-term impacts, which they thought were the dominant contributor to changes in mortality. Two experts (B and G) suggested that a threshold may be more plausible conceptually for short-term mortality impacts of $PM_{2.5}$ than for long-term impacts, based on the postulated mechanisms.

only a proportion of them gets lung cancer or demonstrates an accelerated decline in lung function. He thought that the idea that there is no level that is biologically safe is fundamentally at odds with toxicological theory. He did not think that a population threshold was detectable in the currently available epidemiologic studies. He indicated that some of the cohort studies showed greater uncertainty in the shape of the C-R function at lower levels, which could be indicative of a threshold.

Expert K chose to incorporate a threshold into his C-R function. He indicated that he was 50 percent sure that a threshold existed. If there were a threshold, he thought that there was an 80 percent chance that it would be less than or equal to 5 $\mu g/m^3$, and a 20 percent chance that it would fall between 5 and 10 $\mu g/m^3$.

3.2 RESPONSES TO THE QUANTITATIVE QUESTION

In the final part of the interview, experts were asked to address the following question:

> **What is your estimate of the true percent change in annual, all-cause mortality in the adult U.S. population resulting from a permanent 1 $\mu g/m^3$ reduction in annual average ambient $PM_{2.5}$ across the U.S.? In formulating your answer, please consider mortality effects of both reductions in long-term and short-term exposures. To characterize your uncertainty in the C-R relationship, please provide the 5th, 25th, 50th, 75th, and 95th percentiles of your estimate.**

In addressing this question, the experts first specified a functional form for the $PM_{2.5}$ mortality C-R function and then developed an uncertainty distribution for the slope of that function (the mortality impact per unit change in annual average $PM_{2.5}$), taking into account the evidence and judgments discussed during the conditioning questions. We present the quantitative results provided by each expert below.

3.2.1 SHAPE OF THE CONCENTRATION-RESPONSE FUNCTION

The experts were asked to provide their judgments about the true shape of the C-R function relating mortality with changes in annual average $PM_{2.5}$ concentrations across the specified study range of 4 to 30 $\mu g/m^3$. Experts specified the functional form and indicated if the function included a threshold (i.e., a $PM_{2.5}$ level below which changes in annual average concentration would have no impact on mortality). The experts' responses are summarized in Exhibit 3-7. Eight of the 12 experts specified a log-linear C-R function with no threshold; three specified a non-linear function comprised of two log-linear segments with no threshold. One expert specified a non-linear function comprised of two log-linear segments that included a threshold.[34]

[34] As noted in the section on thresholds (3.1.8), the experts all discussed a theoretical model in which the C-R relationship should exhibit a threshold at the individual level. That is, an individual might have a threshold for a particular cause of death. However, individual thresholds may vary both across individuals and by cause of death for a given individual. Most experts agreed that variation in thresholds among individuals or population subgroups limited the ability of epidemiological studies to detect a population-level threshold. Though one expert's quantitative estimates employed a threshold, none of the experts could cite a study that provided strong evidence of a threshold for $PM_{2.5}$ mortality effects.

EXHIBIT 3-7 EXPERT JUDGEMENTS CONCERNING FORM OF THE C-R FUNCTION

FORM	EXPERT RESPONSES											
	A	B	C	D	E	F	G	H	I	J	K	L
Log-linear, No threshold	x		x	x	x		x	x	x	x		
Piecewise Log-linear, No threshold		x				x						x
Piecewise Log-linear, Threshold											x	

All of the eight experts who specified a single log-linear function for the entire range believed that the available cohort data were consistent with a log-linear interpretation and that existing data did not support the identification of any population thresholds.

The remaining experts specified non-linear functions using two log-linear segments across the study range of $PM_{2.5}$. These experts chose a nonlinear function to characterize what they believed to be the increased uncertainty in mortality effects at lower concentrations seen in the major epidemiological studies. The break point defining "lower concentrations" differed a bit among these experts, as shown in Exhibit 3-8. One expert (K) thought he had insufficient data to support anything other than splitting the study range in half. The other three experts cited the lower ranges of observed data in the cohort studies such as ACS in determining their break point, but acknowledged that data are limited to specify precisely the location where the slope changes and/or where uncertainty in the mortality impact increases.

EXHIBIT 3-8: BREAK-POINTS DELINEATING SEGMENTS OF THE $PM_{2.5}$-MORTALITY C-R FUNCTION

EXPERTS	BREAK-POINT CONCENTRATION
F	7 $\mu g/m^3$
B, L	10 $\mu g/m^3$
K	16 $\mu g/m^3$

The detailed results for these experts will be discussed below in the section on the experts' distributions. In general, the changes expressed at lower PM concentrations by these experts were modest and indicate a shallower slope for the lower range and/or slightly expanded uncertainty.

3.2.2 EXPERT CHARACTERIZATION OF UNCERTAINTY

Following the discussions on the shape of the concentration-response (C-R) function, experts were asked to begin the process of developing the probability distributions that would be used to characterize uncertainty in the PM-mortality relationship. Because all of the experts described the overall C-R function as linear or log-linear over all or portions of the study's concentration range, this step essentially characterized the distribution of possible values for the $PM_{2.5}$ mortality coefficient (i.e., the percent change in mortality per unit change in $PM_{2.5}$ concentration). The experts were specifically asked to estimate the minimum and maximum and the 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} quantiles of their distributions.

The elicitation protocol and accompanying elicitation tools were designed to allow experts some flexibility in developing their distributions. They were given options in 1) the incorporation of their judgments about the likelihood of a causal relationship, and 2) the elicitation of specific quantiles. Experts first decided whether they wanted to integrate their uncertainty about the likelihood of a causal relationship (as elicited in the conditioning questions, described in Chapter 2) directly in their uncertainty distributions, or whether they preferred to characterize uncertainty in the mortality effect conditional on the assumption that the relationship was causal.[35] The uncertainty distributions for the latter group of experts would then theoretically reflect sources of uncertainty other than those affecting the experts' judgments about whether or not the causal relationship exists. Five experts (A, C, F, H, and J) chose to incorporate their causal likelihood estimates directly into their distributions. The remaining experts (B, D, E, G, I and K) preferred that these two judgments be presented separately and combined only if deemed by EPA for the purposes of a benefits analysis.

The Elicitation Team then worked with each expert to elicit individual quantiles of his distribution. The basic approach outlined in the protocol was direct elicitation of individual quantiles, beginning with the minimum and maximum, the 5^{th} and 95^{th} percentiles, and followed by the elicitation of the 25^{th} and 75^{th} percentiles, and finally the median. However, in response to feedback received by the Project Team during the EPA Symposium and the Pre-elicitation Workshop, experts who preferred to work with parametric distributions were given the opportunity to elicit a more limited number of quantiles and to fit a parametric distribution using the Crystal Ball™ software. Experts B, F, H and L developed distributions using the first approach. Seven of the remaining experts decided it was intuitively more straightforward to develop rationales for two or more quantiles, or to estimate a mean and standard error, adjusted for additional uncertainty (Expert E), and to fit a distribution with Crystal Ball. Expert I preferred to characterize uncertainty by choosing three studies whose effect estimates represented different conditions about which he was uncertain, assigning subjective weights to each study, and combining their results via Monte Carlo simulation. Whatever their approach, individual experts were asked to explain the rationale for every quantile elicited and to

[35] In the latter case, the experts were informed that their conditional C-R coefficient uncertainty distribution and their probabilistic distribution for causality would ultimately be integrated when applied by EPA in future benefits analyses.

conduct "reality checks" of the values they selected. These reality checks included using the spreadsheet tools described in Chapter 2 to: 1) allow each expert to visualize his distribution and compare and contrast it with distributions from the key published studies he cited as relevant; and 2) generate example national-scale back-of-the-envelope benefit calculations based on quantiles of the experts C-R coefficient distribution and compare them with data on other major causes of death in the U.S.

Exhibit 3-9 displays the individual values for each of the percentiles requested of the experts.[36] The first column indicates the distributional form of the distributions developed by each expert. The term "Custom" indicates that all of the percentiles were individually elicited. Of the eight who selected parametric distributions, six fit normal (i.e., Gaussian) distributions, one chose a triangular distribution, and one a Weibull. Individually elicited values for each expert are indicated in boldface and italic type; all other values were generated by the elicitation team using Crystal Ball™ statistical software to replicate the expert-specified distributions. Note that the experts who chose normal distributions to characterize their uncertainty distributions chose different percentiles with which to fit their distributions. For the four experts who used piecewise log-linear functions to characterize the C-R relationship over the 4-30 $\mu g/m^3$ range of $PM_{2.5}$, the table displays both of their distributions.

[36] Note that all results presented in this report in Exhibits 3-9 through 3-13 reflect the experts' final judgments after making changes after the Post-elicitation Workshop, if applicable.

EXHIBIT 3-9: SUMMARY OF EXPERT SUBJECTIVE UNCERTAINTY DISTRIBUTIONS FOR C-R COEFFICIENTS

	DISTRIBUTION TYPE	MINIMUM	5TH PERCENTILE	25TH PERCENTILE	50TH PERCENTILE (MEDIAN)	75TH PERCENTILE	95TH PERCENTILE	MAXIMUM	INCLUDES LIKELIHOOD OF CAUSALITY?[3]	CAUSALITY LIKELIHOOD	THRESHOLD SPECIFIED?
A	Normal	*0*	*0.29*	1.1	1.6	2.1	*2.9*	*4.0*	Y	0.95	N
B (4-10 μg/m^3)	Custom	*0.01*	*0.10*	*0.20*	*1.2*	*2.1*	*2.6*	*2.8*	N	0.98	N
B (>10-30 μg/m^3)	Custom	*0.10*	*0.20*	*0.50*	*1.2*	*2.1*	*2.6*	*2.8*	N	0.98	N
C	Normal	*0*	0.40	0.90	**1.2**	1.5	*2.0*	N/A	Y	0.99	N
D[1]	Triangular	*0.10*	0.35	0.66	0.90	1.1	1.4	*1.6*	N	0.95	N
E	Normal	*0*	1.0	1.6	*2.0*	2.4	*3.0*	N/A	N	0.99	N
F (4-7 μg/m^3)	Custom	*0.37*	*0.58*	*0.73*	*0.93*	*1.1*	*1.4*	*1.7*	Y	1.0	N
F (>7-30 μg/m^3)	Custom	*0.29*	*0.77*	*0.96*	*1.1*	*1.4*	*1.6*	*1.8*	Y	1.0	N
G	Normal	N/A	0.70	0.88	*1.0*	1.1	**1.3**	*1.5*	N	0.70	N
H	Custom	*0*	*0*	*0.40*	*0.70*	*1.3*	*2.0*	*3.0*	Y	0.90	N
I[2]	Normal	*0.20*	0.38	0.90	1.3	1.6	2.1	*2.3*	N	0.95	N
J	Weibull	*0*	*0.15*	0.53	*0.90*	1.3	**2.0**	*3.0*	Y	0.99	N
K (4-16μg/m^3)	Normal	N/A	*0.10*	0.28	*0.40*	0.52	0.70	*0.80*	N	0.35	Y
K (>16-30μg/m^3)	Normal	N/A	*0.10*	0.45	*0.7*	0.95	1.3	*1.5*	N	0.35	Y
L (4-10 μg/m^3)	Custom	*0*	*0.20*	*0.57*	*1.0*	*1.4*	*1.6*	*2.7*	N	0.75	N
L (>10-30μg/m^3)	Custom	*0.02*	*0.20*	*0.57*	*1.0*	*1.4*	*1.6*	*2.7*	N	0.99	N

Note: Numbers in bold indicate a percentile value directly provided by the expert. All other numbers were generated by the elicitation team using Crystal Ball for the expert's specified parametric distribution based on percentiles and other information provided by the expert.

Expert K chose to incorporate a threshold into his C-R function. He indicated that he was 50% sure that a threshold existed. If there were a threshold, he thought that there was an 80% chance that it falls between 0 and 5 μg/m^3, and a 20% chance that it falls between >5 and 10 μg/m^3. The elicitation team took this information and created a probabilistic distribution in Crystal Ball with 50% of the weight at zero, 40% of the weight between >0 and 5 μg/m^3, and 10% of the weight between >5 and 10 μg/m^3.

[1] Expert D also provided a most likely value of 0.95 that was used to generate a triangular distribution in Crystal Ball.

[2] Expert I provided three beta coefficients and standard deviations from three epidemiologic studies and had the elicitation team combine these into a single distribution using Crystal Ball (placing equal weights on each study). The resulting data was fit to a normal distribution, as specified by the expert.

[3] Values in this column represent the expert's view on the likelihood that a causal relationship exists between $PM_{2.5}$ and mortality, as described in Section 3.1.7.

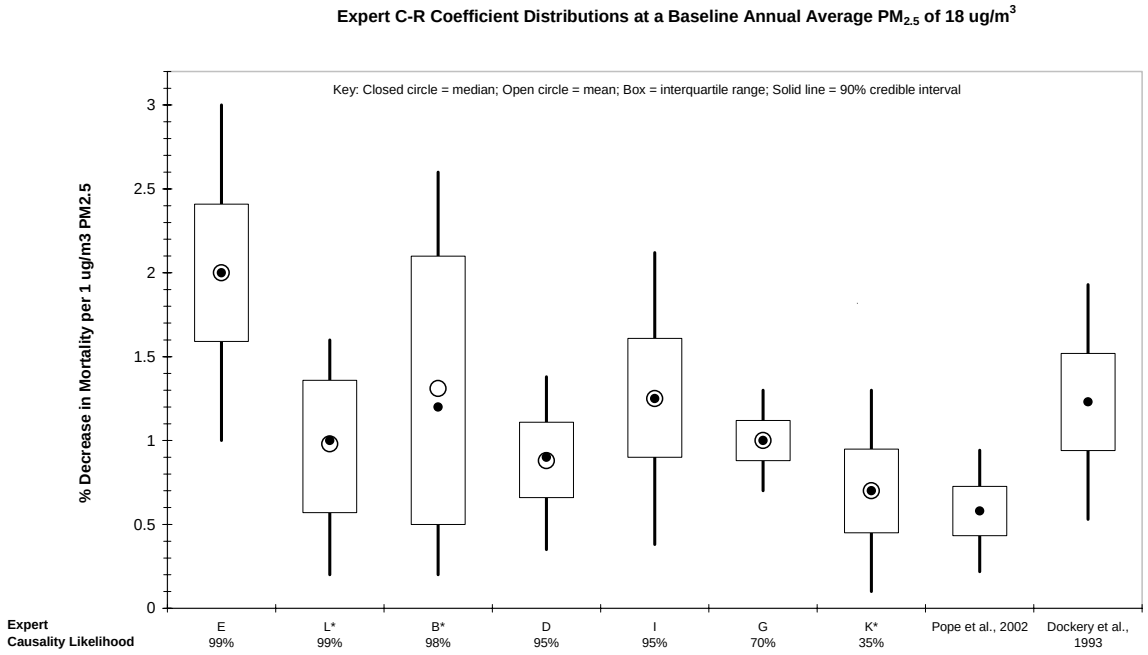
Because of the varied approaches taken by the experts to characterizing their views about uncertainty in the $PM_{2.5}$ mortality relationship, it is difficult to convey and to compare all their results directly in one exhibit. Direct comparison can only be done when their distributions are applied to the same scenario in a benefits analysis. We consequently portray the results in a series of graphs, grouping experts on the basis of whether they incorporated causal likelihoods into their distributions and whether they characterized the C-R function as linear or non-linear over the study range of $PM_{2.5}$.

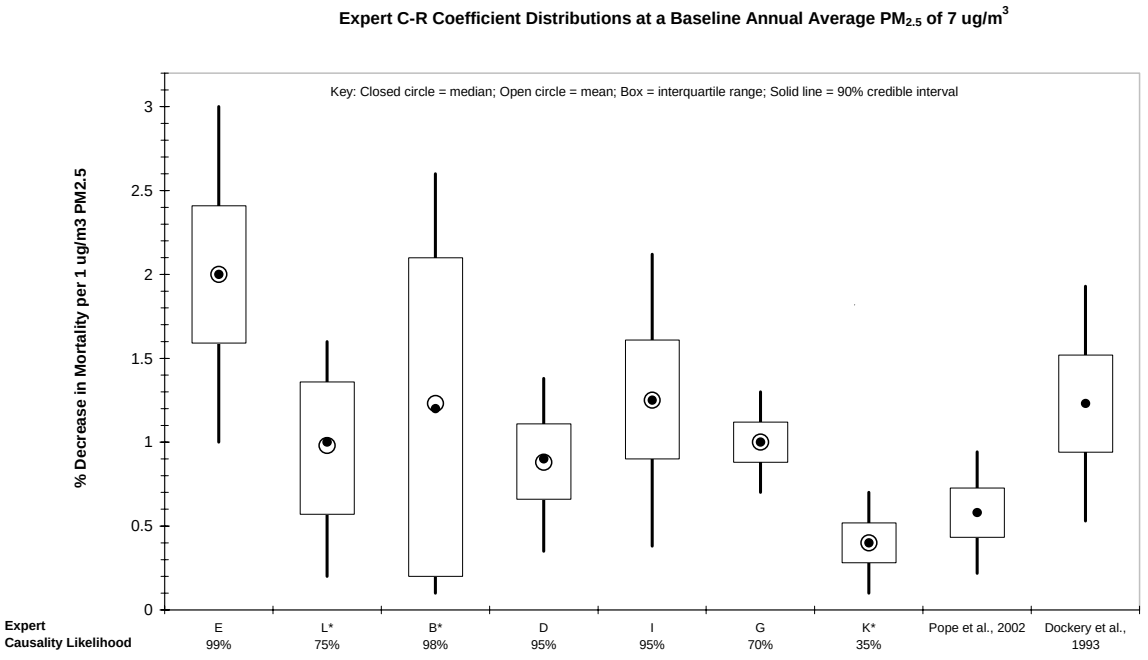
Exhibits 3-10 and 3-11 provide comparisons of the experts' distributions in the form of boxplots. The "whiskers" delimit the lower 5th and upper 95th percentiles. The box represents the interquartile range. The median is represented by a closed circle and the mean, generated by Crystal Ball™, is represented by the open circle. The two boxplots on the far right give the analogous percentiles for the mortality coefficients from the two epidemiological studies often used in benefits analysis to characterize the $PM_{2.5}$-mortality relationship, Pope et al. (2002), and Dockery et al. (1993). The vertical axis is in units of percent decrease in mortality per unit decrease in $PM_{2.5}$ and the horizontal axis displays the individual experts.

Exhibit 3-10 compares the distributions for those experts who preferred to give conditional uncertainty distributions and keep their probabilistic judgment about the likelihood of a causal or non-causal relationship separate (Group 1). Exhibit 3-11 displays the distributions for those experts who chose to incorporate the likelihood of a causal relationship directly into their distributions (Group 2). Both Exhibits 3-10 and 3-11 array the experts in order of decreasing elicited likelihood of a causal relationship, with the experts on the left being most convinced the relationship is causal. The likelihood of a causal relationship elicited from each expert is shown below his letter on the horizontal axis. The two sets of graphs for each group provide comparisons of the experts at two annual average $PM_{2.5}$ concentrations, 18 $\mu g/m^3$ and 7 $\mu g/m^3$, in order to observe the implications of particular experts' assumptions about non-linearities in the C-R function and about differing degrees of uncertainty in the slope of the function across specific ranges of PM.

Exhibit 3-11 illustrates how experts in Group 2 incorporated assumptions about the likelihood of a causal relationship into their uncertainty distributions. For example, Expert H, who was 90 percent sure that the relationship between $PM_{2.5}$ and mortality was causal, has a 10 percent chance that there is no causal relationship (and thus that the C-R coefficient is zero). The 5th percentile of Expert H's boxplot therefore extends to zero.

EXHIBIT 3-10: GROUP 1: EXPERT UNCERTAINTY DISTRIBUTIONS FOR $PM_{2.5}$- MORTALITY C-R COEFFICIENTS; CONDITIONAL ON THE EXISTENCE OF A CAUSAL RELATIONSHIP





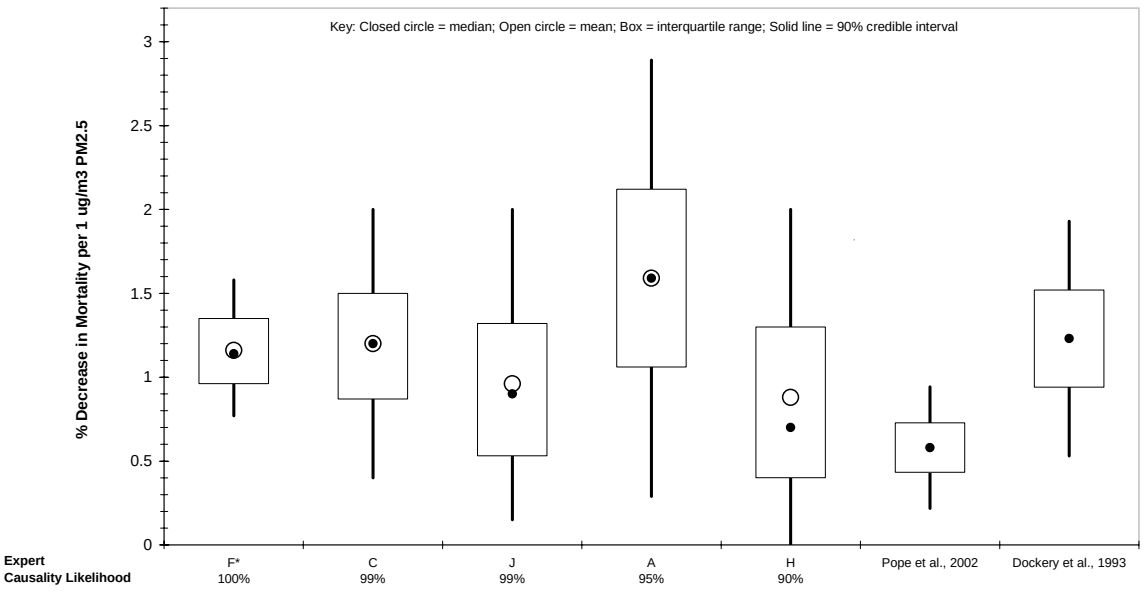
Note: Box plots represent distributions as provided by the experts to the elicitation team. When asked, experts in this group preferred to give conditional distributions and keep their probabilistic judgment about the likelihood of a causal or non-causal relationship separate.

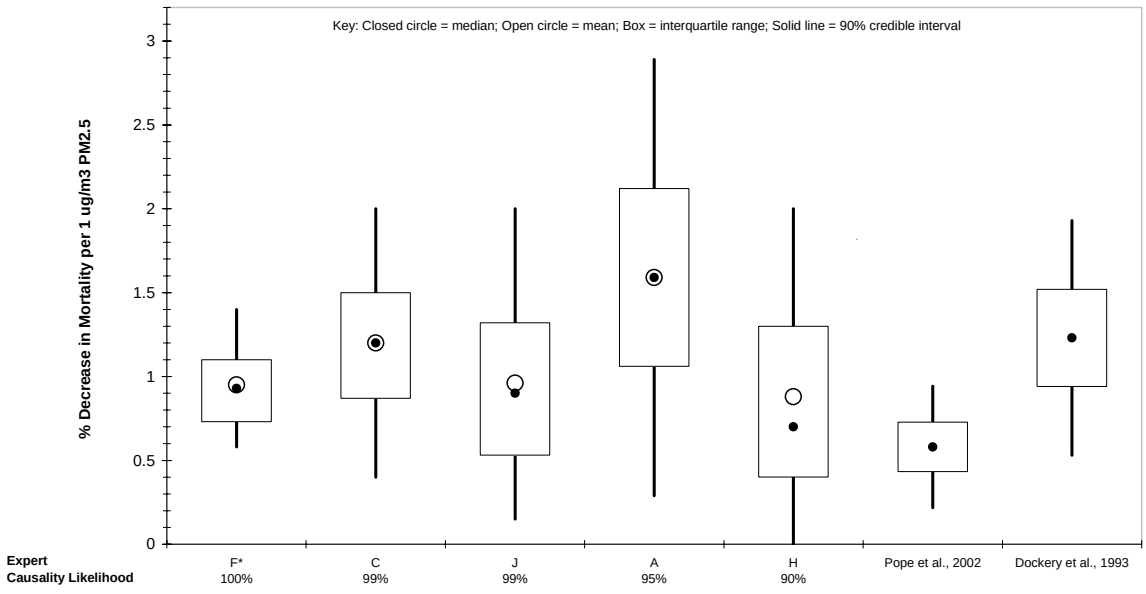
Expert K specified a threshold (not shown).

Expert L provided two different likelihoods of causality for his C-R coefficient distributions at 7 and 18 $\mu g/m^3$, although his distribution appears in the same location in both graphs.

*Experts' C-R coefficient distribution changes between 7 and 18 $\mu g/m^3$.

EXHIBIT 3-11: GROUP 2: EXPERT UNCERTAINTY DISTRIBUTIONS FOR THE $PM_{2.5}$-MORTALITY COEFFICIENT INCORPORATING THE EXPERT'S LIKELIHOOD OF A CAUSAL RELATIONSHIP





Note: Box plots represent distributions as provided by the experts to the Elicitation Team. When asked, experts in this group preferred to give distributions that incorporate their likelihood that the $PM_{2.5}$ mortality association may be non-causal.
* Experts' C-R coefficient distribution changes between 7 and 18 $\mu g/m^3$.

Generally, the results for the experts in each group show substantial variation in the amount of uncertainty expressed, even among individuals expressing similar views on the strength of the causal relationship. For example, in Group 2, Expert F (100% likelihood) predicts a much narrower range of values for the $PM_{2.5}$ mortality coefficient than Experts A or H, whose likelihoods of a causal relationship were 95% and 90% respectively. We see similarity in the variation in Group 1, for example Expert B has a much wider range of values than Expert L, although both have similar causality likelihoods (at 18 $\mu g/m^3$). We do not see dramatic differences in uncertainty between the subjective probability distributions at high versus low concentrations (i.e., the upper and lower panels of Exhibits 3-10 and 3-11).

Exhibit 3-12 displays the box and whiskers plots of the distributions given for each PM range by each of the four experts who favored a non-linear function. The exhibit shows that with the possible exception of Expert K, the changes in the C-R slope and uncertainty expressed by the experts between high and low $PM_{2.5}$ concentrations were very modest:

- Only two experts, F and K, specified a lower median slope estimate for the lower range of their functions. Expert F reduced his median estimate by 18 percent, K by 43 percent.
- Expert B expanded his uncertainty by reducing his minimum, 5^{th} and 25^{th} percentile estimates for the lower range, leaving the rest of his distribution unchanged. The mean of his distribution, which was not directly elicited, moved slightly closer to his median (1.2 percent) as the result of the changes in the lower distribution.
- Expert F specified a distribution for the lower range with slightly lower values for all percentiles elicited. The lower distribution was slightly more skewed towards lower values, but the spread of his two distributions was similar. He thought equally confident in his distributions for both segments, but that the distribution for the lower range would be shifted downwards.[37]
- Expert L's lower and upper box plots appear identical; however, there are two differences in his distributions. First, his minimum value for the lower distribution is zero, as compared to 0.02 for the upper distribution. Secondly, Expert L expressed less confidence in a causal relationship at lower PM concentrations, specifying a 75 percent best estimate likelihood that the relationship is causal in the lower range and a 99 percent best estimate likelihood of a causal relationship in the upper range. The difference in causality could have an impact when Expert L's

[37] He thought that there might be a mechanism operating that would adversely affect health, but believed that at lower levels of exposure that mechanism would operate more slowly, so that some exposed people would be likely to die from another cause first.

judgments are applied in a benefits analysis context where judgments about causality are factored in along with the mortality impact distributions.[38]

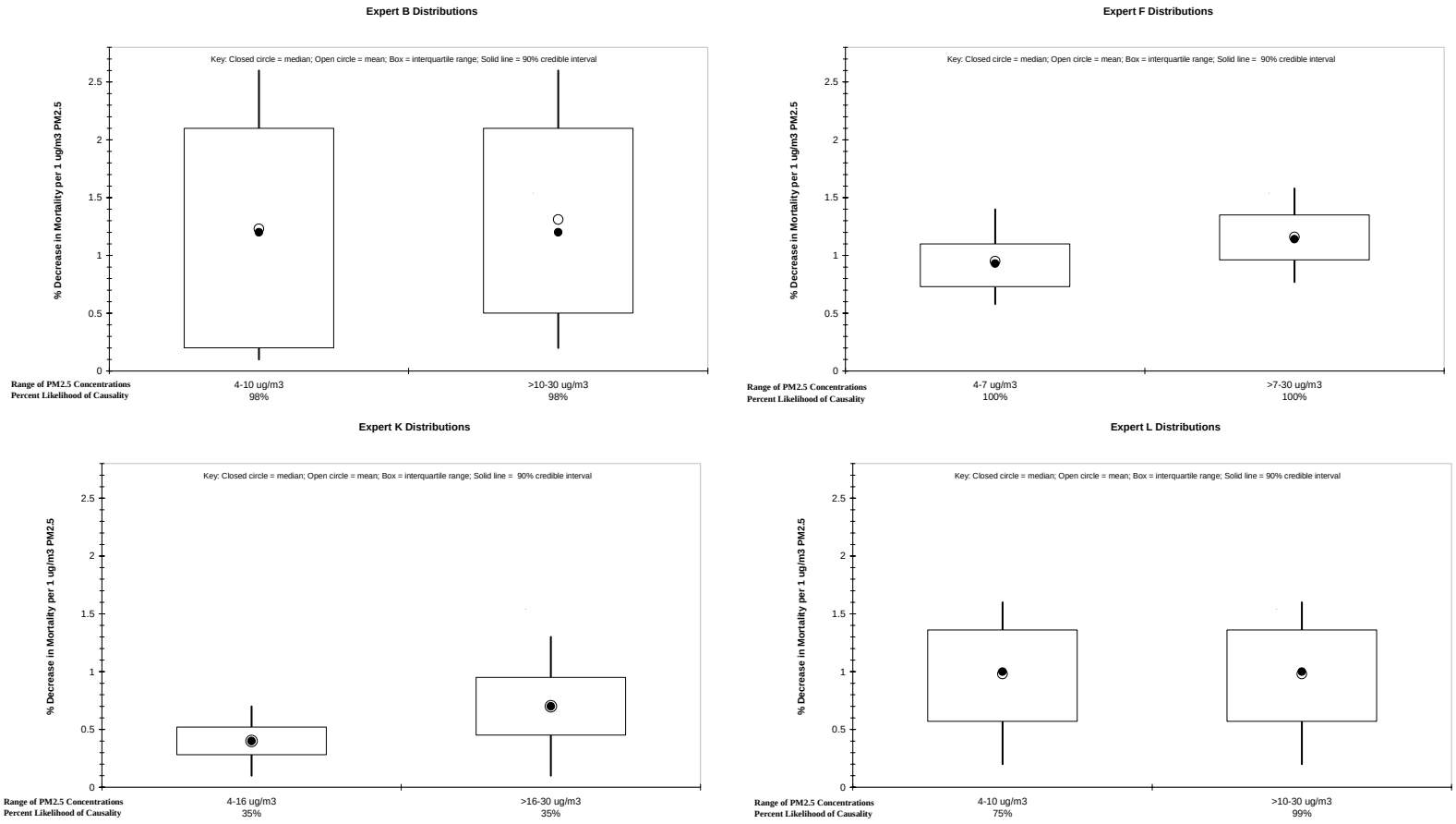
- Expert K's function displays the most significant differences between his upper and lower distributions in values for the C-R slope. Also unlike the other three experts in this exhibit, the spread of his lower distribution is much narrower ("more informative" in expert elicitation terms) than his upper distribution, indicating greater confidence in his lower range estimate than his upper range.

As discussed in Section 3.1.8, Expert K thought that a population threshold could exist at lower PM concentrations, although he was uncertain about where it might fall.[39] Exhibit 3-13 shows the difference in the subjective probability distribution for Expert K with and without his assumptions about a threshold for an example application to a sample of $PM_{2.5}$ annual average concentrations. As one might expect, the impact of incorporating a threshold is to shift the distribution downward to some degree. Ultimately, the implications for estimated mortality benefits of Expert K's views on thresholds as well as of the other experts' judgments about the non-linearities in the C-R function at lower concentrations are best explored as part of a sensitivity analysis. We report on such an analysis in the next chapter. Only with application to a common scenario can all of the experts' quantitative judgments be directly compared.

[38] Assuming Expert L's causality distribution and mortality impact distribution are independent, incorporating the causality judgment as part of the evaluation of a regulatory change in PM would result in lesser benefits being estimated, on average, for individuals whose annual mean $PM_{2.5}$ concentration before the change is 10 ug/m^3 or less. The overall extent of this impact would depend on what proportion of the population in the benefits analysis experiences baseline $PM_{2.5}$ annual average concentrations of 10 ug/m^3 or less.

[39] He indicated that he was 50 percent sure that a threshold existed. If there were a threshold, he thought that there was an 80 percent chance that it would be less than or equal to 5 $\mu g/m^3$, and a 20 percent chance that it would fall between 5 and 10 $\mu g/m^3$.

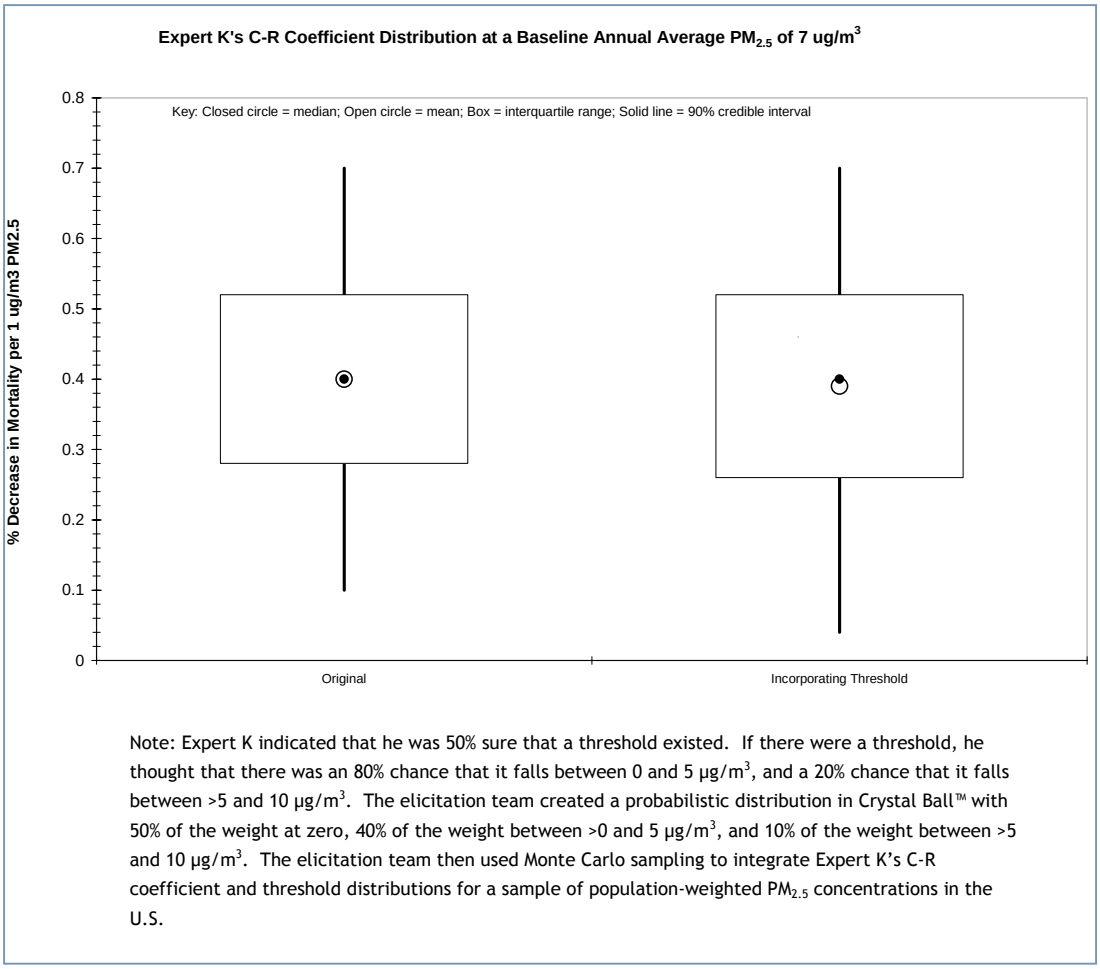
EXHIBIT 3-12: C-R COEFFICIENT DISTRIBUTIONS FOR EXPERTS WHO SPECIFIED NON-LINEAR FUNCTIONS



Note: Box plots represent distributions as provided by the experts to the elicitation team. When asked, Experts B, K, and L preferred to give conditional distributions and keep their probabilistic judgment about the likelihood of a causal or non-causal relationship separate, whereas Expert F preferred to give a distribution that incorporated his likelihood that the $PM_{2.5}$ mortality association may be non-causal.

Expert K specified a threshold (not shown).

EXHIBIT 3-13: THE EFFECT OF INCORPORATING A THRESHOLD INTO THE C-R COEFFICIENT UNCERTAINTY DISTRIBUTION OF EXPERT K



Note: Expert K indicated that he was 50% sure that a threshold existed. If there were a threshold, he thought that there was an 80% chance that it falls between 0 and 5 μg/m^3, and a 20% chance that it falls between >5 and 10 μg/m^3. The elicitation team created a probabilistic distribution in Crystal Ball™ with 50% of the weight at zero, 40% of the weight between >0 and 5 μg/m^3, and 10% of the weight between >5 and 10 μg/m^3. The elicitation team then used Monte Carlo sampling to integrate Expert K's C-R coefficient and threshold distributions for a sample of population-weighted $PM_{2.5}$ concentrations in the U.S.

A visual comparison of the experts' distributions to those from the two primary cohort studies, Pope et al. (2002) and Dockery et al. (1993) suggests that most of the subjective distributions encompass the 90% confidence intervals from both those studies. The medians and/or estimated means fall at or between the median for the ACS study (0.6% per μg/m^3) and the median for the original Six Cities study (1.3% per μg/m^3) for most of the experts. Experts A and E centered their distributions above the original Six Cities results although both their distributions place some probability weight on the ACS results as well.

Exhibit 3-14 helps explain some of the rationales behind both the similarities and the differences. It summarizes the studies that the experts specifically cited in connection with the development of either their median or with the percentiles characterizing uncertainty (minimum, 5th, 25th, 75th, and 95th, and maximum). The closed circle [⊙] indicates those studies from which the experts drew particular effect estimates or which they used to support adjustments to the primary effect estimate on which they relied for the median. The open circle [○] indicates studies mentioned in association with one or more of the percentiles used to characterize uncertainty. More detail on how each expert chose to use data from particular studies may be found in their individual summaries in the Technical Support Document for this study (IEc, under development).

Exhibit 3-14 shows a strong reliance by most experts on the Pope et al. (2002) ACS study, the Jerrett et al. (2005) re-analysis of the ACS cohort in the LA basin, the Dockery et al. (1993) original Six Cities study and finally, the Laden et al. (2006) cross-sectional analysis based on the Six Cities cohort for their median estimates. However, the particular use of individual studies varied from expert to expert.

For many experts, the original ACS study by Pope et al. (2002) and Six Cities study by Dockery et al. (1993) formed an important initial foundation for their judgments. However, questions about the implications of the non-representativeness of the original ACS cohort with respect to educational attainment level (see Pope and Dockery, 2006) and questions about potential exposure misclassification (due to spatial resolution of exposure) raised by the Jerrett et al. (2005) study played critical roles in how the experts weighed the findings from both studies. Several experts thought that the original ACS study results were likely to have been underestimates of U.S. effects due to under-representation of the poorly educated and/or to exposure misclassification and adjusted their median estimates to reflect the impact of educational attainment, exposure misclassification, or both (see also Exhibits 3-4 and 3-5). Others saw the Jerrett et al. (2005) findings as rationale for placing more confidence in the original Six Cities study and its extended analyses than they previously had or to rely more extensively on the Jerrett et al. (2005) study itself. The effect on the central estimates from these two approaches was essentially to converge on a similar range of values. Median effect estimates were lower when experts argued against adjusting the original estimates for one of these factors (e.g., Expert H did not think it appropriate to adjust for educational attainment; Expert K adjusted for neither).

EXHIBIT 3-14: EPIDEMIOLOGIC STUDIES RELIED UPON BY EXPERTS IN CREATING THEIR C-R COEFFICIENT UNCERTAINTY DISTRIBUTIONS

	ACS (Pope et al., 2002)	ACS LA Reanalysis (Jerrett et al., 2005)	Six Cities (Dockery et al., 1993)	Six Cities (Laden et al., 2006 (Cross-Sectional))	ACS (Pope et al., 1995)	Netherlands Cohort Study (Hoek et al., 2002)	Six Cities (Laden et al., 2006 (Change estimate))	Mallick et al., 2002	Willis et al., 2002	NMMAPS (Samet et al., 2000)	Women's Health Initiative[2]	AHSMOG (Abbey et al., 1991 & 1999; McDonnell et al., 2000; Chen et al., 2005)
Expert A	○	⊙	○	⊙			○			○		
Expert B	⊙	⊙			⊙ ○			⊙				
Expert C	⊙	⊙	⊙									
Expert D	⊙		⊙									
Expert E	⊙	⊙		⊙		⊙	⊙		⊙			
Expert F	○	○	⊙	⊙								
Expert G	⊙ ○		⊙ ○									
Expert H	⊙	⊙ ○	○									
Expert I	⊙	⊙		⊙								
Expert J[1]	⊙	⊙ ○	⊙ ○			O				○	○	○
Expert K	⊙		⊙ ○									
Expert L	○	⊙ ○		○						○		
Total ⊙:	9	8	6	4	1	1	1	1	1	0	0	0
Total ○:	4	4	5	1	1	1	1	0	0	3	1	1

⊙ = Expert used the study to inform the median of his C-R coefficient distribution(s).
○ = Expert used the study to inform the uncertainty of his C-R coefficient distribution(s).
[1] Expert J also cited the following short-term studies as support for his uncertainty: Levy et al., 2000; Steib et al., 2002; Anderson et al., 2005; Ostro et al., 2005; Schwartz et al., 1996, Klemm et al., 2000; Burnett et al., 2003).
[2] Study not yet published at the time of the interview.

Another influential factor was the particular statistical model on which experts chose to rely on from the Jerrett et al. (2005) study. The authors explored several different models that varied with respect to the number and nature of variables included in the model. Experts who found the base model with 44 individual covariates more compelling for their median estimates (e.g., Experts A and E) discussed them in the context of the Laden et al. (2006) study and tended to have higher median estimates. Other experts (e.g., Experts D and L) made arguments for use of the Jerrett et al. (2005) model that included parsimonious contextual covariates in addition to the 44 individual covariates.

The Hoek et al. (2002), Mallick et al. (2002), and Willis et al. (2003) studies were cited occasionally in experts' rationales for median estimates but usually in support of the Jerrett et al. (2005) findings or in support of upward adjustments to the original ACS study to account for possible exposure misclassification.

Short term studies, (e.g., NMMAPS (Samet et al., 2000a & b) were cited relatively rarely in the development of experts distributions, even among those experts who generally found the data base for the effects of short term exposures more compelling. They were most often raised to help define the lower limits of the distributions, minimum to 25th percentiles. This use of the studies appears consistent with the views of those experts who argued that mortality from short-term exposures (i.e., those deaths caused primarily by an acute exposure) were likely to make up a small proportion of total deaths from $PM_{2.5}$ exposure.

3.3. POST-ELICITATION WORKSHOP

The purpose of this workshop was to reconvene the panel of experts to present preliminary results of the study and hold discussions on the elicitation process. The workshop consisted of a briefing describing preliminary results followed by several discussion sessions. The discussion sessions were intended to explore topics where expert opinions varied significantly; discuss new scientific evidence that emerged during the course of conducting the 12 interviews; and encourage each expert to critically review his judgments. The workshop was not intended to promote consensus with in the group, but rather to share ideas among experts and discuss evidence for or against such ideas. The workshop also served as an opportunity for experts to provide EPA with feedback on the expert elicitation process.

Topics discussed during the workshop included quantification of effects by PM level, the role of intervention studies in quantifying impacts, exposure error due to the use of central site monitors, effect modification due to educational attainment, and the integration of each expert's judgments on the likelihood of a causal relationship into his overall distribution. There was also an open session at the end where experts could discuss other topics and provide feedback to EPA and the Project Team.

At the end of the workshop, the experts were given instructions on the process by which experts could update their judgments. It required completing a modification form in which the expert was asked to specify the changes he wished to make and provide detailed rationale explaining and supporting the changes.

Although the workshop gave the opportunity to discuss a number of topics where cogent, but conflicting arguments had been made, most of the discussion focused on three topics: exposure misclassification, effect modification, and causality.

3.3.1 EXPOSURE MISCLASSIFICATION

The purpose of this session was to discuss the issue of adjusting effect estimates for exposure misclassification due to the use of central site monitor exposure estimates. This topic was raised primarily because of the variance in opinions expressed by the experts as to the importance of this issue, and the LA ACS results, for developing C-R coefficient distributions. Another reason for discussing it was that during the time the elicitation interviews were being conducted, preliminary results from an analysis by Jerrett et al. of ACS data in New York (NY ACS) were presented at the annual Health Effects Institute (HEI) meeting in April. Like the LA ACS study, the NY ACS study used modeling in an attempt to improve exposure characterization for the cohort. Some of the experts who were interviewed subsequent to the HEI meeting discussed the preliminary NY ACS results, which did not generally corroborate the findings of the earlier analysis of the LA ACS data, when reflecting on the overall evidence for the role of improved spatial resolution. Thus, the Project Team also raised the issue in order to give all experts an opportunity to discuss and critique the newer findings.

The topic sparked a lively discussion that focused less on the Jerrett findings in NY and more about when it is appropriate to cut off the introduction of new information to an elicitation process in general, and what criteria should be used to decide what studies are introduced. The concern expressed by some experts was that new information is always becoming available and that allowing individual studies to be considered on an ad hoc basis could reflect bias. The project team generally agreed that while it made sense to allow experts to use whatever knowledge they have available to them, the project should have some formal cut-off for the introduction of new information.

Experts did discuss some possible reasons for the different results in the NY study, including the smaller range of PM levels over a smaller geographic area, and about uncertainty in the exposure surface model applied to NYC with its street canyons. Ultimately, however, most experts simply thought that the results are too new and too preliminary to give them much weight.

None of the experts chose to alter their quantitative judgments with respect to this topic.

3.3.2 EFFECT MODIFICATION

The purpose of this session was to discuss the issue of adjusting ACS cohort effect estimates for effect modification due to educational attainment, as demonstrated in the HEI reanalysis (Krewski et al., 2000). A number of experts adjusted their effect estimates upward to account for this effect, or relied on studies (e.g., Six Cities) that the experts thought were less affected by it. The discussion centered around whether or not the adjustment of the ACS estimate was reasonable, given the observation during the interviews that the focus of the current elicitation was on the U.S. population today and the fact that the percentage of high-school educated people in the U.S. today is more similar to that in the ACS cohort than in the Six Cities cohort.

One expert argued that re-weighting for education is helpful to see if it explains differences between the Six Cities and ACS study results. Because he believes the education factor really is a proxy for other factors related to socioeconomic status, he was not convinced that changes in education over time would necessarily affect the true factors underlying the difference attributed to education. Others agreed with this interpretation, one calling it an issue of normalizing the results of the ACS and Six Cities studies. One expert added that the HEI reanalysis found effect modification by education in both the original ACS and Six Cities studies (Krewski et al., 2000).

On the subject of the relevance of current educational attainment rates, one expert suggested that a high school degree implied more about education and lifestyle/SES in the past than it does today. Another expert noted that unpublished results from the Women's Health Initiative, a study involving a modern, highly educated cohort, showed some of the largest effect estimates observed to date and no effect modification by education.

One expert, Expert I, made changes to his uncertainty distribution, partially in response to this discussion. He modified his approach (described previously) to include a specific upward adjustment of the 2002 ACS results for educational attainment and revised the subjective weights he had assigned to the three studies he had chosen to combine (adjusted 2002 ACS, LA ACS, and the Six Cities Follow-up) to weight each equally.

3.3.3 CAUSALITY

The most extensive discussion during the workshop focused on the question of how judgments about causal likelihoods were incorporated into distributions characterizing uncertainty in the $PM_{2.5}$-mortality coefficient. As discussed in section 3.2.2, the protocol gave each expert the choice of 1) incorporating his views directly into his uncertainty distribution; or 2) first developing his distribution conditional on the assumption that the relationship was causal, and then having the elicitation team combine that distribution probabilistically with his previously elicited likelihood of a causal relationship. During each elicitation interview, experts who chose the second option (the "disaggregated approach") were shown box plots of both their distribution "conditional on causality" and a joint distribution that incorporated both their judgments about the C-R coefficient and about causality.[40] The joint distribution was presented to help these experts assess the potential impact of their beliefs on causality on their distribution of C-R coefficients. During the Post-elicitation Workshop, these joint distributions were presented along side distributions from experts who chose the first option above (the "aggregated approach"), in order to facilitate comparison across experts.

As discussed in Section 3.1.7, the approach to causality, although a small component of the protocol, engendered an extensive discussion during the workshop. Comments raised during the discussion of causality included the following:

[40] The joint distribution was generated using the spreadsheet tools described in Chapter 2 and Crystal Ball, assuming the two distributions were independent.

- One expert suggested that the quantitative estimates of causality are overly precise and recommended that the experts instead be assigned to qualitative categories based on their responses.
- One expert said the causality question was confusing and poorly defined.
- Others thought that it was unclear during the interview how the causality results were to be combined with their distributions if they were first developed first conditional on a causal relationship. Although this approach had been laid out in the Pre-elicitation Workshop and in the briefing book, not all experts had fully considered the implications of the approach.
- One expert said it was difficult to completely eliminate the concept of causal likelihood when characterizing an uncertainty distribution conditional on the existence of a causal relationship, and therefore, that combining their distribution with causality was potentially double counting. Another expert suggested the probabilities might not be independent and if so, cannot be multiplied.
- One expert argued that his uncertainty distribution was derived from his interpretation of the literature and that it was conceptually distinct from his evaluation of the likelihood of a causal relationship. For example, he noted that his views on the findings of a particular well-designed epidemiological study might not necessarily change over time; however, independent toxicology, mechanisms, clinical studies might change his views of the likelihood that the relationships observed in the epidemiological study are real. He therefore preferred that these concepts be kept separate in the final report.

Overall, the discussion highlighted the difficulties of eliciting conditional distributions from the experts as well as a general unease among the experts with applying the concept of expected value in this context to their mortality effect estimates. Because several expressed discomfort with presenting the joint distributions, we have not included them in this report.

Experts also asked to see a revised causality question. The Project Team agreed to review the causality question following the workshop, but after further review concluded the question was sufficiently clear.

Of the eight experts who originally specified a C-R coefficient distribution conditional on causality, four specified changes to their quantitative judgments following the workshop discussion:[41]

- Expert A changed the interpretation of his distribution from conditional on a causal relationship to NOT conditional, because of the concern over double-counting described above;
- Expert E changed his best estimate of the likelihood of a causal relationship from 95 percent to 99 percent. In explaining the rationale for his change, Expert E

[41] Note that Exhibits 3-9 through 3-14 incorporate the changes made by expert following the post-elicitation workshop.

referred to a statement made by another expert during the workshop that the assigning of some probability of no effect was really an assigning of a finite probability that there was not even a very small effect. After considering his previously cited evidence in light of this statement, he changed his previous assumption that there was a five percent probability of zero effect to there being a one percent probability of there being zero effect;

- Expert L expressed different causal likelihoods for the two segments of his non-linear C-R function. He lowered his best estimate of the likelihood of a causal relationship for the lower PM range (4-10 $\mu g/m^3$) from 99 percent to 75 percent, keeping his estimate for the higher range unchanged at 99 percent. Expert L cited the inherent difficulty of assessing effects at annual average $PM_{2.5}$ levels below 10 given current data. He changed his likelihood of a causal relationship responses for the lower segment of his function to a range of zero to 95 percent with a best estimate of 75 percent;
- Expert K expanded his range of causality estimates from 20 to 50 percent to 5 to 50 percent, but left his best estimate of causality unchanged at 35 percent. Expert K cited concerns about exposure error and unaccounted for potential confounders in extending his lower bound on the likelihood of a causal relationship.

CHAPTER 4 | DISCUSSION

In response to recommendations made in the 2002 National Academy of Sciences (NAS) report, "Estimating the Public Health Benefits of Proposed Air Pollution Regulations," EPA has been exploring ways to improve the characterization of uncertainty in its analyses of the health benefits of proposed or existing regulations affecting air quality. EPA has now conducted two projects, a pilot study (IEc, 2004) and a larger-scale study that has been the focus of this report. Both studies sought to provide a more complete characterization, in qualitative and quantitative terms, of the uncertainties associated with the relationship between reductions in ambient $PM_{2.5}$ and mortality, using judgments elicited from scientific experts.

The previous chapter summarized the results of the expanded study, which were generated though detailed interviews with 12 experts. The experts were selected via a peer nomination process intended to yield a set of highly qualified experts from key scientific disciplines who have extensively studied the health effects of $PM_{2.5}$ exposures. The peer nomination process led to selection of eight epidemiologists, one clinician/scientist, and three toxicologists. Experts were invited to attend a Pre-elicitation Workshop to discuss evidence relevant to the project and review the protocol in detail; to participate in a one-day elicitation session; and to attend a final Post-elicitation Workshop to review results and discuss remaining questions.

As discussed in detail in the preceding chapter, the experts relied on a broad array of evidence from varied scientific disciplines to develop their judgments about the likelihood and magnitude of a relationship between $PM_{2.5}$ and mortality. The epidemiological studies based on the Six Cities and ACS cohorts were central to many discussions, but their results were critiqued in the context of newer evidence from more recent epidemiological studies in the U.S. and abroad, as well as toxicological and clinical studies. The various "intervention" studies, the Jerrett et al. (2005) re-analyses of the ACS cohort in Los Angeles, and the Sun et al. (2005) studies were an important focus of many discussions. Overall, the impact of these more recent studies led to larger mortality effect estimates and greater certainty about a causal relationship than was observed in the pilot.

The remainder of this chapter provides additional perspective on these results. We first present findings from analyses conducted to evaluate the sensitivity of the quantitative judgments to various factors. We next compare the results of this study to those of the pilot study. We conclude with an evaluation of the strengths and limitations of this expert judgment study.

4.1 SENSITIVITY ANALYSIS

IEc conducted a simplified benefits analysis to assess the sensitivity of the results generated to the responses of individual experts and to three factors in the study design: the use of parametric or non-parametric approaches by experts to characterize their uncertainty in the $PM_{2.5}$-mortality coefficient, participation in the pre-elicitation workshop, and allowing experts to change their judgments after the post-elicitation workshop.

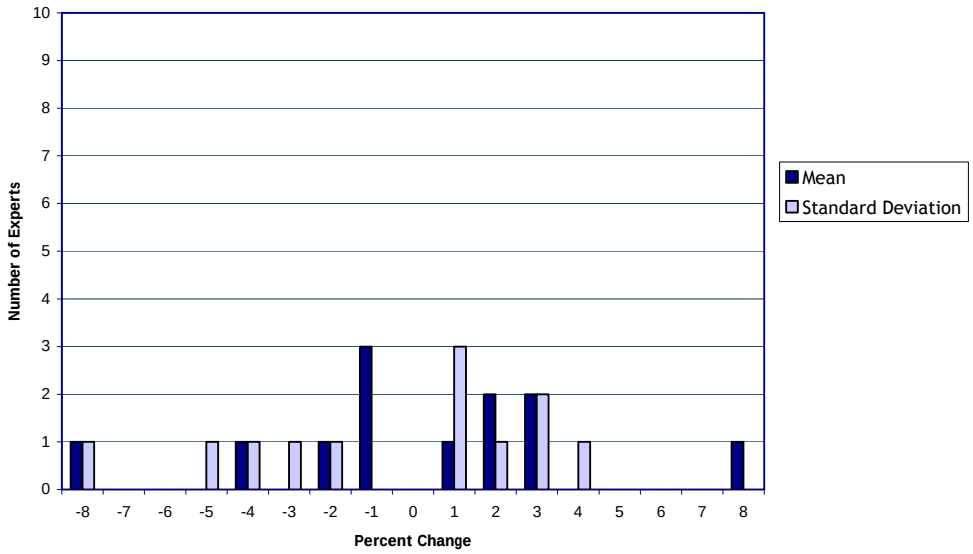
The details of the simplified example benefits analysis can be found in Appendix C. In brief, the individual quantitative expert judgments were used to estimate a distribution of benefits, in the form of number of deaths avoided, associated with a reduction in ambient, annual average $PM_{2.5}$ concentrations from 12 to 11 $\mu g/m^3$. The 12 individual distributions of estimated avoided deaths were then pooled using equal weights to create a single overall distribution reflecting input from each expert. This distribution served as the baseline for the sensitivity analysis, which compared the means and standard deviations of the baseline distribution with several variants.

4.1.1 SENSITIVITY TO INDIVIDUAL EXPERTS

The first analysis examined sensitivity of the mean and standard deviation of the overall mortality distribution to the removal of individual expert's distributions. Individual expert distributions were sequentially removed from the combined distribution, the mean and standard deviation were re-calculated, and the percent change from the full set of experts calculated.

The results of the analysis are shown in Exhibit 4-1 in the form of a histogram of the percent changes in mean and standard deviation resulting from the removal of individual experts. The histogram shows that the mean of the combined distribution shifted less than +/- 4 percent for 10 of the 12 experts and by +/- 2 percent for seven of the 12 experts. Two experts were more influential, with removal of Expert E causing the distribution mean to shift downward by eight percent and the removal of Expert K causing the distribution to shift upward by eight percent. In general, the results suggest a fairly equal split between those experts whose removal shifts the distribution mean up and those who shift it down, and relatively modest impacts of individual experts.

EXHIBIT 4-1: PERCENT CHANGE IN THE MEAN AND STANDARD DEVIATION OF AN EXAMPLE POOLED MORTALITY BENEFIT ESTIMATE BASED ON RESULTS OF THE EXPERT ELICITATIONS, AFTER REMOVING EACH EXPERT



Note: Mortality benefits were pooled of a 1 μg/m^3 decrease in $PM_{2.5}$, beginning at an annual average concentration of 12 μg/m^3. The pooled estimate weighted each expert equally.

The standard deviation of the combined distribution was also not strongly affected by removal of individual experts; Experts E and K again had the greatest influence on the spread of results both causing the overall standard deviation to decline by five percent (K) and eight percent (E) when removed. Overall, these results indicate that no one expert dominates the results.

4.1.2 SENSITIVITY TO CHOICE OF DISTRIBUTIONAL FORM TO CHARACTERIZE UNCERTAINTY

The protocol gave experts some latitude in how to estimate the value of the $PM_{2.5}$ mortality coefficient for different percentiles. They could estimate individual percentiles directly or they could fit parametric distributions either by estimating the parameters of the distribution (e.g., mean and standard deviation) or fitting distributions to individual elicited percentiles (e.g., 50th and 95th percentiles). The elicitation team had some concern that experts fitting parametric distributions might ultimately be more overconfident. While we have no way of knowing ultimately which judgments are closer to the "truth," we can observe how the distributions derived using the two approaches differ.

The sensitivity analysis suggests that the standard deviation of the pooled distribution of experts who specified parametric distributions in fact is slightly larger than that of the pooled distribution of experts from whom all the individual percentiles were elicited (i.e.,

those specifying "custom" or non-parametric distributions). The standard deviation of the pooled estimate for the parametric group was 23 percent greater than that for the non-parametric group. (The means of the two distributions remained close to that of the overall mean for all experts.) It is worth noting, however, that the two most influential experts in the sensitivity analysis above, E and K, are both in the parametric group. When Experts E and K are removed from the parametric group, the difference in the spread of the distributions for the parametric and non-parametric groups falls to less than 3 percent. Thus, this analysis suggests that the use of parametric distributions led to distributions with similar or slightly increased uncertainty compared to experts who provided percentiles of a non-parametric distribution.

4.1.3 SENSITIVITY TO PRE-ELICITATION WORKSHOP PARTICIPATION AND TO CHANGES MADE AFTER THE POST-ELICITATION WORKSHOP

Reviewers of the pilot study suggested that more opportunities for communication and sharing of information between experts, including after the individual interviews, would improve the elicitation process. The goal of such communication is to facilitate sharing of new information and to resolve questions that might have arisen during the elicitation process. However, residual concerns about such workshops remain based on early approaches to expert elicitation, most notably Delphi methods, where sharing of individual results was used to help bring experts to consensus.

Participation in the Pre-elicitation Workshop did not appear to have a significant effect on experts' judgments. The mean and standard deviation for a pooled estimate of avoided deaths for those experts who did not attend the Pre-elicitation Workshop were within 10 percent of the mean and standard deviation of the pooled estimate for Pre-elicitation Workshop attendees.

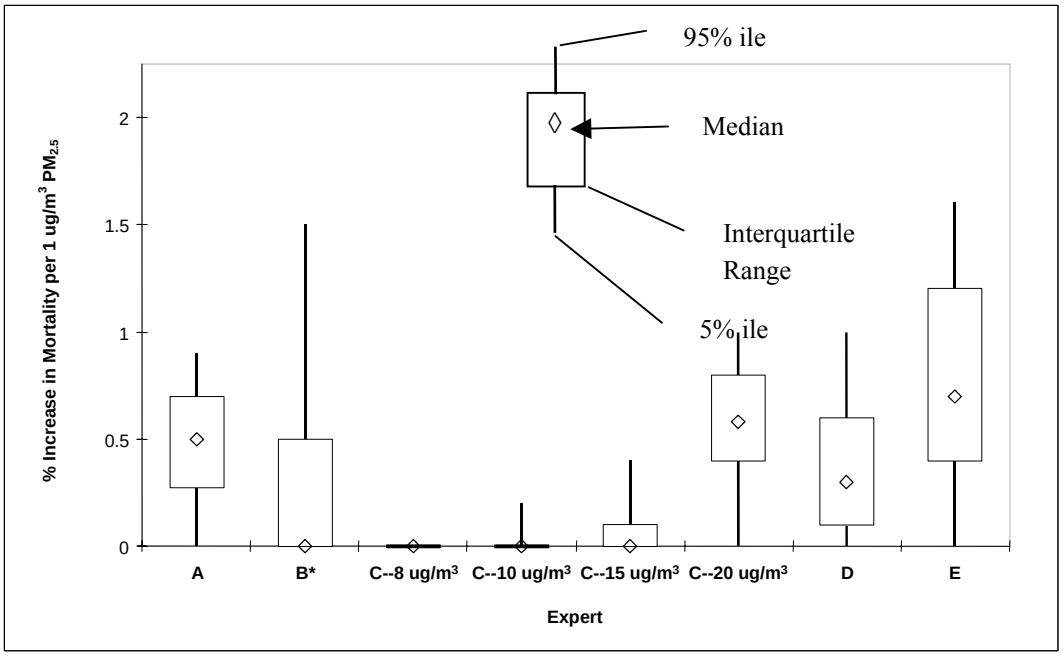
As discussed in earlier sections, we found the Post-elicitation Workshop to be important for discussing some critical points about the design of the study and presentation of results, most notably on the question of causal likelihood. Only four experts chose to make changes to their distributions following our discussions at the Post-elicitation Workshop. Using the same methodology as above, we compared the combined distributions using results prior to and following the Post-elicitation Workshop. We found a minimal impact due to the changes made after the Post-elicitation Workshop. The overall mean number of estimated deaths increased by approximately one percent compared to the estimate prior to the Workshop, and the standard deviation did not change.

4.2 COMPARISON TO THE PILOT STUDY FINDINGS

It is reasonable to ask whether the results of the current study are comparable to those observed in the pilot study (IEc, 2004). The pilot study was conducted in preparation for the current study and thus has many similarities; however a number of changes, both in the basic formulation of the fundamental questions asked, and in the body of available scientific data, make it necessary to be careful about direct quantitative comparisons of individual responses. Some of the differences between the two study designs include:

- The pilot study asked experts to develop separate quantitative distributions for the $PM_{2.5}$-mortality effect for long-term (annual average) and for short-term exposures (24 hour average). The current study asked experts to estimate a <u>total</u> mortality effect integrating effects of long-term and short-term exposures. However, most experts in this study ultimately relied on PM studies that focused on the effects of long-term or intermediate exposures for their quantitative estimates of the total mortality effect, and relied on short-term studies only when estimating possible lower bounds. As a result, the distributions in this study may be most comparable to those elicited for the effects of long-term exposure in the pilot study (See Exhibit 4-2).

EXHIBIT 4-2: PILOT STUDY RESULTS - COMPARISON OF EXPERTS' JUDGMENTS ABOUT THE PERCENT INCREASE IN ANNUAL NON-ACCIDENTAL MORTALITY ASSOCIATED WITH A 1 μg/m³ INCREASE IN ANNUAL AVERAGE EXPOSURES TO $PM_{2.5}$ (U.S. BASELINE 8 TO 20 μg/m³)



*Expert B specified this distribution for the PM-mortality coefficient above an uncertain threshold, which he characterized as ranging between 4 and 15 with a modal value of 12 μg/m³

- The pilot study asked experts to estimate a PM mortality effect assuming an 8-20 μg/m³ concentration range for annual average exposures in the U.S. The current study asked experts to assume a 4-30 μg/m³ concentration range for annual average $PM_{2.5}$. For those experts who assumed linear or log-linear functions, we might not expect the quantitative estimates to change dramatically as a result of expanding the range. Expanding the range on the lower end may have influenced some experts' decision to provide a segmented, non-linear function.

- In the pilot study, the protocol design led experts to incorporate their estimates of the likelihood of a causal relationship directly into their overall distributions. In the current study, the protocol gave experts the option of incorporating their causal likelihoods directly into their distributions or to have them probabilistically combined with their mortality effect distribution conditional on the existence of a causal relationship. Following the post-elicitation workshop, nearly half the experts expressed a preference for displaying their distributions and likelihoods of causality separately. The distributions for these two approaches are therefore presented in separate exhibits, making it difficult to compare across all 12 experts directly and to compare the conditional distributions directly to the distributions in the pilot study.

Despite these differences, some interesting observations can be made about the results of the interviews in the current study (See Exhibits 3-10 and 3-11) relative to the pilot study (Exhibit 4-2).

- While the overall range of median estimates given by the experts in this study is similar to that given by experts in the pilot study, most experts in the current study had higher effect estimates than in the pilot study. In the pilot study, most experts provided median values of 0.5 percent or less for the reduction in mortality per 1 $\mu g/m^3$ reduction in $PM_{2.5}$, and the bulk of most distributions was below 1 percent (only one included a 75^{th} percentile above 1). In this study, most of the experts estimated the median to be near or above 1 percent.
- The experts' distributions in the current study display a similar diversity in the degree of uncertainty expressed as in the pilot study. The distributions range from Expert F whose uncertainty was similar in spread to that of the Pope et al. (2002) study, to Expert A whose uncertainty was 3-4 times broader. However, all of the experts in this study gave distributions that were roughly similar or wider than those published in the ACS and Six City cohort studies.
- In the pilot study, two of five experts incorporated a threshold into their distribution; in the current study only one of 12 experts incorporated a threshold.
- The experts' views on the likelihood of a causal relationship between $PM_{2.5}$ exposure and mortality also appeared to show some marked changes from the pilot study. (Note that the experts in the current study were asked about the likelihood of a relationship between annual average exposures, including the influence short-term or peak exposures, and all-cause mortality, whereas the pilot study asked the experts about long-term and short-term exposures separately.) In general, both the best estimate likelihoods and the likelihood ranges were higher in the current study than in the pilot study. In the current study, two of 12 experts expressed some doubt about the strength of the causal relationship, giving overall likelihoods of 35% and 70%. The remainder expressed probabilities of 90% or above. As indicated in Exhibit 4-3, the experts participating in the pilot study expressed much more skepticism, and were less convinced by the evidence for long-term exposures than for short-term exposures.

EXHIBIT 4-3: PILOT STUDY RESULTS ON THE LIKELIHOOD OF A CAUSAL RELATIONSHIP BETWEEN $PM_{2.5}$ AND NON-ACCIDENTAL PREMATURE MORTALITY

EXPERT	EFFECTS OF LONG-TERM $PM_{2.5}$ EXPOSURE		EFFECTS OF SHORT-TERM $PM_{2.5}$ EXPOSURE	
	LIKELIHOOD (CATEGORICAL)	LIKELIHOOD RANGE[A] (BEST ESTIMATE) %	LIKELIHOOD (CATEGORICAL)	LIKELIHOOD RANGE[A] (BEST ESTIMATE) %
A	Highly likely	85-90 (88)	Highly likely	90-95 (93)
B	Somewhat unlikely	40-50	Somewhat likely	65-80
C	Somewhat unlikely	50[b]	Somewhat unlikely to somewhat likely	50 [b]
D	Somewhat likely	50	Somewhat to highly likely	80-90
E	Likely	80-98 (95)	Likely	80-98 (95)

a. Represents minimum to maximum for categorical likelihood specified. Each expert specified his own quantitative probability estimates to match his qualitative categorical description.

b. Expert C wanted to answer this question separately for different parts of the range. The 50 percent value represents his "average" for the range, with little or no probability of a causal relationship at the low end of the range and a high probability at the upper end of the range.

The differences in results compared with the pilot appear to reflect the influence of new research on the interpretation of the key epidemiological studies that were the focus of both elicitation studies, more than the influence of changes to the structure of the protocol. They may also reflect differences in the composition of the expert panel.

The higher central tendency appears to have been influenced in particular by research that has been published since the pilot study was conducted. More of the experts in the current study made explicit adjustments to their estimates to account for effect modification and exposure misclassification in the ACS cohort studies. While effect modification by educational attainment was raised by several experts in the pilot study as a result of the Krewski et al. (2000) reanalysis, it was viewed there more as a source of uncertainty rather than a rationale for adjustment of study findings. The Pope and Dockery (2006) evaluation of the impact of educational attainment, discussed at the pre-elicitation workshop and by many of the experts during their interviews, appeared to convince several experts that the Pope et al. (2002) mortality effect estimates were likely to be biased low to some degree relative to the effect estimate that might have been observed in a population more representative of the U.S. population. The Jerrett et al. (2005) study, which many experts believed to demonstrate the effect of improved spatial resolution of exposure in reducing exposure misclassification, influenced several experts to think that the Pope et al. (2002) estimates were likely to be biased low and that the higher effect estimates in the Dockery et al. (1993) study might reflect in part better exposure measurement. Results from the various intervention and exposures change studies (Clancy et al., 2002; Laden et al., 2006), which showed higher mortality effects than Pope et al. (2002), were also cited as support for experts' subjective judgments about the magnitude of potential mortality reductions, although experts varied in their views about what types of exposures (long- or short-term) may have contributed to these changes in mortality. The impact of recent literature illustrates how expert views can

evolve in the face of new information and suggests that there may be benefits to following up with experts in the future as the scientific knowledge of this issue continues to advance.

It also appears that experts' increased overall confidence in a causal effect of $PM_{2.5}$ on mortality was driven by the recent contributions to the body of scientific evidence. These were discussed in greater detail in Chapter 3 of the report, but some examples include the Clancy et al. (2002) and Laden et al. (2006) studies which several experts cited as specifically relevant to this elicitation because they showed declines in mortality following declines in particle air pollution. The long-term animal study of Sun et al. (2005) was considered by several experts to provide important support for a causal link between long-term $PM_{2.5}$ exposures and the development of atherosclerosis and increased risk for cardiovascular related mortality. As discussed above, several experts also believed the Jerrett et al. (2005) study lent support not just to the ACS findings but also to the Six Cities study, which had received less weight in the pilot study. The impact of the differences between the current and pilot study in the strength of the causal relationship are not so evident in the displays of the distributions themselves. It will become more evident in benefits analyses where the causal likelihoods will factor directly into the calculations of numbers of deaths.

4.3 EVALUATION OF THE EXPERT ELICITATION PROCESS

We begin our evaluation of the expert elicitation process with a summary discussion of the strengths and potential concerns arising from the particular design of this study. The discussion touches on some of the differences between the design of the pilot study and the current study.

4.3.1 STRENGTHS

Below we cite strengths of the process in the areas of the expert panel size, expert communication, and elicitation protocol and interview process.

Expert Panel Size

This study involved a larger number of experts than the pilot study. The purpose of the larger panel was primarily to achieve broad representation of prevailing scientific opinions about $PM_{2.5}$ and mortality. The availability of a larger panel also permitted some observations about the marginal contribution of additional experts, particularly when they share highly similar views or rely on the same data.

Our study tends to corroborate the work of others (e.g., Clemen et al., 1985) that show when judgments are highly correlated, the marginal influence any individual on the predictive value of the group of judgments is small. As discussed in previous sections, all of our experts relied on many of the same core studies, though their evaluations of those studies were colored by differing experience, perspectives, and knowledge of additional research. We therefore expect that many of these judgments are not independent. We do not have means of calibrating these judgments, individually or collectively, but our sensitivity analysis suggests that none of our individual experts, even those with more

independent views, had a strong influence on the outcome of the simple mortality estimates. This result suggests that adding more experts would add relatively little information. Recall that a broad representation of viewpoints in this context does not mean a statistically representative sample but rather a range of well-reasoned opinions.

Expert Communication

The current study allowed for a substantial degree of inter-expert communication. The EPA Symposium in April 2005 gave a wide variety of interested scientists and policy makers the opportunity to hear about the project and to comment on the proposed protocol. Several changes were made as a result, including the decision to drop separate questions about the role of $PM_{2.5}$ components, and to offer a disaggregated approach to incorporating causal likelihood into the uncertainty distributions.

The workshops held with the experts prior to and following the individual interviews were an important element of the process. The Pre-elicitation Workshop was used to improve the level of preparedness for the interviews by increasing understanding of project objectives, clarifying questions and assumptions in the protocol, sharing and critiquing new data or analysis. The Post-elicitation Workshop was very valuable for re-visiting points of confusion or concern about the protocol, in particular regarding presentation of the causal likelihood information, as well as for providing feedback to the project team. Unlike some elicitation processes where these sessions may be used to bring scientists with divergent views "into the fold," these workshops were explicit that consensus was not the goal. We did see that information presented and discussed at the pre-elicitation workshop was influential in interviews and that clarifications in the post-elicitation workshop led four experts to change their distributions slightly. The sensitivity analysis suggests that these modifications did not lead to substantial changes or movement towards a consensus view. We found that individual expert's views were often deeply held and resistant to challenge.

Elicitation Protocol and Interview Process

Newly-developed, computerized elicitation tools facilitated the development of quantitative distributions in the current study. In response to the varied approaches to characterizing the $PM_{2.5}$-mortality relationship in the pilot study, in particular the specification of thresholds and other non-linear functions, we developed spreadsheets to enable the experts see in "real-time" the implications of some of their stated assumptions or opinions. They allowed the experts the freedom to characterize their uncertainties in the $PM_{2.5}$-mortality effect using a variety of distributional forms, parametric and non-parametric. They allowed for reality checks on the mortality implications of any individual estimate. While we had some concern that parametric approaches might lead to greater reliance on the "anchoring and adjustment" heuristic, the sensitivity analysis suggest that is not the case. The distributions given by the experts who assumed parametric distributions were collectively broader than the non-parametric distributions.

Compared to the pilot study, the protocol for the current study used a more systematic approach to discussing some of key sources of uncertainty in each epidemiologic study the expert relied upon for his quantitative estimates, such as confounding, effect

modification, and exposure measurement. Though some experts were more comfortable with identifying and discussing individual factors than others, the process did give the elicitation team a clearer record with which to revisit specific assumptions during the quantitative step of the protocol.

The protocol for the current study was designed with a focus on a question of particular relevance to regulatory analysis, the impact of annual average exposures (including the long and short-term exposures) on all cause mortality. The experts were faced with developing one uncertainty distribution rather than two (effects of long-term and effects of short-term) as in the pilot study. This approach did add some complexity to the task since experts essentially had to integrate their views on the influence of long-term and short-term exposures into their final responses.

4.3.2 LIMITATIONS

Below we cite potential limitations of the process in the areas of expert selection, protocol development, eliciting probabilities, and generalizability.

Expert Selection

Some experts expressed concerns about the breadth of expertise (as indicated by professional training) and institutional affiliations represented by the current panel. We are aware that the initial expert selection process followed in this study was not as successful at identifying individuals with as broad a range of expertise and institutions as anticipated. The expert selection process did specify a number of criteria for peer nominators to consider that were intended to foster nomination of individuals with different backgrounds. However, it appeared that in some cases nominators overlooked these criteria in favor of choosing names that were most familiar. The additional steps taken with the assistance of HEI to augment the numbers of experts with backgrounds in fields outside epidemiology were essential. In the end, the expert selection process did succeed at recruiting 12 individuals representing key disciplines whose work has been widely-cited and respected in the PM-mortality field.

We do not believe that having multiple experts from the same institution is necessarily detrimental to the study. Merely being colleagues does not necessarily imply that experts share the same views. We do recognize that many of the experts have collaborated on projects together; however, restricting expert selection on this basis would severely limit the pool of qualified experts and restrict our ability to interview top-notch researchers.

The process is labor intensive and requires a substantial commitment by the experts. The level of commitment can and did in some cases impact the willingness of experts to participate in the project, and this can also affect the final composition of the group.

Finally, as we observed in the pilot study, defining "expert" for complex multi-disciplinary questions like those posed in this study is a difficult task. Making judgments about $PM_{2.5}$ mortality requires extensive scientific knowledge spanning epidemiological studies of various designs, in vitro and in vivo toxicological studies, clinical medicine, and exposure analysis, among other fields. Few, if any experts are equally knowledgeable about all these fields. This limitation poses challenges for expert

selection and the design of elicitation protocols, and for interpreting the experts' uncertainty distributions, which reflect these differing degrees of knowledge.

Protocol Development

Assessing the appropriate degree of disaggregation for quantitative questions remains a challenge. As noted above, the current study tried to simplify the process by re-aggregating the quantitative question relative to the pilot but may have made it more challenging to be certain about the elements included in the final response from a few experts. While a disaggregated approach has intuitive appeal, it can sometimes introduce complications by requiring that the exact mathematical relationship between the elements be specified. When experts agree on the basic relationship, it is not a problem. In the case of the discussions on causal likelihoods and how to incorporate them into distributions, we encountered disagreements during the elicitation exercises that were resolved only in the post-elicitation workshop.

Eliciting Probabilities

As in the pilot study, we observed some variability in the ease with which experts could translate their opinions into quantitative probabilities. Expertise in a particular scientific discipline does not always require expertise in probability and, even when it does, subjective probability elicitations can be challenging. Yet, one of the key challenges facing the acceptance of expert subjective judgments is whether experts have successfully represented their knowledge and expertise in the form of distributions. This is an issue that is common to all subjective probability elicitations and one which has been addressed in some studies by incorporating a calibration step to evaluate the expert's judgments relative to known values or by combining the judgments for use in analysis.

Generalizability

It is important to recognize the constraints on the generalizability of the results from this study. As with any study, this one was conducted with a particular objective in mind – to estimate the effect on all-cause mortality of a 1 $\mu g/m^3$ decline in $PM_{2.5}$. It required a number of simplifying assumptions (e.g., regarding baseline conditions in the U.S., the impact of regulation on co-pollutants and components, and the distribution of susceptible individuals in the U.S. population, among many others), as described in the protocol in Appendix A, which should be reviewed when considering the applicability of these results.

CHAPTER 5 | CONCLUSIONS

Several conclusions can be drawn from the results of this expert elicitation project.

- The experts' distributions in the current study display a similar diversity in the degree of uncertainty expressed as in the pilot study. Compared to the pilot study, experts in this study were in general more confident in a causal relationship, less likely to incorporate thresholds, and reported higher mortality effect estimates. The differences in results compared with the pilot study appear to reflect the influence of new research on the interpretation of the key epidemiological studies that were the focus of both elicitation studies, more than the influence of changes to the structure of the protocol. They may also reflect differences in the composition of the expert panel.
- Experts in this study tended to be confident that $PM_{2.5}$ exposure can cause premature death. Ten of twelve experts believed that the likelihood of a causal relationship was 90 percent or higher. The remaining two experts gave causal probabilities of 35 and 70 percent. Recent research in both epidemiology (e.g., Jerrett et al., 2005, Laden et al., 2006) and toxicology (e.g., Sun et al., 2005) significantly contributed to experts' confidence.
- Among the experts who provided distributions that were conditional on the existence of a causal relationship, medians ranged from 0.4 to 2.0 percent (Exhibit 3-10). For five of the seven experts in the group, the effect of integrating causality in a benefits assessment would be small, given the high likelihoods of causality they expressed. Median benefits estimated using the distributions of two of the experts in this group would show more significant declines.
- Among the experts who directly incorporated their views on the likelihood of a causal relationship into their distributions (Exhibit 3-11), the central (median) estimates of the percent change in all-cause mortality in the adult U.S. population that would result from a permanent 1 $\mu g/m^3$ drop in annual average $PM_{2.5}$ concentrations ranged from 0.7 to 1.6 percent
- Distributions of uncertainty around these median estimates varied across experts, but in almost all cases exceeded the statistical uncertainty bounds reported by any one epidemiologic study, suggesting that the expert elicitation process was successful in developing more comprehensive estimates of uncertainty for the $PM_{2.5}$ mortality relationship.
- Only one of 12 experts explicitly incorporated a threshold into his C-R function. The rest beloved there was a lack of empirical and/or theoretical support for a population threshold. However, three other experts gave differing effect estimate

distributions above and below some cut-off concentration. The adjustments these experts made to median estimates and/or uncertainty at lower $PM_{2.5}$ concentrations were modest.

- Although the quantitative question asked experts to consider mortality changes due both to short-term and long-term $PM_{2.5}$ exposures, all experts based their median effect estimates on effects due to long-term exposures. Short-term exposure effects were sometimes used to derive lower-bound effect estimates.
- A sensitivity analyses conducted using a simplified benefits analysis demonstrated that no individual expert's distribution of effect estimates had more than a plus or minus eight percent impact on an overall, pooled distribution of effects. The influence of individual experts appeared symmetrically distributed (see Exhibit 4-1).
- Confounding of epidemiological results tended to be a minor concern for most experts. Only one of 12 experts expressed substantial concern about confounding as a source of error in the key literature on $PM_{2.5}$ and premature mortality.
- Experts relied upon a core set of cohort epidemiology studies to derive their quantitative estimates, mainly those associated with the ACS and Six Cities cohorts. The Six Cities results tended to be weighted more highly by experts in this study than in the pilot study. The greater emphasis on Six Cities appeared to result from corroborating evidence in the recent Six Cities follow-up (Laden et al., 2006) and from concerns about potential exposure misclassification issues and/or effect modification in the ACS cohort.
- Experts' concerns regarding potential negative bias in the ACS main study results due to effect modification (see Pope and Dockery et al., 2006) and/or exposure misclassification (Jerrett et al., 2005; Willis et al., 2003; Mallick et al., 2002) led many experts to adjust the published results upwards when considering the percentiles of their distribution.

REFERENCES

Abbey, D. E., P.K. Mills, et al. (1991). Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California 7th-Day-Adventists. *Environmental Health Perspectives* 94: 43-50.

Abbey, D.E., N. Nishino, et al. (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American Journal of Respiratory and Critical Care Medicine* 159(2): 373-382.

Avol, E.L., W.J. Gauderman, et al. (2001). Respiratory effects of relocating to areas of differing air pollution levels. *American Journal of Respiratory and Critical Care Medicine* 164:2067-2072.

Brauer, M., J. Brumm, et al. (2002). Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects. *Risk Analysis* 22(6): 1183-1193.

Brook, R.D., J.R. Brook, et al. (2002). Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation* 105(13): 1534-1536.

Cakmak, S., R.T. Burnett, et al. (1999). Methods for detecting and estimating population threshold concentrations for air pollution-related mortality with exposure measurement error. *Risk Analysis* 19(3): 487-496.

Chen, L.H., S.F. Knutsen, et al. (2005). The association between fatal coronary heart disease and ambient particulate air pollution: Are females at greater risk? *Environmental Health Perspectives* 113: 1723-1729.

Clancy, L., P. Goodman, et al. (2002). Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet* 360(9341): 1210-4.

Clemen, R.T, R.L. Winkler (1985). Limits for the Precision and Value of Information from Dependent Sources. *Operations Research* 33(2): 427-442.

Cooke, R.M. (1991). Experts in uncertainty: opinion and subjective probability in science. (New York: Oxford University Press).

Costa, D.L., U.P. Kodavanti (2003). Toxic responses of the lung to inhaled pollutants: benefits and limitations of lung-disease models. *Toxicology Letters* 140-141: 195-203.

Creason, J., L.M. Neas, et al. (2001). Particulate matter and heart rate variability among elderly retirees: the Baltimore 1998 PM study. *Journal of Exposure Analysis and Environmental Epidemiology* 11(2): 116-122.

Devlin, R. B., A. J. Ghio, et al. (2003). Elderly humans exposed to concentrated air pollution particles have decreased heart rate variability. *European Respiratory Journal* 21: 76S-80S.

Diaz-Sanchez, D., A. Tsien, et al. (1999). Effect of topical fluticasone propionate on the mucosal allergic response induced by ragweed allergen and diesel exhaust particle challenge. *Clinical Immunology* 90(3): 313-22.

Dockery, D.W., C.A. Pope, et al. (1993). An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* 329(24): 1753-1759.

Dockery, D.W., H. Luttmann-Gibson, et al. (2005). Association of air pollution with increased incidence of ventricular tachyarrhythmias recorded by implanted cardioverter defibrillators. *Environmental Health Perspectives* 113(6): 670-674.

Dominici, F., A. McDermott, et al. (2003). Airborne particulate matter and mortality: timescale effects in four U.S. cities. *American Journal of Epidemiology* 157(12): 1055-1065.

Dominici, F., R.D. Peng, et al. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association* 295(10): 1127-1134.

Enstrom, J.E. (2005) Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. *Inhalation Toxicology* 17: 803-816.

Evelson, P., B. Gonzalez-Flecha (2000). Time course and quantitative analysis of the adaptive responses to 85% oxygen in the rat lung and heart. *Biochimica et Biophysica Acta* 1523(2-3): 209-216.

Gauderman, W.J., E. Avol, et al. (2004). The effect of air pollution on lung development from 10 to 18 years of age. *New England Journal of Medicine* 351(11): 1057-1067.

Ghio, A.J., C. Kim, et al. (2000). Concentrated ambient air particles induce mild pulmonary inflammation in healthy human volunteers. *American Journal of Respiratory and Critical Care Medicine* 162(3 Pt 1): 981-988.

Ghio, A.J. (2004). Biological effects of Utah Valley ambient air particles in humans: a review. *Journal of Aerosol Medicine* 17(2): 157-164.

Gilmour, P.S., I. Rahman, et al. (2002). Histone acetylation regulates epithelial IL-8 release mediated by oxidative stress from environmental particles. *American Journal of Physiology. Lung Cellular and Molecular Physiology* 284(3): L533-40.

Godleski, J. J., R. L. Verrier, et al. (2000). Mechanisms of morbidity and mortality from exposure to ambient air particles. Research Report 91, Health Effects Institute, Boston, MA.

Gold, D.R., A. Litonjua, et al. (2000). Ambient pollution and heart rate variability. *Circulation* 101(11): 1267-1273.

Goto, Y., J.C. Hogg, et al. (2004). Exposure to ambient particles accelerates monocyte release from bone marrow in atherosclerotic rabbits. *American journal of physiology. Lung cellular and molecular physiology* 287(1): L79-85.

Guigueira, S.A., J. Lawrence, et al. (2002). Rapid increases in the steady-state concentration of reactive oxygen species in the lungs and heart after particulate air pollution inhalation. *Environmental Health Perspectives* 110(8): 749-755.

Hawkins, N.C., J.D. Graham (1990). Expert scientific judgment and cancer risk assessment: a pilot study of pharmacokinetic data. *Risk Analysis* 8: 615-625.

Hedley, A.J., C.M. Wong, et al. (2002). Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. *Lancet* 360(9359): 787-788.

Hill, A.B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine* 58: 295-300.

Holmstead, J. (2005). EPA, Office of Air and Radiation. Letter to Robert O'Keefe, Health Effects Institute. July 11.

Hora, S.C. (1992). Acquisition of expert judgment: Examples from risk assessment. *Journal of Energy Engineering* 118(2): 136-148.

Industrial Economics, Inc. (2004). An expert judgment assessment of the concentration-response relationship between $PM_{2.5}$ exposure and mortality. Prepared for the Office of Air Quality Planning and Standards, Research Triangle Park, NC.

Industrial Economics, Inc. (2006). Briefing Book Binder. January.

Industrial Economics, Inc. (Under Development). Technical Support Document for Expanded Expert Judgment Assessment Of The Concentration-Response Relationship Between $PM_{2.5}$ Exposure And Mortality: Expert Interview Summaries. Prepared for the U.S. EPA Office of Air Quality Planning and Standards.

Jerrett, M., R.T. Burnett, et al. (2005). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16(6): 1-10.

Kahneman, D., P. Slovic, et al. (1982) Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, UK.

Krewski, D., R.T. Burnett, et al. (2000). Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Special Report to the Health Effects Institute, Boston MA, July.

Kunzli, N., S. Medina, et al. (2001). Assessment of deaths attributable to air pollution: should we use risk estimates based on time series or cohort studies? *American Journal of Epidemiology* 153(11): 1050-1055.

Kunzli, N., M. Jerrett, et al. (2005). Ambient air pollution and atherosclerosis in Los Angeles. *Environmental Health Perspectives* 113(2): 201-206.

Laden, F., J. Schwartz, et al. (2006). Reduction in Fine Particulate Air Pollution and Mortality: Extended Follow-up of the Harvard Six Cities Study. *American Journal of Respiratory and Critical Care Medicine* 173: 667-672.

Lipfert, F. W., H. M. Perry, et al. (2000). The Washington University-EPRI veterans' cohort mortality study: Preliminary results. *Inhalation Toxicology* 12: 41-73.

Lipfert, F.W., H.M. Perry Jr., et al. (2003). Air pollution, blood pressure, and their long-term associations with mortality. *Inhalation Toxicology* 15(5): 493-512.

Lipfert, F.W., R.E. Wyzga, et al. (2006). Traffic density as a surrogate measure of environmental exposures in studies of air pollution and health effects: Long-term mortality in a cohort of U.S. veterans. *Atmospheric Environment* 40: 154-169.

Mallick, R., K. Fung, et al. (2002). Adjusting for measurement error in the Cox Proportional Hazards Regression Model. *Journal of Cancer Epidemiology and Prevention* 7(4): 155-164.

McDonnell, W. F., N. Nishino-Ishikawa, et al. (2000). Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. *Journal of Exposure Analysis and Environmental Epidemiology* 10(5): 427-436.

Morgan, M.G., S.C. Morris, et al. (1984). Technical uncertainty in quantitative policy analysis - a sulfur air pollution example. *Risk Analysis* 4: 201-216.

Morgan, M.G., M. Henrion, (1990). Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. (New York: Cambridge University Press).

M. Granger Morgan, L.F. Pitelka, et al. (2001). Elicitation of Expert Judgments of Climate Change Impacts on Forest Ecosystems. *Climate Change* 49: 279-307.

Murphy, A.H, R.L. Winkler, (1992). Diagnostic verification of probability forecasts. *International Journal of Forecasting* 7(4): 435-455.

National Research Council (NRC). (2002). Estimating the public health benefits of proposed air pollution regulations. The National Academies Press: Washington, D.C.

O'Keefe, R. (2005). Health Effects Institute. Letter to Jeff Holmstead, EPA Office of Air and Radiation. August 15.

O'Neill, M.S., A. Veves, et al. (2005). Diabetes enhances vulnerability to particulate air pollution-associated impairment in vascular reactivity and endothelial function. *Circulation* 111:2913-2920.

Otway, H., D. von Winterfeldt (1992). Expert judgment in risk analysis and management: process, context, and pitfalls. *Risk Analysis* 12(1): 83-93.

Peters, A., E. Liu, et al. (2000). Air pollution and incidence of cardiac arrhythmia. *Epidemiology* 11(1): 11-7.

Peters, A., M. Frohlich, et al. (2001). Particulate air pollution is associated with an acute phase response in men; results from the MONICA-Augsburg Study. *European Heart Journal* 22(14): 1198-1204.

Plopper, C.G., M.V. Fanucchi (2000). Do urban environmental pollutants exacerbate childhood lung diseases? *Environmental Health Perspectives* 108(6): A252-253.

Pope, C.A. (1989). Respiratory disease associated with community air pollution and a steel mill, Utah Valley. *American Journal of Public Health* 79(5): 623-628.

Pope, C.A. (1991). Respiratory hospital admissions associated with PM10 pollution in Utah, Salt Lake, and Cache Valleys. *Archives of Environmental Health* 46(2): 90-97.

Pope, C. A., M. J. Thun, et al. (1995). Particulate air-pollution as a predictor of mortality in a prospective-study of U.S. adults. *American Journal of Respiratory and Critical Care Medicine* 151(3): 669-674.

Pope, C.A. (1996). Particulate pollution and health: a review of the Utah valley experience. *Journal of Exposure Analysis and Environmental Epidemiology* 6(1): 23-34.

Pope, C.A., D.W. Dockery, et al. (1999). Oxygen saturation, pulse rate, and particulate air pollution: a daily time-series panel. *American Journal of Respiratory and Critical Care Medicine* 159(2): 365-372.

Pope, C.A. (2000). Epidemiology of fine particulate air pollution and human health: biologic mechanisms and who's at risk? *Environmental Health Perspectives* 108(Suppl 4): 713-723.

Pope, C. A., R. T. Burnett, et al. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association* 287(9): 1132-1141.

Pope, C.A., R.T. Burnett, et al. (2004). Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation* 109(1): 71-77.

Rhoden, C.R., J. Lawrence, et al. (2004). N-acetylcysteine prevents lung inflammation after short-term inhalation exposure to concentrated ambient particles. *Toxicological Sciences* 79(2): 296-303.

Riediker, M., R.B. Devlin, et al. (2004). Cardiovascular effects in patrol officers are associated with fine particulate matter from brake wear and engine emissions. *Particle and Fibre Technology* 1(1): 2.

Saldiva, P.H., R.W. Clarke, et al. (2002). Lung inflammation induced by concentrated ambient air particles is related to particle composition. *American Journal of Respiratory and Critical Care Medicine* 165(12): 1610-1617.

Samet, J. M., F. Dominici, et al. (2000a). The National Morbidity, Mortality, and Air Pollution Study Part I: Methods and Methodologic Issues. Research Report 94, Health Effects Institute, Boston, MA.

Samet, J. M., S. L. Zeger, et al. (2000b). The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality from Air Pollution in the United States. Research Report 94, Health Effects Institute, Boston, MA.

Schwartz J. (1993) Particulate air pollution and chronic respiratory disease. *Environmental Research* 62: 7-13.

Schwartz, J., D.W. Dockery, et al. (2000). Harvesting and long term exposure effects in the relation between air pollution and mortality. *American Journal of Epidemiology* 151:440-448.

Schwartz J., S.K. Park, et al. (2005). GSTM1, Obesity, Statins, and autonomic effects of particles: gene by drug by environment interaction. *American Journal of Respiratory and Critical Care Medicine* 172:1529-1533.

Seaton, A., A. Soutar, et al. (1999). Particulate air pollution and the blood. *Thorax* 54: 1027-1032.

Sun, Q., A. Wang, et al. (2005). Long-term air pollution exposure and acceleration of atherosclerosis and vascular inflammation in an animal model. *Journal of the American Medical Association* 294(23): 3003-3010.

Suwa, T., J.C. Hogg, et al. (2002). Particulate air pollution induces progression of atherosclerosis. *Journal of the American College of Cardiology* 39: 935-942.

U.S. Environmental Protection Agency. (1997). The benefits and costs of the Clean Air Act, 1970 to 1990. Prepared for U.S. Congress by U.S. EPA, Office of Air and Radiation/Office of Policy Analysis and Review, Washington, DC

U.S. Environmental Protection Agency. (1999). Benefits and costs of the Clean Air Act: 1990-2010, EPA Report to Congress. U.S. EPA, Office of Air and Radiation and Office of Policy. Washington, DC. Document No. EPA-410-R-99-001. November.

U.S. Environmental Protection Agency. (2000). Regulatory impact analysis for the heavy-duty standards/diesel fuel rulemaking. Prepared by: Innovative Strategies and Economics Group, Office of Air Quality Planning and Standards, Research Triangle Park, NC December.

U.S. Environmental Protection Agency, Science Advisory Board. (2004). 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. Advisory by the Health Effects Subcommittee of the Advisory Council on Clean Air Compliance Analysis. EPA-SAB-COUNCIL-ADV-04-002.

U.S. Environmental Protection Agency. (2004). Final regulatory impact analysis: control of emissions from non-road diesel engines. Office of Transportation and Air Quality. EPA-420-R-04-007.

U.S. Environmental Protection Agency. (2005a) Regulatory impact analysis for the final Clean Air Interstate Rule. Office of Air and Radiation. EPA-452/R-05-002.

U.S. Environmental Protection Agency. (2005b). Regulatory impact analysis for the final Clean Air Visibility Rule or the Guidelines for Best Available Retrofit Technology (BART) Determinations under the Regional Haze Regulations. Office of Air and Radiation. EPA-452/R-05-004.

Walker, K.D., P. Catalano, et al. (2003). Use of expert judgment in exposure assessment: part 2. Calibration of expert judgments about personal exposures to benzene. *J Expo Anal Environ Epidemiol.* 13(1): 1-16.

Walker, K.D. (2004). Industrial Economics, Inc. Memorandum to Jim Neumann, Henry Roman, and Tyra Gettleman, Industrial Economics, Inc. Appropriate Number of Experts for the Particulate Matter Expert Judgment Project. November 11.

Wellenius, G.A., B.A. Coull, et al. (2003) Inhalation of concentrated ambient air particles exacerbates myocardial ischemia in conscious dogs. *Environmental Health Perspectives* 111(4): 402-408.

Willis, A., M. Jerrett, et al. (2003). The association between sulfate air pollution and mortality at the county scale: an exploration of the impact of scale on a long-term exposure study. *Journal of Toxicology and Environmental Health, Part A* 66(16-19): 1605-1624.

Winkler, R.L., R.M. Poses (1993). Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Management Science* 39(12): 1526-1543.

Winkler, R. L., T.S. Wallsten (1995). An assessment of the risk of chronic lung injury attributable to long-term ozone exposure. *Operations Research* 43(1): 19-28.

Wolff, S.K., N.C. Hawkins, et al. (1990). Selecting experimental data for use in quantitative risk assessment: an expert judgment approach. *Toxicology and Industrial Health* 6:275-291.

Yamawaki H., N. Iwai (2006). Mechanisms underlying nano-sized air-pollution-mediated progression of atherosclerosis: carbon black causes cytotoxic injury/inflammation and inhibits cell growth in vascular endothelial cells. *Circulation Journal* 70(1): 129-140.

Zeger, S.L., F. Dominici, et al. (1999). Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology* 10(2): 171-175.

Zeka, A., A. Zanobetti, et al. (2005). Short term effects of particulate matter on cause specific mortality: effects of lags and modification by city characteristics. *Occupational and Environmental Medicine* 62(10): 718-725.

Zelikoff J.T., L.C. Chen, et al. (2003) Effects of inhaled ambient particulate matter on pulmonary antimicrobial immune defense. *Inhalation Toxicology* 15(2): 131-150.