IEc

Technical Support Document for the Expanded Expert Judgment Assessment of the Concentration-Response Relationship Between $PM_{2.5}$ Exposure and Mortality: Expert Interview Summaries

September 30, 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

Table of Contents

INTRODUCTION

As part of the *Expanded Expert Judgment Assessment of the Concentration-Response Relationship Between $PM_{2.5}$ Exposure and Mortality* (IEc, 2006), the elicitation team conducted day-long personal interviews with 12 health experts who have conducted research on the relationship between $PM_{2.5}$ exposures and mortality. This technical support document consists of summaries of the discussions that took place during each of these interviews. The summaries provide additional detail and context that may be useful for understanding each expert's judgments.

Exhibit 1 presents the list of the experts interviewed. As described in the main report, experts are identified by a randomly assigned letter from A to L in order to preserve confidentiality. The section headings in the interview summaries correspond to those found in the elicitation protocol (See IEc, 2006, Appendix A). Each expert was given the opportunity to review his summary and confirm that it accurately captured the views he expressed in the interview. In addition, if an expert opted to change any of his responses to the questions in the protocol following the June 2006 Post-elicitation Workshop, the changes are reflected in a "Modification to Expert Judgment" form, appended to the end of his interview summary.

EXHIBIT 1: 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

IEc prepared these summaries to document both the views expressed by each expert in his interview and the process by which he arrived at the quantitative distribution presented in the main report (IEc, 2006).[1] As part of documenting the process of developing an expert's final quantitative distribution, the summaries may include additional distributions generated by the IEc elicitation team during the interview that were shown to the expert to assist him in evaluating his responses. For example, if an expert indicated during the interview that his distribution was conditional on the existence of a causal relationship between PM2.5 and mortality, the elicitation team showed the expert a distribution generated by statistically combining his distribution with his stated probability of a causal relationship, assuming that the probabilities are independent. All of these additional distributions are marked as "IEc-generated" in the summaries and do not reflect the expert's final judgments, which are presented in the main report. We also note that content in the interview summaries may differ from that presented in the main report if an expert modified his judgments following the Post-elicitation Workshop. As noted above, these changes are documented in a modification form appended to the end of the summary of each expert who revised his judgments.

[1] IEc prepared all of the interview summaries, with the exception of Expert D. Expert D wrote his own summary, to which IEc added the tables and graphs and a short introductory paragraph.

Expert A
Interview Summary

Interview Summary
Expert A

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects from Exposure to $PM_{2.5}$

Expert A thought that there were separate mechanisms for mortality associated with short-term and long-term exposures to $PM_{2.5}$. However, he did feel that there was some overlap between them. He began by discussing mortality related to short-term exposures, which he felt to be the most compelling. He then discussed potential mechanisms of long-term exposures, and ended with a discussion of mechanisms for intermediate length exposures. He felt that the intervention studies (Clancy et al. (2002) in Dublin and Pope et al. (1996) in Utah Valley), involving intermediate time scales of months up to a year, actually provided more insight into how the effects measured in short- and long-term studies might be tied together and help understand the total effect of reducing particulate air pollution on mortality.

Short-Term Exposures

For short-term exposures, Expert A thought that the major causes of death, in order of importance were: acute cardiac events; respiratory disease; stroke; and lung cancer deaths (where people are frail from lung cancer and die from acute exposure to air pollution).

Expert A argued that the most compelling evidence for an association between particles and acute cardiac events came from the time-series cardiovascular studies. In particular, he thought the pooled studies (e.g., the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), Air Pollution and Health – A European Approach (APHEA)) provided the strongest evidence because they are not as subject to confounding by co-pollutants and weather as other study designs. He cited as supporting evidence the studies involving implantable defibrillator discharges (e.g., Peters et al., 2000 and Rich et al., 2005 & 2006) and case-cross over studies of myocardial infarction survivors (Peters et al., 2001 and Foresteri et al., 2005 in Rome). He thought their strengths include unambiguous outcome measures and the fact that by "…looking at the same individual you eliminate by design possibilities for confounding by individual characteristics, such as smoking. You can also largely eliminate seasonal effects, temperature. You can eliminate factors that potentially confound time-series studies. On the other hand, there are some other subtle issues that enter with the selection of case and control periods, and there is a loss of power in case-crossover compared to a time-series analysis." He noted that a weakness of both types of studies is that they have a selected population that is not representative of the general population (e.g., frail, high risk of cardiac instability). Other case-crossover studies have not supported a relationship between myocardial events and air pollution (e.g., Sullivan et al., Checkoway et al., 2003, and Levy et al., 2001). Though these studies had good outcome measures, he was not convinced that the nephelometry measures of particles were good measures of air pollution.

In terms of mechanistic evidence for a role of particles in acute cardiac events, he was not sure of the mechanisms. He noted that several mechanistic pathways had been proposed (e.g., "inflammatory responses, change in autonomic function, release of hormones, cytokines, chemokines, and other chemicals delivered to the heart"). Experimental evidence had not ruled any out any pathways. He cited experimental studies in animals (Godleski et al., 2000) showing changes in cardiovascular responses in dogs; Watkinson and others at EPA showing electrical abnormalities and dysfunction in rodents; Devlin et al. chamber studies measuring intermediate endpoints, such as changes in heart rate variability and other electrical disturbances in human subjects.

The next three causes of death related to short-term exposures (respiratory disease, strokes, lung cancer) were discussed in less detail and the expert did not feel the evidence was as strong. Expert A cited respiratory disease as the next most important, though he did not discuss specific evidence. He noted that death certificate coding would tend to place acute cardiac death as the primary cause of death even if the individual had underlying chronic obstructive pulmonary disease (COPD). Therefore respiratory effects might not show up as associations in mortality studies.

Expert A was less certain about strokes as a cause of death related to particles. Strokes can be hemorrhagic or ischemic and it is often difficult to determine the type from death certificates. He would not expect to see an association with hemorrhagic stroke, but thought that ischemic strokes might follow a similar pathway as ischemic heart attacks, and therefore are more plausible. He suggested that most of the evidence for stroke associations with air pollution comes from Asian countries where the underlying rates of stroke deaths are much higher than in the U.S.

The final category of death that Expert A discussed was lung cancer. "People with lung cancer are more likely to respond to air pollution and die acutely." However, he noted that there were no short-term studies that have actually looked at lung cancer specifically.

Long-Term Exposures

For long-term exposures, Expert A used the analogy of smoking and environmental tobacco smoke (ETS) as a conceptual model. Airborne particles might impact human health through "changes in lung function, deterioration of lung volumes, development of chronic obstructive pulmonary disease, and a long-term debilitation, loss of respiratory reserve"; development of atherosclerosis, plaques and increased risk of myocardial infraction (MI); vascular changes; and cancer. However, he noted that "in terms of air pollution, [there is] somewhat limited evidence for those effects… The prospective studies [including new Adventist Health and Smog (ASHMOG) data] have suggested that you can see some changes in pulmonary function, but I don't think the evidence is very compelling … [A]ir pollution at the levels we're talking, is even less than the typical exposures from environmental tobacco smoke. So I don't think this is a likely pathway for major disability."

He cited some evidence suggesting the development of atherosclerosis as a pathway for the influence of chronic exposure to particles on increased risk of mortality. "There are new studies coming out, the Kunzli study for example, that are looking at carotid IMT, intima-media thickness. The carotid IMT, looking at deposition of plaques in the carotid arteries, shows more plaque in those people living in the higher air pollution levels." Expert A thought that the recent work by Sun and Lippmann in atherosclerotic mice has provided experimental evidence for the cardiac pathways. At the same time he expressed some skepticism about this pathway contributing to major increases in risk or death from atherosclerosis, stating, "there are other, larger factors."

Intermediate Exposures

Expert A thought that the intervention studies (the Pope et al. Utah Valley study and Clancy et al. Dublin study) tie the concept of the smoking analogy and cumulative exposures to air pollution together as well as point to larger mortality effects in the intermediate time scale than have been typically captured in time-series or cohort studies. He thought that the time-series studies show effects on the same day, as well as two to four days later "[b]ut if you … go out to longer time periods, you can still see residual associations from the short-term exposures. And when you add up all those effects, they're much larger than … short-term effects. Given that we're normally exposed to repeated days of events, I think that these air pollution events are not independent … Usually we'll see two, three, four, maybe a week of air pollution. I think that's where the action is." As with smoking cessation, he thought most of the intermediate effects are in reduction of deaths from cardiovascular disease with a smaller impact on respiratory and cancer deaths, which are results of longer-term exposures.

3.2 Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

When asked to describe his conceptual framework for the relationship between types of death caused by air pollution, he indicated that he would prefer if the Kunzli framework incorporated an intermediate effect. He indicated that categories A and C constitute short-term effects but would include larger effect windows up to weeks or months. He thought categories A and C constitute the bulk of the mortality effect. He indicated that category B, the mortality effect from long term exposures, was relatively small compared to A and C.

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

Expert A thought that, to the extent that short-term exposures are correlated with long-term exposures, prospective cohort studies provide information on the effects of both types of associations, although they would not capture deaths advanced by just a few days. He thought time-series studies are restricted to the short term. Given the focus of our question on the impact of a change in $PM_{2.5}$ concentrations, he thought the intervention studies (Utah Valley, Dublin, Hong Kong) would be most informative

because he thought most of the mortality effects occur in weeks to months. However he noted that the re-analysis of the Six-City data (Laden et al., 2006), which looks over the last 10 years, might capture some long-term effects as well.

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

Specifically, Expert A relied on the estimates from the original Six Cities cohort study (Dockery et al., 1993) and the extended follow-up (Laden et al., 2006), a reanalysis of the American Cancer Society (ACS) cohort study in Los Angeles (Jerrett et al., 2005), and the NMMAPS (Samet et al., 2000) for his quantitative estimates.

3.5 Confounding

Expert A generally felt that most confounders had been adequately controlled in the epidemiologic literature and therefore this did not have a large effect on his quantitative estimates. He noted that the Krewski re-analysis had done a better job of controlling for confounding than the original ACS and Six Cities analyses, and that it strengthened the findings. However, he did list some factors that might not have been fully accounted for in particular studies. The table below lists the factors he discussed.

SUMMARY OF KEY POTENTIAL CONFOUNDERS IN PM MORTALITY EPIDEMIOLOGICAL STUDIES

Study (author, date)	Potential Confounder	Rationale	Score (1-3)[a]
Overestimated RRs			
Six Cities Study			
Dockery et al., 1993	Ethnic differences between cities		1
Laden et al., 2006	Ethnic differences between cities	Results of this study suggest that ethnicity is unlikely to affect the results.	1
Intervention Studies			
Laden et al., 2006 (Six Cities)	Time-varying trends in healthcare		1
Clancy et al., 2002 (Dublin)	Time-varying trends in healthcare	Includes only a single study and because pollution levels dropped, but did not rise again.	2
Pope et al., 1996 (Utah Valley)	Time-varying trends in healthcare		1
American Cancer Society			
Jerrett et al., 2005 (Los Angeles)	Co-pollutants		1
Pope et al. (1995)	Occupational Exposures		2
Uncertain Direction of Bias			
American Cancer Society	Community characteristics		
	Socioeconomic Status (SES)		
	Healthcare		
	Contextual		

[a] The scores are defined by the magnitude of their effect on the published estimates: 1 = minimal effect; 2 = medium effect; and 3 = major effect.

3.6 Effect Modification

Expert A discussed the potential for effect modification by race, educational level, and housing characteristics. While he thought there might be an argument for effect modification by race (due more to underlying differences in SES and health status), he ultimately did not think that race intrinsically affected susceptibility to air pollution and that therefore the effect estimates in the cohort studies were likely to be representative. He noted that educational attainment per se was not an effect modifier but likely to be an indicator for other factors. He thought the Six Cities studies had a representative sample so it was not an important issue. He thought that there could be some minimal effect modification by housing characteristics (e.g., air conditioning) in the cohort studies leading to a possible underestimate of the effect estimate.

3.7 Exposure Issues

Expert A thought that exposure misclassification (i.e., problems with use of central site monitors to characterize individual exposures to ambient $PM_{2.5}$) was the biggest issue of concern. He thought that evidence suggests that exposure misclassification causes the published effect estimates to be biased downwards. "The evidence we have from the epi studies is that when we have improved measures of $PM_{2.5}$ exposures they get much stronger associations. Specifically comparing the Six Cities study to the [original and extended] ACS [Pope et al., 1995 & 2002], comparing to the Jerrett [et al., 2005] study to the [original and extended] ACS studies, you see substantially higher effect estimates.

When asked if variation in composition affected published effect estimates, he did not think so. He basically argued that compositional differences do not seem likely to be resulting in an overestimate of mortality effects. He thought similar mortality effects had been estimated in various areas of the country with different PM compositions (e.g., sulfates in the northeast and Utah Valley, nitrates in Los Angeles) as well as in intervention studies where the PM composition changes within a community over time, suggesting PM is at worst a good surrogate for whatever the toxic agent might be. If the agent were known, Expert A thought it would more likely underestimate, rather than overestimate the relative risks attributable to that agent.

He thought that the measurement method in the Laden et al. (2006) extended analysis of the Six Cities data, in which the $PM_{2.5}$ is estimated rather than measured directly, might be an important limitation of the study. He thought it might introduce some bias though he was unable to define a direction of bias.

3.8 Causality

Expert A discussed a number of the standard criteria for supporting a causal relationship. He particularly focused on the consistency of results across studies and across different study designs (e.g., time series, prospective cohort, intervention). He thought the intervention studies are important because they are semi-experimental designs. In addition, he thought clinical and toxicological studies have provided supporting evidence.

Expert A specified a range of values for the likelihood of a causal relationship between $PM_{2.5}$ and mortality of 70 – 99.5 percent, with a most likely value of 95 percent (1 in 20). He selected the low end of the range based on the concern that there is a lack of specificity in the mechanism (i.e., every pathway proposed seems to have an association) and possibility that there just might be some alternate explanation for the effects, though it is difficult to imagine what it might be. The upper end of the range was based on Expert A's confidence in there being a mortality effect of $PM_{2.5}$ exposures. "There's lingering doubt, but not as much as there was in [the past]."

3.9 Thresholds

While Expert A felt strongly that individual thresholds exist (i.e., concentrations of particles that are sufficient to overcome and individual's defense mechanisms), he also argued that there is a continuum of individual responses, based on variety of genetic, environmental, and SES factors. Therefore, he did not expect to see evidence of a population threshold in epidemiologic studies. "That's a construct of experimental studies, but it's not something I expect to see in epidemiology." He indicated that epidemiological studies, not clinical or toxicological, are the appropriate tool for exploring population thresholds because they allow investigators to look at the full range of susceptible individuals in a population.

Expert A indicated that the epidemiologic literature has shown no evidence of a threshold in the C-R function. In addition, he thought that study designs were becoming increasingly sophisticated, and therefore more able to detect effects at lower levels, further supporting the lack of threshold in the C-R relationship. He thought that arguments for thresholds that focus on lack of statistical significance for relationships at low concentrations were confusing detectability with the likelihood that a threshold exists. He thought that the population data indicate a linear relationship at low doses.

3.10 Other Influential Factors

Expert A thought that an additional source of uncertainty arises from his concern that there is a lack of independent statistical and epidemiological expertise to provide adequate external criticism of the particulate matter (PM) studies. He thought that many working in this field are very self-critical and are better at identifying weaknesses in their studies than the critics. Nonetheless, it concerns him that they are all "moving in the same direction … There actually are quite a few studies coming along, replicating [the Six Cities and ACS studies]. That's where the scientific certainty, I think, comes from, that is from other people really challenging thee results [in other] populations."

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert A thought that the C-R function was linear (expressed in percent change per 1$\mu g/m^3$) and consistent over the entire range of annual average $PM_{2.5}$ concentrations that were the focus of the study (4-30 $\mu g/m^3$).

Expert A chose to provide a C-R function that was conditional on the existence of a causal relationship (with a 95 percent likelihood of a causal relationship). The elicitation team then combined his elicited distribution with his likelihood of causality using Monte Carlo simulation to create a final distribution.[2]

Expert A preferred to discuss estimates in terms of mortality impact per 10 $\mu g/m^3$ change in $PM_{2.5}$. The elicitation team later scaled these estimates to a 1 $\mu g/m^3$ change.

Expert A began by thinking about the studies he thought were most informative. " [T]he benchmarks that I would use [are] the Six Cities original analysis, which would give me a 13 percent increase per 10 micrograms; the Jerrett study, which gives me 17 percent increase per 10 micrograms; the extended Six Cities analysis (Laden), gives me 16 percent; the ACS extended analysis, which gives me 6.2 percent and the Six Cities change analysis, which gives me actually 37 percent. So these kind of define the upper bounds. I suppose also there's another bound, which might be the NMMAPS study, which was about half a percent, 0.4 percent, approximately, for 10."

Expert A iterated through the process of developing his distribution, relying on results from these studies to guide him. He initially identified the NMMAPS estimate, 1.004, and the 1.37 relative risk from the Laden et al. 2006 as plausible lower and upper bounds, respectively.[3] He chose the NMMAPS estimate, 1.004, as a lower bound, noting that the NMMAPS is probably over controlled. He argued for using the Jerrett (2005) study relative risk estimate (1.17), supported by the (1.16) extended analysis for the Six-Cities study, as the basis for the central estimate. He chose these studies because he thought they did not suffer from the biases created by non-representativeness, educational attainment levels, and exposure errors inherent to the ACS original study and re-analysis. Using the natural log of these values, he then estimated a standard deviation based on an assumed normal distribution. The median and standard deviation then were converted into percent change per 1 $\mu g/m^3$ change in $PM_{2.5}$. He then decided to increase his standard deviation because he was not satisfied with the spread of the distribution. He settled on a median of 1.6 percent increase in mortality per 1 $\mu g/m^3$ change in $PM_{2.5}$ and a standard deviation of 0.78. From these values, he calculated 5^{th} and 95^{th} percentile values of 0.29 and 2.9, respectively. The elicitation team entered these values into Crystal Ball and using a normal distribution, calculated the remaining percentiles. He ultimately set his minimum value at zero and his maximum at 4 percent per 1 $\mu g/m^3$ change in $PM_{2.5}$. His final distribution in Exhibit 1 incorporates his estimate that there is a 95 percent likelihood that the relationship is causal, as noted above.

---

[2] Using Monte Carlo simulation in Crystal Ball™, 95 percent of the iterations take a value drawn from the elicited distribution and 5 percent are assigned a value of zero.

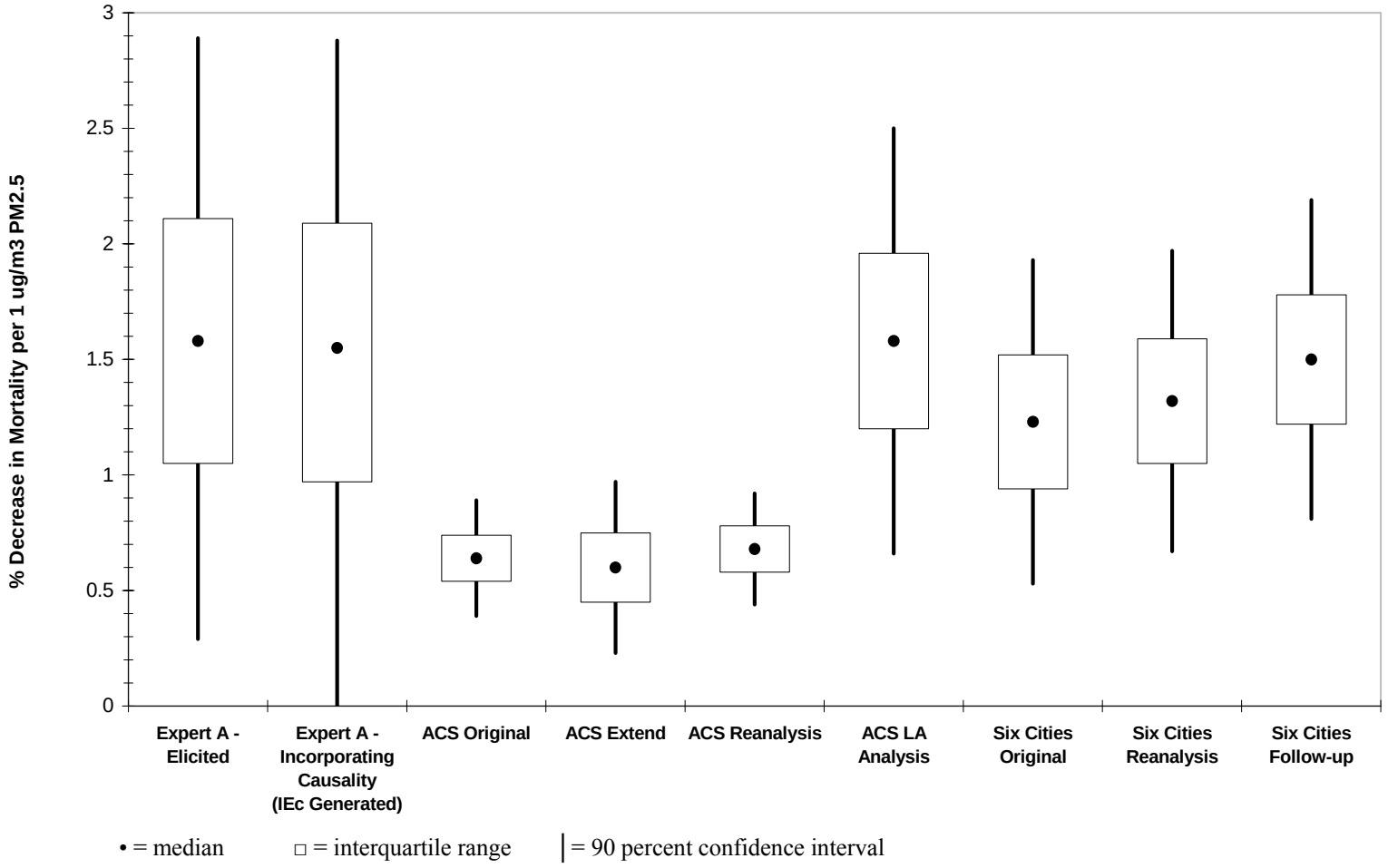
[3] Expert A briefly considered using as his upper bound the upper 95 percent confidence limit (1.75) on the 1.37 RR from Laden et al, 2006, but rejected it as implausibly large.

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations

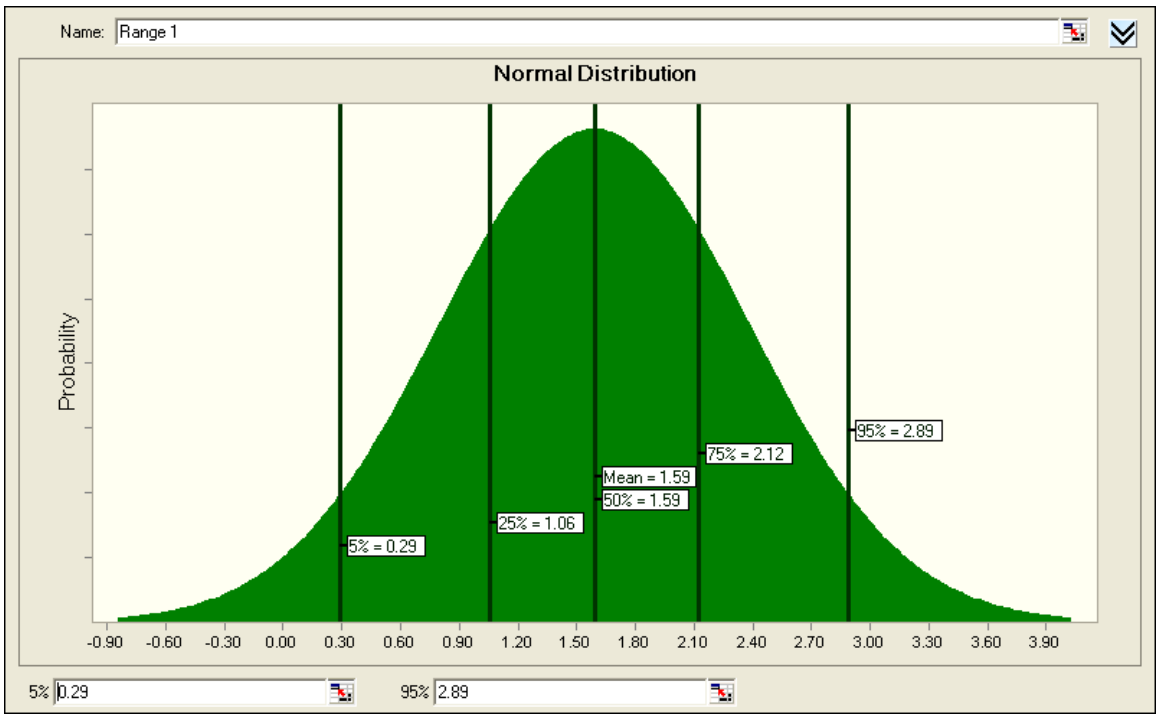
Percentile	Percent Change in Mortality Elicited Distribution[a]	Percent Change in Mortality Incorporating Causality (IEc Generated)[b]
Minimum	0	0
5th	0.29	0
25th	1.1	0.97
50th	1.6	1.6
75th	2.1	2.1
95th	2.9	2.9
Maximum	4.0	4.0

a. Assuming or conditional on a causal relationship
b. Incorporating 5 percent likelihood of a non-causal relationship using Monte Carlo simulation.

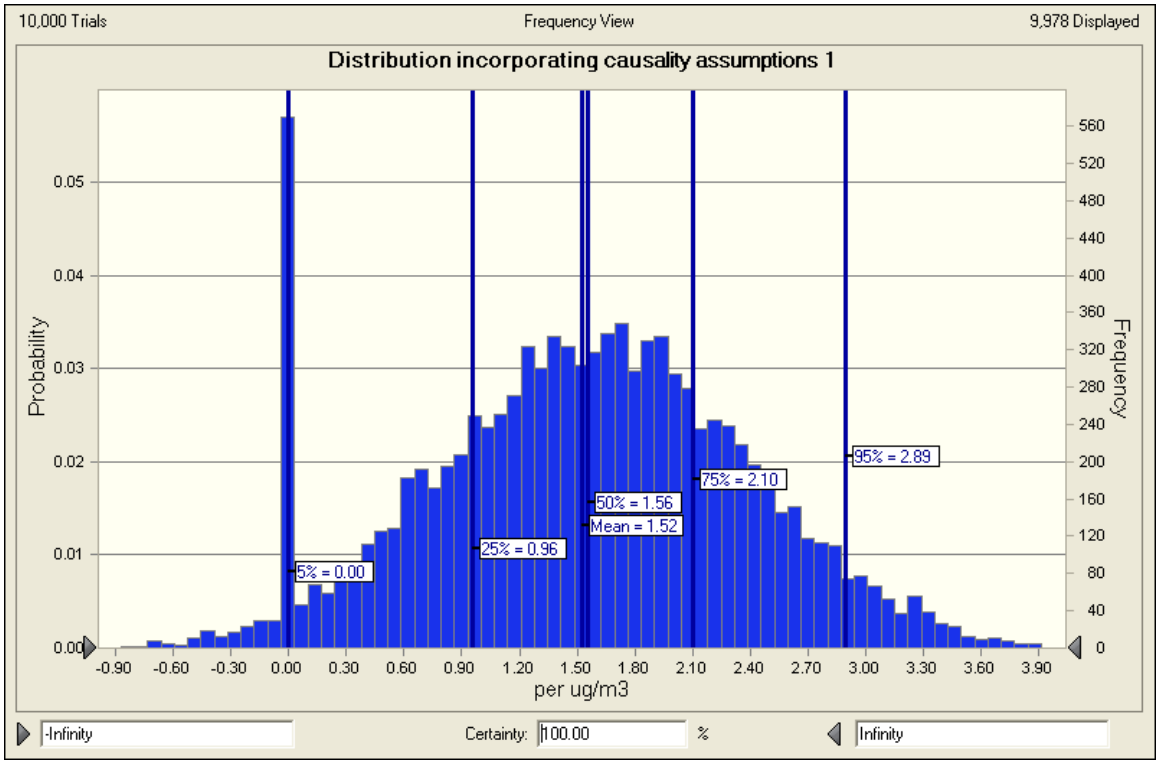
Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 percent Confidence Intervals for Various Studies to Distributions from Expert A



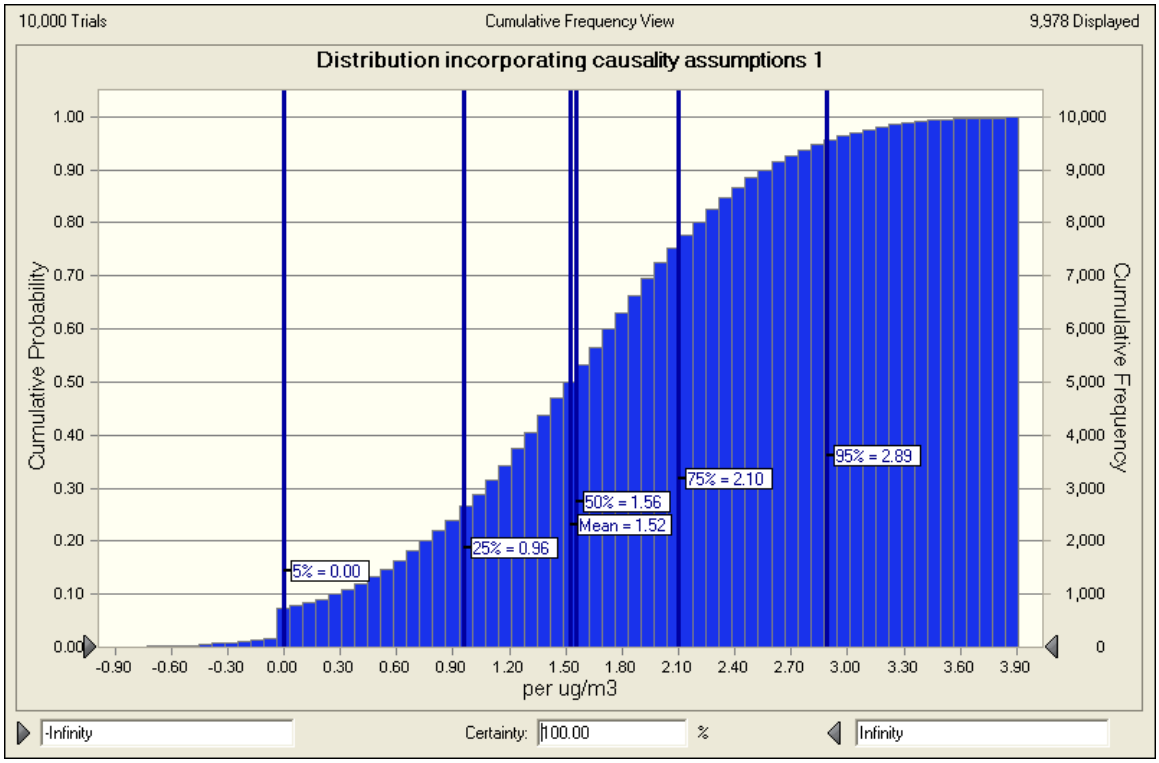
Elicited Distribution



Distribution Incorporating Causality - Probability Density Function (IEc Generated)



Distribution Incorporating Causality - Cumulative Density Function (IEc Generated)



U.S. EPA EXPERT ELICITATION STUDY OF THE CONCENTRATION-RESPONSE RELATIONSHIP BETWEEN ANNUAL AVERAGE $PM_{2.5}$ EXPOSURE AND MORTALITY

Modification to Expert Judgments

Expert A

Date: 11 July 2006

Section of Protocol Affected (Section Number and/or Title):

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Description of Change (e.g. to a specific percentile, or to a qualitative opinion or statement of belief):

C-R function is NOT conditional on the existence of a causal relationship

Expert B
Interview Summary

Interview Summary
Expert B

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert B discussed the biological mechanisms for short-term and long-term exposures separately, preferring to begin with long-term.

Long-term Exposures

Expert B thought that the two main causes of death associated with long-term exposures were lung cancer and cardiopulmonary disease. He cited evidence from studies of concentrated ambient particles (CAPs) showing the mutagenicity of these complex mixtures (e.g., the 1992 International Programme on Chemical Safety (IPCS) study in which CAPs from Washington, D.C. were analyzed (Krewski et al., 1992)[4]) as the basis for a hypothesis of a relationship between exposure to airborne particles and cancer. The extended analysis of the American Cancer Society (ACS) cohort (Pope et al., 2002) then provided clear epidemiologic evidence for increases in lung cancer associated with $PM_{2.5}$ exposures. For cardiopulmonary mortality, he noted that as recently as 1998, when the National Research Council (NRC) set forth research priorities for particulate matter, the mechanism for cardiopulmonary mortality was "a mystery." However, he indicated that he now felt that inflammatory responses to particles (and release of cytokines affecting endothelial cells) leading to accelerated formation of atherosclerotic plaque is now the prime hypothesis for the mechanism for this cause of death, citing the paper by Pope et al. published in Circulation (2004), and toxicological studies in rabbits conducted at the University of British Columbia. While he thought cardiopulmonary mortality is the dominant cause of mortality, he indicated that the weight of evidence was stronger for the causal mechanism for cancer mortality. He noted respiratory mortality as a third cause of death, but he thought that less epidemiological evidence exists for the respiratory effect, and the evidence that is available is weaker.

Short-term Exposures

Expert B did not think that there was sufficient evidence to discuss the specific mechanisms underlying the causes of death associated with short-term exposures, although he did feel that there were some deaths associated with short-term exposures that were not captured by the long-term studies. He did not feel that the time-series studies allowed for a clear understanding of the specific role of $PM_{2.5}$. The issue of harvesting remained uncertain in his mind. "There's a lot of conversation about this harvesting hypothesis, frail individuals or in poor health, and you hit them with this little additional insult and it just is enough to push them over the edge. But if you look at studies that Mark Goldberg has done in Montreal, where you examine the association

[4] Krewski et al. (1992) Mutation Research/Reviews in Genetic Toxicology 276: 33-59.

between mortality and pre-existing co-morbidities, you don't see a real strong predictive effect there … If harvesting was a real phenomenon, you should be able to show that people with pre-existing co-morbidities are at greater risk than the general population." He did not feel that toxicological or clinical studies shed light on the mechanisms for mortality related to short-term exposures.

3.2. Conceptual Framework for Mortality Effects of Short-term and Long-Term $PM_{2.5}$ Exposures

Expert B thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures, although he did not think it adequately captured the influence of other co-factors and he would alter the relative magnitude of the circles. He thought the biggest contribution of $PM_{2.5}$ to mortality would be from long-term exposures, and that the overlap between long- and short-term exposures, if any, would be very small.

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

Expert B thought that a combination of cohort, time-series, case-crossover, case-control, and intervention studies were appropriate for capturing the mortality effects of changes in annual average $PM_{2.5}$ concentrations. The effects captured by each study design are shown in the table below:

Study Design	Type of Effects Captured (e.g., short-term, long-term, or both)
Cohort Studies	Long-term, and possibly some acute effects that are the end stages of long-term exposure. Missing true acute effects related to peak pollution episodes.
Time-Series Studies	Short-term
Case-Crossover Studies	Short-term
Intervention Studies	Long-term

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert B thought that the following characteristics would be part of an ideal epidemiologic study to characterize the $PM_{2.5}$-mortality relationship in the U.S. population:

- Individual exposure measurements (ideally, each participant would have a personal dosimeter that could distinguish between ambient and indoor exposures, though a dosimeter that gives average annual ambient exposures might be sufficient);
- Large number of individuals; and

- Collection of information on a series of co-factors (e.g., indoor air pollution, gaseous co-pollutants, and confounders such as diet and occupation). "Any of them that had significant temporal heterogeneity should be longitudinally measured."

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert B focused primarily on the estimates from the original ACS cohort study (Pope et al., 1995) and the reanalysis in Los Angeles (Jerrett et al., 2005), and the Six Cities Cohort Study (Dockery et al., 1993).

"[N]umber one would be Arden Pope's 1995 first analysis of the ACS cohort, which had as the two key strengths the very large sample size, both in terms of number of people, about a half a million it's less than that in the $PM_{2.5}$ cohort … But more importantly … [t]here were about 50 cities in the $PM_{2.5}$ cohort. The big limitation of that first [ACS] analysis was the ecologic measure of the exposure. You've got one measure of exposure for each of those 50 cities …. The second most important study is Mike Jerrett's 2005 study of intra-urban variation in Los Angeles, where there were some 30-odd pollution monitors around the city. People were allocated to the closest monitor. And the risk estimates went up by a factor of two- to three-fold when you got a more localized measure of exposure."

Expert B also mentioned a "what if" analysis by Mallik et al. (2002) of the Six Cities data that found essentially the same effect. He indicated that he authors "postulated a certain level of exposure misclassification by using ecologic rather than personal dosimeters [and by] adjusting for it analytically, … actually got risk estimates that were two- to three-fold higher."

His third choice was the Harvard Six Cities study, which has the limitation of only having six cities and … "there's such a co-linearity among all the pollutants measured that you get almost the same concentration-response function for any pollutant that you look at." He also did not believe it was demographically representative of the U.S. (e.g., it does not represent Washington D.C., which is more white collar and less industrially based).

Expert B also briefly discussed the Adventists Health and Smog (AHSMOG) and the Veteran's cohort (VA) studies but he was not familiar with the methodological details. He also discussed Moolgavkar's criticisms that the cohort studies and their re-analyses have focused too heavily on $PM_{2.5}$ as a prior, and have therefore potentially missed effects of other gaseous co-pollutants. Expert B argued that the Health Effects Institute (HEI) re-analyses did study other co-pollutants but only found effects associated with SO_2. He felt that Moolgavkar's views are based on time-series studies primarily and argued that "it's harder to disentangle the effects of multiple pollutants in the acute studies because typically you're focusing on a much narrower geographic area, where you don't have the heterogeneity of the pollution mixes that you get in the ACS cohort."

Expert B felt the most influential published studies of the effects of short-term exposures to date were the multi-city studies (Canadian studies of 8-11 cities by Burnett et al. (2000

& 2003), the National Morbidity, Mortality and Air Pollution Study (NMMAPS) (Samet et al., 2000), and the Air Pollution and Health – A European Approach (APHEA) study).

3.5 and 3.6. Confounding and Effect Modification

Expert B preferred to discuss confounding and effect modification together. He discussed the following factors as possible confounders or effect modifiers of the relationship between $PM_{2.5}$ and mortality:

Cohort Studies

- **Confounding by Occupational Exposures:** Expert B did not think that this was a confounder because of extensive analysis done on the ACS (Siemiatycki et al., 2003) and Six Cities studies examining this factor.
- **Confounding by SO_2:** Expert B thought SO_2 was correlated with both $PM_{2.5}$ and mortality. "Another key limitation … of the ACS study, was this powerful effect of SO_2, which can almost negate, almost but not quite, the $PM_{2.5}$ effect … There's no mechanism; obviously it's a marker for something else. So at this point I don't have any … clear explanation for why that SO_2 effect is so strong, but it is a very strong effect that does require some kind of investigation." He felt that failing to account for SO_2 could lead to a "major" (i.e., a score of 3) overstatement of the $PM_{2.5}$ effect.
- **Effect Modification by Educational Attainment:** Expert B thought effect modification by educational attainment was a "major" (i.e., a score of 3) issue for the ACS study and to a lesser extent for the Six Cities Study. He thought that individuals with less than high school education were under-represented in the ACS study and effect estimates for this group were nearly six times as high as the group with greater than a high school education. He felt that educational attainment was an indicator for a complex set of lifestyle factors (e.g., exercise, access to health services, diet, and occupational exposures). A similar type of effect, though smaller and non-significant, was also found in the Six Cities study.

He indicated that he has not seen clear statistical evidence of effect modification by other factors that might be expected to modify risk, such as gender, smoking status, or pre-existing co-morbidities.

Short-Term Studies

Expert B thought that confounding by gaseous co-pollutants (NO_x and SO_x, especially) could lead to a "moderate" (i.e., a score of 2, or on the order of a 25 percent) overestimate of the relative risk for $PM_{2.5}$ in these studies. He noted that time-series studies examining mortality and reproductive effects have shown strong associations with both PM and primary gaseous pollutants. "We don't have a good scientific basis, in my view, to distinguish between the two." However, he thought "if the dominant circle in the Künzli diagram is the long-term mortality and the acute effects are much smaller, then really our understanding … is much better in the long-term effects. So we're capturing much more

of the population health impact, and the uncertainty caused by lack of ability to disentangle co-pollutants in the short term effect contributes less to the overall uncertainty in terms of overall mortality." He thought effect modification by weather or seasonal effects is possible, which is why he thought these variables should be included in the time-series models, but he thought it would be minimal (i.e., a score of 1) for $PM_{2.5}$.

3.7. Exposure Issues

Expert B focused on two issues 1) exposure misclassification; and 2) temporal variation in exposure. He also discussed instrument measurement error and the role of indoor versus outdoor measurements of exposure, but thought these were trivial issues.

Exposure Misclassification

Expert B thought that exposure misclassification, particularly differences between central site monitoring exposure estimates and actual individual exposures, could cause the published effect estimates to be biased downward. He saw evidence for this in a series of studies: "[First,] when you take one monitor or the average of all monitors in an urban area to represent individual exposure from everybody in a city of hundreds of thousands or several million people, just intuitively doesn't seem like a good thing to do. And when we actually do find associations with that very crude ecologic measure, it's first of all quite surprising … Second is [the Willis et al. (2003) reanalysis of the] ACS cohort,[5] actually had some data within urban areas at the county level … for about half the cohort … [The analysis] showed that the risk estimates were approximately double when you use ecologic measures on a county level as opposed to an SMA level … So that was the second chapter in the story, having gone [from] an association with such a crude measure of exposure, to showing that we get stronger associations going down to the county level. Third chapter was [a paper by Mallick et al. (2002)] … which was this sort of hypothetical, what if the exposure misclassification from using one fixed-site monitor representing all of an urban area on actual personal exposures of such and such a magnitude?" [She] gauged that by looking at inter-monitor variation in cities, where you have multiple monitors, and then did a regression calibration kind of adjustment, and the risk estimates went up by a factor of two- to three-fold. And then the fourth and final piece was Mike Jerrett's Los Angeles analyses, where a totally different way of adjusting for exposure misclassification [than the] Mallick paper [was taken] and got almost the same effect."

Expert B thought this type of exposure misclassification was more of an issue in the ACS cohort study (Pope et al., 1995 & 2002) than in the Six Cities study, which had better spatial resolution of exposure.

[5] Willis, A.J. et al. (2003) Journal of Toxicology and Environmental Health 66: 1605-1624.

Temporal Variation in Exposure

Expert B thought that temporal variation in exposure was the second area of concern; that is to understand "what the long-term effects of air pollution are from exposure in the last year, the last five years, the last ten years – When in the past does exposure actually contribute most to mortality? And if we believe the accelerated atherosclerosis hypothesis for cardiopulmonary effects …you would expect maybe 5, 10, 15 years ago would be the most important window for current mortality."

Expert B believed we currently do not know the answer to these questions. He indicated that Villeneuve et al. (2002) attempted to look at temporal variation in exposure in the Six Cities study taking into account people who moved, but few of the original cohort moved and the author was not "able to tease out much of a temporal effect of exposure on risk." He did not think that the Pope et al. (2002) extended analysis of the ACS cohort really revealed significant differences between effect estimates based on exposures from 1980 versus the later exposures up to 2000, because of correlation between exposures for the two time periods. He indicated that work is being done on a nutritional sub-cohort of the ACS study that will take into account both mobility and temporal changes in PM, but data are not available yet from that study.

3.8. Causality

Expert B felt a causal relationship would be best supported by a "strong, statistically significant, and robust association. Robust in the sense that adjustment for confounders and effect modifiers still leaves the association intact … replication of the results in independent studies … a plausible biological mechanism by which those effects could occur … a clear exposure response gradient."

"[I]f we look at long-term exposure, I think we have strong, statistically significant, robust associations that have been replicated in the Harvard Six Cities study, the American Cancer Society cohort, and the Seventh Day Adventist. [These associations are r]obust against adjustment for a whole host of factors." Expert B noted that the Health Effects Institute in their reanalysis "has looked at the ACS data just about every way imaginable, and have had extensive opportunity to control for 140 covariates. [W]e do have a plausible … biologic mechanism, by which particles can cause cardiopulmonary mortality and lung cancer mortality." He thought the exposure response gradient had been observed in both the ACS studies and the Six Cities study.

Expert B thought there was less strength of evidence for causal effects of short-term exposures "even though there are only a handful of chronic studies and lots of short-term studies." For example, he pointed out that there are not statistically significant results for all 90 cities in NMMAPS.

Expert B specified a range and most likely value for the likelihood of a causal relationship for long-term and short-term exposures separately. For long-term, he initially specified a range of 80-95 percent with a most likely value of 95 percent. "The main factors that would [contribute to the 80 percent] are the SO_2 effect on $PM_{2.5}$, which

is quite substantial as we discussed earlier" and to a lesser extent the ecologic measures of exposure. For short-term, he initially specified a range of 70-90 percent with a most likely value of 90 percent. After further consideration of his views on the overall strength of the evidence, he expressed greater confidence in the strength of the evidence for a long-term impact, and ultimately provided a range of 90-99 percent and a most likely value of 98 percent for the likelihood of a causal relationship for long-term and short-term exposures combined. This estimate reflects a greater emphasis on causality for long-term effects, which he thought were the dominant contributor to effects seen for changes in annual average $PM_{2.5}$.

3.9 Thresholds

Expert B summarized his views on thresholds with four main points "Number one, conceptually, looking at mechanisms of lung cancer and cardiopulmonary mortality, particularly for lung cancer, I think there's a strong argument to be made that conceptually there is no population threshold. Number two, for acute effects … without knowing those mechanisms, just drawing analogies with other causes of acute mortality due to environmental exposures, I think there's a stronger argument to be made for the existence of a threshold. Number three, observationally, for both long-term and short-term exposures, I don't think we have much of a chance of demonstrating a threshold because of the limits of resolution of epidemiologic data at low exposure levels. We see in the chronic studies, as we go down the concentration-response curve, the uncertainty bands widening to the point where you get to the lower end of that 4 to 30 $\mu g/m^3$ range, you really can't tell definitively if there is an increased risk or not, unless you believe that trend line holds even with confidence bounds at the possibility of no risk at the lower end of the range." He indicated that Cakmak et al. (1999) did an analysis published in Risk Analysis showing empirically how difficult it would be to identify a threshold even in acute studies.

He thought the population exposure-response relationship might flatten out some at lower exposures, including levels below $4\mu g/m^3$. In essence, he thought it was more an issue of some non-linearity at low doses rather than a threshold concept.

He concluded: "[A]nother overarching observation here is that the question of thresholds may not be a question that's really worth trying to resolve because of the limits of observational studies and defining that threshold, because any conceptual arguments that you can put forward will remain conceptual unless you've worked out the biological mechanism by which particles cause mortality in full detail and embed a sensitive marker that you can trace right down to the very low exposure levels. And I don't think we're there yet."

3.10 Other Influential Factors

Expert B discussed additional sources of uncertainty that were not part of the protocol. He thought that more research into biological mechanisms was warranted. In addition, he thought that errors in outcome ascertainment (misinterpretation of cause of death, ascertainment of vital status) could add uncertainty to the concentration-response (C-R)

function (although all-cause mortality would not be affected by miscoding of cause of death). Expert B did not think publication bias was an issue. Nor did he think ecologic community covariates were an alternative explanation for the mortality effect seen in the ACS study, in particular.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert B thought the C-R relationship was largely linear on the basis of observational data. He noted that work by Abramowitz et al. (2003) using flexible exposure-response curves with no constraints on the prior form of the distribution showed a largely linear C-R relationship.

Expert B wanted to characterize his C-R function in a piece-wise linear fashion, specifying two C-R functions over the range of annual average $PM_{2.5}$ concentrations that were the focus of the study (4-30 $\mu g/m^3$), one that applied to concentrations of 4-10 $\mu g/m^3$ (hereafter, "Range 1"), and one that applied to concentrations of >10-30 $\mu g/m^3$ (hereafter, "Range 2"). He did so to reflect his greater uncertainty about the shape of the concentration response function in the lower concentration range.

Expert B chose to provide C-R functions that were conditional on the existence of a causal relationship. The elicitation team then combined his conditional distributions with his percent likelihood of causality specified in Section 3.8 (98 percent likelihood of a causal relationship). In addition, the two distributions were applied to a distribution of population-weighted annual average $PM_{2.5}$ concentrations in the U.S. from EPA's BenMap model to create a combined distribution (hereafter, "Example Applied Distribution"). He stated that his effect estimates represented primarily the effects of long-term exposure (on the order of less than 10 percent attributable to effects of short-term exposure).

Expert B specified the same 50^{th}, 75^{th}, 95^{th} and maximum values for both Range 1 and 2. Expert B estimated his 50^{th} percentile by adjusting the original ACS relative risks (1.06 per 10 $\mu g/m^3$) upward by a factor of two to account for exposure misclassification using as justification the Mallick et al. (2002) analysis and the Jerrett et al. (2005) studies. The 95^{th} percentile was based on adjusting the ACS cohort study estimates to account for the influence of a non-representative population with respect to socioeconomic status (SES) (educational attainment), for uncertainty about the true basis of SO_2 effect, and for some statistical uncertainty. His 75^{th} percentile was chosen largely as an intermediate value between the 50^{th} and 95^{th} percentiles. Finally, his maximum values for both ranges were based on allowing for uncertainties that he had not already taken into account, such as measurement error in the monitoring science, unidentified covariates, a larger SES adjustment, and model uncertainty.

Expert B specified different values for the minimum, 5^{th}, and 25^{th} percentiles for Ranges 1 and 2. Expert B began with the upper range (Range 2) and set his minimum value at 0.1 percent per 1 $\mu g/m^3$. For the 5^{th} percentile, Expert B adjusted the 50^{th} percentile downward first by 0.4 to account for an "appropriate" confidence limit. This took the estimate from 1.2 to 0.8. He then decreased this by 75 percent to account for an SO_2

effect, ending up with 0.2 percent. For the 25th percentile in Range 2, he again adjusted the median to 0.8 for the confidence limit and then reduced it 37.5 percent (halfway between 25 and 50 percent) to account for an SO_2 effect, ending up with a value of 0.5. For Range 1, he decreased the minimum to show his increased uncertainty about the C-R relationship at low concentrations. He reduced the 5th percentile to account for the fact that there may be no acute effects at these concentrations, and the potential flattening out of the C-R function for cardiopulmonary deaths. He reduced the 25th percentile effect estimate to account for the potential absence of acute effects at these concentrations, and the potential flattening out of the C-R function for cardiopulmonary deaths.

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Range 1 (4 – 10 μg/m^3)

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	0.01	0
5th	0.10	0.07
25th	0.20	0.19
50th	1.2	1.2
75th	2.1	2.1
95th	2.6	2.6
Maximum	2.8	2.8

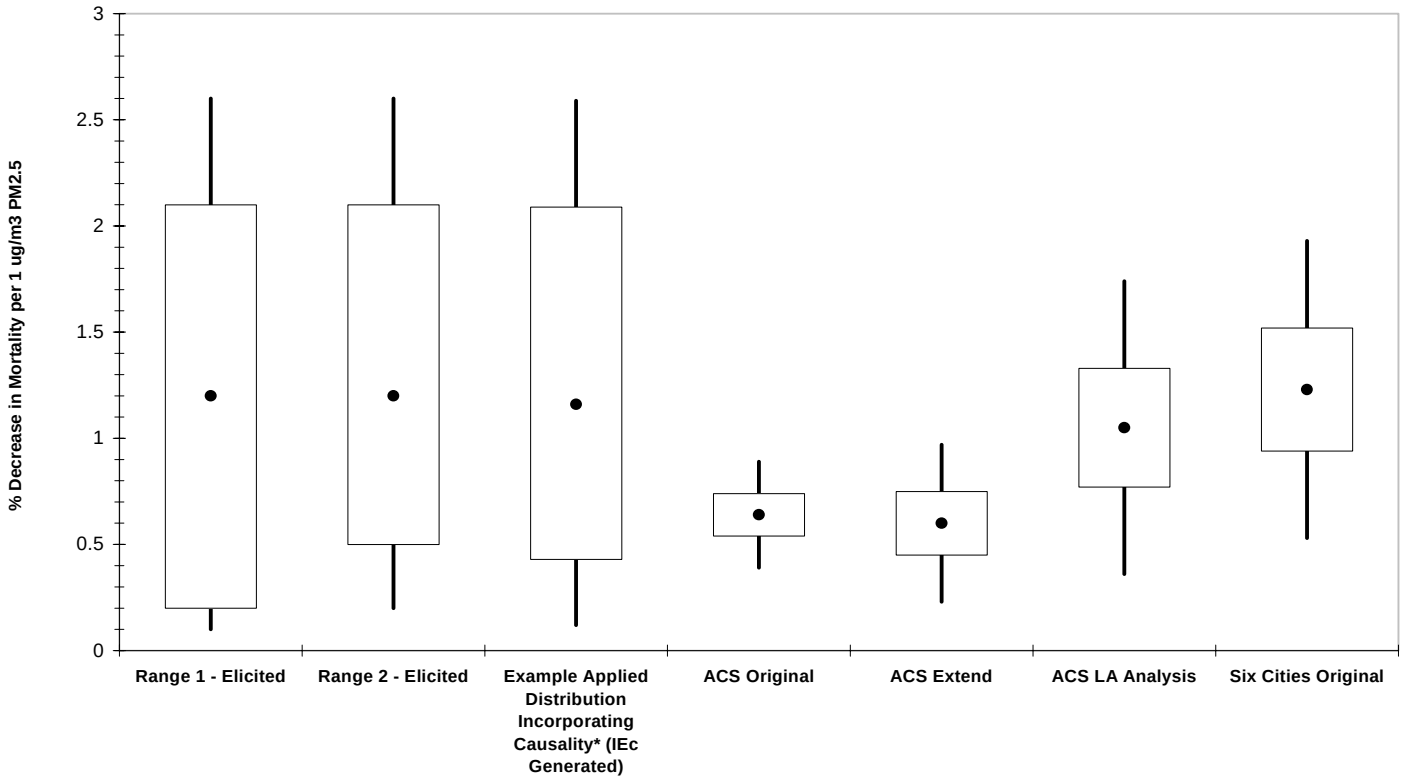
Exhibit 2: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Range 2 (>10 – 30 μg/m^3)

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	0.1	0
5th	0.2	0.16
25th	0.5	0.47
50th	1.2	1.2
75th	2.1	2.1
95th	2.6	2.6
Maximum	2.8	2.8

Exhibit 3: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations After Incorporating Causality and Applying the C-R functions to the Population-Weighted Annual Average $PM_{2.5}$ Concentration Distribution in the U.S. from BenMap - Example Applied Distribution (IEc Generated)

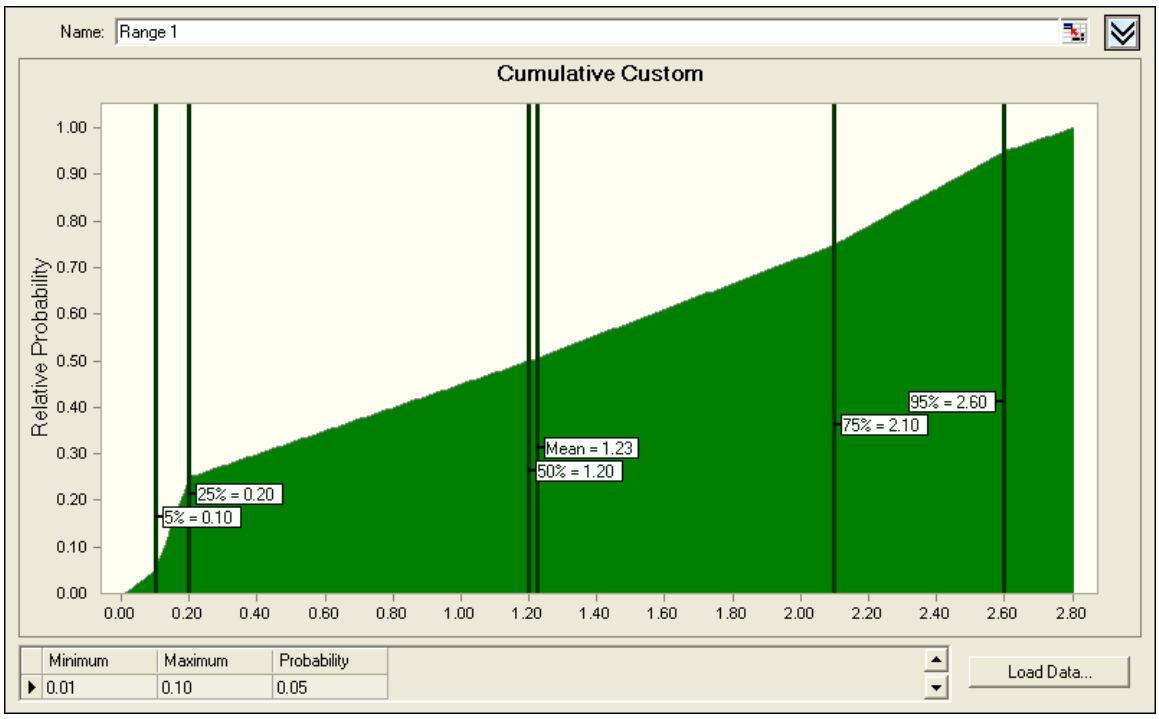
Percentile	Percent Change in Mortality
Minimum	0
5^{th}	0.12
25^{th}	0.43
50^{th}	1.2
75^{th}	2.1
95^{th}	2.6
Maximum	2.8

Exhibit 4: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 percent Confidence Intervals for Various Studies to Distributions from Expert B

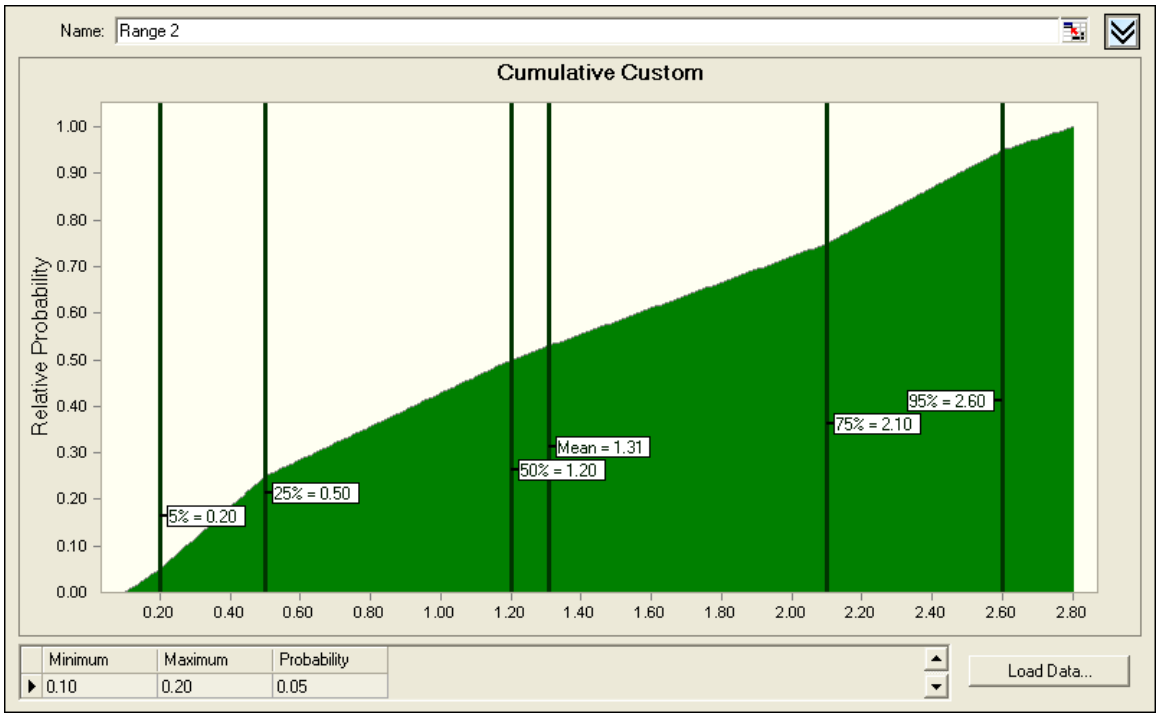


* = Distribution incorporating causality and applying the C-R functions from Ranges 1 and 2 to a 2002 population-weighted annual average $PM_{2.5}$ concentration distribution in the U.S. from BenMap.
• = median □ = interquartile range | = 90 percent confidence interval

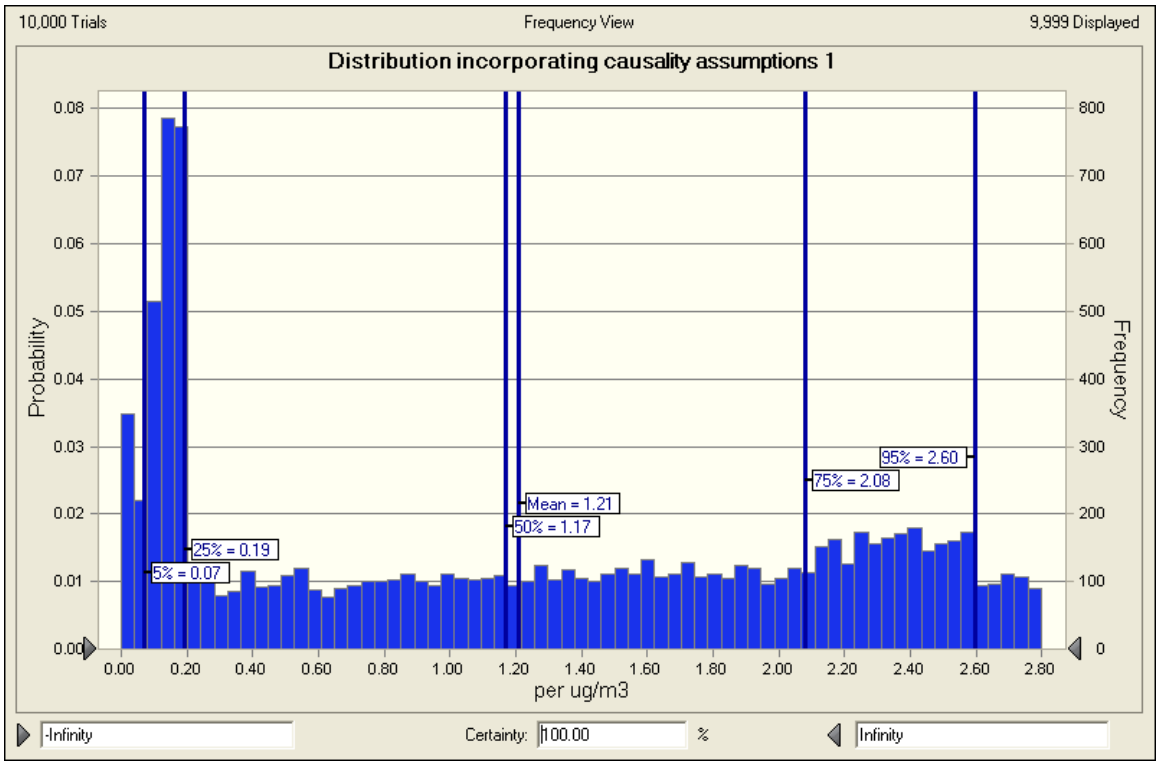
Elicited Distribution – Range 1



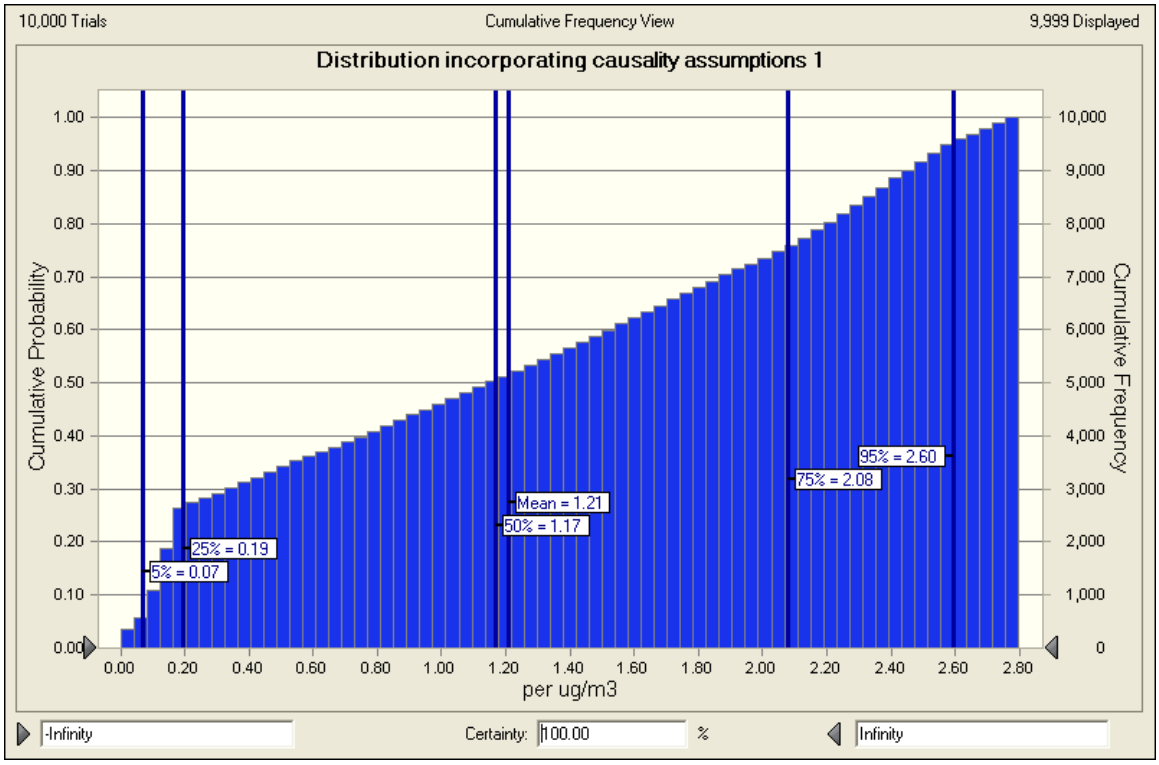
Elicited Distribution – Range 2



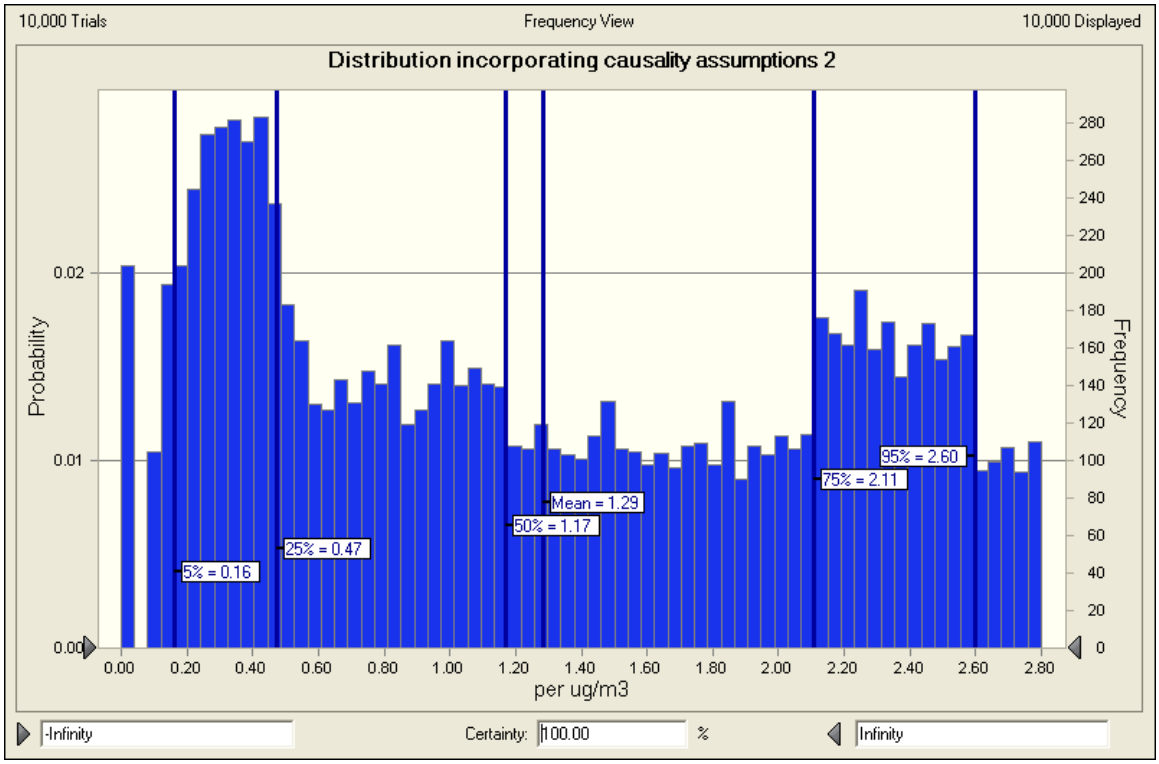
Range 1 Incorporating Causality - Probability Density Function (PDF) (IEc Generated)



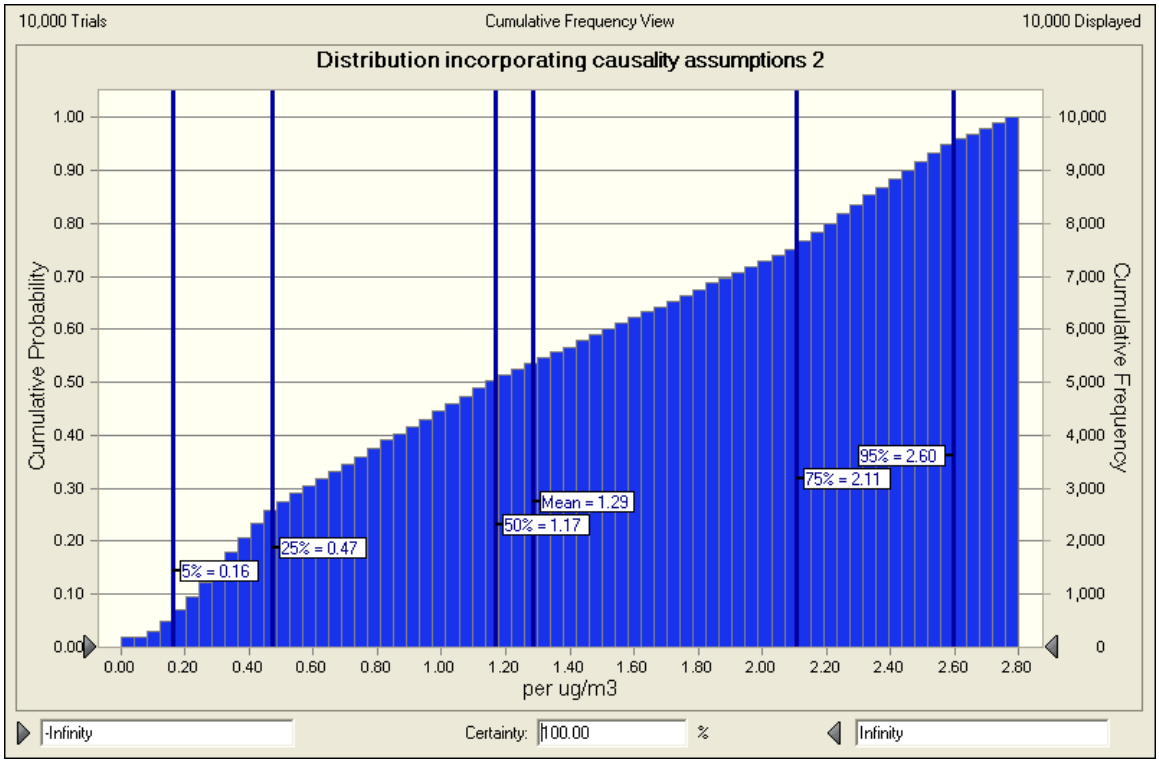
Range 1 Incorporating Causality - Cumulative Density Function (CDF) (IEc Generated)



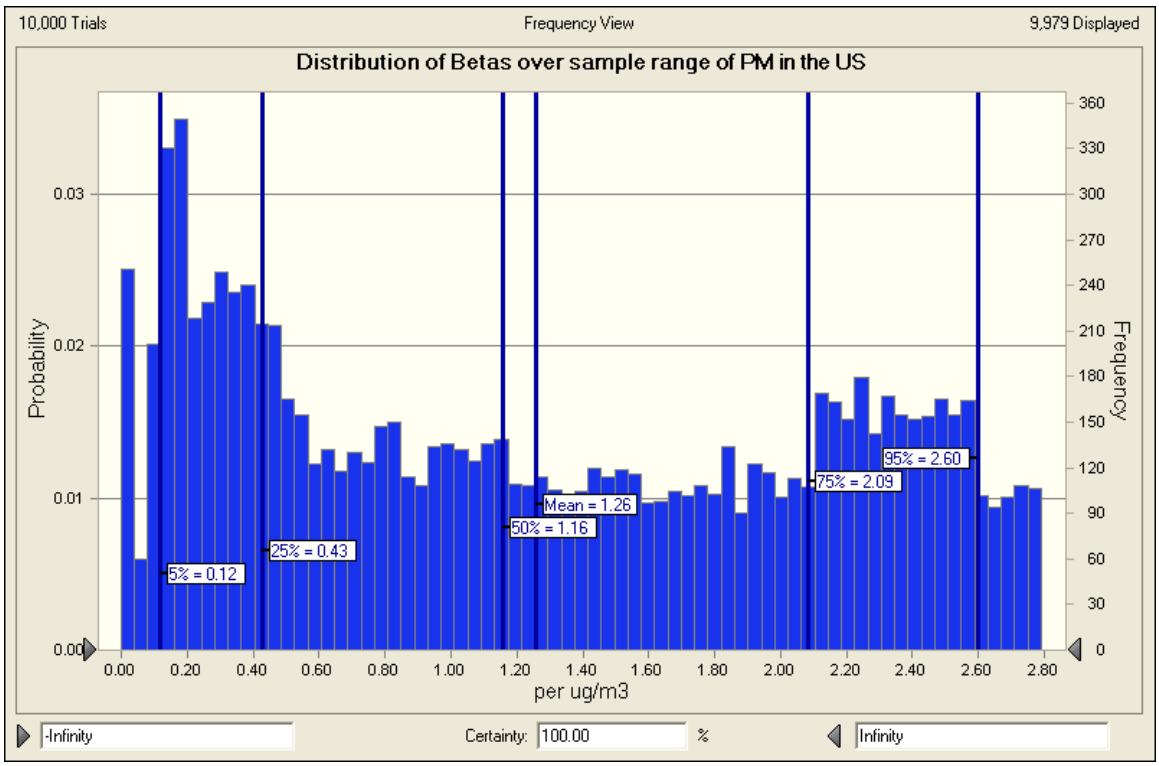
Range 2 Incorporating Causality – PDF (IEc Generated)



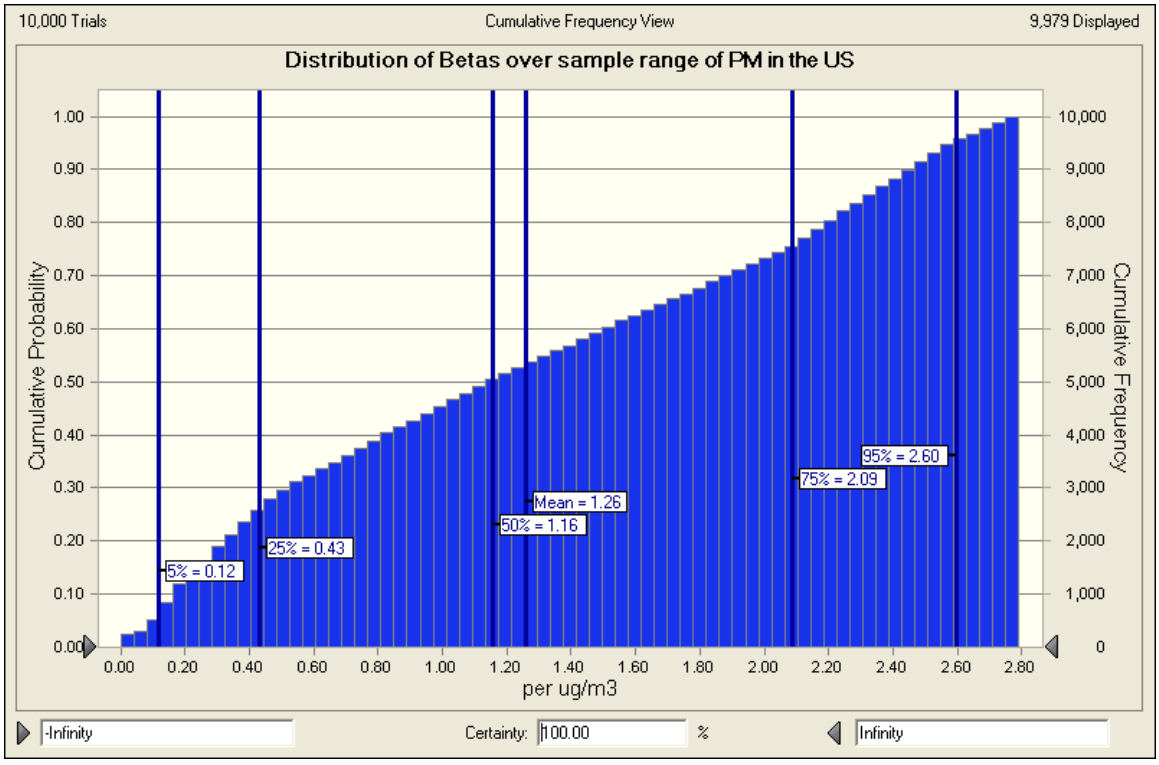
Range 2 Incorporating Causality – CDF (IEc Generated)



Example Applied Distribution – PDF (IEc Generated)



Example Applied Distribution – CDF (IEc Generated)



Expert C
Interview Summary

Interview Summary
Expert C

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure $PM_{2.5}$

Expert C discussed the biological mechanisms for the effects of long-term and short-term exposures separately, focusing first on long-term.

Long-Term Exposures

He thought that the two main causes of death associated with long-term exposures were cardiac disease and lung cancer. In particular, he cited the Pope et al. 2004 paper in *Circulation* as influential in showing that most of the excess mortality due to PM exposure was related to cardiac disease. He described the cardiac mechanism as an accelerated aging process brought on by the contribution of the body's natural defense mechanisms to inflammation, to long-term functional changes, and ultimately to changes in viability. He thought that the exact intermediate steps for the mechanism "are pretty well undefined" but are believed to involve deposition of the particles into the lung, which stimulate mediator release, leading to endothelial inflammation. He cited the Sun et al. (2005) work (in *JAMA*) and the nine papers in the Mar.-Apr. 2005 issue of *Inhalation Toxicology* describing earlier CAPs inhalation research conducted at NYU. These studies showed histological evidence of accelerated arteriosclerotic lesions in ApoE-/- mice exposed to concentrated air particles (CAPs) as important evidence for a mechanism for cardiac disease. Although the exposures to mice involved high concentrations (~100 $\mu g/m^3$) for 6 hours a day, 5 days a week, he thought that it was the cumulative average exposure (on average 15-17 $\mu g/m^3$) that was important for the effects seen.

For lung cancer, he thought that the Pope et al., 2002 study showed a significant excess of cancer; this work has been supported by other cohort studies which have shown consistent elevated risk, though the others were not statistically significant. He thought the mechanisms here are not certain but may be related to mutagenic or carcinogenic particle components and whether they are initiators or promoters. He noted that there is some excess risk for pulmonary disease but since pulmonary disease is less prevalent, the total impact on numbers of deaths is smaller than for cardiac disease and cancer.

Short-Term Exposures

For short-term exposures, he thought that PM acted analogously to ozone and could cause an irritant response (broncho-constriction and inflammation in the airways resulting from stimulation of epithelial cells to release mediators). He noted that, unlike ozone, particles can remain in the lung for a long time and may contribute to chronic as well as acute conditions. He thought that deaths from short-term exposures could occur in people with little reserve capacity because underlying pulmonary disease (inability to ventilate

effectively) can contribute to cardiac deaths. He also cited evidence from Annette Peters' epidemiological work and from animal studies showing declines in heart rate variability, a factor in viability, which could be leading to acute cardiac deaths.

3.2 Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

Expert C generally found the Kunzli framework very useful and the definitions of the categories of death relevant, though he thought it could be enhanced to include different causes of death. He would want to change the size of the ovals to show effects of long-term exposure to be much greater than the effects of short-term exposure.

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

Expert C thought the cohort studies are the best way to estimate total mortality, represented by Categories B (long-term effects) and A (mixed effects) in the Kunzli framework. He though total short-term effects are best captured by the time-series studies with lags of up to 40 days (e.g., as shown by Joel Schwartz and his colleagues).

The intervention studies (Sugiri et al., 2005; Oglesby et al., 2005; Avol et al., 2001; Clancy et al., 2002; and Hedley et al., 2002) suggest an intermediate effect on daily or longer-term functional changes and mortality. He noted that the Clancy et al. study saw reductions in both respiratory and cardiac deaths; the Avol et al. study showed changes in lung growth related to changes in PM exposures. He did not have a clear sense of what the dominant mechanisms were for these changes, although findings of inflammation and atherosclerosis development in the Sun et al (2005) six month mouse study (discussed earlier) may provide part of the explanation. He thought the composition of the particles may turn out to be relevant and discussed unpublished work suggesting the importance of nickel and vanadium in the mortality changes observed in intervention studies.

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

Expert C thought an ideal study for characterizing the U.S. concentration-response (C-R) function for $PM_{2.5}$ and mortality would resemble the Six Cities study but with a broader range of cities that would be more geographically representative of the U.S. He would want the study to be prospective, to involve a representative sample of the U.S. population, to include measurements of multiple air quality parameters (e.g., acid vapors, organics, elemental carbon, $PM_{2.5}$), and to collect data on personal risk factors both on enrollment and on follow-up.

Expert C made the argument that ACS (Pope et al., 2002) and Six Cities studies (Dockery et al., 1993 (original) Laden et al., 2006 (follow-up)) were essentially a "complementary pair of studies which found differences in [the] magnitude [of effects that] can be explained. … [We] can decide some of their biases and correct [for] them."

Expert C thought that the ACS study (Pope et al., 2002) was the most useful for answering the causal question because it is geographically more representative of the whole U.S., though it is "less than ideal" in terms of its measurement of pollutants (use of single monitors for broad areas, lack of detail on other pollutants than those included for compliance monitoring, lack of information about where members of the cohort lived after enrollment), and the non-representativeness of its population with respect to socioeconomic status (SES). He thought the Jerrett et al., 2005 ACS reanalysis in Los Angeles (LA) used a more precise measure of individual exposure.

He thought the Six Cities Study had a more representative population with respect to SES and better exposure measurements (use of investigator controlled monitoring stations, located to be more representative of populations). However he was concerned that the six cities were not geographically representative of the U.S.

Expert C thought the intervention studies were very important but "they're most useful for … plausibility … that the effects of the particles are substantial and real, rather than in terms of quantitation." His point was that they each use different measures of particle pollution, and do not directly address $PM_{2.5}$. The Hong Kong study (Hedley et al., 2002) looked primarily at SO_2, but Expert C views SO_2 as a surrogate for fine PM in this study; the Dublin study used black smoke and the correlation with $PM_{2.5}$ was uncertain; and the Utah Valley study used PM_{10} – the results could be corrected for $PM_{2.5}$ though he was not aware that they had been.

Expert C also discussed the Adventist Health and Smog (AHSMOG) and Veteran's studies. He thought the AHSMOG study was generally supportive of a PM effect though he was concerned about the non-representative population and the indirect construction of the $PM_{2.5}$ metric. He characterized the Veteran's study as difficult to understand based on available publications and he was concerned about its non-representative population.

3.5 Confounding

Focusing on the ACS and Six Cities studies, Expert C discussed a number of theoretically influential confounders and then evaluated whether or not he thought they had been dealt with adequately in the two studies. The initial discussion identified the following:

- Criteria pollutants;
- Smoking;
- Occupation;
- Other pollutants (fresh motor vehicle exhaust, ultrafines, products of incomplete combustion, others depending on the source (e.g., wood smoke));
- Differential migration;
- SES factors (e.g., access to health care, financial status); and
- Secular trends in obesity.

Of these, he thought both the ACS and the Six Cities studies had done a good job of exploring the impacts of the first three. His view was that neither smoking nor occupation was influential in biasing the results. As a result, he thought even the early

cross-sectional studies needed to be re-considered. Nor did he think differences in criteria pollutants made a "demonstrable difference." "[The time-series data] … show the same $PM_{2.5}$ coefficient across high and low ozone, NO_2, SO_2 and CO communities." When asked about the SO_2 effect shown in the ACS study, Expert C thought that it was not likely to be a "true confounder, in that it is a precursor … [with a] significant correlation [with fine PM], but that it's probably not causal." Similarly, "[s]ulfate has been an extremely good index of fine particles over years, but is it because it's causal or because it's a surrogate for something else, like perhaps nickel and vanadium? Or other pollutants from coal [or oil] combustion."

The remaining issues were ones that he saw as more hypothetical concerns, ones that he lacked information to characterize more definitively.

We discussed whether other, currently unmeasured, pollutants might be confounders for the PM effect. Expert C did not think that there was likely to be a differential across cities in terms of indoor-generated pollutants. He thought wood smoke "could be a confounder, but probably not in terms of large population influence." He thought the ACS and Six Cities studies reflect largely urban populations with relatively little wood burning.

He thought that a recent series of studies looking at traffic as a risk factor raised the question of possible confounding by "fresh" motor vehicle exhaust (which could include ultra-fines and/or products of incomplete combustion). Ultimately, this factor contributed to his uncertainty about the $PM_{2.5}$ effect on mortality (he was uncertain of the direction of bias) and he gave this factor a score of 2 (which he equated to plus or minus 20-30 percent).

Expert C engaged in a discussion of obesity (and with it diabetes) as a risk factor. During this discussion Expert C made comments indicating that it might be a confounder of the mortality effect (e.g., higher rates of obesity in low $PM_{2.5}$ areas, as in the rural south, which could lead to an underestimate of the $PM_{2.5}$ relative risk), and that it could function as an effect modifier of the mortality risk estimate ("… if the relative risk is greater in obese people, then [the effect estimate] would be understated if we didn't acknowledge that obesity was increasing [in the U.S. population]." He gave obesity a score of 2.

Finally, Expert C discussed differential migration and its impact on the PM-mortality effect estimate. He thought it might bias the estimate toward the null because he thought, on average, people might be more likely to move from the "Rust Belt" to the "Sun Belt" and therefore from higher to lower exposures. He cited the Roosli et al. and Schwartz et al. papers that suggest that, "it's the most recent years of exposure which determine mortality rates … [and if the population] moved to a cleaner environment more than three or four years ago, that would … bias it downward." He ultimately thought it would have a minimal effect on the $PM_{2.5}$ relative risk (score = 1 which he characterized as <10 percent change).

3.6 Effect Modification

Expert C discussed two potential effect modifiers, race and SES (as indicated by educational attainment). Race as an effect modifier was discussed extensively; Expert C did not think that race, per se, would necessarily predict response to air pollution but that it might be a proxy for other risk factors (e.g., SES, access to health care, diet) for early mortality. He noted that the Six Cities and ACS cohorts were largely white. Expert C felt it was largely beyond his expertise to assess the impact of excluding other races from these cohorts, although his intuition was that it would tend to bias the PM effect estimate downward. However, given that race was not reported in the Krewski re-analysis of the ACS as a significant risk factor, he thought the impact might be minimal (score = 1).

Expert C's views on the effect of educational attainment (as a proxy for SES factors) on the relative risks for $PM_{2.5}$ were influenced by and expert's discussion at the pre-elicitation workshop, in which he presented estimates showing that different approaches for adjusting for educational attainment increased the relative risks by 30-50 percent. Expert C consequently thought this factor merited a score of 3.

3.7 Exposure Issues

Expert C initially raised several exposure issues that might impact the $PM_{2.5}$-mortality effect estimate:

- Exposure measurement error;
- Compositional variation across the country;
- The time course of exposure, relevant exposure period for the mortality effect; and
- Air conditioning usage.

Expert C thought exposure measurement error, resulting from use of single compliance monitors for a large population was an important potential source of bias for the ACS study. He thought the Six Cities Study did a better job of exposure measurement. Although he thought the Jerrett et al. (2005) study's improvement in exposure measurement was at least partly responsible for the increased $PM_{2.5}$-mortality effect estimate relative to the original ACS study, he was not certain that it was all attributable to improved exposure metrics. He thought a difference in components of the $PM_{2.5}$ in the Los Angeles area might also have played an important role. He scored the exposure measurement error issue a 3.

Expert C cited the National Morbidity, Mortality and Air Pollution Study (NMMAPS) publications with daily mortality showing that the mortality coefficients vary by region of the country (e.g., higher in the northeast) which suggests to him a role for the composition of $PM_{2.5}$. For example, he thought that there is more acidic aerosol in the northeast than in the rest of the country. However, he recognized that this "doesn't necessarily mean it has an influence on annual mortality" although he thought the Jerrett et al. paper might suggest, in part, the influence of more toxic $PM_{2.5}$ in southern

California. For the Six Cities Study, he argued that the PM mixture "is not that different from a lot of the U.S. and Europe, and therefore it's a reasonable estimate for … the U.S. effect … There probably is some bias, but it's indeterminate."

Expert C also acknowledged that the relevant period exposure for estimating the $PM_{2.5}$-mortality effect was not known. He thought the intervention studies, as in Hong Kong, suggest that more recent exposure is more relevant to mortality risk, rather than the previous 20-40 years, although he found that judgment somewhat counterintuitive. He indicated that one rationale might be that the mortality risk is higher among the older, frailer population, than in the young.

Noting unpublished work [by George Thurston] showing higher effect estimates between areas where air conditioning use is low (e.g., San Francisco) compared to where air conditioning use is more prevalent (e.g., Houston), Expert C thought air conditioning could be an important factor in determining effect size by reducing the correlation between central site monitoring data and personal exposures. Furthermore, he noted that use of air conditioning has been increasing nationwide over time. He thought it might be influential in the ACS study and could be reducing the PM effect in southern cities. Expert C expressed the view that air conditioning would have led, therefore, to a negative bias in the overall ACS effect estimate as compared to a true national PM effect. But it appears that his "true national effect" in this case was for a hypothetical U.S. population that did not use air conditioning. Expert C was not sure that the distribution of air conditioning use in the ACS study was un-representative of the US distribution of air conditioning usage. Thus the issue of whether the ACS effect estimate was biased in relation to the actual U.S. adult population remained unclear. He went on to suggest that air conditioning was likely to be less prevalent among the poor and that, because the ACS study was non-representative of the U.S. with respect to educational attainment, air conditioning use and its influence on exposure may underlie some of the effect modification by educational attainment observed there. For the Six Cities study he thought that there was minimal bias but that it was in the direction of overestimating the mortality risk in the U.S. since there is probably less air conditioning use in those six cities compared to the U.S.

3.8 Causality

Expert C argued that both animal studies and human epidemiological data are important for establishing plausibility of the $PM_{2.5}$/mortality association. The recent experiments (discussed under mechanisms) with the atherosclerotic mouse model (the $ApoE^{-/-}$ mouse) showing accelerated development of atherosclerosis in mice exposed to CAPs as well as "prompt and substantial changes in heart rate variability and heart rate" add a "new increment of plausibility for particles being causally related to cardiac disease endpoints." For epidemiological data, "the best evidence is the so-called 'found' experiments, where there have been sudden changes in pollution [Dublin, Hong Kong, European studies of children showing changes in respiratory function]" corresponding to changes in mortality. Expert C felt that the fact that studies all around the world have demonstrated an excess daily mortality associated with fine particles, an association that has not been

successfully challenged, establishes plausibility. He thought that the ACS cohort, Six Cities cohort, the Netherlands cohort (Hoek et al., 2002), AHSMOG, and even the Veterans cohort all "provide essentially the same message" for annual mortality.

He thought both relationships with short-term and long-term exposures were very likely to be causal though he thought it was stronger for long-term because of the greater amount of mortality associated with long-term exposures. He was certain in both cases that the relationships were likely to be causal, comparable to his certainty that smoking increases the risk of mortality.

In developing his final uncertainty distribution, he ultimately settled on 99 percent likelihood of a causal relationship, which he felt accommodated his views on the strength of the evidence and was consistent with the probabilistic distribution of effect estimates he specified.

3.9 Thresholds

Expert C thought cohort studies based on large, representative populations, not clinical or animal studies would, in theory, be most suitable for detecting a threshold if one existed, but that it would require such a large population and so much clinical follow-up that it would be prohibitive.

While Expert C thought that there are "clearly" thresholds for individuals, he would not expect to see a population threshold, even if one existed, given the population variability in sensitivity (due to genetic differences or pre-existing conditions, for example), in the general population of humans. He did not think there was a conceptual basis for arguing that a population threshold might exist in the 4-30 $\mu g/m^3$ range for this study. Although he noted that the ACS study was directly informative about exposures only down to around 10 $\mu g/m^3$, and that a "modest extrapolation, even down as far as 4, is not unreasonable…[compared to] orders of magnitude as you do in carcinogen risk assessment."

His opinion on thresholds did not vary for short-term and long-term exposures. He did not elect to incorporate a threshold into his C-R function.

3.10 Other Influential Factors

As discussed to some extent earlier, Expert C thought that PM composition was likely to be influential. He thought the operating assumption of the protocol was that "composition made no difference, but that clearly can't be correct. Not that I can tell you in specific terms." He stated that recent studies have shown that composition might be influential on mortality estimates. He thought that future regulatory scenarios would likely be designed to deal with source-related or individual components as data improves.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert C chose to base his judgments on a linear, non-threshold (log concentration, linear response) function over the entire range of concentrations included in the study (4-30 $\mu g/m^3$). "I'm not absolutely sure that it is truly linear, but the linear approximation is the best estimate that can be made, in my judgment at this time. Clearly it no longer is linear as you get to much higher concentrations." He cited the Pope et al. (2002), and Laden et al. (2006) comparison of the original and extended relationships as influential. Based on our earlier discussion, the likelihood of a causal relationship is assumed to be 99 percent.

Although we began by trying to establish the maximum or upper bound for the $PM_{2.5}$-mortality effect, Expert C found beginning with the extremes very difficult. He asked to start with a midpoint and work out to the extremes considering the uncertainties we had discussed.

He approached his estimation of the midpoint in two ways; one beginning with the central estimate from the Six Cities study (Dockery et al., 1993) and the other beginning with the Pope et al. (2002) ACS study. He began with the estimate of 13 percent per $10\mu g/m^3$ from the original Six-Cities study rather than the ACS study, whose estimate he thought was low given the non-representative population and degree of exposure misclassification. He was concerned that the Six Cities study might be biased a little high relative to a national estimate given its focus on the northeast where he thought the toxicity of PM might be greater and because of likely less use of air conditioning. But after some discussion of these issues, he did not consider them large enough to change his estimate, so he chose to maintain his central estimate at around 13 percent based on the Six Cities study.

As an alternative approach to deriving the median estimate, based on the ACS study estimate of 6.2 percent per $10\mu g/m^3$, he argued that it should be increased to about 9 percent as a crude adjustment for educational attainment as discussed in the workshop. Adjusting further for exposure misclassification, he arrived at about 13 percent again, noting that although the Jerrett et al. (2005) study reported a measure of 17 percent, he thought that composition played some role in that effect measure.

Considering both studies and the inexact rationale for various adjustments, he settled on a value of 12 percent per $10\mu g/m^3$ for his 50^{th} percentile.

He then worked on developing an upper 95^{th} percentile, which he estimated by not quite doubling his median estimate to about 20 percent. These were converted to values per $1\mu g/m^3$ change in $PM_{2.5}$. He then asked that a Gaussian distribution be fit to these two percentiles, and the elicitation team generated the remaining percentiles of his uncertainty distribution using Crystal Ball™. He allowed the distribution to be truncated at zero, leaving about 1 percent probability of a non-causal relationship. He felt this was largely trivial and consistent with his views on the strength of the $PM_{2.5}$-mortality relationship.

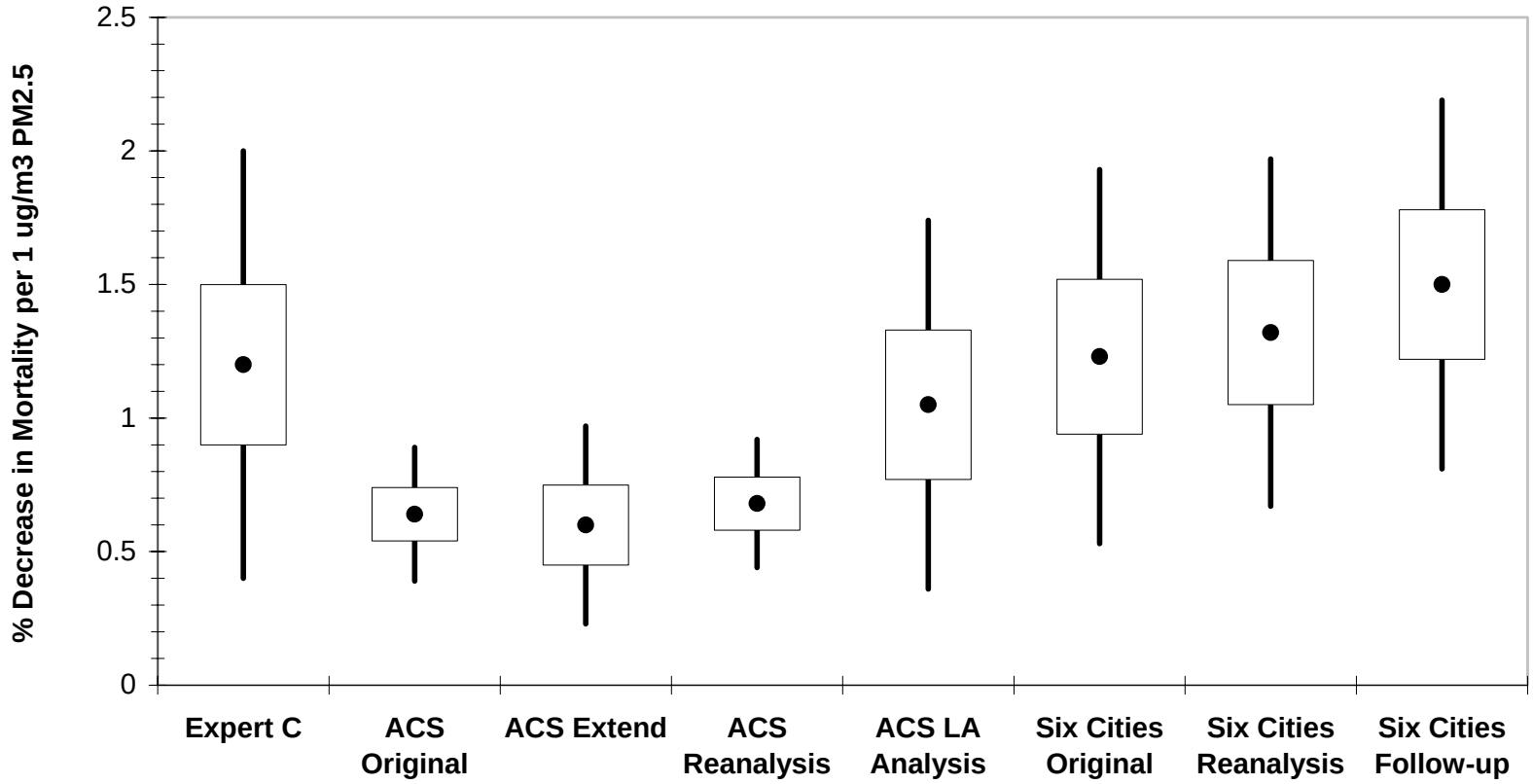
Expert C chose not to specify a maximum value for his distribution because the Gaussian distribution, on which his C-R function distribution was based, has an asymptotic maximum.

Comparing his distribution to some of the other studies on which he relied, he felt that specifying an upper 95th percentile that was comparable to those based on statistical error for the Six Cities Study was justifiable. "You could be more certain. No reason why weighing all the evidence doesn't shift you to be more or less certain."

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations

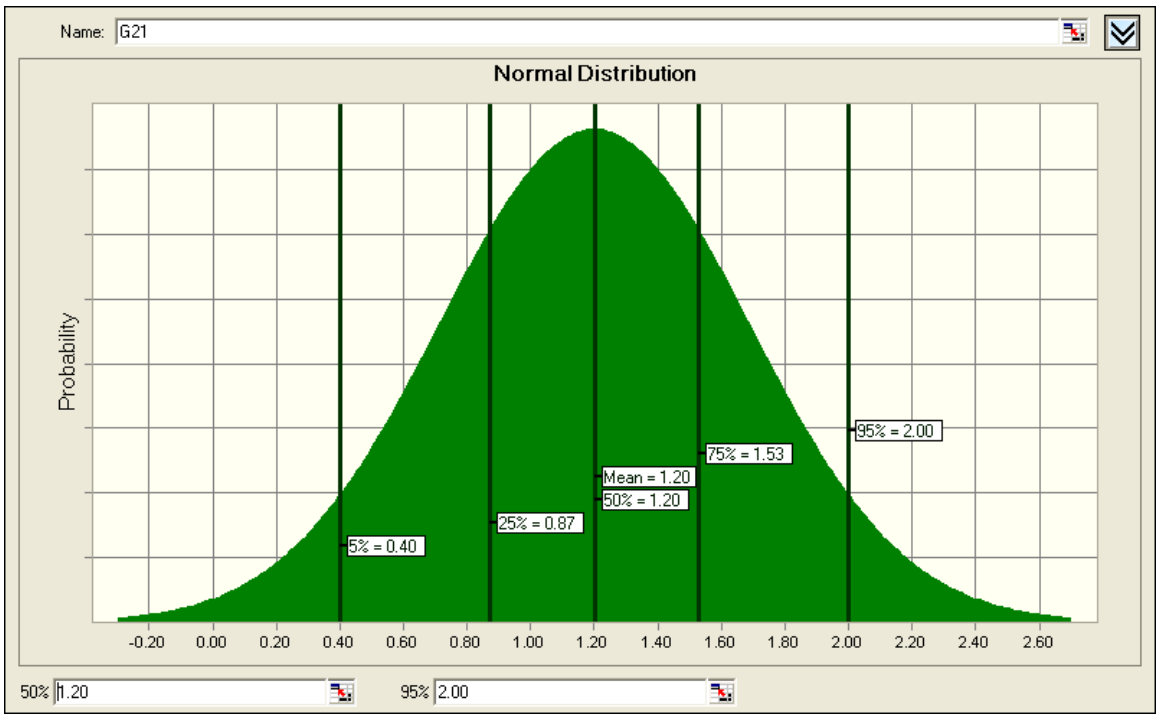
Percentile	Percent Change in Mortality
Minimum	0.0
5th	0.40
25th	0.90
50th	1.2
75th	1.5
95th	2.0
Maximum	-

Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 percent Confidence Intervals for Various Studies to Distribution from Expert C



• = median □ = interquartile range | = 90 percent confidence interval

Expert C Distribution - Probability Density Function



Expert D
Interview Summary

Interview Summary
Expert D[6]

Expert D began our discussion by providing a conceptual framework, informed by evidence, for explaining the possible relationship between mortality and exposure to fine particles. He essentially described a microenvironmental model in which individuals spend varying amounts of time in different microenvironments defined broadly by differing proximity to roadways. It was within this context that Expert D discussed remaining uncertainties in the PM mortality relationship.

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

A. Epidemiological estimates considered most influential to my decision:

1. Six Cities Study (Dockery et al., 1993; Laden et al., 2006; Krewski et al., 2000)
   Six single citywide $PM_{2.5}$ data for first study ('74-'89)
   Estimated $PM_{2.5}$ for follow-up ('90-'98)
   Only white adults
   $PM_{2.5}$ ↓ from first to follow-up
   RR ↓ from first to follow-up (using numbers in Laden)
   1.7 to 1.3 %/µg

2. ACS Study (Pope et al., 1995, 2002; Krewski et al., 2000)
   50 metropolitan area $PM_{2.5}$ from zip code for '82-'89
   Estimated $PM_{2.5}$ for extension to '98
   Enrolled adults
   $PM_{2.5}$ ↓ in all areas, with much greater ↓ for areas over 20 µg/m^3
   RR ↑ from first to follow-up (using numbers in Pope '02)
   0.4 to 0.6 %/µg
   RR higher if adjusted for education level

3. L.A. ACS cohort (Jerrett et al., 2005)
   L.A. by zip code (267 zip codes interpolated from 23 area monitors)
   RR was lowered by progressive controls in model
   1.1 %/µg after controlling for the most variables
   No apparent freeway proximity effect (??)

4. Elderly in L.A. (Enstrom et al., 2005)
   Two periods: '73-'82, and '83-'02
   RR ↓ between periods
   0.4 to 0.0 %/µg

[6] Expert D provided his own summary. We have supplemented it, where necessary to provide details for comparison with other experts.

B. Epidemiological studies providing useful supporting information:

1. Pope et al., 2004
   ACS nationwide – cardiovascular
   RR for CV + diabetes (?) higher in former and current smokers vs. never smokers
   1.2, 1.3, and 1.9 %/µg, respectively

2. Lipfert et al., 2006
   Men – VA – nationwide
   RR was higher for **traffic proximity** than for $PM_{2.5}$

3. Pope et al. Utah Valley (1989; 1991)

C. Toxicological support for quantitative estimates of RR for human mortality

None suitable for quantitative estimates

D. Toxicological support for plausibility and mechanisms (among many)

1. Utah Valley Studies (reviewed in Ghio, 2004)
   Coherence of toxicological and epidemiologic studies
   Inflammation
   Importance of soluble transition metals

2. Dog model of myocardial ischemia (Wellenius et al., 2003)
   Boston CAPs @ 345 µg/m^3 x 6 hr x 5 d
   Alterations in ST segment
   Multivariate analysis > Si

3. CV effects in ApoE mice (Sun et al., 2005)
   Regional NY CAPs @ 85 µg/m^3 x 6 hr x 5 d x 6 mo
   Increased HR and atheromatous vascular changes

4. LRRI studies
   a. DOE studies of PM and co-pollutants in engine emissions
      Both PM and non-PM fractions are important to several effects
   b. NERC repeated exposures to combustion emissions at relevant levels:
      Altered cardiac electrophysiology (diesel, gasoline)
      Oxidative vascular injury (gasoline ± PM)
      Enhanced atheromatous responses (gasoline ± PM)
      Altered systemic immune competence (diesel and wood smoke)
      Altered resistance to respiratory infection (diesel and gasoline)

E. Slope of exposure-response function

1. Best sources of information

EPA PM Staff Paper (2005), Figure 3-4, p 3-57
Pope and Dockery (2006), Figure 1

2. Shape and assumptions

No biologist would believe that the D-R curve is actually linear throughout its length.

Current epidemiological data do not provide enough information to select any other than a linear function in the range of current annual concentrations.

If anything, the slope is estimated to steepen in the low range, but the uncertainty expands too much in the low range to estimate a non-linear function.

F. Major Sources of Uncertainty

1. Responsibility of co-pollutants for a portion of effects ascribed to PM

Only a few co-pollutants are measured. Epidemiologists have no way of accurately distinguishing effects of pollutants that are not measured. An indirect assessment can be made by assessing "PM" effects in locations having different co-pollutant levels. There are some data that may satisfy this strategy for co-pollutants that are typically measured (e.g., NAAQS pollutants), but I am not aware of studies that have proposed to have done this for unmeasured co-pollutants.

Evidence from toxicological studies, strongly suggests to me (but does not prove) that unmeasured co-pollutants are responsible for a portion of effects attributed to PM. These effects include inflammation, atherogenic vascular responses, and resistance to respiratory infection, and may include immune/allergic effects and developmental effects.

This view is based almost exclusively on studies of fresh, or relatively fresh traffic emissions. Because I believe that current evidence points toward engine emissions PM as the most toxic PM overall (toxicity x prevalence of exposure), the most important exposure to unmeasured co-pollutants occurs concurrent with the most important exposure to PM. This concurrence adds to the plausibility that co-pollutants cause a portion of the effects ascribed to PM.

Current data do not allow the confident parsing of effects among different unmeasured co-pollutants, measured co-pollutants, and PM. However, based on current data, I consider gas and vapor-phase organics (VOCs, SVOCs) to be very likely among the culprits.

It is true that many VOC and SVOC species will migrate into the PM population with distance from source; moreover, the organics will be changed

by oxidation, chemical reactions, etc. Some portions of this material will become more toxic and some will become less toxic. We don't know enough yet to deal with this in a quantitative manner. Regardless, it is also true that all exposures in close proximity (on-road or near roadways) to engine emissions include exposures to these unmeasured non-PM emissions. There will be a declining concentration of these emissions with distance from the roadway, which parallels the declining concentration of PM (and especially ultrafine PM). This concurrence of distribution lends itself to ascribing the effects of unmeasured co-pollutants to PM.

Credible epidemiologists (e.g., Samet, Krewski) acknowledge that existing epidemiological data cannot completely separate effects of PM from co-pollutants
Credible groups (e.g., NRC committees, HEI) acknowledge need to consider exposures as mixtures, reflecting an eroding confidence that one can confidently ascribe effects to single pollutants (based on current data), or disentangle effects among the hundreds of physical-chemical species that people actually breathe.

Elicitation ground rule assume no change in nature of co-pollutants. This is interpreted to mean that the proportional "silent" contribution of co-pollutants will remain the same, and that any overestimate of PM effects due to this factor will remain the same. Of course, the nature of both PM and co-pollutants will change.

2. Accuracy of PM exposure estimates

None of the predominant epidemiological studies actually measured personal exposures

Portions of the data are not even direct measures of $PM_{2.5}$.

It is usually assumed that all personal PM exposure is proportional to, or identical to, the average PM concentration at the place of residence.

The <u>most exposure</u> to the <u>most toxic PM</u> is probably in or near traffic, and that is likely to differ substantially from exposure at the residences of people living more than 100-300 meters from a major roadway.

The <u>most exposure</u> to <u>co-pollutants most likely to add to the effect</u> (silently, because they are unmeasured) would also occur on or near roadways.

The "traffic effect" is believable with a fairly high level of confidence. The Lipfert (2005) study, and many others (e.g., multiple studies from the Brunekreef group, the Peters MI study, etc.), support the notion that <u>traffic emissions are likely to be key</u> to many health effects associated with PM. Most of these studies either did not attempt to single out the effects of PM, or had low power to do so. Nonetheless, the importance of proximity to traffic has been convincingly demonstrated

To the extent that the effects are indeed due to PM, and if most of the most important PM exposure (C x T x toxicity) occurs on or near roads and is thus underestimated by exposure at residence, and further, if that exposure parallels the estimated exposure at the residence in a roughly linear fashion, then the "unit risk" has been overestimated because the most important exposure has been underestimated. The proportional change in mortality with decreasing PM could still be approximately correct, even if the unit risk, and thus the body count, attributed to PM is overestimated.

No epidemiology studies of $PM_{2.5}$ have been able to deal seriously with PM composition; incrimination of traffic PM is almost exclusively by location, not composition. My (unproven) view is that good composition data for personal exposure would support my view of the importance of traffic PM.

3. Changes in risk with time

Both the 6-cities and ACS studies, as well as the Enstrom study, indicate that unit risks have changed with time.

Risk went down in 6-cites and Enstrom
Risk went up in ACS

Regardless of the real, or if the direction has been different in different locations, this adds to uncertainty about impact of future changes

The nature of $PM_{2.5}$ will likely change with time, which could change risk; however, the elicitation ground rules preclude this consideration.

4. Plausibility (as defined by identification of potential mechanisms)

There is not much of a roadblock now to believing that PM may contribute to the effects associated with it by epidemiology – at least superficially. Findings with time are providing increasing, rather than decreasing, mechanistic plausibility.

Experimental exposures of humans tend to support, rather than detract from, plausibility. The problem is that we can't conduct the studies that would be most informative – exposure of the most sensitive people to the most toxic PM, under the most hazardous conditions, using the most discriminating endpoints.

Toxicology is providing increasing mechanistic plausibility, but still very little "dose plausibility". The biggest problem with plausibility today is confidence in high dose to low dose extrapolation and confidence in animal model to human extrapolation. The former can be fixed by improved study design. The latter will probably improve, but will never be eliminated. Overall, we have enough confidence in the general utility of the models to do informative work - if investigators would extend the dose

range downward. The problem of course, is that when that is done, effects are very hard, and sometimes impossible to demonstrate - especially inexpensively (e.g., short exposures, simple atmospheres, few measured variables, small groups of animals, rapid-response models for chronic human effects, lack of methodical comparisons among exposures). There may be a message here – in fact, it is probably true that some (perhaps many) of the effects that are published in the toxicology literature do not occur in either animals or humans under realistic exposure regimes. Of course, the alternate outcome is also possible; it could be that repeated exposures of large numbers of animals of diverse susceptibility would, in fact, reproduce the epidemiological findings under realistic exposure conditions.

Expert D characterized the likelihood of a causal relationship as ranging between 90 and 100% with a most likely value of 95%.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

G. My answer: % change in all cause mortality associated with a 1.0 $\mu g/m^3$ change in PM

It all boils down to selecting the studies in which I have the most confidence and generating a distribution from those results using my biases regarding their likelihood of having overestimated or underestimated the actual value.

I pick the 6-cities and ACS bodies of work as the most likely to approximate the truth, although both involve uncertainties. I pick these because:

- Multiple locations in U.S.;
- Large numbers of people;
- Credibility of investigators;
- Repeated analysis with extended follow-ups; and
- Withstanding lots of scrutiny by other credible analysts.

Other studies add texture to our understanding and expansion of updated methods (e.g., more localized exposure estimates, further parsing of the population, better characterization of regional differences in PM and response slopes) may yield different results. However, I choose to answer on the basis of existing data, not by imagining where evolving knowledge will take us.

I start with a "most likely" value derived by a simplistic average, as shown below. These values have some numerical justification because they are derived from published results, and they are the most recent from the two bodies of work that I have selected as most informative. There are arguments from more recent published work for higher values, but I avoid going higher for two reasons. First, the more recent results are not derived from studies of comparably large populations in multiple U.S. locations. Second, and perhaps most reflecting my bias, I believe that the effect of PM has been overestimated across the entire epidemiological database, for the reasons I state above. On the other hand, I don't

have sufficient scientific basis for selecting an arbitrary "most likely" value that has no basis in published work. I do not have high confidence in this number.

6-cities	(most recent)	1.3
ACS	(most recent)	0.6
Mean	(most likely value)	0.95

I next provide the following perspectives on the distribution around the above "most likely" value:

Both of the 6-Cities and ACS values above must be plausible; i.e., within the bounds of 90-95% confidence limits.

If the real value is not the "most likely" value of 0.95, it is more likely below that value than above it; thus the distribution of probability must be weighted between the "most likely" value and the lower bound of the distribution. I base this on my bias that the current data most likely miss-assign at least a portion of the effects of co-pollutants to PM, and that there is as much, if not more, evidence that the unit risk for PM is decreasing with time, rather than increasing.

The lower bound of the actual value could be very low. On the basis of present knowledge, it could not plausibly be zero, but I don't exclude the possibility that it could be near zero. I arbitrarily select a lower bound of 0.1.

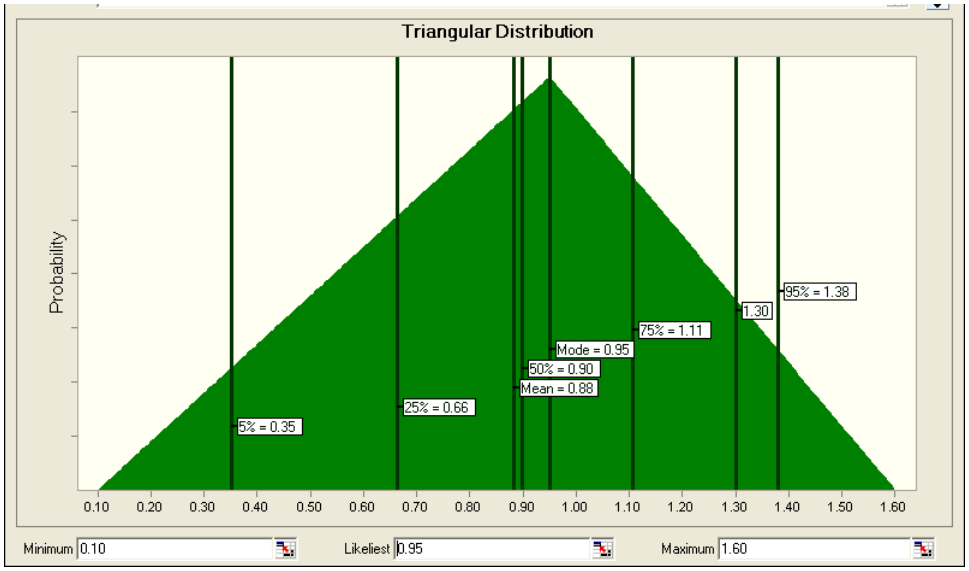
In keeping with my biases regarding current data, I don't think that the upper bound of the actual value could be a great deal higher than the "most likely" value I selected. I do not have a solid rationale for picking an upper bound. I arbitrarily select a value of 1.6, which is near, but slightly below, the original 6-cities estimate of 1.7.

I next selected a probability distribution that best fits the above criteria.

I relied on the elicitation staff to provide a range of distribution functions that might fit my criteria. Most standard distributions do not fit the criteria very well. Although my criteria are sufficiently arbitrary that I could have elected to change them to better fit a standard distribution function, I did not do so.

All things considered, a simple triangular distribution fit my criteria best, using the "most likely" value as the mode and fitting to the upper and lower bounds.

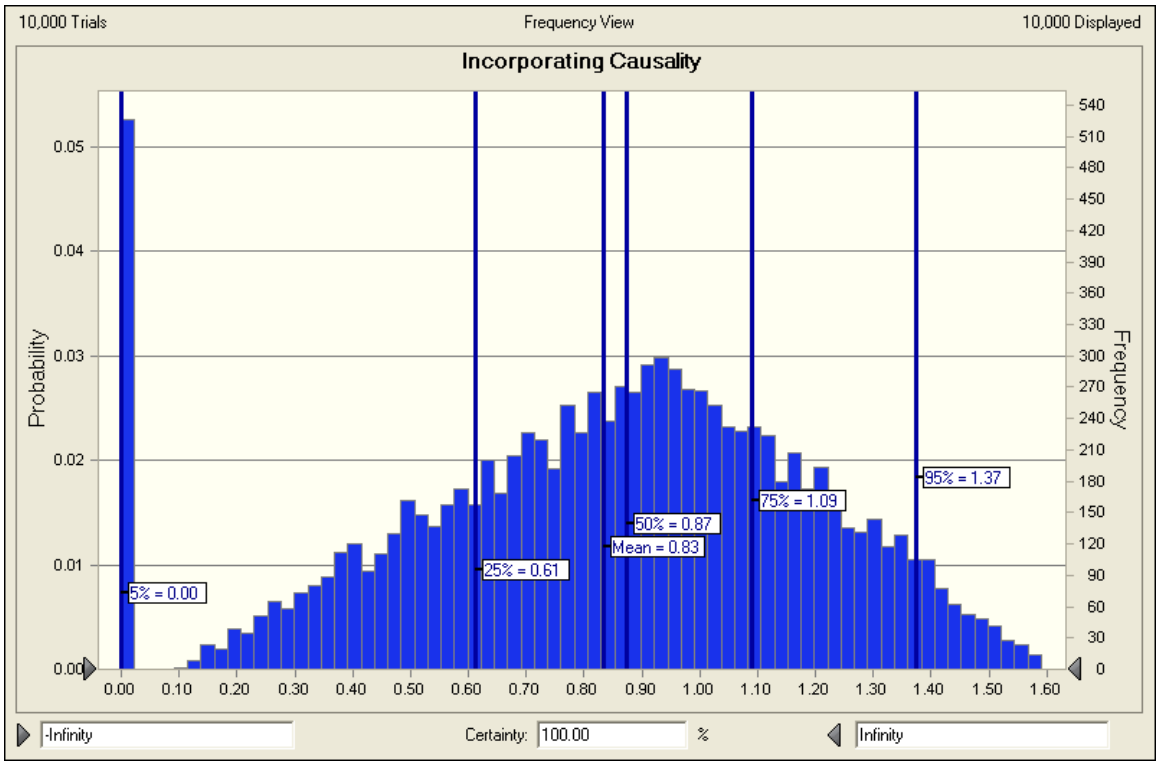
Elicited Distribution



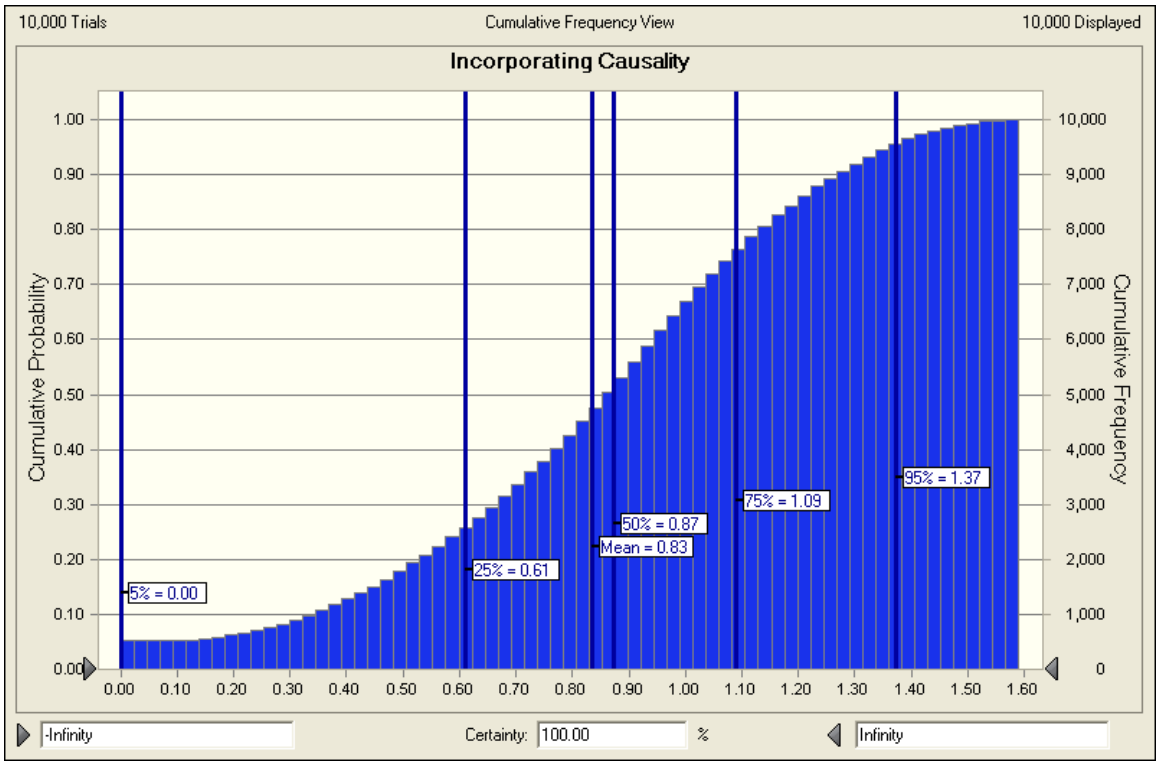
Min	5th %ile	25th %ile	50th %ile	75th %ile	95th %ile	Max
0.10	0.35	0.66	0.90	1.1	1.4	1.6

Expert D chose to provide a C-R function that was conditional on the existence of a causal relationship. The elicitation team then combined his conditional distributions with his percent likelihood of causality (95% likelihood of a causal relationship). A probability density function (PDF) and a cumulative density function (CDF) of Expert D's distribution incorporating causality as well as the are below:

Distribution Incorporating Causality –PDF (IEc Generated)



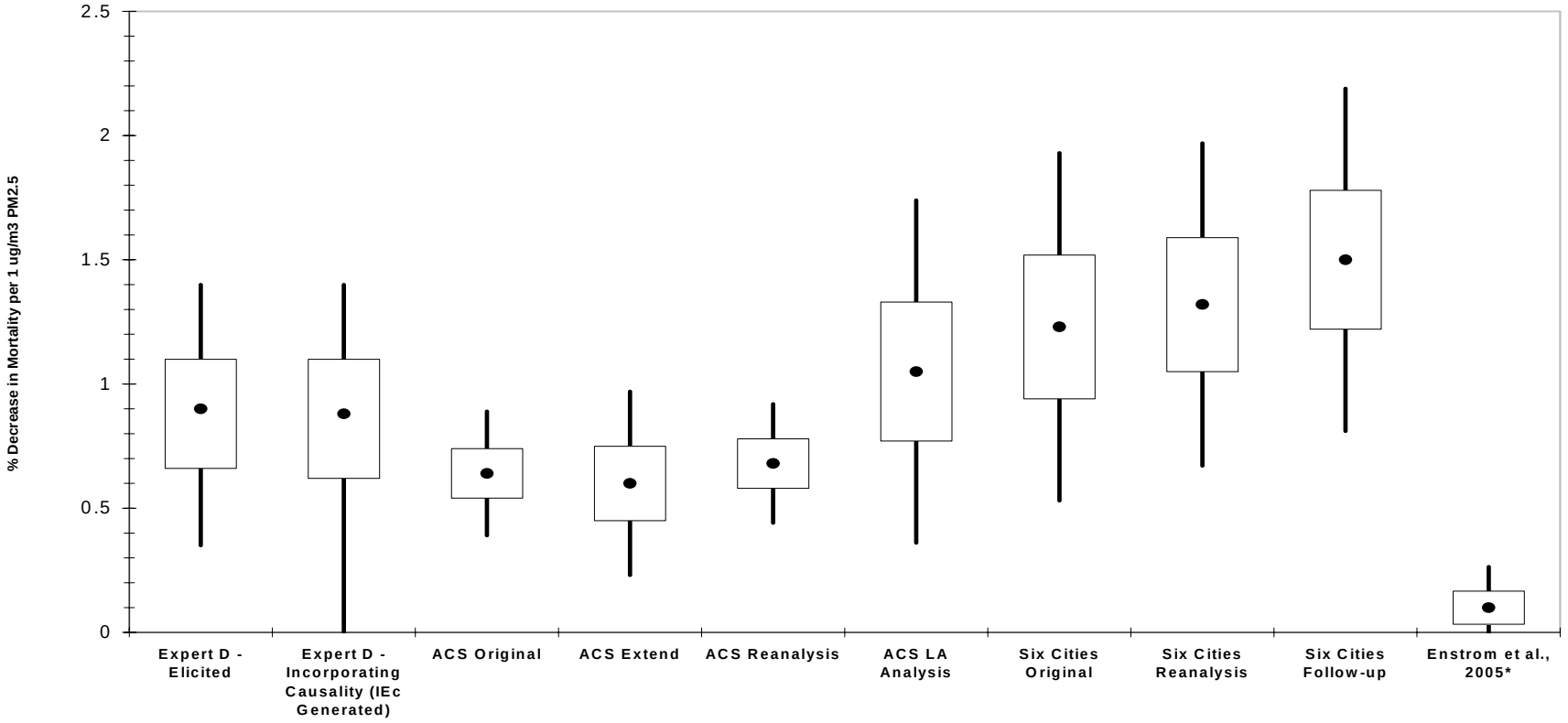
Distribution Incorporating Causality – CDF (IEc Generated)



Percentiles of Distribution Incorporating Causality (IEc Generated):

Min	5th %ile	25th %ile	50th %ile	75th %ile	95th %ile	Max
0	0	0.62	0.88	1.1	1.4	1.6

Exhibit 1: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 % Confidence Intervals for Various Studies to Distributions from Expert D



• = median □ = interquartile range | = 90 % confidence interval

* Estimate for entire follow-up period.

Expert E
Interview Summary

Interview Summary
Expert E

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert E discussed the biological mechanisms for short-term and long-term exposures together because he thought that it was difficult to separate them, since the physiology is not well understood enough to do so. He first discussed cardiovascular disease as a cause of death from particulate matter (PM). He thought the main mechanisms related to this cause of death were oxidative stress and inflammation leading to atherosclerosis, changes in autonomic function, and arterial reactivity. He also thought chronic obstructive pulmonary disease (COPD) and lung cancer could be potential causes of death.

Cardiovascular Disease

He first discussed a series of recent studies that he thought lent biological plausibility for a relationship between particle exposure, oxidative stress and inflammation, and ultimately, atherosclerosis and plaque instability. He cited an animal study by Gurgueira et al. (2002) that involved exposing rats to concentrated ambient particles (CAPs) and showed increases in reactive oxygen species in the lungs and heart. He cited another study by Evelson & Gonzalez-Flecha (2000) that placed animals who had been exposed to typical ambient particulate levels (~11 $\mu g/m^3$) in filtered chambers, thereby removing their PM exposure, and found that the reactive oxygen species levels decreased. In addition, he cited *in vitro* and *in vivo* studies showing that "part of [the body's] defense mechanism against oxidative stress, seem to modify the effects of particles [i.e., reduction of pro-inflammatory cytokines, lung inflammation]" (Gurgueira et al., 2004; Rhoden et al., 2004). He indicated that "to me, that makes endothelial dysfunction plausible; it makes myocardial infarction plausible; it makes arrhythmias plausible." He thought that a study that found in populations with lower defenses against reactive oxygen species (e.g., individuals with genetic polymorphisms of genes that impair the ability to defend against oxidative stress and obese individuals), there are higher effects from PM (Schwartz et al., 2005) further supported the reactive oxygen species theory.

Expert E thought that recent work by Kunzli et al. showing a relationship between PM exposure and intima-media thickness suggests another connection to oxidative stress and a risk factor for cardiac mortality. He indicated that there is evidence from studies in rats that suggests that PM–related oxidative stress contributes to inflammation in the endothelium (the lining of the arteries), which could lead to atherosclerosis (Rhoden et al., 2004). As further evidence of PM's inflammatory role, he cited a CAP study in rats by Dvonch et al. (2004) that showed plasma asymmetric dimethylarginine (ADMA) levels to be significantly elevated in rats exposed to CAPs versus those exposed to filtered air.

Expert E indicated that recent studies of brachial artery reactivity, a measure of the ability of the arteries to dilate in response the body's demand for blood flow, suggests a connection between both long- and short-term exposures with cardiac mortality. The more "stiff and impaired the arteries get," the less responsive they can become. A recent study in *Circulation* (O'Neill et al., 2005) found decreases in brachial artery reactivity related to the previous couple of day's air pollution (both black carbon and sulfates). He indicated that the percent increase in arterial dilation was less in diabetics and individuals with coronary artery disease.

Expert E added that studies by Van Eeden et al. exposed animals with atherosclerosis to PM and found that short-term exposures lead to decreased stability of plaques. He thought that this, coupled with thickening of the arteries, could increase the risk of a myocardial infarction. On a related note, he indicated that there was increasing evidence for associations between ischemic stroke and PM. He said that there were multiple studies done in Korea where there is a higher incidence of stroke than in the U.S., so there is more power to see associations in that population. He cited studies by Zeka et al. (2005), Wellenius et al. (2005), and Dominici et al. (2006), all of which found significant associations between PM and stroke. "Ischemic stroke is the same thing as a heart attack, only in the brain. Something's blocking some arteries, say a plaque got ruptured, and it cuts off the supply of blood to some of the brain tissue."

He also thought that both animal and human studies have shown that PM could cause changes in heart rate variability (HRV). In animals PM seems to increase HRV (rather than decrease, as in humans), but he thought there was evidence that it occurs through the inflammatory pathway described above; when animals are simultaneously given N-acetylcysteine, a precursor to the anti-oxidant glutathione, no changes in HRV are observed. He indicated that he thought the changes in HRV could be connected to lung inflammation, in that HRV is controlled by the autonomic nervous system, which has nerve endings in the lungs that feed back to the control of the heart.

<u>COPD</u>

Expert E thought that, "the fact that particles are associated with increased reactive oxygen species in the lungs seems to make sense for COPD. COPD is chronic lung inflammation, and if you have more inflammation going on in the lung, you're going to tend to get mucus hypersecretion. You're going to tend to get structural damage. And so, generating reactive oxygen species, increasing inflammatory cells and cytokines, it all seems like it would help contribute to develop COPD." He cited studies that have shown epidemiological associations between particles and COPD, including the Adventist Health and Smog (AHSMOG) study, a 1993 study by Schwartz examining National Health And Nutrition Examination Study (NHANES) data, and a Chinese study. He also mentioned the Children's Health Study that found when children moved to a less polluted area, their lung function growth accelerated and when they moved to a more polluted area, their lung function decreased (Avol et al., 2001). In addition, he cited a study by Paolo Saldiva and John Godleski (2002) that showed that the capillary walls in the lungs of the animals that were exposed to CAPs were thicker than those that were unexposed.

Lung Cancer

Expert E discussed lung cancer as a potential cause of mortality related to PM. He indicated that the cohort studies show increased relative risks with lung cancer mortality, but with large standard errors. He also indicated that, "there is a long history of studies showing that there's an urban-rural gradient in lung cancer among non-smokers, which [could be due to] … air pollution." He also mentioned studies looking at measures of air pollution in cities that looked at the distance from the inner city and from industrial zones in the city and risk (Barbone et al., 1995; Biggeri et al., 1996). He thought that the evidence was pointing towards the traffic particles and diesel exhaust as causal agents of lung cancer. He also thought that, "inflammation is definitely a promoter. So, you would think that even when other things are causing the fundamental mutations, that in the presence of chronic lung inflammation and oxidative stress, it's going to develop faster."

3.2. Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

Expert E generally thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures. "I think it's a fine conceptual view to think about how air pollution affects mortality. What's less clear is whether there are regression coefficients that come out of epidemiological studies that … relate to [individual categories] or whether they're all mixed." In reality, he thought it was difficult to see a clear distinction between categories (A) and (C) because "there's no one who's completely unexposed." He thought one concept not captured by the diagram is the possibility that long-term exposures to PM could modify the effects of short-term exposures. In theory, one might expect that individuals who are chronically exposed to high levels of PM would be more susceptible to short-term effects than those with lower long-term exposures. Paradoxically, the empirical evidence to-date points towards *bigger* short-term PM effects in cities with *lower* long-term PM exposures.

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

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 looked directly at mortality changes. "The general pattern that … you see in the cross-sectional studies is buttressed by similar findings for other health endpoints, like lung function and chronic respiratory symptoms, where similar study designs have been done" (e.g., in Germany (Heinrich et al., 2002) and Switzerland (Oglesby et al., 2000)). He indicated that these types of studies not only measure what we are asking for in the elicitation study (risks associated with a change in exposure) but also eliminate confounding by factors that vary across cities. However he noted the weakness is that potential confounding can still exist if some risk factors changed from one period to another.

He thought the next best study design for answering the question posed by the protocol was cross-sectional cohort studies. This study design is less optimal than the previous studies because it requires making the inference that differences in mortality between cities with high and low exposure also translate into similar mortality changes given temporal declines in air pollution. He thought the Laden et al. study (2006), having both temporal and cross-sectional designs, provided evidence that this was a reasonable inference.

He thought that neither the cohort studies with temporal changes in air pollution nor cross-sectional studies would capture deaths "brought forward by a relatively short period of time … but if the people who die today would have lived three years had they not had their heart attack today, then I would think that the cohort study would … capture that."

When asked about the Utah Valley (Pope et al., 1996) and Dublin (Clancy et al., 2002) "intervention" studies, Expert E's view was that they were supporting information but did not provide the coefficients relevant to estimating the impact of long-term changes in PM. He indicated that the Utah Valley study included a dummy variable for year the steel mill closed; "the fact that it was significant tells you that we're not just looking at short-term harvesting and that if you change annual average pollution levels, something happens, but it doesn't tell you what happens to the equilibrium mortality rate after you change pollution for a long time." The Dublin study looks at a longer period of time and so gives you a "more useful coefficient," but "unlike the cohort studies," lacked control for risk factors that might change over time.

He then mentioned time-series studies with long distributed lags (1-2 months). He thought that these did not provide useful quantitative coefficients to estimate the effects of long-term exposure, but show that "acute effects aren't all short-term harvesting … that they persist over months." Finally, Expert E discussed time-series studies with short lags (1-2 days). He thought that these provide "qualitative support" for the hypothesis that "air pollution can kill people" but because they do not capture "any longer-term effects," and therefore, the coefficients are underestimating the total mortality effects.

Expert E's views on the mortality effects captured by each study design are shown in the table below:

Study Design	**Type of Effects Captured (e.g., short-term, long-term, or both)**
Cohort studies looking at changes over time	Long-term
Cross-sectional cohort studies	Long-term
Intervention studies (Utah Valley, Dublin)	Supporting evidence (Intermediate 1+ years)
Time-series studies with long distributed lags	Intermediate (1-2 months)
Time-series studies with short lags (1-2 days)	Short-term

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert E thought that the following characteristics would be part of an ideal epidemiologic study to characterize the $PM_{2.5}$-mortality relationship in the U.S. population:

- Examines changes in long-term exposure to PM in multiple locations with different magnitude of changes, "so that your delta exposure is variable";
- Controls for potential confounding;
- Intermediate time scale ("more than a year or so, but less than 20 years, where there are fewer things that might have changed over time");
- Representative geographic sampling of the U.S.;
- Information about individual risk factors and modifiers of exposure (e.g., window air conditioning);
- Population that is representative of the general U.S. population (e.g., follow-up on NHANES); and
- Collects information on genotype to understand mechanisms and susceptibility.

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert E discussed the following studies:

Study (author, date)	Key findings	Strengths	Limitations
Six Cities Follow-up (Laden et al., 2006)	Significant association of change in long-term exposure with change in risk of dying in the cohort	• Directly addresses question posed by the protocol • Eliminates across cities confounding by design • Less exposure misclassification than cross-sectional cohort studies	• Potential confounding by factors that may have changed over time within the cities • Not geographically representative (does not include west coast)
American Cancer Society (ACS) Los Angeles (Jerrett et al., 2005)	Examined zip code levels exposure estimates and found a larger slope than original ACS study	• Better measure of exposure than original analysis	• Only included Los Angeles, which has a different PM mix than the rest of the U.S. (higher fraction of secondary organic aerosol and nitrates – traffic related)
The Netherlands Cohort study (Hoek et al., 2002)	Examined mortality effects of traffic particles using geocoding to enhance exposure assessment. Found larger coefficients than the original Six Cities study.	• Better measure of exposure than the ACS or Six Cities studies	• Uses black smoke and is a measure of traffic particles, rather than $PM_{2.5}$ • The Netherlands is homogenous for non-traffic particles
Willis et al., 2003	Used the ACS data but restricted it to people	• Better measure of exposure than the original ACS analysis	• Did not provide results for $PM_{2.5}$ (only sulfates)

Study (author, date)	Key findings	Strengths	Limitations
	who lived in the same county as the monitor. Coefficient for sulfates increased two-fold.		
AHSMOG	Qualitatively sees signals with mortality and particles as the mortality in the cohort has increased over time	• Provides qualitative support for mortality effects of PM	• Low mortality (healthy cohort) which leads to low statistical power • Non-representative population • Manipulations of data decreases the applicability of results

Expert E thought the key strength of the Six Cities study is that the population is a random sample of the area of study. He thought that this was important because epidemiology contrasts mortality experiences of samples in different locations with pollution concentration differences. If the sampling scheme produced nonrandom samples, then we need to be assured that it did so in a uniform way, so that it is reasonable to ask whether pollution contrasts across locations are associated with survival. In the case of the Six Cities study, this is not necessary, as the samples were representative. In the case of the ACS study, the sample was not representative because it was a convenience sample with a bias towards high socioeconomic status (SES), and possible other biases, as they were recruited by volunteers. Fortunately, Pope et al. have compared the results of the ACS analysis to a result comparing age standardized mortality rates by city against the pollution differences by city, and found comparable results. This means that the sample is the same subsample in each location, and the relative mortality risks are preserved compared to the general population. Without this, we cannot consider it reasonable to look for pollution as an explanation of differences, since the differences would reflect sampling differences across community, rather differences in risk across community. This supports the use of those two studies. The Six Cities study has the added advantage that it was designed to look at air pollution, and therefore sampled people in defined neighborhoods of each community, with monitors within each neighborhood, and generally within a few miles of each subject.

Given the different aerosol mix in the Jerrett study in Los Angeles (LA) (2005), Expert E argued that, "while directionally it suggests that as you reduce exposure error, you get higher coefficients, it would be [difficult] to take that coefficient and say that that should be applied nationally." On the other hand, he thought that work by Willis et al. (2003) (published in *Toxicology and Environmental Health*) on sulfates was supportive evidence for the exposure error hypothesis. This study found that when you include only 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 compared to including all individuals in the metropolitan-area. He noted that this study did not include the full ACS cohort, but was a more nationally representative sample than the Jerrett et al. (2005) study.

Expert E was also asked about what reliance he might have on the AHSMOG study, the Lipfert et al. series of studies and the recent Enstrom et al. (2005) study. He indicated that he would not want to rely on coefficients from the AHSMOG study to represent the U.S. population because that cohort was highly non-representative (i.e., non-smokers, healthier dietary patterns). In addition, he thought the exposure measure was non-standard, the study size is small, and not many deaths have been observed. He noted, however, that as the number of deaths has increased, the mortality coefficients have increased. He concluded by saying, "I think it's qualitatively supportive."

When asked about the cohort studies by Lipfert et al., Expert E indicated that he did not give any weight to the results; he felt strongly that the industry funding sources were likely to bias the findings and that they are not published in mainstream epidemiologic journals (where they would get rigorous peer-review). In addition, he indicated that the Lipfert study differs substantially from the Six Cities study in the sample design issue. It is clear that the relative ranking of mortality rates across cities for members of the cohort is quite different from the relative ranking of the general population. For example, Lipfert has reported a relatively high mortality rate in Salt Lake City, which in general has a low mortality rate compared to other cities. This violation of the principle that the sampling frame has to sample the same way in each city means that one cannot contrast the community exposure against the mortality experience of biased samples not representative of the communities. A second major concern Expert E mentioned is that the cohort was basically recruited from hypertensive subjects and did not control for cardiovascular medication. There is documented wide variability in the aggressiveness of the use of cardiovascular medication by region of the country, and by deliberately sampling a cohort that maximizes the potential for confounding exposures that are defined geographically, it could lead to bias, unless medications are controlled in the study. Hence findings of no association with sulfates, which are driven by the lower than average mortality rates in the Northeast in this cohort, could be entirely due to the well documented more aggressive use of cardiovascular medications in the Northeast.

He was not as familiar with the Enstrom et al. (2005) work but had similar concerns as for the Lipfert work. He noted that given that Jerrett has reported almost as much variation in $PM_{2.5}$ exposure within Los Angeles county as exists across the US in the ACS study, it is clear that the use of county level averages represents much greater exposure misclassification in California (CA) than elsewhere in the US. For example, there is little difference between $PM_{2.5}$ concentrations in the county north of Boston and the county to the south, and there are fairly similar levels in Philadelphia and Washington, DC. Given the topography in CA and the dominance of traffic as opposed to long range transported particles, he felt that this study moves in the exact opposite direction as that suggested by the Six Cities study, Willis et al., and Jerrett et al.

3.5 Confounding

In general, Expert E thought that all of the potential confounders he discussed had a minimal impact on published effect estimates (less than a 20 percent change) and were well-controlled in the studies that he relied upon for his quantitative estimate. He thought a key point in this discussion is that many people first focus on whether or not a factor is

correlated with outcome but in order for their to be confounding, the factor also has to correlated with exposure. He did not think that there is much evidence that many potential confounding factors actually meet this second criterion. His discussion of confounders in the ACS and Six Cities cross-sectional studies included the following:

- **Age:** Expert E thought this could be an important factor if it is not controlled for adequately. He thought that "in general, cities with older populations, other than in Arizona and Florida, tend to be older industrial cities, where young people are moving out … so it's plausible that exposure and age are correlated." He thought this was potentially more of a problem in the Six Cities study than the ACS study. He indicated that not controlling for age would result in an overestimate of the pollution effect. He thought that the Six Cities and ACS studies adequately controlled for this factor.
- **Smoking:** He indicated that "[s]moking is … a pretty good predictor of mortality, but there's not much evidence in these cohort studies of a correlation between smoking and exposure … and it didn't look like, in either the Six Cities study or the ACS study, the controlling for smoking did much to the PM effect."
- **SO_2**: Expert E indicated that when SO_2 was added to the model in the ACS study, the particle coefficient changed significantly. However, he thought it was unlikely to be contributing to mortality ("concentrations of $S0_2$ in the U.S. are extremely low, and 90 percent of it is stripped out in [the] nasal passages, … [it's] a somewhat reactive gas, the indoor-outdoor ratio is small.") He thought that "SO_2 could only be standing for something else, and the most plausible thing it's standing for is particles. Because SO_2 and particles are highly correlated. I just view that as a model that had two different indices of particles in it."
- **Socioeconomic Status:** He thought that socioeconomic status (SES) could be a potential confounder in the cross-sectional cohort studies, but not in the "change" studies. He said that most epidemiologic studies control for SES by using a variable for education. He thought that family income might be a useful measure to include in the models. He also thought that examining the census block groups, which are "extremely homogeneous on socioeconomic position" would be informative.
- **Diet/BMI:** Expert E indicated that diet was not included in the Six Cities analysis, which was a "major failing" although he noted that it did include alcohol and body mass index (BMI), which does a "pretty good job." He thought that there could be some residual confounding by diet, even after controlling for BMI, but he was not sure how it could be correlated with exposure.

In general, he thought that residual confounding by SES and age in the Six Cities and ACS studies, if it existed, would be more likely to contribute to over estimates of the mortality effect coefficient, but thought it was a minimal impact (score ~1; < 20 percent).

He did not think that occupational history or pre-existing disease were potential confounders in the cross-sectional cohort studies.

Expert E then discussed potential confounding in the Six Cities follow-up (Laden et al., 2006). The first factor he mentioned was smoking, since the study did not record smoking status during the most recent follow-up period. He thought that, most likely, people are quitting rather than starting to smoke, and it's "not obvious that people would quit smoking more or less in places with more or less change in air pollution, although it's always possible." He thought SES and diet were probably not confounders in this study because they would have to change over time. He indicated that it was possible in this type of study design to have confounding if there were changes in medical care and life expectancy over time. He said that the Laden et al. study controlled for a period two versus a period one effect, which attempted to capture all of these factors. Expert E pointed out, however, "if the changes were different in the different cities, and those differences happened to be correlated with the differences in air pollution, then we have a potential for confounding." In general, he thought the direction of confounding in the Laden et al. paper was uncertain and not very large (i.e., that these factors might cause the coefficient to shift by 20 percent one way or another but not by as much as 50 percent).

3.6 Effect Modification

Expert E began his discussion of effect modification using acute studies which he felt provided mainly qualitative insights, particularly about the influence of underlying susceptibilities on PM-related mortality (diabetes, obesity, systemic inflammatory disease, genetic susceptibility). He thought short-term studies have shown effect modification by diabetes (e.g., hospital admissions, flow-mediated dilation studies: Goldberg et al. in Montreal, Rome case-crossover studies), and some evidence of effect modification by obesity. In general, Expert E thought that there are a number of biological phenomena affecting susceptibility (including other systemic inflammatory diseases such as rheumatoid arthritis) that track with age so that age is a possible surrogate for these factors.

Expert E thought that genetic polymorphisms related to susceptibility for developing atherosclerosis (e.g., GSTM1-null mutation) were potential effect modifiers. He noted that the existing cohorts did not have genotypes for the population, but that in the future, "we're going to find that there's actually a fair amount of genetic heterogeneity in the population with respect to things that may turn out to modify the effects of $PM_{2.5}$." He also indicated that, "there are genetic differences between races" and that one would logically expect race to be a surrogate for these, although he acknowledged that short-term studies have not really detected any evidence of racial differences (Zeka et al., 2006 and Zanobetti et al., 2000). He did not think race was something that could be examined in the cohort studies given that the Six Cities study cohort, including the Laden et al. follow-up, did not include minorities and the ACS cohort underrepresented minorities. In theory, he thought that this would lead to an underestimate in the effect estimates, although there was limited evidence that this was occurring.

Expert E discussed potential effect modification by education, which was found particularly in the in the ACS study and to a lesser extent in the Six Cities study. He said

he was "a little skeptical" about this factor coming out of the extensive Krewski et al. reanalysis (2000) in part because of the multiple comparisons problem but also because the Six Cities extended follow-up found effect modification by education for some types of mortality and not others. Still, he thought that it was plausible to some degree because "poor people have a lot of stresses … poor diet, … less air conditioning. So, there are both exposure-related differences and host susceptibility-related differences." He thought that this was more of an issue in the ACS study than the Six Cities study.

Expert E discussed effect modification by air conditioning, citing a paper by Jeremy Sarnat (2000) that found "a significant association between personal $PM_{2.5}$ and outdoor $PM_{2.5}$, in all three groups [low, medium, and high ventilation in the home], with a factor of two difference in the slope between the high ventilation and the low ventilation group." He also indicated that studies in Atlanta often did not find positive associations between exposure and ill-health, which he thought was due to air conditioning use (studies that were part of the Aerosol Research and Inhalation Epidemiology Study (ARIES) and by Paige Talbert's group). He thought this was a large effect modifier that would have a "factor of two impact on what the true health effects are."

Expert E thought that underlying disease, such as diabetes or COPD could be an effect modifier in the long-term studies. He thought that those with COPD have difficulty breathing and therefore, would be more susceptible to the effects of particles. In addition, he said that those with COPD have "ventilation/perfusion inhomogeneity," meaning that parts of the lung are exposed to larger volumes of air than others. This could lead to enhanced particle deposition in parts of the lung that get higher airflow. He thought the evidence for effect modification from COPD was mixed, which in part could be due to the variability in the diagnosis.

The following table summarizes the discussion on potential impacts of effect modifiers:

Study (author date)	**Effect Modifier**	**Score (1-3)***
Six Cities	SES Obesity Other susceptibilities (diabetes, COPD)	1 (underestimated RR) 1 (underestimated RR) 1 (uncertain whether these groups were under or over represented in cohort)
ACS	Educational attainment Obesity Poverty Race	2 (underestimated RR)** 2 (underestimated RR)** 2 (underestimated RR)** 2 (underestimated RR)**

* 1= minimal, 2= moderate, 3= major

**Expert E was not sure whether these all had independent effects of similar magnitude. He thought they might be correlated and be represented essentially by educational attainment (collectively account for ~50 percent underestimate, crudely estimated)

3.7. Exposure Issues

Expert E thought that exposure misclassification due to central site versus individual exposures was a significant exposure issue. He indicated that "the gradient in effect size estimate from … neighborhood-level measurements, to the county [level] to the metro area measurements … suggests that more geographically localized assessment of exposure results in a larger effect size estimate." He thought the ACS study results were likely to be biased downwards due to exposure error, based on the results from Jerrett et al. (2005) and Willis et al. (2003), which both showed higher effect estimates than the original study with increased spatial resolution of exposure assignment. While the Jerrett study showed some effect estimates that were higher by a factor of three or more depending on the model, he acknowledged that other factors like particle composition might account for some of the difference. He thought that the ACS study estimates should be increased by a factor of two to account for this issue.

Expert E next discussed differential toxicity of particles. He thought that the Six Cities study was not representative geographically of the U.S. in that it only spanned from Topeka to the east. Given that time-series study effects tend to be smaller in the west (National Morbidity, Mortality, and Air Pollution Study (NMMAPS)), whereas the Jerrett paper found higher cross-sectional effects in LA, he thought uncertainty exists about the representativeness of the effect estimates in the Six Cities study. He was not sure of the direction of the uncertainty caused by this issue, but assigned it a score of 2. He did not think that this affected the ACS study, because it included cities all across the U.S.

He thought that another exposure issue was whether the exposure in the cohort studies was measured at the appropriate time to see the full mortality effects. He thought the ACS study exposures were somewhat inadequate because they only included a few years of data and "that creates some additional uncertainty as to how relevant that exposure was to the mortality follow-up." He indicated that the Six Cities study's measure of average exposure was significantly better than the ACS study because "for the first follow-up, they monitored for most of the time throughout the period … and therefore presumably, [there was] less measurement error, less downward bias in the slope." However, for the Laden et al. (2006) follow-up period, there were no $PM_{2.5}$ monitors, and the concentrations were actually predicted by a regression model that included PM_{10} and airport visibility. He still thought this was an improvement over the ACS study in that there were continuous exposure estimates throughout the follow-up, but acknowledged that direct $PM_{2.5}$ measurements would have been superior, had they been available for the extended follow-up. He indicated that the use of estimated $PM_{2.5}$ would probably have added random measurement error, which could have caused a downward bias in the Laden et al. (2006) estimates.

When asked about differences in the slope over time, Expert E thought it wasn't necessarily just the timing of exposure measurements relative to outcome, but could be other factors, such as: 1) "confidence intervals may in fact overlap and may just be noise"; 2) differences in particle composition; and 3) air conditioning prevalence increases over time. He indicated that in the Six Cities study, "when [the investigators]

looked at difference between period one and period two … [there was] a bigger coefficient … for the change." He thought "that the Dublin study says that you've got to see some of [the mortality effect] next year, and [the Laden study] says you're basically seeing all of it within ten years."

3.8. Causality

Expert E thought that there was a "web of evidence" supporting a causal relationship. He indicated that there were many different study designs with different strengths and weaknesses and that the evidence was consistent across all of them. He mentioned epidemiologic studies, chamber studies, animal studies, and cell culture studies. He indicated that the animal studies (Sun et al., 2005; Van Eeden et al.) show changes in oxidative stress, inflammation, destabilization of atherosclerotic plaques, and changes in electrocardiogram patterns, all of which contributes to the biological plausibility. He thought the chamber studies in people (though under-powered and over-controlled) provided evidence of changes in fibrinogen levels, systemic inflammation, and endothelial function. He thought these studies are consistent with adverse effects on mortality and with the mechanistic pathways that are suggested by the animal studies. In addition, he mentioned epidemiologic studies on PM and mortality, including time-series, cross-sectional cohort, and change cohorts all have different limitations. "Individual studies have individual flaws, but that's the advantage of different study designs with different kinds of vulnerabilities. You can now build this web and say the web is a lot less vulnerable." He did not think that there could be another factor to explain the consistency across all of the study designs.

Expert E specified a range of the likelihood of a causal relationship of 80 – 99 percent with a most likely value of 95 percent. He set his upper limit at 99 percent because "I'm pretty convinced … but nothing's for certain." Maybe "a substantial amount of the effect is flowing down pathways that haven't been identified." He noted some inconsistencies in studies of inflammatory markers (e.g., C-Reactive Protein) for example. He noted, however that the medical community accepted a causal relationship between smoking and cancer long before many of the data gaps (e.g., an animal model) were filled in.

In order to get at his lower bound and most likely value, he drew an analogy with environmental tobacco smoke (ETS). He thought that, "the evidence for PM is stronger than for environmental tobacco smoke" because the evidence for ETS is based mostly on case-control studies, which he thought could be biased because of the difficulty in selecting appropriate controls. He also thought that PM was a more randomly assigned exposure than ETS. He thought for ETS and lung cancer, his causal likelihood would range between the high 60 percents to the low 90 percents with a most likely value of 80 percent. Since he thought the evidence for PM was greater than ETS, he selected his minimum based on the most likely value for ETS, and his most likely value for PM based on the upper bound for ETS.

3.9 Thresholds

Expert E thought that conceptually, the idea of an individual threshold in animal studies (i.e., that a specific dose causes death or not) is too deterministic. He thought that the probability of an animal dying from a given dose at a given time follows a Poisson distribution. Therefore, he thought that the fact that an animal died in a study from being administered a specific dose was not entirely based upon their individual threshold or susceptibility level, but rather was a compilation of several different stochastic factors. With regard to a population threshold, he explained that, "in a population that's genetically diverse, and diverse with respect to all these effect modifiers that we spoke about … I think it's mathematically impossible. So the probability that something is mathematically impossible could happen, is extremely low. So I think that there isn't a threshold in a population. That doesn't mean the slope can't change. That doesn't mean the probability of dying, on average over the population as a function of dose, has to be linear. But I don't think it's going to be zero for something that at some dose can kill you."

He mentioned that epidemiologic studies, such as Pope et al., 2002 that plotted the concentration-response (C-R) function and suggested a steeper curve at lower levels. He also discussed an expert's presentation at the pre-elicitation workshop of a model averaging approach, the results of which put most of the weight on the linear term and the overall curve looked linear down to about 5 $\mu g/m^3$. Expert E concluded that "in this range [4 to 30 $\mu g/m^3$], it looks pretty linear to me from all the data" and therefore he did not chose to incorporate a threshold into his C-R function.

3.10 Other Influential Factors

Expert E discussed additional sources of uncertainty that were not part of the protocol. He did not think that methodology was major issue. He indicated that "Cox proportional hazards models are pretty standard for cohort studies. There's actually a lot to be said for fitting a parametric survival model, because there's not huge evidence of a need to go non-parametric for the underlying survival curve. A two-parameter Weibull model is pretty flexible to get it fitting any shape you want." He also did not think that spatial auto-correlation had much impact on the analyses. Expert E did not think there were any other outstanding issues not already covered by the protocol.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert E thought the C-R relationship was linear over the entire range of concentrations that were the focus of the study (4-30 $\mu g/m^3$) based on the current evidence. He indicated that the Pope et al., 2002 and some acute studies suggest that the C-R function is steeper at lower levels. However, he thought that, "the evidence for a deviation from linearity is very weak."

Expert E chose to provide a C-R function that was conditional on the existence of a causal relationship. The elicitation team then combined his conditional distributions with

his percent likelihood of causality specified in Section 3.8 (95 percent likelihood of a causal relationship).

Expert E began with the 50th percentile value because he felt it offered the advantage of thinking about different studies and integrating them. He based his estimate largely on three studies: the change estimate from the Six Cities extended follow-up by Laden et al. (2006) of 3 percent per 1 $\mu g/m^3$; the Six Cities cross-sectional cohort study (about 1.5 percent per $\mu g/m^3$); and the ACS study (0.6 percent per $\mu g/m^3$). He adjusted the ACS estimate first by doubling to account for exposure misclassification, to get 1.2 percent. He based the doubling on the paper by Willis et al., 2003 that found coefficients for sulfate from the ACS data that were twice as high as the original analysis with greater spatial resolution of exposure assignment (he assumed that the exposure error in $PM_{2.5}$ would be at least as high as with sulfate). He then increased the ACS estimate 50 percent more (based on Pope's estimates at the pre-elicitation workshop) to 1.8 percent to account for under-representation of low SES, diabetes and obesity within the ACS study population. He then indicated that qualitatively, the Jerrett et al., 2005 estimate and the Hoek et al., 2002 study estimates support higher estimates with finer geographic resolution. He took all of this information and combined it roughly into an "average" estimate of 2 percent for the 50th percentile.

To get at the distribution about the median estimate, Expert E proceeded by thinking about how the variance between studies and within studies might be accounted for in coming up with an overall distribution. He reasoned that the overall variance would be the sum of stochastic within-study variance and the variance due to heterogeneity across studies. He further reasoned that these two variances were roughly equal in magnitude. For the stochastic portion of the variance, he indicated that the variance of a meta-analysis estimate is the sum of 1 over the weights. The weights in turn are 1 over the individual within-study variances. Assuming that the within-study variances were all the same, and if there were five studies contributing to his estimate, he calculated that the variance of the average is "going to be roughly half the variance of the typical study." Based on his earlier assumption that the heterogeneity variance is about the same size as a typical within-study variance, it follows that the total variance is about 1.5 times the typical within-study variance (1 part for the heterogeneity variance and 0.5 part for the meta-analytic stochastic variance). Taking the square root, he came up with an estimate of roughly 1.3 (times the typical within-study SE) for the standard error around his central effect estimate. He then discussed the concept that since he had increased the estimate to correct for measurement error, he would need to inflate the standard error further because "I've added some additional uncertainty by doing this correction factor. And so, the standard error of my now unbiased estimate of beta is bigger than the standard error of the biased estimate, using the exposure that was measured with error." He estimated that this would inflate the standard error by an additional 30 percent. But this applies only to the stochastic "within-study" standard error (0.3). Multiplying that by 1.3 to account for this additional source of uncertainty yielded a new stochastic error of 0.4, and a total SE of 1.4.

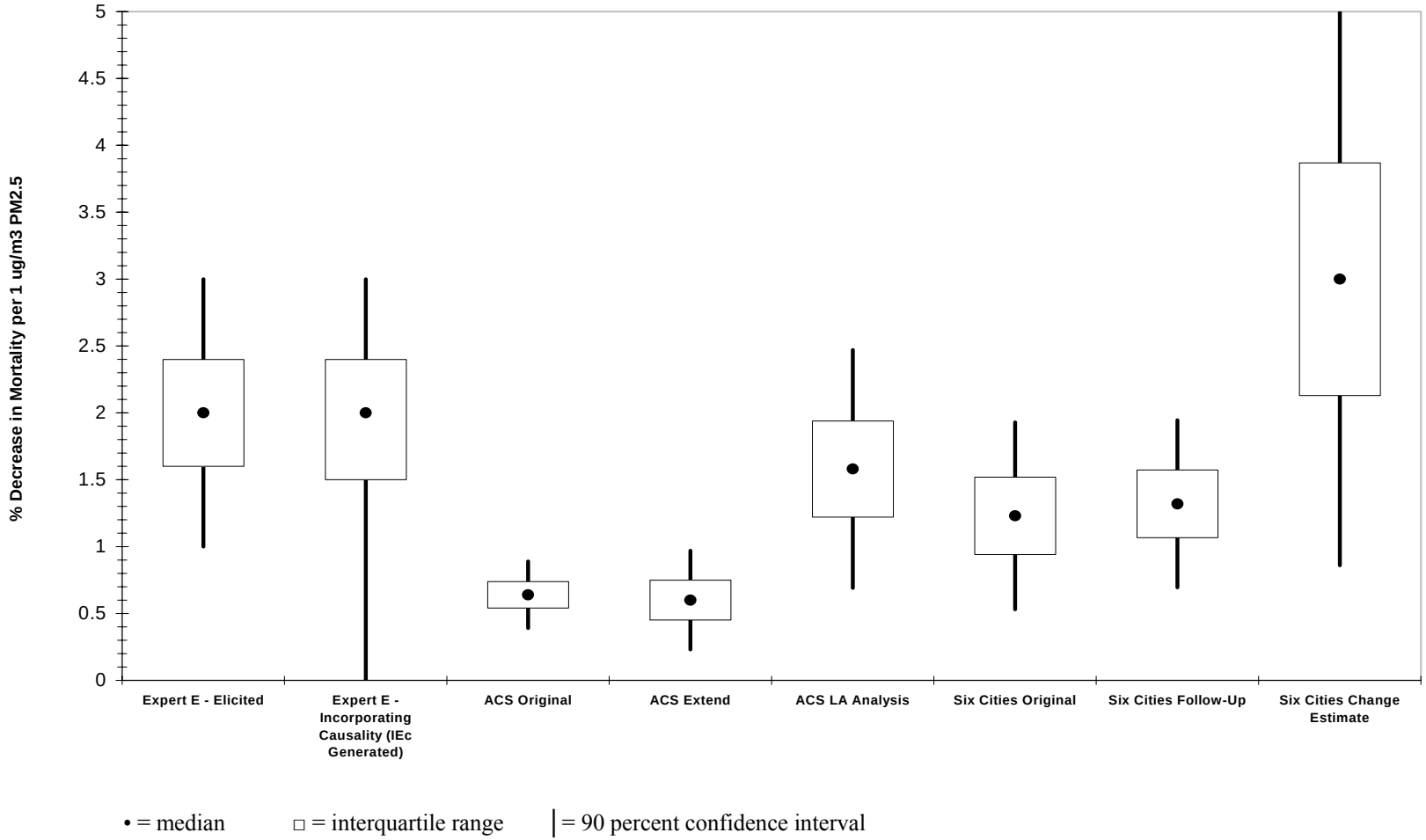
After combining these estimates of variance, he concluded "we might want to take our typical standard error from these studies and multiply it by 1.4 to generate an estimated range of uncertainty." He decided to use the Pope et al., 2002 standard error to represent a "typical" standard error from a study. Therefore, he took the standard error from the Pope study (0.0023) and multiplied it by 1.4 to get 0.32 percent. He then adjusted the standard error further to take into consideration the adjustment for under-representation of lower education in the ACS cohort by multiplying the standard error from Pope by 1.5, instead of 1.4 to get 0.345 percent.

Expert E then stipulated that a normal curve be fit to a distribution with a 50^{th} percentile at 2 and a SE of 0.345. After the elicitation team plotted this curve, he decided that the estimate from Laden et al., 2006 was not receiving enough probability weight. Therefore, he altered the normal distribution so that the 50^{th} percentile was at 2, and a 95^{th} percentile was at 3. When asked to reconcile placing the Laden change estimate at the 95 percentile given how directly he thought it addressed the question, Expert E explained that he had some concerns about the size of the estimate relative to all the other studies. The interquartile range was estimated from the distribution. He truncated his distribution at zero but did not specify a maximum value.

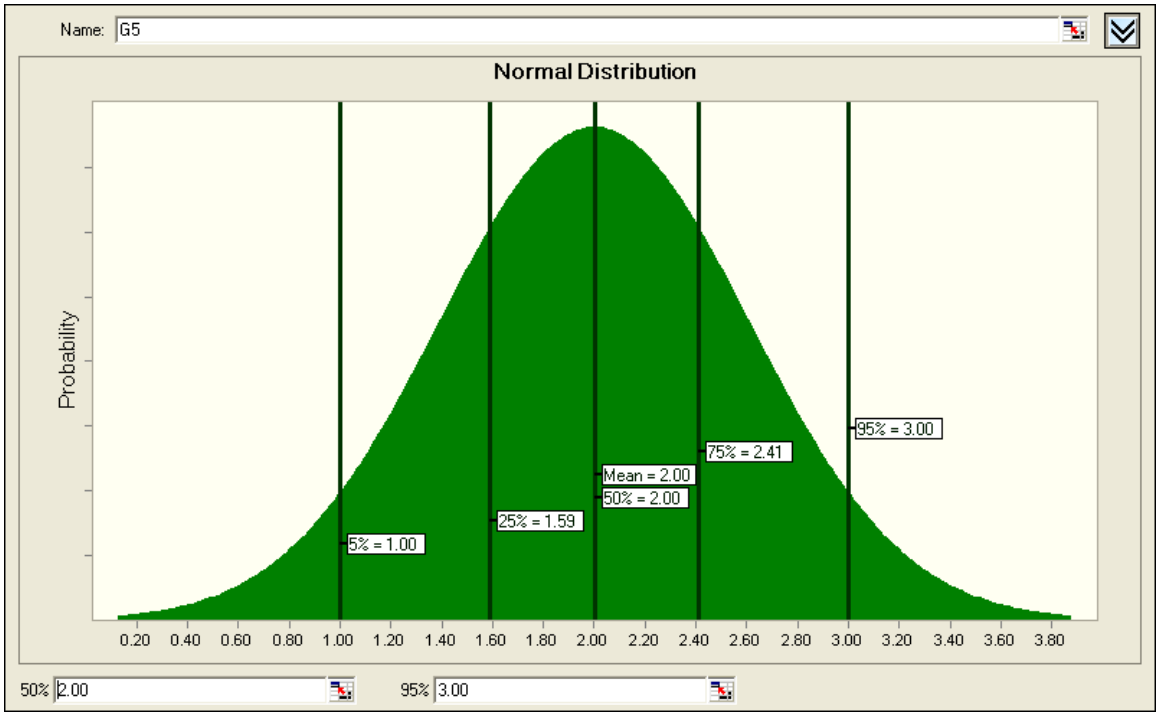
Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Expert E

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	0	0
5^{th}	1.0	0
25^{th}	1.6	1.5
50^{th}	2.0	2.0
75^{th}	2.4	2.4
95^{th}	3.0	3.0
Maximum	-	-

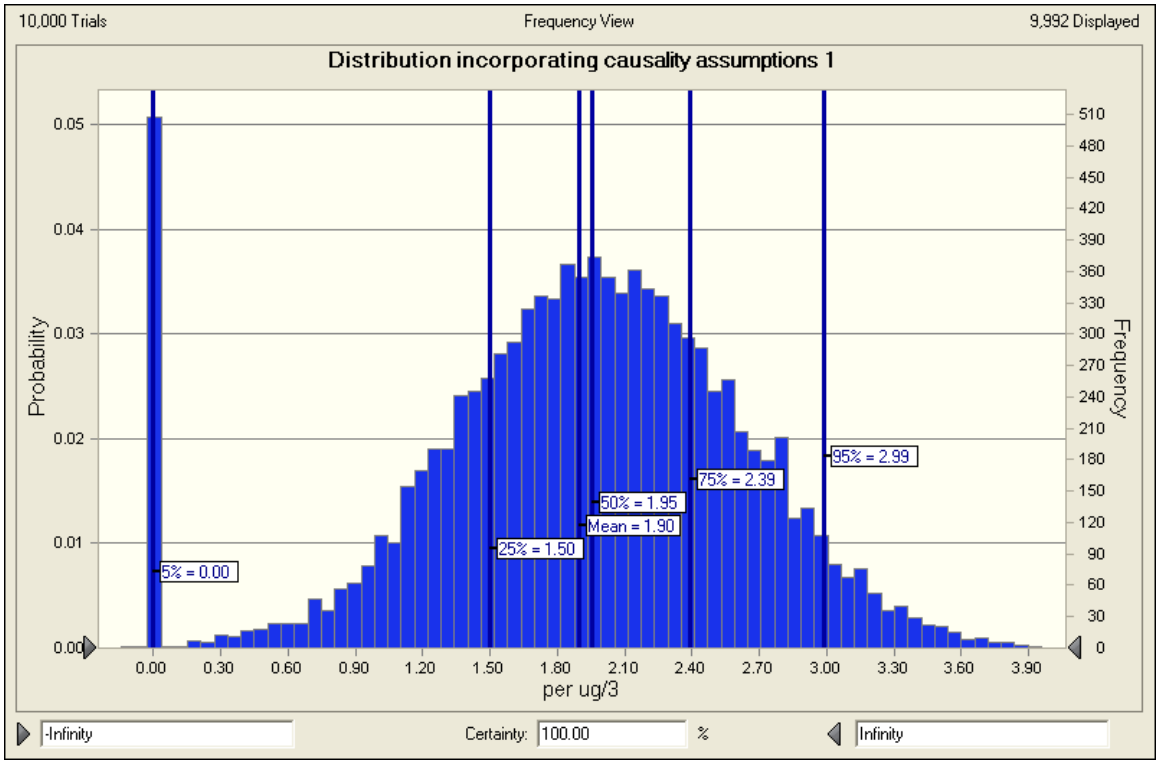
Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distributions from Expert E



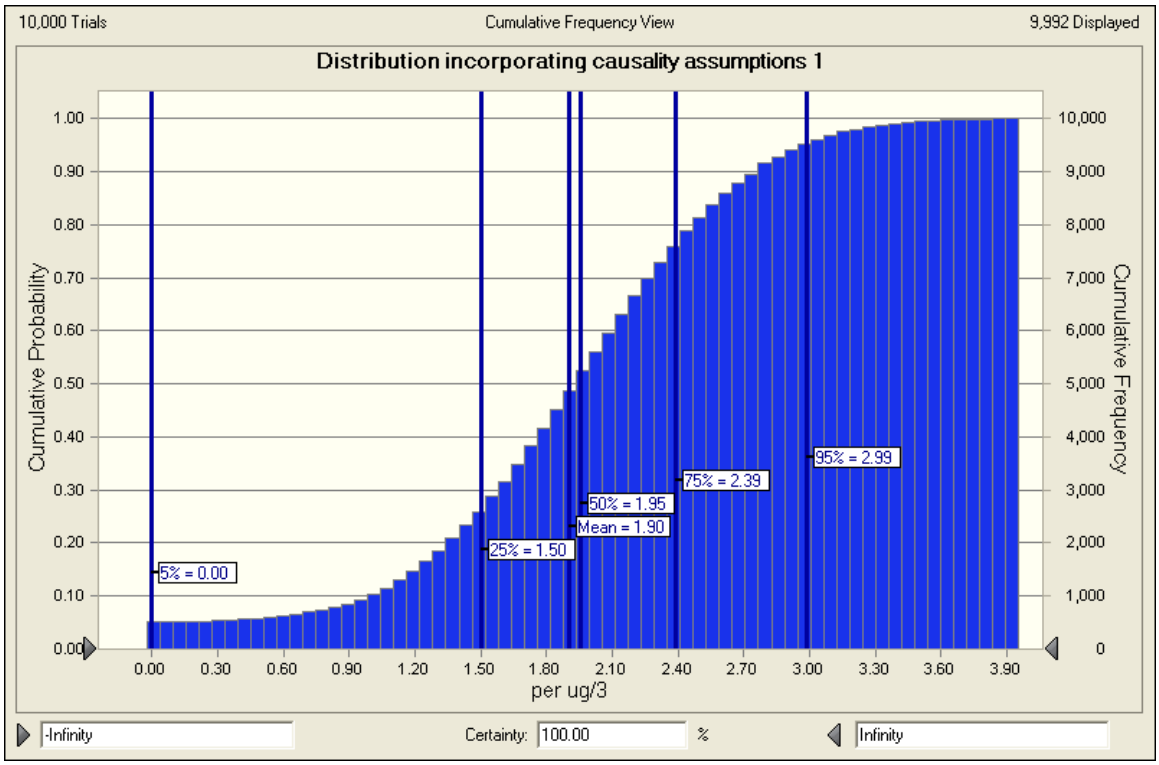
Elicited Distribution



Distribution Incorporating Causality - Probability Density Function (IEc Generated)



Distribution Incorporating Causality - Cumulative Density Function (IEc Generated)



U.S. EPA EXPERT ELICITATION STUDY OF THE CONCENTRATION-RESPONSE RELATIONSHIP BETWEEN ANNUAL AVERAGE $PM_{2.5}$ EXPOSURE AND MORTALITY

Modification to Expert Judgments

Expert E

Date: July 7, 2006

Section of Protocol Affected (Section Number and/or Title):

Causality—dose response

Description of Change (e.g. to a specific percentile, or to a qualitative opinion or statement of belief):

In the post elicitation workshop I was struck by an expert's statement 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, that the size and probability interacted. In thinking about this, I would like to change my previous assumption that there was a 5 percent probability of zero effect to there being a 1 percent probability of there being zero effect. I would leave all the other assumptions about dose-response and its uncertainty the same.

Expert F
Interview Summary

Interview Summary
Expert F

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.2. Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

Expert F wanted to discuss the effects of short-term and long-term exposures separately, and chose to begin with Kunzli's framework as a way of structuring the discussion. Expert F generally agreed with the Kunzli framework although he differed on the relative magnitude of the effects. He agreed that most chronic long-term exposure effects fall into Category B and that deaths due acute exposures largely fall into Category C. He argued that chronic exposures account for the majority of effects; he estimated the ratio of long-term to short-term effects to be about 3:1. Expert F thought that Category A, where air pollution is both the underlying cause of frailty and the acute cause of death was likely to be a much smaller percentage (e.g., around 3 percent of all long-term effects, 10 percent of short-term effects). This judgment was based on a view that the "acute effects … are largely independent of the long-term [effects]" and that the majority of individuals who get sick over the long-term are not getting sick due to air pollution, and that therefore the joint probability of both B and A is likely to be very small.

In essence, Expert F viewed "estimates from the long-term studies and estimates from the short-term studies to be basically, not completely … additive." In other words, he thought the long-term studies do not capture all of the short-term effects, partly because of the statistical argument given above but also because of the underlying mechanisms, and the fact that the life-shortening of the acute PM exposure effects was likely much shorter than for the effects of long-term exposure.

3.1 Mechanisms for Effects from Exposure to $PM_{2.5}$

Expert F thought that the mechanisms for the effects of short-term and long-term exposures were "relatively" independent. He began with a discussion of long-term exposures.

<u>Long-Term Exposures</u>

In order of contribution to overall deaths from $PM_{2.5}$, he discussed deaths from cardiovascular disease, respiratory disease, and cancer. He saw the lung as being primarily the "portal" through which other systemic effects occur, rather than the primary site for most of the mortality effects.

Expert F thought the mechanistic pathway for cardiovascular disease involved chronic lung inflammation, leading to release of cytokines and other indicators of oxidative stress, which lead to increased plaque formation in the vascular system, which can ultimately increase the likelihood of thrombosis and death. He cited the Sun et al. (2005)

paper in atherosclerotic mice (presented at the pre-elicitation workshop) as influential in providing evidence for this pathway. He also thought animal studies by Costa and Kodavanti, (2003) and other work done at EPA showing the oxidative stress and damage to tissues caused by exposure to acids and transition metals were also supportive. He argued that such studies point to the importance of fossil fuel combustion as a "dominant source of the problem."

Impact on the respiratory system was the second cause of death Expert F thought could be associated with air pollution, although more weakly. Expert F suggested that the overall mechanism involves both direct damage to the lung and impacts on lung development that in turn contribute to increased risks of respiratory infections. He thought damage can occur via different routes; particles initiate chronic inflammation which itself can damage lung tissue (Ghio et al., 2000) and macrophages that take up particles may then die releasing enzymes that can cause scarring in the lung. He cited a study by Plopper and Fanucchi (2000) that showed altered immune defenses in animals where the lungs did not develop both physically (e.g., development of the alveoli and size of airways) and immunologically. He pointed to some epidemiological evidence of increases in death from influenza and pneumonia related to air pollution as supportive of this mechanism. In general, he noted that there are fewer deaths from respiratory disease, so the statistical power is weaker to assess this association.

The third cause of death Expert F discussed was lung cancer, which has been associated with $PM_{2.5}$ in epidemiological studies (Pope et al., 2002; Jerrett, 2005). He did not think that the same degree of mechanistic evidence has been established in animals for lung cancer as it has for cardiac and respiratory effects. His argument for the mechanism was largely based on general cancer principles linking together pieces of evidence; particles contain known carcinogens (e.g., benzo[a]pyrene and other organics) which have been shown to be mutagenic in Ames-type assays and which act as initiators; metals and sulfur can then act as promoters by contributing to oxidative stress and cell damage and cell turnover, as shown in animal studies. He did not know of any long-term animal studies that have addressed the question of the carcinogenicity of inhaled particles.

Short-Term Exposures

Expert F expressed the opinion that the primary causes of death associated with short-term exposures to $PM_{2.5}$ were related to cardiac and respiratory mechanisms. He thought the mechanisms for cardiac deaths included not only ischemic heart attacks (as for long-term) but also effects related to arrhythmias or changes in heart rate variability (HRV). Expert F also discussed evidence that PM exposures lead to acute changes in blood coagulation (Devlin et al., 2004) leading to increased risk of coronary thrombosis, and to increases in C-reactive protein (CRP) (Peters et al., 2001; Ruckerl et al., 2006). He cited work by Annette Peters (2000, 2001), a study by Robert Devlin (2003) in elderly subjects exposed to CAPs, and a study by William Penn Watkinson (2000) in animals as influential evidence of the impact of PM on reducing HRV. Expert F speculated that the impact on the autonomic nervous system might be mediated by the oxidative stress and

release of fibrinogen (Schwartz et al., 2001 National Health and Nutrition Examination Survey (NHANES) study) and of CRP (Seaton, 1999; Riediker et al., 2004).

Expert F thought the arguments for respiratory-related deaths was more compelling for short-term exposures to PM than for long-term exposures. He cited studies by Gilmour et al. (2002) and Zelikoff et al. (2003), which showed increased mortality rates in animals exposed to both streptococcus and particles compared to those in animals exposed to streptococcus alone. The rate of inactivation of the bacterium was also reduced suggesting impairment of the immune system by exposure to high levels of particles. Expert F thought acute inflammation by high exposures to particles was a separate mechanism contributing to respiratory-related deaths, primarily in individuals suffering from asthma or other chronic respiratory diseases. Expert F did not think cancer was a cause of death related to short-term exposures, but indicated that cancer patients could be more susceptible to effects from acute exposures to particles.

However, he noted the biggest (but not unusual or prohibitive) limitation in the evidence for the mechanistic pathways is that much of it relies on animal studies; the mechanisms are shown to be biologically plausible, but cannot be absolutely proven in human subjects because of the practical limitations on such human studies.

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

Expert F thought that prospective cohort studies represent a "gold standard" for estimating the effects of long-term exposures. These were closely followed, in his view, by the cross-sectional cohort study designs, which he thought provided qualitative support for the quantitative effect estimates reported in prospective cohort studies. In both types of studies, he estimated that roughly 97 percent of the mortality reported was due to long-term exposures with the remaining 3 percent attributable to cumulative short-term exposure effects.

For estimating the mortality effects of short-term exposures to $PM_{2.5}$, he thought the time-series and case-crossover designs were most useful. He thought the panel studies were important for demonstrating that the effects that are observed in population studies are also observed on an individual level.

When asked about intervention studies or "natural experiments" (such as Pope's Utah Valley study, the Clancy study in Dublin, and the Hedley Hong Kong study), he thought that they were not as directly appropriate to the quantification of mortality effects, though they do speak to the issue of cessation lag; such studies suffer from questions about temporal confounding (i.e., changes in other factors that could also affect mortality).

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

Expert F's ideal epidemiologic study included the following attributes:

- Representative of the U.S.
    - Appropriate sampling to include important subgroups (e.g., age, race, education level, socioeconomic status (SES)); and
    - Geographically.
- Collection of individual level data at both beginning and follow-up (e.g., personal risk factors related to cardiac, respiratory disease, and lung cancer).
- Exposure assessment
    - Representative of lower and higher exposures;
    - Personal exposures to outdoor $PM_{2.5}$, not personal exposure overall; and
    - Includes information on mobility.

The specific studies that Expert F thought would be most informative for his quantitative estimates were the extended American Cancer Society (ACS) cohort studies (Pope et al., 2002 & 2004), the reanalysis in Los Angeles (LA) (Jerrett et al., 2005), and the original Six Cities cohort study (Dockery et al., 1993) as well as the follow-up (Laden et al., 2006). He thought the strengths of the ACS studies included their large study population that is inclusive of an age range relevant to the elicitation, more nationally representative, detailed individual level data on risk factors, and exposure measurements at the beginning and end of study that were well correlated. Expert F noted that limitations of the ACS studies related primarily to effect modification by education and exposure misclassification (this topic is discussed in greater detail below). He indicated that strengths of the Six Cities study were that it was specifically designed for investigating the relationship between air pollution and health, that its exposure estimates did not rely on central site monitors only (they have some personal-level monitoring), inclusion of more co-pollutant information, data on potential confounders, and a population that is more representative of the general U.S. population (regarding educational level, industrial, and non-industrial America). He thought its limitations included relatively small sample size, restriction to the eastern half of the U.S. (i.e., no representation of the west, especially California), and non-representative racial composition (mostly white).

Expert F thought the Jerrett (2005) study of the LA component of the ACS study improved measures of exposure within LA and, by focusing on one metropolitan statistical area (MSA), eliminated some of the issues with between-city comparisons that existed in the parent study. He thought potential limitations of the Jerrett analysis were that it is representative only of one city, that spatial correlations might exist between SES variables and more polluted parts of LA, and that there may be within-MSA mobility impacts on exposure. He argued that the Jerrett study essentially trades one exposure measurement problem for another, yet the fact that it finds similar results provides evidence that the $PM_{2.5}$ mortality association is robust. He thought the Laden 2006 follow-up to the Six Cities study has similar strengths and weaknesses of the original Six Cities study, though it has the benefit of longer follow-up period.

Finally, Expert F noted that the ecological studies (e.g., Ozkaynak and Thurston; Evans, Tosteson, and Kinney, 1984) showing an air pollution effect on mortality have been recently exonerated by all the subsequent cohort studies. He thought the concerns about

confounding that were raised about the ecological studies have not turned out to be substantiated.

Expert F briefly discussed papers that he felt were less conclusive about the relationship between $PM_{2.5}$ and mortality, such as the Lipfert studies and the Enstrom (2005) study, neither of which Expert F wanted to place much emphasis on. He basically argued that, although Lipfert's later study found results similar to some of the studies above when he used more comparable measures of exposure, his studies lack credibility given the funding source (Electric Power Research Institute (EPRI)), and that the papers were not published in a journal where they were likely to be peer-reviewed by epidemiologists (e.g., *Inhalation Toxicology* and *Atmospheric Environment*). He had similar concerns about the Enstrom (2005) paper in *Inhalation Toxicology*. Although not fully familiar with this study, he had concerns with the way exposure was measured (at point of residence as opposed to an area average), the highly educated population, and large loss to follow-up. He thought if one assumes that women spend more time at home than men, then the exposure error might not be so great and might explain why they saw effects in women but not men.

When asked about evidence provided by Moolgavkar papers against the hypothesis of a PM-mortality effect, he stated simply that he did not believe the work was credible.

Although Expert F initially thought his estimates of the total mortality effect would have to include some additional increment from the short-term studies, he ultimately decided to rely only on the cohort studies. He felt that it was difficult to know what the protocol question was really asking for. Nevertheless, he argued that even though strategies to reduce the annual average $PM_{2.5}$ concentration could affect peaks as well, annual average measures of mortality would not pick up much acute mortality and that any additional contribution from acute mortality to the cohort estimates would be small, relative to the chronic exposure reduction benefits.

3.5 Confounding

As a backdrop to the discussion on confounding, Expert F discussed the results of an analysis conducted by Zidek et al. (1996) entitled, "Causality, measurement error, and multi-colinearity in epidemiology" in *Envirometrics*. He indicated that the study basically found that if both the predictor variable and the outcome variable are measured with error, the impact of confounding is diminished. In order to have substantial influence on the regression estimates by a confounder, the degree of correlation between the confounder and the predictor variable had to be very high (e.g., around 0.9). Expert F's conclusion was that although confounding is something to be concerned about and to control for, the fear that modest correlations with confounding variables could fully account for effects measured in epidemiological studies does not appear to be warranted.

The few variables that Expert F thought might be of some residual concern in the studies he planned to rely on for his quantitative estimates are summarized in the table below:

Study	Confounders potentially leading to overestimate of the true RR	Score (1-3)*
ACS	Contextual SES	2 (lower end)
	Gaseous co-pollutants, occupational, smoking	1
Six Cities (original and follow-up)	Contextual SES	1
	Occupational	1-2
	Gaseous co-pollutants, smoking	1
Jerrett, 2005	Contextual SES	2 (higher end)

* Expert F defined the scores in the following way: 1= <10 percent, 2= 10-15 percent, 3 = >20 percent

Expert F generally did not think that gaseous co-pollutants (CO, NO_x, SO_2) were "big confounders to the PM effect." The reason for this was that "these pollutants are correlated because they're coming from the same sources and not because they cause the effect." He predicted that, as we begin to investigate exposures by source categories, CO and NO_x would be related to traffic particles and SO_2 would be related primarily to coal combustion. He did not think there were plausible arguments for mechanisms by which these co-pollutants could affect mortality over the long-term (he thought CO might have a plausible mechanism for affecting mortality during acute exposures). Although the ACS study and subsequent re-analyses found a strong SO_2 effect, Expert F did not think there is a plausible explanation for how SO_2 actually causes the mortality effects (i.e., it does not get deep in the lung, does not cause increases in CRP or plaque development). He thought it is more likely that SO_2 is acting as an effect modifier, by making the particles more toxic. He similarly thought smoking had been well-controlled for in the ACS studies (examined using different smoking variable in different models) and Six Cities studies though he could not rule out some minor residual confounding. The concern he raised here was as to whether measures of smoking and risk fit the proportionality assumptions of the Cox Proportional Hazard model.

He thought there could be some residual confounding by "contextual SES" variables (i.e., that there is some correlation between living in a more polluted location and being poor), particularly in the ACS study. He did not think confounding by contextual SES variables was likely to be as great in the full ACS study which averaged variables across whole metropolitan areas, as in the Jerrett, 2005 study of Los Angeles which looked within a particular metropolitan area.

He expected confounding by occupational exposure to be minimal as well in all of these studies although it might rise to the level "moderate [2]" for the Six Cities study given the statistical influence of Steubenville, originally a highly industrial city, on the mortality effect estimate.

3.6 Effect Modification

Expert F discussed primarily the ACS and Six cities study in his review of effect modification. He discussed the potential for effect modification by SES status (i.e., as indicated by educational attainment in the ACS study), SO_2, and smoking. With the exception of SES status in the ACS study and possibly SO_2 in the Six Cities study, he thought it unlikely that effect modification would have more than a minimal effect on the mortality effect estimates:

Study	Effect modifiers	Score (1-3)*
ACS	SO_2	1 (overestimated RR)
	Education (SES)	3 (underestimated RR)
	Smoking	1 (uncertain direction of bias)
Six Cities (original and follow-up)	SO2	2 (overestimated RR)
	Smoking	1 (uncertain direction of bias)

* Expert F defined the scores in the following way: 1= <10 percent, 2= 10-15 percent, 3 = >20 percent

Expert F thought the most influential effect modifier was the educational effect observed in the ACS study where most of the mortality effect was observed in the portion of the cohort with less than a high school education. Since the ACS cohort was more highly educated than the U.S., Expert F thought the relative risks were probably underestimated by as much as 50 percent or more in the ACS study (in part based on discussion from the pre-elicitation workshop).

He thought effect modification by SO_2 might be a larger problem in the Six Cities study than in the ACS study. His argument was that higher mortality has been observed in Steubenville in the Six Cities study (as it has been in the Pittsburgh-Steubenville area for the ACS study) and that these higher mortality rates may relate to effect modification of the SO_2 on the PM effect because SO_2 adsorption on the surface of PM might well make the particles more toxic. He thought this modification effect might be absent in other areas of the country (e.g., Seattle where SO_2 levels and coal combustion are lower), which are not as well represented in the Six Cities study and, to a lesser extent, in the ACS study. As a result, he thought the relative risks (RRs) might be somewhat overestimated in each of these studies. However, the influential weight of the Steubenville data in the Six Cities studies discussed earlier led Expert F to believe that effect modification by SO_2 might be greater in those studies than for the ACS studies. He did not think this was an issue for the Jerrett study though noted that recent source apportionment studies in the LA area have seen increasing evidence of exposure to sulfates from cargo ships in the port of LA.

While he thought there might remain some potential for effect modification by smoking in the ACS and Six Cities studies, he thought it was likely to be minimal (a score of 1) and he was uncertain which way it might influence the effect estimates if at all.

3.7 Exposure Issues

The types of measurement error Expert F focused on in his discussion of the ACS and Six Cities studies (including the follow-up paper) included the impacts of mobility of the population on measures of exposure, central site versus individual site monitors, and choice of the appropriate monitoring time period given historical trends in exposures. In general, Expert F thought the ACS study was more affected by exposure measurement error, in particular the first two issues listed above than the Six Cities studies:

Study	Exposure Issues	Score (1-3)*
ACS	Mobility (coupled with effect modification by education)	3 (underestimated RR)
	Central site v. individual exposures	2 (underestimated RR)
	Historical trends	1 (if use correct time period) (underestimated RR)
Six Cities, original	Mobility	1 (underestimated RR)
Laden et al. (2006) Six Cities follow-up	Central site v. individual exposures	1 (underestimated RR)
	Estimation method for $PM_{2.5}$ central site value	1 (underestimated RR)

* Expert F defined the scores in the following way: 1= <10 percent, 2= 10-15 percent, 3 = >20 percent

Expert F thought the issue of mobility of the study population is essentially one of measurement error. If people move after their exposure has been assigned, and their new location and exposure is unknown and cannot be adjusted for (as in the ACS study), there is more measurement error. The impact of the error could be different depending whether people, on average, moved to a cleaner or a more polluted area, and might differ by economic status if people of lower economic status are less likely to move out of the study area, as census data suggest. Expert F thought this factor could be influential in underestimating the relative risks for the ACS study, as well as contributing the reported effect modification by education level.

Expert F thought the degree to which the central site monitors reflected individual ambient or outdoor exposures was the other important source of measurement error for the ACS study. He thought it could contribute to underestimate of the RR in the study; he assigned it a score of 2 ("at most"). He did not think that the Jerrett study provided a clean measure of the impact of more precise geographic estimates on the effect estimates (i.e., he thought the higher effect estimate in the Jerrett study might also be due in part to more spatial confounding by SES).

He discussed the relative importance of outdoor, indoor, and personal exposures as exposure metrics. Basically Expert F believed that outdoor air is what predicts indoor air

quality and it is outdoor air that is ultimately of interest for regulatory purposes. Expert F did not consider differential ability of particles to penetrate indoors in different climates to be a big issue because he did not think air conditioning was a very efficient way to remove particles.

Expert F thought that choice of the exposure period used to estimate mortality effects could influence effect estimates, given historical changes in particle concentrations with time. He thought there exists the potential to "underestimate the effect size when you use [exposures from] the beginning of the study because you're overestimating the exposure." He noted how the reanalysis of the ACS study estimated the mortality effects using two exposure periods, early and late in the study, and that the effect estimates were higher when the later exposure period was used. He thought the average of the two periods made the most sense. However, he also pointed out that Roosli et al.'s work suggests that the last 5-10 years before death are the most influential for the effect estimate. He was not sure how this finding could be reconciled with the mechanisms for developing cardiac disease that involve chronic inflammation and accumulation of atherosclerotic plaques. For cancer development, he thought the longer-term exposures would be more important. He concluded that if the correct exposure period were used (e.g., using the average of the two exposure periods in the ACS re-analysis study), the effect estimates would not be biased or mis-estimated, giving larger estimates than those generally used from the ACS study (i.e., using those based on the 1980's exposures, as reported in the manuscripts' abstracts).

3.8 Causality

Expert F would ideally have the following types of evidence to support a conclusion of a causal relationship:

- Specificity (e.g., would like to see an impact on cardiovascular and, with sufficient follow-up, lung cancer, but not other causes);
- Consistency across studies that are done well;
- Plausible biological mechanisms that are consistent with the epidemiological evidence;
- Concentration-response relationship;
- Coherence across different types of study designs; and
- A temporal relationship (as observed in intervention studies).

Expert F's views on the strength of the causal relationship differed for long-term and short-term exposures. He felt that the likelihood of a causal relationship was 100 percent for the effects of long-term exposures; it would take his ideal study, done well and showing no relationship between particles and mortality to convince him otherwise. For the effects of short-term exposures, Expert F thought that, under the least optimal conditions (if one believed that weather and ozone were confounding the relationship) the likelihood of causal relationship for between short-term exposures and mortality was about 60 percent. At best, believing all other relevant factors to be controlled or analyzed

for, he could be 100 percent certain of a causal relationship. His best estimate was that it was highly likely, around 95 percent.

3.9 Thresholds

In theory, Expert F could imagine that a threshold might exist at a natural background level (if one could eliminate all fossil fuel combustion). At the same time, conceptually, there exists an "infinite pool" of susceptible individuals, who could be "knocked off" at various concentrations.

He thought that the most appropriate types of studies to assess the existence of population thresholds would be long-term studies in large populations experiencing a wide range of pollutant levels. They should look at indicators of disease development (e.g., intima-media thickness (IMT) of arteries, as in Kunzli's latest work). He thought toxicological studies were useful for studying mechanisms, but not for determining population thresholds or effect estimates.

He did not think that a threshold is detectable in any of the epidemiological studies currently available. He did not want to incorporate a threshold into his characterization of the concentration-response (C-R) relationship.

3.10 Other Influential Factors

Expert F briefly discussed other factors that were of residual concern to him. He mentioned again some uncertainty about how well the Cox proportional hazards model deals with potential confounding by smoking.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert F chose to characterize the C-R function as piece-wise linear across the range of annual average $PM_{2.5}$ concentrations that were the focus of the study (4-30 ug/m^3). He estimated a distribution for the first slope that applied to concentrations of 4-7 ug/m^3 (hereafter, "Range 1"), and one that applied to concentrations of >7-30 ug/m^3 (hereafter, "Range 2"). Although he indicated that there was not enough data to specify other than one linear range, he thought that there was probably more uncertainty about the slope in the lower portion of the range (i.e., below the range of observed $PM_{2.5}$ data in the ACS study).

For purposes of comparison to other experts who specified a single distribution for the whole PM range for this study, the elicitation team simulated a combined distribution (hereafter, "Example Applied Distribution") by linking his two distributions via a 2002 distribution of population-weighted annual average $PM_{2.5}$ concentrations in the U.S. from EPA's BenMap model.

Expert F began by specifying the percentiles for Range 2 (note all values were initially given per 10 μg/m^3 change in $PM_{2.5}$ and converted later to a 1 μg/m^3 change). His

minimum value was based on an estimate that lay between the effect estimates from the ACS earlier and later exposure periods (around 6 percent). His 5th percentile was based on an estimate from the ACS study for those with less than a high school education (around 8 percent - estimated from Figure 4 in Pope et al., 2002). Expert F's 50th percentile was based initially on estimates from the original Six Cities study (Dockery et al., 1993) as well as the (Laden et al., 2006) follow-up (around 14 percent), but he adjusted it downward to 12 percent to account for concerns about residual confounding by occupation and effect modification by SO_2.

Expert F's 95th percentile was based on an estimate from the Jerrett et al., (2005) (17 percent per 10 $\mu g/m^3$ based on the model with 44 individual covariates). He argued that it was appropriate to think of this estimate as an upper 95th percentile because, although he thought the Jerrett study estimates represented improvements in exposure measurement, he thought they also might still be confounded somewhat by contextual SES factors. To generate his maximum value, he adjusted this estimate upwards by an additional 3 percent from his 95 percentile estimate to account for additional uncertainty, because he could not see the "true" effect estimate of all anthropogenic PM effects for the U.S. being greater than 20-25 percent of total mortality. Expert F calculated his interquartile range by interpolating roughly between the 50th and the 5th percentile values, as well as between the 50th and the 95th.

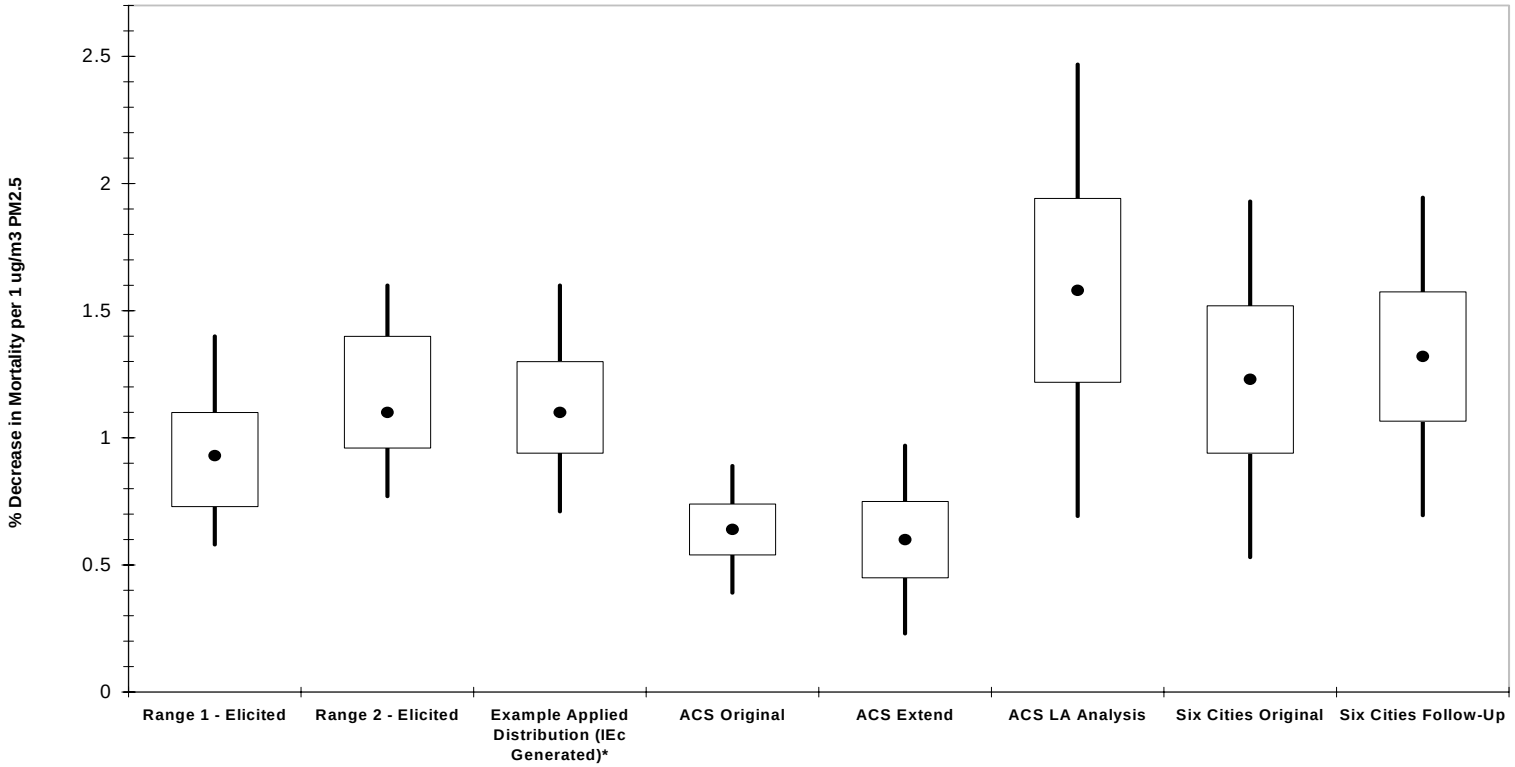
To calculate the percentiles for Range 1, Expert F reduced the percentile values from Range 2 by 25 percent to indicate that he thought slope might be lower. He then made adjustments to the 50th, 75th, 95th, and maximum to ensure that the two distributions had overlapping interquartile ranges and to express greater uncertainty about the magnitude of the slope in the lower range.

He was satisfied that his Example Applied Distribution essentially encompassed the distributions from the main studies on which he relied, even though the 90 percent confidence intervals were not much wider. He thought they represented essentially a mental meta- analysis and that would reduce, not increase the uncertainty distribution.

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1 ug/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations

Percentile	Percent Change in Mortality (Range 1)	Percent Change in Mortality (Range 2)	Percent Change in Mortality -Example Applied Distribution (IEc Generated)
Minimum	0.37	0.49	0.37
5^{th}	0.58	0.77	0.71
25^{th}	0.73	0.96	0.94
50^{th}	0.93	1.1	1.1
75^{th}	1.1	1.4	1.3
95^{th}	1.4	1.6	1.6
Maximum	1.7	1.8	1.8

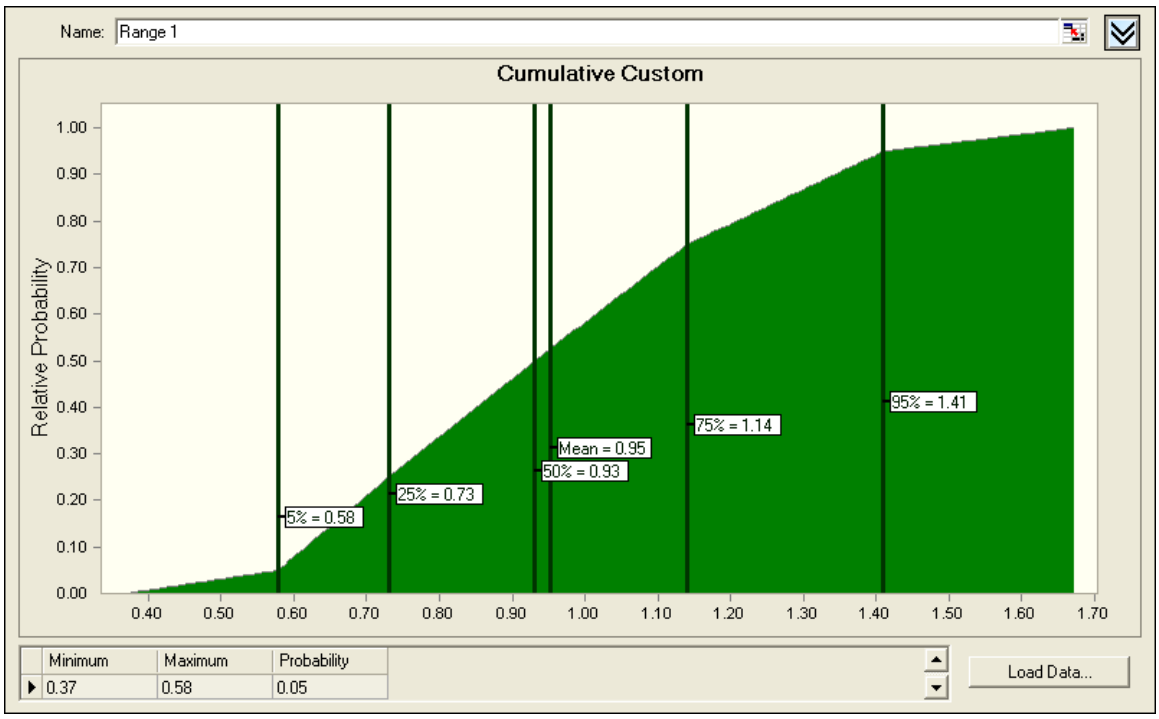
Exhibit 2: Percent Change in Annual Mortality Associated with a 1 ug/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distributions from Expert F



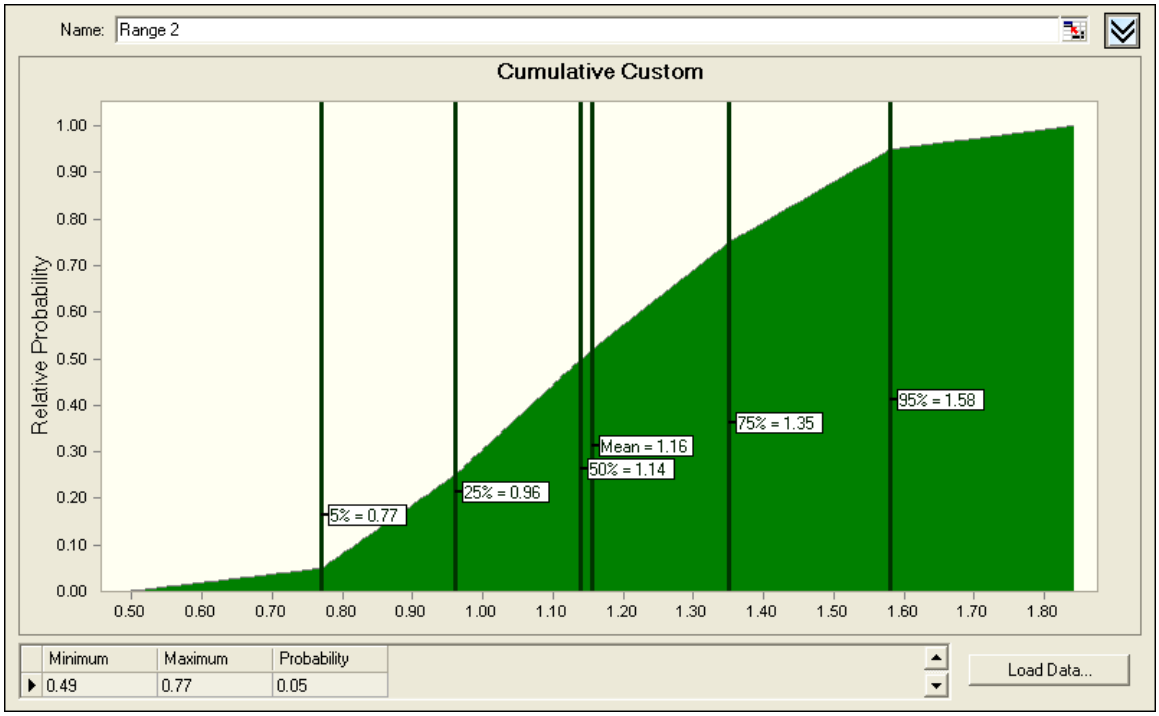
* = Distribution incorporating causality and applying the C-R functions from Ranges 1 and 2 to a 2002 population-weighted annual average $PM_{2.5}$ concentration distribution in the U.S. from BenMap (IEc Generated).

• = median □ = interquartile range | = 90 percent confidence interval

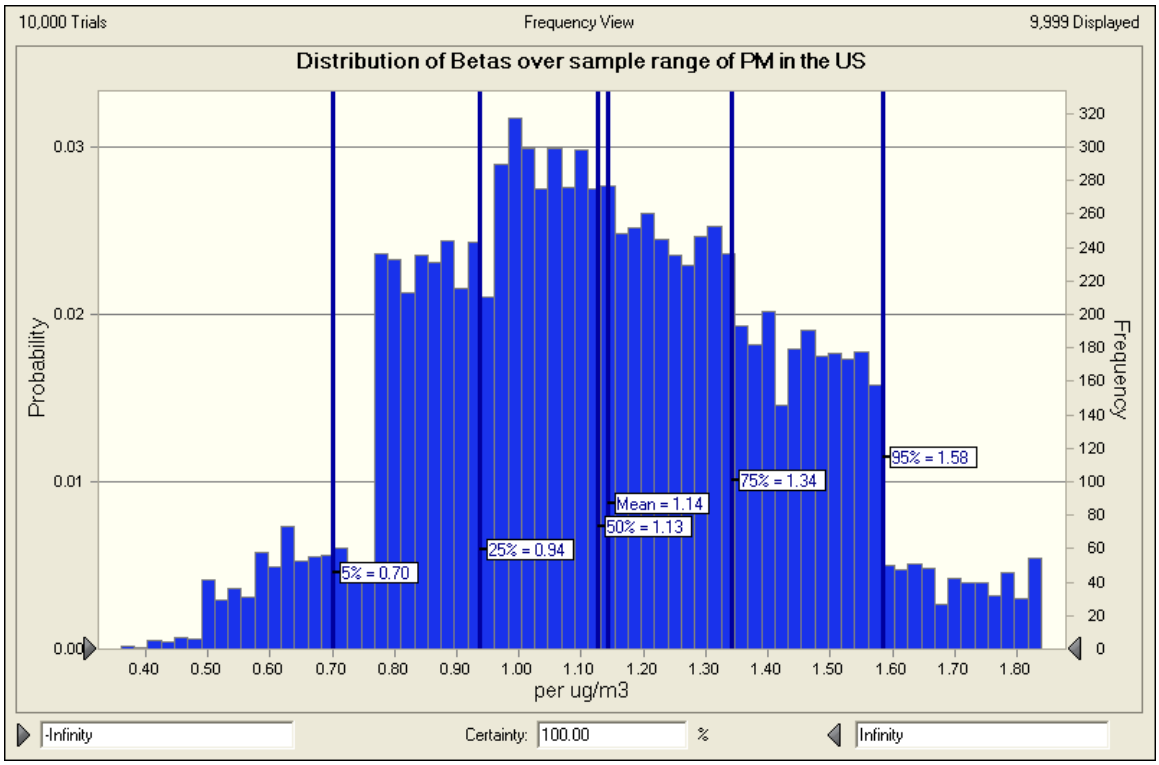
Elicited Distribution – Range 1



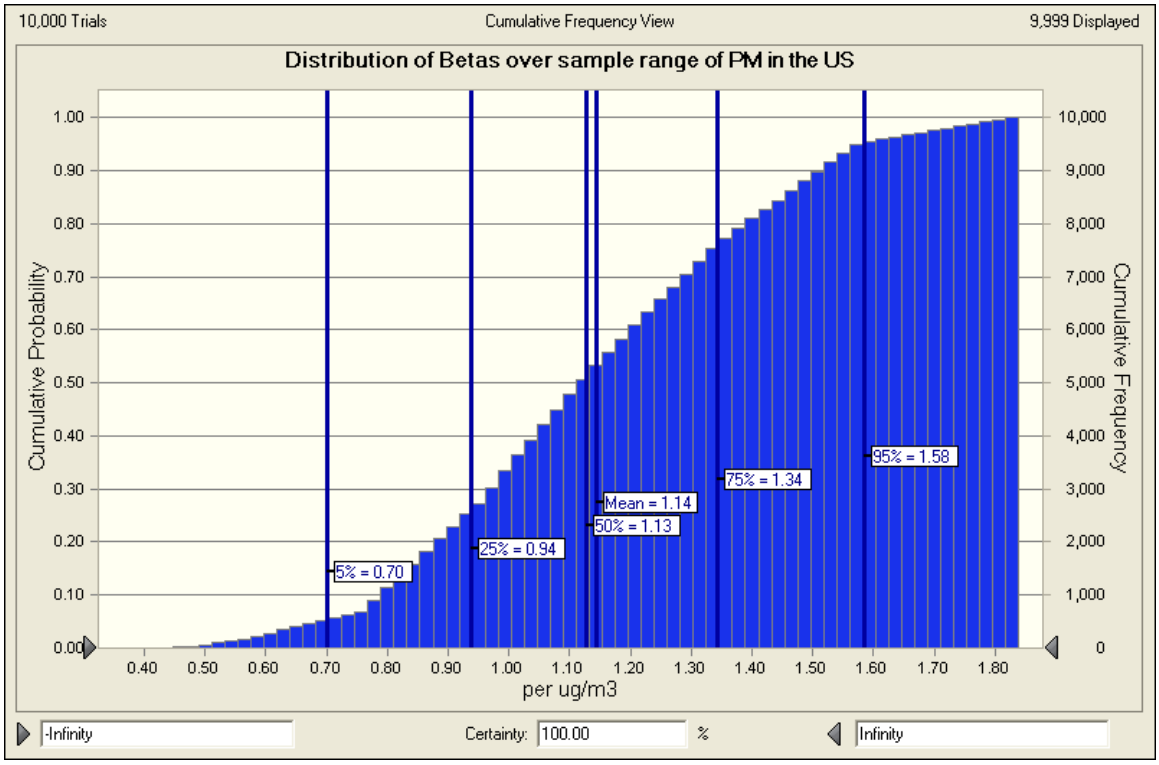
Elicited Distribution – Range 2



Example Applied Distribution - Probability Density Function (IEc Generated)



Example Applied Distribution - Cumulative Density Function (IEc Generated)



Expert G
Interview Summary

Interview Summary
Expert G

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert G discussed the biological mechanisms for short-term and long-term exposures separately, beginning with short-term exposures.

Short-Term Exposures

Expert G primarily discussed two types of effects in the context of short-term exposures to PM: cardiovascular effects and acute respiratory infection.

Expert G thought that cardiovascular effects were the "most compelling in terms of [the ability of short-term exposures] to cause mortality." He focused on "cardiovascular dysfunction," which he defined as involving arrhythmia or other changes in heart function, "including changes in the ability of the heart to respond to nervous system stimulation." He thought that this effect would be more significant in susceptible populations such as "[p]eople with pre-existing cardiac or respiratory disease, and elderly." He thought that a body of evidence provides support for the mechanism underlying this cause of death. "Looking at the overall database … epidemiological studies … looking at populations that have changes in certain aspects of EKG, changes in heart rhythm, changes in heart rate variability, which suggest effects on the nervous system control … [and] I look at the toxicology studies, which have shown similar effects … in terms of change in heart rate variability or changes in heart rhythm … this consistency between the epidemiology and the toxicology." He thought that limitations of this database included differences in dosimetry and between ambient exposures and controlled experimental exposures.

When asked about how the cardiovascular mechanism for deaths related to PM worked, Expert G indicated that there are two hypotheses 1) that the particles "get out of the respiratory tract and actually end up in the cardiovascular system"; and 2) "the oxidative stress hypothesis … the chemical nature of the particles is such that it causes some sort of biochemical effect in the lungs, such as release of certain mediators, like interleukins … which will then travel out of the respiratory tract, go systemic, and then affect the heart." He also thought that "pro-inflammatory mediators can cause tissue damage, which can alter organ susceptibility to other physiological influences and/or cause structural damage … hypothetically that can affect an organ's ability to responds to nervous system stimulation." But he thought there was "still a gap [in understanding what the] structural, biochemical changes within the tissue [are] that result in its inability to respond normally to nervous system input."

Expert G thought the second most important cause of death from short-term exposures is acute respiratory infection, such as bronchitis or pneumonia in individuals "already

stressed either by age or pre-existing disease" because their defense mechanisms may be compromised. He thought this mechanism was supported by evidence from epidemiology as well as toxicology (Zelikoff's work, studies from NYU).

Expert G also briefly discussed another mechanism for cardiovascular impacts involving PM's impacts on blood coagulation ability. He noted that there were a few studies that seem to support this pathway. When asked about stroke as a potential cause of death from short-term exposures, Expert G thought this was related to this cardiovascular mechanism in that changes in blood coagulation ability could also cause strokes.

Long-Term Exposures

Expert G indicated that he viewed effects from long-term exposures as "increasing biological stress levels on a long-term basis … [s]o you have cumulative damage over the years … [which] appears to reduce life span." "The effects on organs, especially the heart and perhaps the lung as well, may be similar to aging, in that the additional stress induced by sub-clinical release of pro-inflammatory mediators could result in the inability to handle additional stress to a certain point, and that would cause death either by heart attack or maybe lung disease." He discussed the Children's Health study in California that "infers that [children] have compromised lung function development," and reasoned that it might contribute to increased risk of mortality in adulthood.

For cardiovascular disease, Expert G discussed the Sun et al. (2005) study experiments at NYU with the apoE-/- atherosclerotic mouse model. While he noted that this research showed that PM accelerates the development of atherosclerosis, his critique of this work was that the apoE-/- mouse is a sensitive model and he questioned how representative it might be of human experience. Expert G also discussed Künzli et al. (2005) study that found changes in intima-media thickness of the carotid arteries associated with PM. He thought that this could be related to "inflammatory effects causing damage to the wall, which then gets repaired by fibrotic tissue, which is thicker than normal tissue."

Overall, Expert G thought the body of evidence was more complete for short-term exposures than for long-term exposures: "I think that ... there is a coherence between the epidemiology and the controlled exposure studies, which include human clinical and the tox, in terms of acute responses to PM mechanisms, potential mechanisms, that would explain mortality. [For] long-term exposures … there is a lack of chronic studies in controlled exposure. [The Sun et al. study] at NYU is … the longest-term controlled exposure study and is showing responses that could explain long-term exposure effects. But ... it's still only a six month study … [T]he results are consistent with what one would think would be going on in the population exposed chronically, but the data set isn't as strong as it is for the acute, yet."

Expert G discussed lung cancer as a potential cause of death from long-term exposures. He thought that there was some epidemiological support for this. For a plausible mechanism, he drew a parallel to chronic cigarette smoking. He thought PM contains carcinogens and the combination of years of exposure to the carcinogens and a genetic

pre-disposition to cancer make it plausible that PM could cause cancer. However, he noted that cancer did not appear to be the largest risk associated with PM.

3.2. Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

Expert G thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures. He elaborated on the definition of Category B, long-term effects, to say that he thought that the long-term effects of air pollution could result, not just from long-term average exposures as is commonly assumed, but from long-term exposures to a series of peak PM levels, "but the effects my be different and affect different people." In the peak model, short-term peak exposures could lead to sub-clinical development of disease over time. "Some of those peaks might also kill people … Quantitating the overlap [between the categories] is the question."

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

Expert G's views on the mortality effects captured by each study design are shown in the table below:

Study Design	Type of Effects Captured (e.g., short-term, long-term, or both)
Cohort Studies	Long-term and some short-term (Categories B and A, respectively)
Intervention Studies	Short-term and mixed (Categories A and C)
Time-Series Studies	Short-term and mixed (Categories A and C)

Expert G thought that he would most want to rely on the cohort studies for quantifying the effects of long-term exposures. He led an extensive discussion to assess to what extent the cohort design could distinguish whether increased mortality from long-term exposures was due to average levels or to a series of peaks. He ultimately concluded that they could not make such a distinction. After initially thinking cohort studies would pick up purely short-term effects (C), he concluded that they would not necessarily pick them all up.

Expert G thought the intervention studies were an important "component of a portfolio of epidemiology studies used to develop some sort of quantitative risk assessment." He indicated that they show the effect of an intervention and ideally, if the difference in components were known, it might be possible to understand more about which components of the PM are producing the change in mortality. He thought they would largely capture short-term effects.

Expert G thought time-series studies would largely capture the effects of short-term or acute exposures. Ultimately, Expert G thought estimating the mortality effects of

changes in annual average $PM_{2.5}$ concentrations might require a combination of results from cohort, intervention and time-series design. He found it difficult, however, to determine what the basis might be for the relationship between the magnitudes of the effect estimates seen using different study designs and thus, was unsure of how they should be combined.

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert G's ideal study is one that would allow him answer the question he had raised in section 3.3, "what's the relationship between acute versus long-term exposure in terms of mortality?" At this point, he was not considering the representativeness of the study.

Expert G thought that to characterize the total mortality effects from long-term and short-term exposures, he would want to conduct a cohort and time-series study in the same population, to "control for … genetic differences" between populations. He would conduct the studies in the same three cities, one with a low PM level that's constant, one with a high PM level that has peaks, and one with an intermediate level. Ideally, all three cities would have PM with similar physical and chemical properties to avoid the difficulty of comparing across studies in areas with different PM component mixtures. In terms of exposure measurement, he would want the placement of monitors and types of measurements to reflect what the study population is being exposed to.

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert G first indicated that a limitation of PM/mortality epidemiologic studies in general is that PM is correlated with other pollutants, and therefore the investigators "can't separate out the PM [effect]."

Expert G thought that the American Cancer Society (ACS) study by Pope et al., the Six Cities study (Dockery et al., 1993) as well as the Krewski et al. (2000) reanalyses of these studies were the most informative for his quantitative estimates. He also discussed intervention studies (Utah Valley (Pope et al., 1996); Hong Kong (Hedley et al., 2002)) but thought these were more important for determining effects of different PM components, rather than for quantifying mortality effects in the U.S.

Expert G indicated that the strengths of the ACS study were its large population size, geographical distribution of the population, and its examination of the effect of co-pollutants. He thought the limitations were uncertainty related to the "monitoring sites versus where the population is" as well as downward bias due to effect modification by education and race. He thought the Six Cities had more representative exposure monitoring but was less geographically representative.

3.5 Confounding

Expert G chose to discuss confounders that affect the cohort studies generically, rather than describing the effect on individual studies.

Expert G first discussed confounding by gaseous co-pollutants, in particular ozone, NO_2, SO_2, and semi-volatile organics. He thought the gaseous co-pollutants were problematic because they are often correlated with PM and they "move like PM," making them difficult to control. However, he acknowledged that "[i]t's an open issue as to what role the gaseous co-pollutants do have in PM toxicology, if any." He expressed the view that SO_2 could be a surrogate for PM. He thought that correlation of PM with gaseous co-pollutants could cause the published effect estimates to be overestimated, and assigned this confounder a score of 3. He cited a similar problem with semi-volatile organics that can adsorb to particles making difficult to isolate the PM effect.

He then discussed other potential confounders that could affect study estimates. He thought that weather could be an issue in short-term studies because it "clearly will increase mortality in certain populations" and "PM does tend to go up when it's hot." However, he thought that this had been reasonably controlled in the epidemiologic studies.

He thought socioeconomic status (SES) could be a confounder, indicating that those living in inner cities could have higher exposures and lower SES. A related issue he raised was urban stress (e.g., related to living in close quarters, noise, job/money problems), which could reduce immune competence, could also be greater in inner cities. He was uncertain about whether it had been dealt with in epidemiologic studies using education as a proxy. He assigned SES a score of 2. In addition, he thought pre-existing health status could be a confounder as also assigned it a score of 2.

Expert G thought smoking, alcohol, and diet were well-controlled in the cohort studies.

3.6 Effect Modification

Expert G discussed potential effect modification in the ACS study (Pope et al., 2002). After extensive discussions, he stated that he thought that this study population underrepresented non-whites, low educational attainment, and low SES although he thought educational attainment might be capturing the other two factors to some degree. Therefore, he thought that the published effect estimates might be biased downwards.

Expert G thought that ozone could potentially be an effect modifier, in that those exposed to both PM and ozone would have higher mortality than PM alone. However, he could not cite specific evidence from the ACS study to support ozone as an effect modifier. He was not sure how much adjustment should be made, if any, to the published relative risk estimates, or whether any differences might already be captured in the statistical uncertainty of the estimates or whether he might need to expand the confidence interval.

3.7. Exposure Issues

Expert G thought that the differences in exposures measured by central monitors and actual exposures to the study population could be a major (score = 3) source of uncertainty. He thought that local sources, such as traffic or point sources could cause individual exposures to differ from exposure measured at a regional central site. He

thought that this could bias the results in either direction, depending on the placement of the central site and the types of local sources. He also thought that there could be differences in PM composition at the central site and individual exposures. Therefore, he concluded that a greater number of monitors or more representative monitors would improve study estimates. He thought that this could be a major factor, however, he did not feel that he had sufficient data to determine how this would affect the published relative risks. Although he was not familiar with the details of the Jerrett et al. (2005) LA study, for example, he generally thought that its findings might not necessarily be representative of findings in other parts of the country.

Expert G discussed the issue of whether the studies had measured the exposure period relevant to the mortality effects observed. He thought that using historical exposure estimates could cause an overestimate in the mortality effect. He also indicated that he was unsure of what the relevant exposure period should be, which contributed to overall uncertainty in the effect estimates, but to a lesser extent (score = 1).

3.8. Causality

When asked what type of evidence he relied upon to evaluate the strength of the causal relationship between PM and mortality, Expert G answered, 1) "reliable epidemiological studies having minimal confounders and be able to adjust for any confounders or effect modifiers and other factors to the best of their ability; … [2)] biological plausibility, meaning are there reasonable or believable mechanisms that could underlie the causes of death reported in the epidemiologic studies?"; and 3) coherence between epidemiology and toxicology, which includes both human controlled studies and animal studies as well as in vitro studies. "And even if there's a dosimetry issue between … the epidemiology and the controlled exposure … as long as there's a mechanism [to explain] how the PM could be doing this in human." He indicated that he did not view epidemiology as causal, but rather a "statistical relationship."

Expert G thought that the current state of science "strongly suggests a causal relationship between PM exposure … and increased mortality." This position represents a change in his views since the first Irvine colloquium in 1994 when the relationship had only been based on epidemiology and had little credibility in his view. He indicated that he put most weight on recent toxicological studies using concentrated ambient particles (CAPs), such as those published by Godleski et al. at Harvard, Devlin's group at EPA, and sub-chronic studies from New York University (NYU) by Lippmann and Chen. He indicated that the exposure levels are higher than ambient levels, but not by as much.

Expert G chose to discuss causality separately for short-term and long-term exposures because "I have more confidence in the short-term than in the long-term." Expert G thought that the likelihood of a causal relationship between PM and mortality for short-term exposures fell between 50 and 80 percent, with a most likely value of 70 percent. For long-term exposures, he did not feel comfortable providing a range of values because "at some point, I think the chronic turns into acute. I think the acute effects may override the chronic if levels are above a certain concentration," but provided a most likely value

of 60 percent. When asked what evidence he would need to make him 100 percent certain, he suggested that we would like to have more controlled exposure studies conducted at ambient concentrations, perhaps in sensitive animals, that demonstrated effects on mortality. He used smoking and its impact on heart disease and lung cancer as an example of a relationship that was 100 percent likely or causal.

When asked for an overall likelihood of causality for a change in annual average PM exposures, Expert G specified a most likely value of 70 percent, based on the "whole dataset together." He later provided a range of values around this estimate from 60 to 80 percent reflecting his uncertainty in the relative contribution of short-term and long-term exposures to overall mortality.

Expert G's discussion of the likelihood of a causal relationship was somewhat complicated by his belief that the strength of the causal relationship might differ within the range of 4-30 $\mu g/m^3$ and by the sense that the likelihood of a change in mortality might differ if one asked about a larger incremental change in PM concentration than 1 $\mu g/m^3$.

3.9 Thresholds

Expert G first discussed what evidence he would like to have to determine whether there is a threshold in the PM/mortality relationship. "You can do epidemiological studies, which can show an exposure-response relationship at different exposures, and then develop a dose-response curve as low as you can go. And you can look at animal and/or human controlled exposure studies to see with specific materials or … CAPs, to look at the dose-response there. So … for the population study, it needs to be a big enough group … to be representative of reality … [to include the] hypersensitive and hyposensitive [individuals]." He thought that animal studies could be informative "as to sensitivity of different populations to PM in sensitive groups within the population, and slopes of dose-response curves, but they probably won't be able to determine whether in reality there is a threshold for human populations exposed to ambient air."

On a conceptual basis, Expert G thought that "if there's some damage that the body is always able to repair, the repair mechanisms will come into play to fix things … what's disturbing from a mechanistic standpoint is that no matter how low you go, that there doesn't seem to be the ability to repair whatever is going on, to the extent that it mitigates any mortality."

When asked if his views on threshold differed for long-term and short-term exposures, Expert G thought there could be different thresholds. He felt there was a greater likelihood that there is no threshold for long-term effects than short-term effects "because … repeated … sub-clinical damage by PM, which could occur at low concentrations, would just build up and be cumulative … [w]hile for short-term exposure, I found it hard to believe, say, 2 $\mu g/m^3$ is going to kill somebody … So, conceptually, the no-threshold model makes sense to me for the long-term. There probably would be a threshold for the short-term."

He did not think that thresholds have been detected in epidemiologic studies, for example the ACS and Six Cities studies. "In the tox studies there are thresholds, but … [y]ou can't directly apply the toxicology studies to the epidemiology." Expert G did not elect to incorporate a threshold into his concentration-response (C-R) function.

3.10 Other Influential Factors

Expert G discussed additional sources of uncertainty that were not part of the protocol. When asked whether he thought that there was any publication or investigator bias, he answered that he thought that some investigators were biased, but he did not think that this affected the studies that he relied upon for his quantitative estimates. Expert G did not think there were any other outstanding issues not already covered by the protocol.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert G thought the C-R relationship was linear over the entire range of concentrations that were the focus of the study (4-30 $\mu g/m^3$). He elected to provide a C-R function that was conditional on the existence of a causal relationship. The elicitation team then combined his most likely causal likelihood (70 percent) with the conditional C-R function.

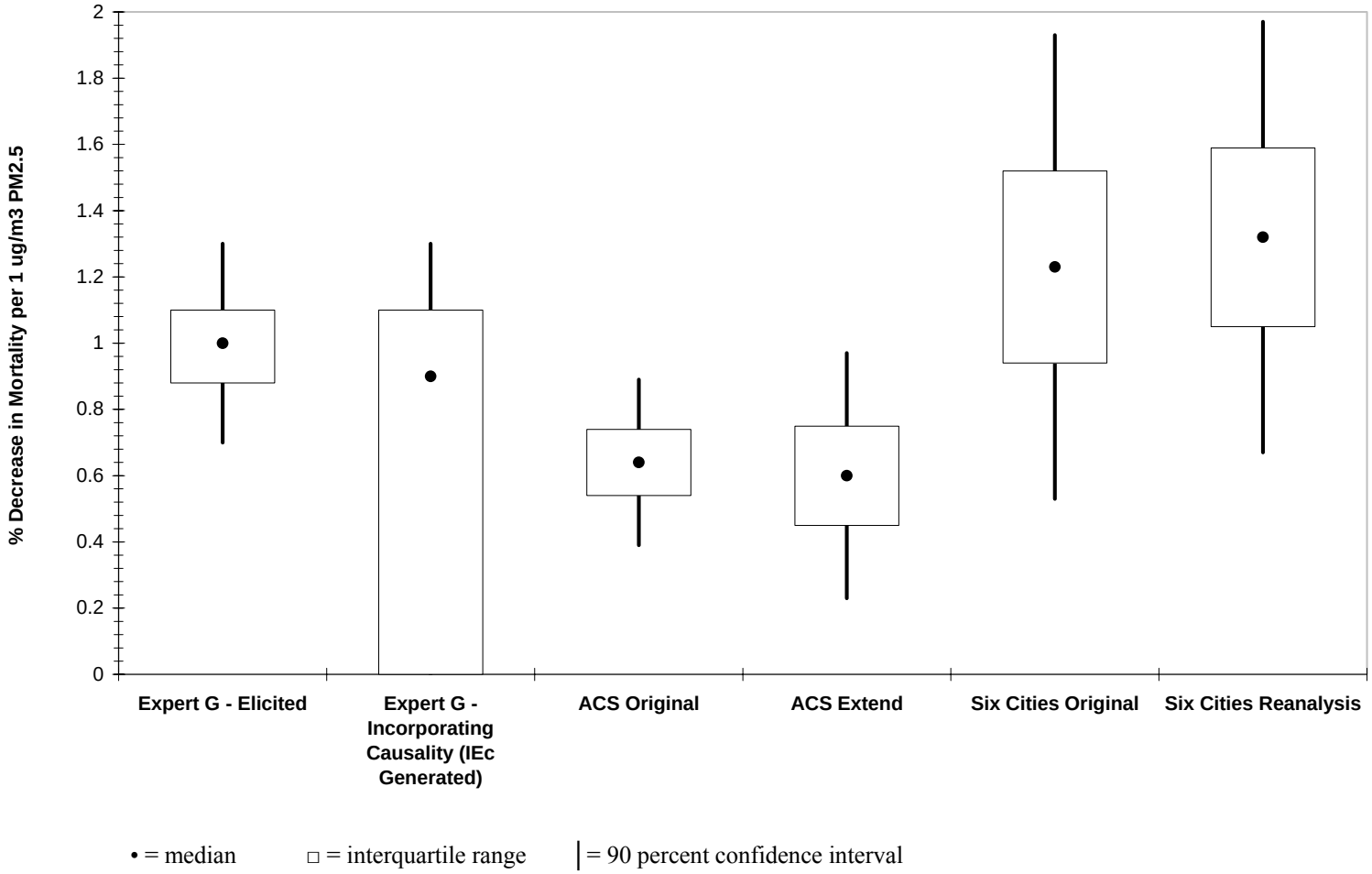
Expert G began by specifying his maximum. He specified a value of 1.5 percent based on a rough interpolation between the upper confidence intervals on the ACS and Six Cities studies, weighted on their respective strengths and weaknesses. He then specified a 95^{th} percentile value of 1.3. He chose a value relatively close to his maximal value, because he assumed that the C-R function was normally distributed and that the max and 95^{th} percentile would be close together. Expert G then specified a 50^{th} percentile value of 1.0 percent based on his view that the Six Cities estimate is biased upward due to confounding by co-pollutants and the ACS study was biased low due to effect modification of education and race. Although the issue of exposure misclassification was again discussed, Expert G remained unconvinced that the evidence supported a need for additional adjustment of the effect estimate. He suggested it be discussed at the post-elicitation workshop.

He then had the elicitation team fit the 50^{th} and 95^{th} percentiles to a normal distribution in Crystal Ball™ to determine his 5^{th}, 25^{th}, and 75^{th} percentiles. Expert G did not specify a minimum value. This distribution, which is conditional on the assumption of a causal relationship, is shown in the first column below. The second distribution is the probabilistic combination of the first with Expert G's likelihood of a causal relationship.

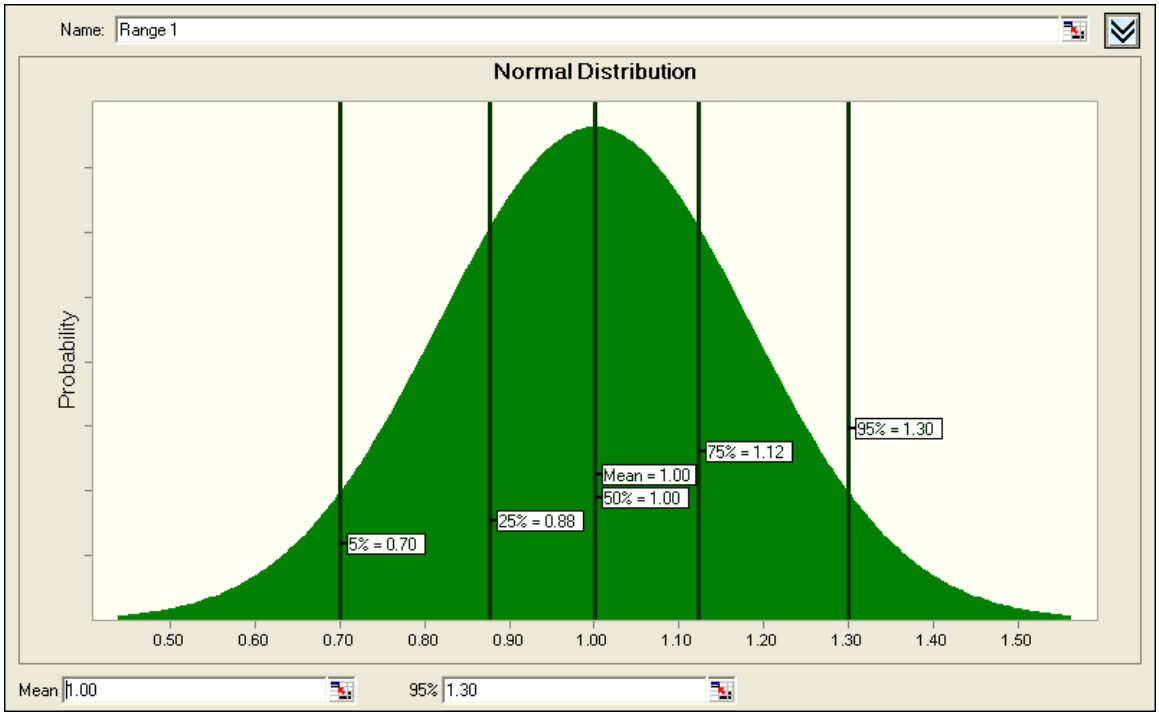
Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	-	0
5th	0.70	0
25th	0.88	0
50th	1.0	0.90
75th	1.1	1.1
95th	1.3	1.3
Maximum	1.5	1.5

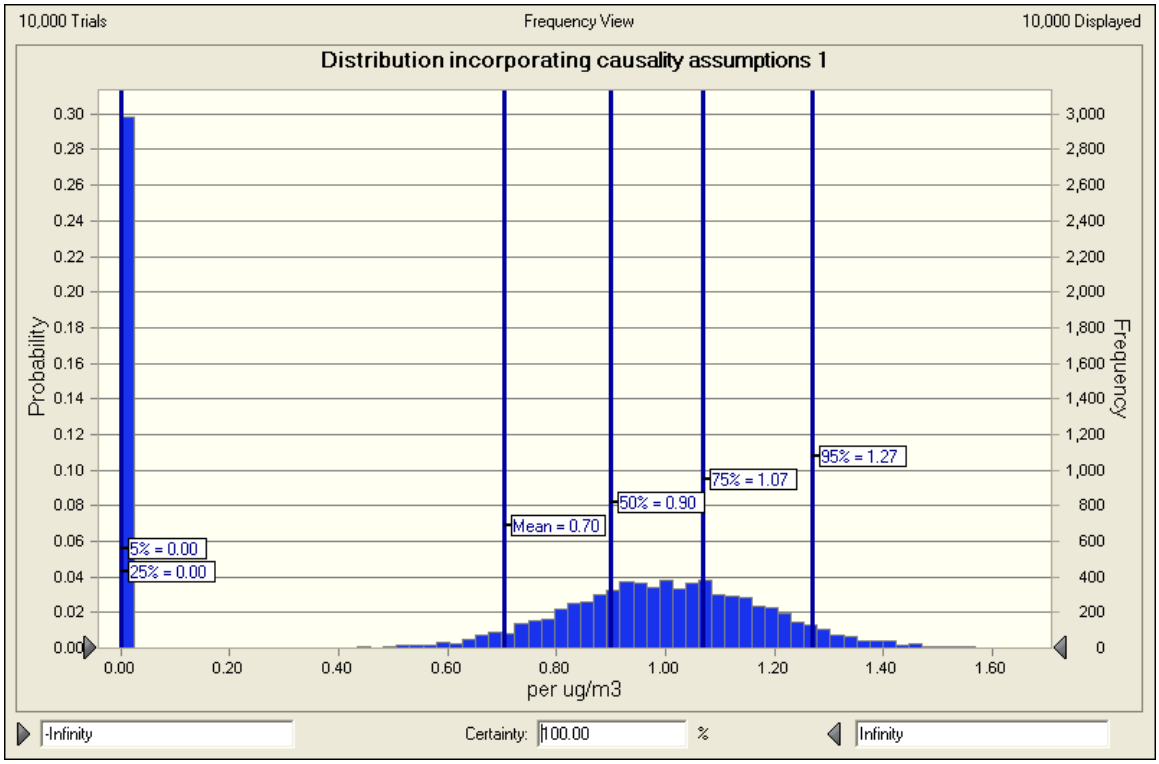
Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distributions from Expert G



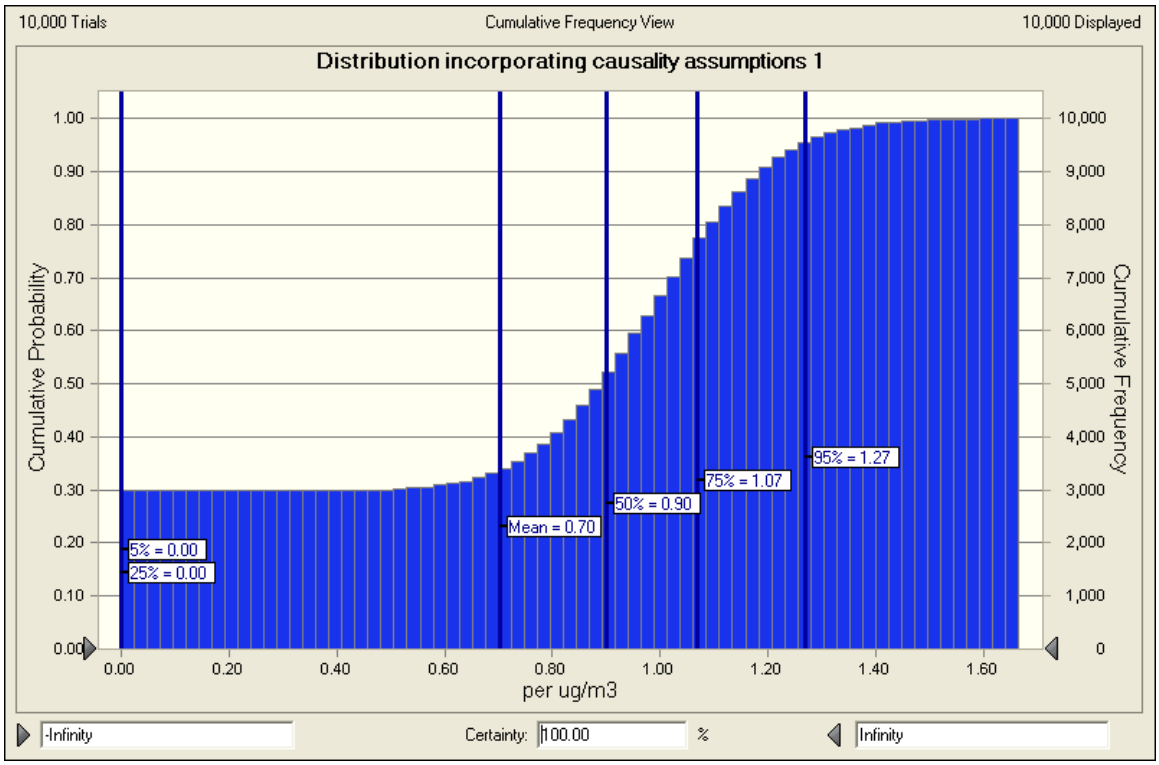
Elicited Distribution



Distribution Incorporating Causality - Probability Density Function (IEc Generated)



Distribution Incorporating Causality - Cumulative Density Function (IEc Generated)



U.S. EPA EXPERT ELICITATION STUDY OF THE CONCENTRATION-RESPONSE RELATIONSHIP BETWEEN ANNUAL AVERAGE $PM_{2.5}$ EXPOSURE AND MORTALITY

Modification to Expert Judgments

Expert G

Date: 7/11/06

Section of Protocol Affected (Section Number and/or Title):

Part 4 – Elicitation of Quantitative Judgments

Description of Change (e.g. to a specific percentile, or to a qualitative opinion or statement of belief):

The causality factor should be disassociated from the final numbers I gave you.

Expert H
Interview Summary

Interview Summary
Expert H

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert H discussed the biological mechanisms for short-term and long-term exposures separately, but he acknowledged that, "they may overlap."

Long-Term Exposures

He began by discussing causes of death from long-term exposures, indicating that the main causes of death were cardiovascular disease, respiratory disease, and lung cancer.

Expert H first discussed possible mechanisms for cardiovascular deaths related to PM. He thought that particles could cause systemic inflammation leading to increased arterial intima-media thickness (IMT) and atherosclerosis. He cited a combination of studies to support this theory, including an animal study by Sun et al. (2005), an epidemiologic study by Kunzli et al. (2005) of $PM_{2.5}$ and carotid artery IMT, and a study of environmental tobacco smoke (ETS) and IMT by Howard et al. (1990). He noted that before the Sun et al (2005) study, "I didn't see a very strong connection … between toxicological evidence and cardiovascular long-term mortality because many of the toxicological, human studies appeared to be more about acute response." Expert H mentioned that one complication with the Sun et al. study was that detection of increased IMT in mice with high fat diet suggests a major effect modifier and he wasn't sure the ACS or Six Cities studies had really looked at diet (other than using body mass index (BMI) as an indicator).

Expert H also thought that chronic exposures to PM could cause respiratory deaths, "even though it's not a major fraction of [overall] mortality." He cited a study by Ghio et al. (2000) that examined humans exposed to concentrated particles and found respiratory inflammation indicated by increased blood fibrinogen. He thought that chronic lung inflammation could eventually lead to respiratory mortality. He noted, however, that, "respiratory mortality is not … showing up [strongly] in long-term epi studies." He thought this could be due to the small numbers of deaths attributed to respiratory disease or that PM could be related only to short-term respiratory mortality.

He also thought it was plausible that particles could cause lung cancer although he did not think there were specific studies connecting $PM_{2.5}$ to lung cancer, except perhaps Pope et al. (2002). He drew an analogy to cigarette smoking and lung cancer, indicating that it was "conceivable … that combustion-related $PM_{2.5}$ may be as toxic or even more toxic than cigarette smoke." However, he also acknowledged that he was uncertain about how similar $PM_{2.5}$ is to cigarette smoke. He thought that occupational studies have also shown positive associations between lung cancer and exposure to diesel exhaust, although he was less clear about this.

Short-Term Exposures

Expert H thought that short-term exposures to PM were related mostly to cardiovascular deaths, specifically with outcomes such as heart failure, myocardial infarction (MI), and arrhythmia. He cited the defibrillator study by Peters et al. (2000), studies linking changes in heart rate variability (HRV) to changes in PM (e.g., Gold et al., 2000), and studies on increased blood viscosity by Peters et al. (1997). He thought that the strength of these studies was that they use human subjects; however, he thought their limitation was that the findings were not always specific to $PM_{2.5}$. For example, in the Peters (2000) he thought that the strongest association was with NO_2 or perhaps black smoke, which he still thought was indicative of $PM_{2.5}$. Similarly, he thought the Dockery et al. (2005) follow-up of that study also showed associations with sulfates, CO and SO_2, implicating both stationary and traffic-related sources. He thought the HRV studies were "less convincing" than the defibrillator studies because they had less consistent results.

3.2. Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

Expert H thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures. He thought that cohort studies would capture categories A, B and C but that not all of the short-term effects were captured by the cohort studies. In particular, he thought that, "cohort stud[ies] can miss some of [Category] C, and that is the small … portion of the short-term association that [could potentially account for] harvesting." He considered harvesting to include deaths pushed forward by just a few days and that it may be concentration dependent.

Although the notion of harvesting was conceptually appealing, he did not think that there was evidence in the literature to support it. He cited an analysis by Zeger et al. (1999) that simulated the potential effect of harvesting on a dataset. Zeger et al. then analyzed actual data and could not demonstrate the harvesting effects he anticipated. In addition, he indicated that papers by Joel Schwartz (2000) found increasing coefficient size with longer time window, which "is the opposite of [the] harvesting situation."

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

Expert H's views on the types of mortality effects captured using different study designs is summarized in the table below.

Study Design	Type of Effects Captured (e.g., short-term, long-term, or both)
Cohort Studies	Long-term and short-term minus harvesting
Cross-Sectional Studies	Long-term, and short-term minus harvesting
Intervention Studies	Short-term and some long-term
Time-Series Studies	Short-term

He thought cross-sectional studies would likely capture similar types of mortality as the cohort studies but expressed concerns about their ability to deal with individual level confounding.

He stated that time-series and case-crossover studies cover mortality effects of short-term exposures only. However, because he thought the short-term was largely captured by the cohort studies, he found the real value of the time-series studies to be their contribution to establishing a causal relationship. He indicated that they have less confounding to deal with and demonstrate the temporal relationship between exposure and effect.

When asked about the intervention studies, he thought they captured long-term effects but over a shorter period of time than the cohort studies. He thought of the intervention studies as a "special type of time-series study where you are estimating sort of long-term trends that usually [are not] estimated in time series [studies]," although he thought they had the potential limitation that there may be changes in certain factors that are concurrent with the change in air pollution (e.g., in behaviors) that could confound the PM/mortality relationship.

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert H thought that the following characteristics would be part of an ideal epidemiologic study to characterize the $PM_{2.5}$-mortality relationship in the U.S. population:

- Geographically representative of the U.S.;
- Population that is representative of the general U.S. population with respect to race and socioeconomic status (SES);
- Large sample size; and
- Exposure assessment/monitor placement appropriate to source types. "I think … [it's important to] have monitors that can possibly capture several major source types."

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert H first discussed the American Cancer Society (ACS) cohort study by Pope et al. (2002). He thought the strengths of this study were the long follow-up period, which allowed for a large number of deaths, the use of $PM_{2.5}$ monitoring data, large study size, and geographic coverage of the U.S. He thought the main limitation of the study was in the recruitment of subjects and that the population was more highly educated than the general U.S. population. He indicated that the ACS cohort had about 12 percent with less than high school education versus 28 percent in Six Cities.

He next discussed the Six Cities study (Dockery et al., 1993). Expert H thought that the strengths of this study were that the study authors carefully selected the six cities to include a wide range of mean PM exposure levels, that they placed their monitors in

central locations and in such a way as to avoid proximity to major point sources. He also thought that, although the limited geographic representation of the study could be seen as a limitation, it could also be seen as a strength because it provides "continuous geographic coverage of a certain area in the U.S." and thus "eliminate[s] factors that can actually confound" if they had chosen individual cities from different regions. He thought the Krewski et al. (2000) reanalysis of this paper resolved many of the questions surrounding its statistical analysis.

When asked to discuss evidence against a hypothesis of PM related mortality, Expert H argued that the debate is more "about good design versus bad design which is different from having several [well-designed] studies [with evidence] for and against." He briefly discussed two papers that he thought were not well-designed, the Adventist Health and Smog (AHSMOG) study and the Veterans (VA) study. He thought the AHSMOG study was well designed to look at the population included in that specific cohort (non-smoking non-Hispanic white Seventh Day Adventist in California), but not useful for quantifying the C-R function for the general U.S. population. He indicated that the cohort in the Veteran's study was too specialized (mild to moderate hypertensives, with a large fraction of former and current smokers) to be useful, not just for this elicitation project, but even for the general public health research agenda, and in that sense, it was not well designed.

He thought that the information from the intervention studies could be qualitatively useful as support for effects responding to a reduction in exposure, but he did not rely on them for his quantitative estimates.

3.5 Confounding

Expert H's review of confounding applied to both the ACS and Six Cities studies. He indicated that in the Krewski et al. (2000) reanalysis of the ACS and Six Cities studies, no individual level confounders were found to significantly affect the original quantitative estimates. However, some ecologic or contextual variables did change the estimates when included in the models. In particular, he thought that "population change" and spatial auto-correlation could be potential confounders.

He indicated that "population change" occurs because survival times of individuals may be more similar in closer distances than those who live farther apart. This creates spatial (positive) autocorrelation. If this pattern is created by some unmeasured risk factors that also coincide with spatial autocorrelation in the PM levels, then confounding may occur. However, he thought both the Krewski et al.'s (2000) re-analysis of ACS data and Pope et al.'s (2002) extended analysis addressed this issue by fitting spatial smooth functions and the results did not change much. This led him to believe that confounding by this factor was small. He assigned both population change and spatial auto-correlation a score of minimal (1) for both the ACS and Six Cities studies because although he thought they could be correlated with PM, he could not see how they would be able to cause death, and therefore were not true confounders. Although he also discussed the fact that SO_2 affected the mortality estimates in the Krewski et al. (2000) analysis, he thought SO_2 was actually a precursor of, and thus a surrogate for $PM_{2.5}$, rather than a confounder.

On a more speculative basis, Expert H also thought that high fat diet and environmental tobacco smoke (ETS) could be potential confounders that would lead to an overestimate of the PM effect in the ACS study. He assigned a both a score of minimal (1) given his perceived lack of strong evidence.

Expert H also discussed the potential for residual confounding in the Jerrett et al. (2005) paper. He speculated that the Jerrett et al. study might be affected by residual confounding because "I see additional uncertainty in the Jerrett et al study, in comparison to the Krewski et al. (2000) re-analysis or the Pope et al. (2002)." He provided the following rationale:

(1) The geographic scale in the Jerrett et al study is much smaller than that for the nationwide ACS studies, and while this scale has the advantage of examining the within-city variation of $PM_{2.5}$, judging from Figure 1 in the Jerrett et al study, there appears to be strong spatial autocorrelation in the $PM_{2.5}$ data in this locale at this scale;
(2) Jerrett et al noted that in models with only individual covariates and $PM_{2.5}$, some residual spatial autocorrelation was present, but inclusions of ecologic (contextual) variables eliminated residual spatial autocorrelation. To me, this suggests that the ecologic variables do have spatial autocorrelation that coincide with the spatial autocorrelation in the survival times of the cohort; and
(3) In the Jerrett et al study (Table 1), the addition of ecologic variables did result in reductions of $PM_{2.5}$ risk estimates (e.g., from 1.17 with individual covariates only to 1.11 with ecologic variables). In contrast, in Krewski et al's sensitivity analysis of nationwide ACS data, the addition of similar ecologic variables did not seem to affect $PM_{2.5}$ risk estimates.

Expert H concluded that, "I cannot think of a major confounder that would influence this estimate. I guess [the] more important question to answer is about effect modifiers."

3.6 Effect Modification

Expert H thought that the main effect modifier of the PM/mortality relationship in both the ACS and Six Cities studies was educational attainment. He indicated that Krewski et al. had examined a number of different variables but education was the only one that exhibited strong effect modification. He indicated that education was probably a surrogate for other factors related to healthier lifestyles, such as diet, access to medical care, and housing characteristics.

He thought that the Six Cities cohort overrepresented those with less than a high school education (leading to an overestimate) and the ACS study underrepresented this group (leading to an underestimate) in comparison with the general U.S. population. He assigned a score of 2 for this factor in each of the two studies.

Expert H indicated that the size and importance of this adjustment really depended on what the real statistics are about educational attainment in the study populations versus the U.S. population, as well as the relevant measurement period. The Six Cities study had 28 percent with less than a high school education, whereas the ACS study had 12 percent. Expert H found a statistic of 20 percent for the U.S. population from the U.S. Census and the elicitation briefing book had a graph from the U.S census bureau showing that in 1980, about 32 percent of the population aged 25 and over had less than high school education. By 2003, only 15 percent had less than high school education. If the relevant point in time is the educational level of the cohort at enrollment, then the Six Cities study is more representative of the U.S. at that time. If the U.S. average for 1980 (or whatever the right comparison time is) is closer to 20 percent, both studies are off but in opposite directions. He also noted that there was about 45 percent with less than a high school education in Steubenville versus 12 percent in Topeka raising the possibility of confounding rather than effect modification.

He also discussed the possibility of effect modification by SO_2. However he indicated that studies by Joel Schwartz (2000) and the National Morbidity, Mortality and Air Pollution Study (NMMAPS) examined this issue and did not find supporting evidence. In addition, he thought that in the short-term "it's conceivable that the areas where you have certain components of $PM_{2.5}$, you may see a strong association of $PM_{2.5}$," but he did not think it would affect the overall effect estimates.

3.7. Exposure Issues

Expert H thought that differences between central site concentrations versus individual exposures could be an important exposure issue. He indicated that the ACS cohort had been reanalyzed by Jerrett et al. (2005) and had found that in Los Angeles (LA), better spatial resolution of exposure lead to increased effect estimates. However, he thought that it was "not straightforward" to interpret the increased effect as being attributable solely to the central site versus individual exposure issue because of the potential pollutant mix and because these results were specific to one city. He noted that the as yet unpublished Jerrett work in New York found mixed results: no association for all-cause mortality and an association as large or larger than the LA study for cardiovascular mortality. Overall, he thought this type of exposure misclassification would lead to an underestimation of mortality effects in both the short-term and long-term studies. However, given the disparity in the ACS results in the two cities he felt that he could not determine by what magnitude the effects might be off. He did state that the LA study might provide an upper limit on the magnitude of the downward bias and that it might be an overestimation. He assigned a score of 1 to 2 to the ACS study for exposure misclassification but a score of 1 to the Six Cities study because he thought the investigators selected more representative monitoring sites in the latter study.

When asked about the relevant time window of exposure, Expert H expressed the view that for cardiovascular deaths, the critical window would be 10 years, and it would be longer for cancer. He thought that since "the ranking of PM levels don't change much

over time across these cities, I don't think it's much of a problem." He thought it would be more of a problem if the critical period is 40 years or so before, but he thought the evidence was more consistent with a 10-20 year period. When asked whether the intervention studies suggested a shorter critical time window of a few years, he said that he did not think so. He thought it was difficult to discern "how much of the intervention drop is accumulation of short term process [or] long-term effects."

He did not think any of the other exposure issues listed in the protocol were influential.

3.8. Causality

In general, Expert H thought a causal inference would be supported by epidemiological studies that include testing of all kinds for confounders and other factors. He would want to see strong and consistent associations, not just for mortality, but also for subclinical markers like IMT. He would like to see evidence demonstrating an appropriate temporal relationship between exposure and mortality effects as in the intervention studies. Ideally, he would like clinical or other studies that demonstrate a plausible biological mechanism. "But I don't think it's absolutely necessary."

He thought the ACS and Six Cities studies, in particular were important evidence of a causal relationship between long-term $PM_{2.5}$ exposures and mortality.

Expert H thought it was easier to think about a causal relationship for short-term exposures and mortality than for long-term exposures. In particular, Expert H felt intervention studies provide potential support for a causal relationship for short-term exposures and mortality. He indicated that, "I think it gets more difficult with [the] longer period [of time] you try to estimate risk reduction for." He also mentioned a study by Clancy et al. (2002) that examined changes in mortality in Dublin after a coal ban, and a study by Hedley et al. (2002) in Hong Kong. Although in the Hong Kong study the effects appeared to be more related to SO_2 than PM, he thought that SO_2 might be serving as an indicator for something else in the source emissions, in particular metals. He cited a poster by Hedley showing that nickel levels had been reduced along with SO_2 and inferred that nickel might be the important agent. He also discussed a laboratory study that exposed cells to ambient PM samples in the Utah Valley during the time when area steel mills were functioning and found an inflammatory response that was not present for particles from the mill closure period (Dye et al., 2001). Transition metals are believed to play a role in the inflammatory response.

He indicated that the studies that he discussed when answering questions in Section 3.1 regarding biologic mechanisms were also supportive of a causal relationship (Sun et al. (2005); Howard et al. (1990); Ghio et al. (2000); and Kunzli et al. (2005).

Expert H specified a range of the likelihood of a causal relationship of 70 – 95 percent with a most likely value of 80 percent. His lower bound was based on "[j]ust the fact that there's no counterpart to intervention studies for the long-term effects. It's hard to establish direct, convincing proof." His upper bound was based on "the results from the

sensitivity analysis of the ACS and Six Cities studies [which] alone give me a lot of confidence." However, in later characterization of his uncertainty in the magnitude of the C-R relationship, he did not think that an 80 percent likelihood of a causal relationship was consistent with his views that there are no major confounders of the PM/mortality relationship in the key epidemiological studies. He adjusted his likelihood upward to 90 percent for his most likely value, and changed his range of values to 80 – 95 percent.

3.9 Thresholds

Expert H thought, "[c]onceptually, you can come up with [a] threshold for some … very narrowly defined sub-group that [has a] uniform level of frailty [or] sensitivity … [b]ut when you are talking about a population that is diverse in terms of susceptibility, effect modifying factors, confounding factors, I think you are bound to come up with something that is monotonic … I cannot think of any reason why we should see a population threshold in this range." His views were the same for short-term and long-term exposures.

When asked what types of evidence would be informative for assessing a threshold level, Expert H thought that, "it would be easier to find a threshold in toxicological studies, only because you can have uniform levels of susceptibility." However, he did not think that the results of toxicological studies could be used to support a U.S. population threshold.

He did not think that a population threshold was detectable in any of the studies currently available.

Expert H did not opt to include a threshold in his C-R function.

3.10 Other Influential Factors

Given his emphasis on the cohort studies, Expert H did not think there were any other influential factors within those studies that would change the estimates. He did mention that it would be helpful in the future to examine the factors that are involved in the effect modification by education, since it was still somewhat vague. He indicated that he would have been more skeptical about the statistical modeling in the epidemiologic studies if it were not for the Health Effects Institute's extensive sensitivity analysis of the cohort studies' data.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert H thought the C-R relationship was log-linear over the entire range of concentrations that were the focus of the study (4-30 $\mu g/m^3$) based on the statistical model used in the ACS and Six Cities studies and because it was consistent with his view that there is no threshold.

Expert H based the values for his uncertainty distribution on a subjective weighting of the ACS (Pope et al., 2002) and the Six Cities Study (Dockery et al., 1993) and adjustment

for exposure misclassification and for differences in educational attainment relative to the U.S. population. He began by specifying his 50th percentile because he first wanted to take into account the education and exposure issues in his midpoint before characterizing the other percentiles. He started with an estimate from Pope et al. (2002) of 5.9 percent per 10 $\mu g/m^3$. He then decided to adjust it upward to account for exposure misclassification, which he supported with the Jerrett et al. (2005) ACS LA analysis. This study found an effect estimate of 11 percent per 10 $\mu g/m^3$ for the model including 44 individual covariates and parsimonious contextual covariates. However, given that the study was conducted in just one city and that he thought there was epidemiological evidence suggesting all-cause mortality is higher in LA than, for example, in the Midwest, he was reluctant to adjust his basic estimate up by a factor of 2. He also thought that even this model could be overstating the effect because of some residual confounding or for spatial representativeness or "spatial correctional factor taking into consideration it was LA where educational attainment is not the same as U.S. general." He therefore moved his central estimate up to 7 percent per 10 $\mu g/m^3$.

Expert H raised an interesting question about the adjustment for effect modification by education. Initially, he thought that educational attainment could be causing the Pope et al. estimates to be underestimated. But, after learning that the U.S. Census bureau[7] reported in 2003 that 15 percent of the U.S. population aged 25 and older had less than a high school diploma, a percentage which does not differ drastically from the percentage in the ACS cohort (12 percent), Expert H decided not to adjust for this factor. Therefore, his 50th percentile value remained at 7 percent per 10 $\mu g/m^3$ (0.7 percent per 1 $\mu g/m^3$).

Expert H then specified his 75th percentile value at 1.3 percent per 1 $\mu g/m^3$ based on the estimate from the original Six Cities study. When thinking about his 95th percentile value, he relied on the estimate from Jerrett et al., 2005 (adjusted for 44 individual covariates) of 1.7. He thought it could be slightly higher and set the value at 2. He then set his maximum at 3 percent to ensure that his distribution would be smooth.

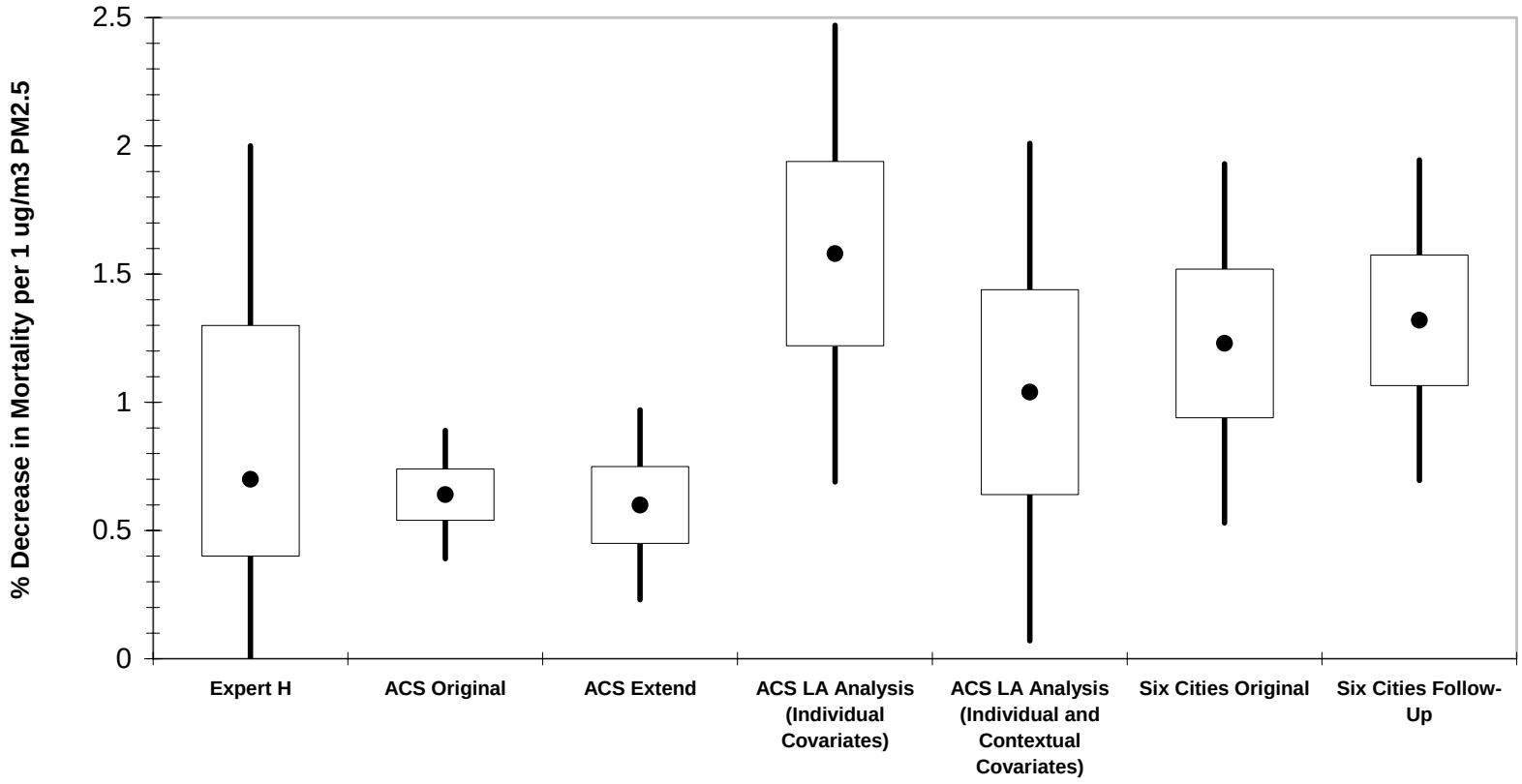
After reviewing his assumptions on causality, and electing to change his most likely value for the likelihood of causality from 80 percent to 90 percent as discussed in an earlier section, Expert H set his minimum and 5th percentile values at zero and his 25th percentile at 0.4 percent per 1 $\mu g/m^3$.

[7] U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, 2003 (http://www.census.gov/rod/2004pubs/p20-550.pdf).

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Expert H

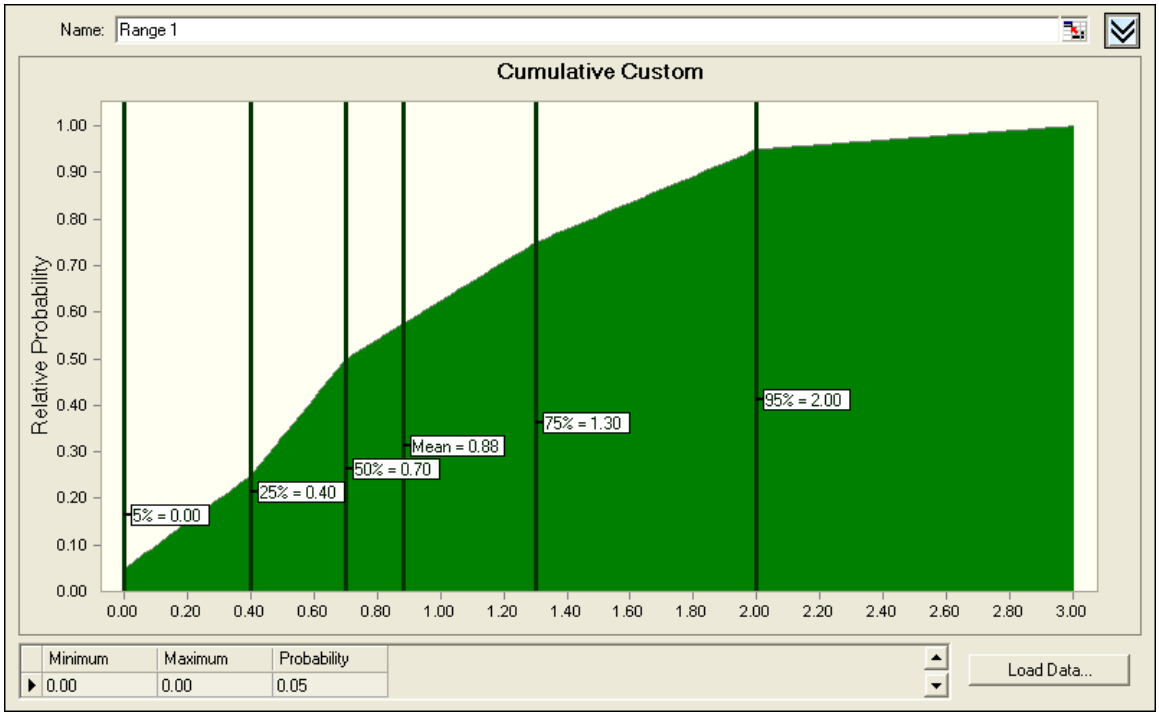
Percentile	Percent Change in Mortality
Minimum	0
5^{th}	0
25^{th}	0.40
50^{th}	0.70
75^{th}	1.3
95^{th}	2.0
Maximum	3.0

Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distribution from Expert H



• = median □ = interquartile range | = 90 percent confidence interval

Expert H Distribution



Expert I
Interview Summary

Interview Summary
Expert I

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert I discussed the biological mechanisms for short-term and long-term exposures together because he thought that, "there are … similar mechanisms involved." Expert I thought that the main cause of death from particulate matter (PM) was cardiovascular disease. He thought the key mechanisms related to this cause of death were oxidative stress and inflammation leading to atherosclerosis, changes in autonomic function, and other markers of heart disease.

He first discussed oxidative stress and inflammation leading to plaque development and atherosclerosis, stating that this mechanism was first brought to light in a study by Seaton et al. (1995) that showed "inflammation and blood coagulation relating to the accumulation of particles as a potential pathway." He thought papers by Pope et al. (2004) and Jerrett et al. (2005), although they do not shed light on specific pathways and mechanisms, provide supporting evidence for mortality from various cardiovascular disease outcomes as a result of long-term exposures. Additionally, he cited epidemiologic studies that give insight into mechanistic pathways, such as Kunzli et al.'s (2005) study on atherosclerosis, and animal studies such as Wellenius et al. (2003), showing ischemic heart disease in dogs as a result of fine particles, and Sun et al. (2005), that found plaque development, inflammation, and mortality in mice.

He then discussed autonomic function effects. He indicated that this mechanism was supported by studies that have found changes in heart rate variability (HRV) associated with PM, such as Pope et al. (1999 & 2000), Gold et al. (2000), Creason et al. (2000), and Devlin et al. (2003). He indicated that, "heart rate variability, the different measures, both the time domain and the frequency domain, are good predictors of congestive heart failure and sudden death." While some have argued that acute changes in HRV are really transient effects, he thought a recent study by Schwartz (2005) provided evidence for a longer term role; the study essentially looked at HRV in individuals with different genotypes that make them more or less susceptible to oxidative stress. Individuals with genotypes more likely to experience oxidative stress had greater changes in effective HRV than those who did not.

Expert I thought that studies examining other markers of heart disease provided evidence for a link to air pollution, although they have not been as well-studied. For instance, he cited a paper by Zanobetti et al. (2004) that found blood pressure changes, and a paper by Brook et al. (2002) examining arterial vasoconstriction, a study by Schwartz et al. (2005) looking at changes in inflammatory markers in the blood such as fibrinogen and white blood cells, and the Peters et al. (2000) implantable defibrillator studies.

He thought that respiratory deaths from the progression of Chronic Obstructive Pulmonary Disease (COPD) and lung cancer were secondary causes of death from PM. He indicated that the Six Cities cohort study extended analysis (Laden et al., 2006) did not find a separate effect for respiratory deaths. However, he thought there was some evidence for a mechanism by which inhalation of particles could have effects on the respiratory system. He drew an analogy with smoking, which involves higher concentrations but over shorter periods of time than air pollution, which is a more constant exposure. He cited studies by Gaudermann et al. (2004) that showed long-term effects on lung function in children exposed to $PM_{2.5}$ and a paper by van Eeden et al. (2005) whose results indicated that, "effects on the COPD from long-term exposures could also have cardiac effects." He indicated that epidemiologic evidence for lung cancer as a result of PM is "quite reasonable and robust" specifically citing the American Cancer Society (ACS) cohort study (Pope et al., 2002) and the Adventist Health and Smog study (AHSMOG) (Abbey et al., 1999).

Expert I discussed whether these mechanisms were related to long-term exposures, acute exposures or both. He thought that both types of exposure could contribute to mortality, noting a parallel with smoking. He thought that the pathway involving inflammation and build-up of atherosclerotic plaque leading to increased risk of death from heart attack was likely a long-term process. However, the evidence of shorter-term exposures on changes in HRV and "increased markers of inflammation, like fibrinogen [and] white blood cells … [suggests that acute exposures could affect individuals who are] already compromised with heart disease, whether the disease is brought on from exposures to air pollution or not." He thought the development of COPD and lung cancer was related to long-term exposures. He thought the mechanism for PM's impact on lung cancer was analogous to the mechanisms for smoking and lung cancer.

3.2. Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

Expert I thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures. "I'm willing to accept this diagram. I think it describes adequately the possible scenarios." He thought that not all of the short-term effects were captured by the long-term studies, but was not certain enough of the exact amount that he would want to add in additional deaths related to short-term effects to those reported in the long-term studies. He thought it was a small percentage, no more than ten percent.

We had a lengthy discussion of the challenges to estimating how many deaths fall into different exposure windows, the related years of life lost, and what deaths are ultimately captured by various study designs. Expert I suggested these topics would be good ones for discussion at the Post-elicitation Workshop.

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

Expert I thought that a combination of cohort, cross-sectional, intervention, and short-term studies looking at distributed lags of various lengths were appropriate for capturing the mortality effects of changes in annual average $PM_{2.5}$ concentrations. The effects captured by each study design are shown in the table below:

Study Design	Type of Effects Captured (e.g., short-term, long-term, or both)
Cohort Studies	Long-term, and potentially some short-term
Cross-Sectional Studies	Long-term, and potentially some short-term
Intervention Studies	Short and Intermediate-term
Time-Series Studies with Distributed Lag	Short (multi-day) and Intermediate-term (days to weeks)

Expert I also expressed the opinion that there was need for a better understanding of the biological explanation for the different mortality effect sizes observed in short-term, intermediate-term (intervention), and long-term exposure studies. Regarding the intervention studies, for example, "they're telling us that something other than same-day exposures or same week exposures seem to be important, but it's not clear whether it's several months or several years."

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert I thought that the following characteristics would be part of an ideal epidemiologic study to characterize the $PM_{2.5}$-mortality relationship in the U.S. population:

- Includes a cohort population that is representative of the national population;
- $PM_{2.5}$ measured over the lifetime of the cohort members;
- Personal monitors (home, workplace, everywhere) in addition to central site monitors with which the personal monitors could be related;
- Measures of fine, coarse, and ultrafine particles, gases, benzenes, and benzo-pyrenes;
- Measures several confounding factors such as body mass index (BMI), age, gender, race, alcohol, tobacco, smoking histories, occupational exposures, socioeconomic status (SES), extreme weather events, diet, psychological measures, hormone replacement, and neighborhood effects;
- Periodic follow-ups including residential histories; and
- Large samples to allow for stratification on several different factors.

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert I discussed the following studies:

Study (author, date)	Key findings	Strengths	Limitations
ACS Cohort (Pope et al., 2002 & 2004)	• Effects on cardiopulmonary disease and lung cancer • Effects for cardiovascular disease-specific subsets (Pope et al., 2004)	• National sample (includes variety of particle mixes) • Controlled for many confounders • Replication by Krewski et al. (2000) • Large sample size • Historical data on pollution • Range of exposures • Well-established cohort	• Not representative of the general U.S. population, with respect to lower-income, minority, and education • Gaps in exposure estimates • Exposure assessment problems
Jerrett et al., 2005	With improved exposure metric, found associations with cardiopulmonary disease and lung cancer	• Improved exposure metric • Careful consideration of spatial auto-correlation • Less inter-city variation and confounding	• No historical pollution data • Possibly not representative of national mixes of PM • Not representative of the general U.S. population
Laden et al., 2006	Found both long-term and relatively recent exposures to fine particles associated with cardiovascular disease	• Monitors set up for the purpose of the study and well-followed • Representative population • Random population design • Longer-term and recent set of data on pollution	• Only included 6 cities • Limited geographic representation of the U.S. • Extrapolation from PM_{10} to $PM_{2.5}$
Krewski et al., 2000	Replicated original findings of the ACS and Six Cities studies with extensive sensitivity analysis. Found that results were robust.	• Objectivity in approach and analysis • Extensive analysis including multiple confounders, such as smoking and occupation.	• Potential for multiple comparisons (i.e., over manipulation of the data)
AHSMOG (Abbey et al., 1999; McDonnell et al., 2000)	Generally inconsistent effects, often gender-specific	• Less confounding from alcohol, smoking, and environmental tobacco smoke (ETS) • Frequent follow-up allows for a complete exposure profile • McDonnell et al. analysis of this cohort had improved exposure estimates	• Very small sample • Not representative of the general U.S. population • Used airport visibility as exposure measure
Enstrom et al., 2005	Effect found in the younger portion (43-65 at enrollment) of the cohort. Found smaller effect estimates than other cohort studies.	• Moderate sample size	• Potential healthy survivor effect • Elderly cohort (many 65-99 at enrollment) • Questionable objectivity • Potential confounding (ETS, smoking)
European studies (Hoek et al., 2002; Filleul et al., 2005)	Found effects related to either traffic or to measured pollutants	• Long-term follow-up	• Relatively small sample sizes • PM mixture potentially differs from that in the U.S. • Different SES/demographics factors than in the U.S. (more central city, less suburban) • No direct measure of $PM_{2.5}$ • Small sample size

Study (author, date)	Key findings	Strengths	Limitations
Intervention studies (Clancy et al., 2002 – Dublin; Pope et al., 1996 – Utah Valley; Hedley et al., 2002 - Hong Kong)	Associations between relatively short-term exposures and mortality. Effect estimates higher than short-term studies.	• Large change in exposures, and exposure conditions very clear	• Not capturing all of the long-term changes (intermediate between short-term and long-term)
Veteran's Cohort (Lipfert et al., 2000 & 2006)		• Used population at risk (hypertensive) so more power to detect associations	• Population not representative of the general U.S. population • Small sample size • Unclear and perhaps inappropriate statistical analysis

3.5 Confounding

Expert I thought that, based on Krewski's reanalysis of the ACS and Six Cities studies (2000) and the Jerrett et al. (2005) ACS L.A. reanalysis, many of the factors commonly thought to be confounders did not substantially affect the estimates. He thought that there were three remaining issues of "particular interest":

- **SES:** Expert I thought that SES could be a marker of a "constellation" of other confounders, such as alcohol, diet, occupational exposure, smoking, and exercise. He thought that SES could be a proxy for exposure, because those with higher SES are not likely to live near a major roadway or other pollution sources. "When we do find the effect modification of SES … the SES could be picking up pure income [or] health habits, [but] it also could just be modeling exposure."
- **Pre-Existing Health Status:** Expert I thought that conceptually, "people with pre-existing health status [e.g., cardiovascular disease] should probably have a large effect, because they're on the pathway." He indicated that Krewski et al. (2000) had examined this issue and did not find higher effect estimates in people with pre-existing health issues. However, he did point out that those with pre-existing health problems could be susceptible to short-term exposures, so the effects might not be easily captured by the long-term studies.
- **Temporal Trends:** Expert I thought that exposure changes over time could potentially affect the published estimates, although the implications for the mortality estimates are not entirely clear. On the one hand, one might expect that early childhood exposures might be very important. However, Jerrett's analysis of this issue for the L.A. portion of the ACS cohort, as well as some calculations done by Expert I, convinced him that this potential source of error would not make a big difference in the effect estimates even under the most extreme assumptions. He indicated that results from Laden et al. (2006) found that the previous one or two years of exposure were the most influential for effects. He

was unsure of the disease process that could explain the bigger effects found when a more recent exposure period is used.

In addition to those three topics discussed above, Expert I also touched on the question of co-pollutants and contextual ecologic variables.

- **Co-pollutants:** Expert I thought the issue of co-pollutants had been handled statistically, though "the explanations have not been developed yet." He was concerned that it was difficult to measure co-pollutants well in long-term settings with central site monitors because they are too localized in terms of exposure. He discussed only SO_2 specifically. He thought that SO_2 might be a confounder (but representing something other than pure SO_2). That is, he did not think there was a plausible biological mechanism linking SO_2 and mortality. His understanding was that it was unlikely to get into the deep lung and cause the types of inflammation that we discussed as part of the mechanisms for health effects. He said that clinical studies have shown health effects to occur only in asthmatics that are exercising while exposed to high levels. He thought it might be more of an effect modifier. He thought SO_2's role would be a good subject to discuss in the Post-elicitation Workshop.
- **Ecologic Variables:** Expert I thought that ecologic covariates, or factors outside the individual level, such as neighborhood effects or air conditioning use, were potential confounders. However, he thought that the issue of how much to control for them remains unclear. He thought there is a strong likelihood of over-control if these factors are correlated with pollutants. He also discussed the possibility that unmeasured spatial variables cause mortality rates to correlate among closely spaced cities. He indicated that Jerrett et al. (2005) found reduced effect estimates after controlling for spatial auto-correlation in LA, but that the Pope et al., 2002 study found that it did not affect results for the full cohort.

Ultimately, Expert I did not want to assign scores, or indicate the degree of adjustment he might make as a result of any of these issues, indicating that he would rely on effect estimates from studies that had adjusted or accounted for these factors.

3.6 Effect Modification

Expert I thought that the only variable for which empirical evidence existed for effect modification was educational attainment. He indicated that the effect estimates in the ACS study were biased downwards because the cohort population overrepresented highly educated people. He discussed the possibility of effect modification by criteria co-pollutants. He did not think that there was evidence for effect modification or interaction effects (e.g., SO_2 with particles). The one exception he cited was the Air Pollution and Health – A European Approach (APHEA) study of short-term exposures that found greater effects for PM in cities with high NO_2. He thought that the NO_2 effect has been interpreted to be a proxy for traffic and diesel, meaning that "that in those areas where particles are dominated by traffic, the effects might be higher."

3.7. Exposure Issues

Expert I discussed three issues: 1) relevant time period of exposure; 2) exposure misclassification related to location of monitors (i.e., central site versus averages of monitors versus monitors closest to the home); and 3) pollutants measured (e.g., PM_{10}, $PM_{2.5}$, PM_1, elemental (EC), organic carbon (OC), metals)

Expert I discussed potential exposure misclassification resulting from monitoring $PM_{2.5}$ during time windows that may not be most relevant for the effect. He thought that if childhood exposures were important and past concentrations were higher, risk would be overestimated based on monitoring in the years immediately preceding death. On the other hand, he though that if very recent exposures were the critical window and an annual average or several year average was used, risks could be underestimated. He thought that this was not a major influence, and thought it was "a neutral bias, slightly weighted toward overestimation." He then went on to quantify the influence of this factor, stating that the effect estimates could be overestimated by about 5 to 10 percent, based on information from the Jerrett et al. (2005) paper.

Expert I then discussed the impact of location of PM monitors and techniques for estimating local exposures. Based on the "Hoek [2002] study and the Jerrett [2005] study, which used localized measures [of exposure], there's some reason to believe that the estimates from the other studies are too low, based on measurement misclassification." He thought that this was more of an issue in studies that use few monitors for a large geographic area, such as the ACS study. He thought that this could be a potentially large bias, and published effect estimates from the ACS study could be underestimated by a factor of two (for example, the ACS estimate is 6 percent but the Jerrett study finds 15 percent per 10 $\mu g/m^3$).

Finally, Expert I discussed the role of the pollutant mix, specifically the potential that we are not measuring the causal pollutant. He was not sure whether this is a big factor in estimating the mortality effect. On one hand, if the mis-measured causal pollutant is correlated with PM, he did not think there would be an issue. He thought that it could be an issue if that pollutant's concentration relative to PM varies by region. He cited studies of short-term exposures by Laden et al. (2003) and Mar et al. (2000 & 2003) that suggest, "traffic-related particles seem to be a little worse than oil residual particles." He noted the difficulty of knowing, for example whether particle constituents (or traffic related particles) were a possible explanation for the higher effect estimates measured in the Jerrett (2005) study in L.A. However, he indicated that overall, the evidence on the relative toxicity of PM components was limited and he would not feel comfortable adjusting his quantitative estimates of the mortality effect related to long-term exposures based on this issue.

3.8. Causality

Expert I thought that a causal relationship would be best supported by "replicated epi studies that are well designed and carefully done." In addition, he said it is useful but not

necessary to have good mechanistic hypotheses. He also mentioned temporality, exposure response, and plausible quantitative results.

"First of all, based on my own criteria, there has been replication in different forms of the original studies[;] the fact that [the ACS and Six Cities studies] were replicated by Krewski … is a very powerful evidence for a real effect … And then second, when all the different sensitivity analyses were conducted, the findings were basically upheld … Then, the existence of the other epi studies that have now been conducted support it. The parallels between [PM-mortality] and the models of smoking and ETS and the fact that you see similar types of endpoints, cardiovascular and lung cancer effects, help support causality. And then I think the toxicological studies that have been conducted over the last five years (most recently the Sun studies and Lippmann studies) showing the effects in mice in terms of plaque development are a help, as well as some of the other previous studies. The Godleski studies and the Ghio studies and the van Eeden studies and the Seaton studies and some of these other studies, both in humans and in animals, [support] causality. Even the short-term exposure studies, where there's much less of a likelihood of confounding, and where you see effects on mortality and morbidity, [support my belief in] causality… [S]ome of the short-term studies and … the endpoints that they found, like MIs and defibrillations and changes in blood plasma and inflammatory markers [are also supportive]. [So,] observational studies as well as controlled human studies, and animal studies also give me some evidence. And finally I would say that, regarding my last criter[ion], when [one] examines what the quantitative implications are of these studies, [they're] entirely plausible in terms of the magnitude of the effect[s] that are predicted."

Expert I specified a range of the likelihood of a causal relationship of 80 – 100 percent with a most likely value of 95 percent. His lower bound was based on having some residual doubt. "You have these monitors, and who knows what those are representing? And you've got all sorts of competing risk factors and nothing is ever fully measured, and you've got people commuting all over the county and who knows what they're exposed to? And you get different penetration rates. And you get tons of pollution driving in your car, and in your bus. And you get people taking all sorts of medications or not … I factor all that in and how high could be my doubt? I could say as high as … 20 percent." Expert I's upper bound was based on the strength of the evidence such as cohort studies, and the addition of more recent toxicological studies (e.g., Sun et al., 2005 study in mice), and epidemiologic studies providing evidence on mechanisms (e.g., studies of IMT (Kunzli et al., 2005), genetic predisposition to oxidative stress, myocardial infarctions (MI's), and work by van Eeden and Seaton).

3.9 Thresholds

Conceptually, Expert I thought "that [al]though individuals can have thresholds, for such a huge distribution of individuals and such heterogeneity across individuals, it's unlikely at the population level to have a particular threshold." The types of studies he thought would be most informative about the quantitative level of a threshold included a

combination of epidemiological studies, clinical studies of markers of inflammation, studies of disease with long-term follow-up, and long-term primate studies.

However, Expert I thought "epi evidence … from both the mortality studies and morbidity studies and lung function studies, failed to show any kind of threshold or even non-linearity in most of the cases. So I think the scientific [evidence] at this point does not suggest any kind of threshold concentration." Further, he noted that there has been a simulation study in *Risk Analysis* by Brauer et al. (2002) with short-term exposures showing that if there is significant exposure measurement error, thresholds would be difficult to detect, even if one existed. However, he did think that there was the possibility of non-linearity at the low end of the concentration range because limited data exists for the C-R function at concentrations below 8-10 $\mu g/m^3$ and the uncertainty bounds are wider. He concluded that his best estimate was that the C-R function shape was linear based on the existing evidence and did not elect to incorporate a threshold into his quantitative response.

3.10 Other Influential Factors

Expert I did not think there were any other outstanding sources of uncertainty not already covered by the protocol. He did not think that publication bias was an issue because all of the existing cohorts are well-known and followed. He also thought that the statistical modeling used in published studies is "appropriate." He indicated that the use of Cox proportional hazard models for survival analysis was "well established theoretically and empirically." He noted that Krewski's re-analysis evaluated other statistical models and found the results to be similar.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert I thought the C-R relationship was log-linear over the entire range of concentrations that were the focus of the study (4-30 $\mu g/m^3$) based on the current evidence. He indicated that, "those who have done specific statistical tests, chi-square tests … for linearity, seem to be able to rule out non-linearity."

Expert I felt the most comfortable basing his C-R function distribution on empirical estimates from the literature. He indicated that rather than attempt to adjust for specific factors (e.g., confounding, effect modification, exposure issues), his approach to incorporating uncertainty would be by weighting different studies whose estimates reflect different conditions or approaches about which he is uncertain. He selected three estimates and their associated standard errors and assigned each one a subjective weight.[8] The elicitation team then combined the estimates in Crystal Ball™ using a Monte Carlo simulation, according to the assigned weights. The trial values from the Monte Carlo run were then fit to a Beta distribution, which was ranked the most highly by Crystal Ball™ based on the Anderson-Darling goodness-of-fit test. The elicitation team then combined the Beta distribution with Expert I's estimate for the likelihood of a causal relationship.

[8] The estimates and standard errors from the three studies were each fit to a normal distribution.

For his first modeling approach, Expert I chose the first estimate from Pope et al., 2002 (β (SE) - 0.0583 (0.02157)) because it has a large sample size and the cities includes are geographically representative of the US. However, he thought the population was not representative of the entire U.S. because of oversampling of highly educated people and therefore, the estimates were biased downward. The second estimate chosen by Expert I was from the ACS L.A. reanalysis by Jerrett et al., 2005, in (0.14 (0.0574)) based on the model using principal components analysis for the ecological covariates. He chose this study because it has "better modeling of exposure, less potential inter-city-related confounding, regional confounding." He thought the limitations of this study were that it only included the "L.A. basin, so the pollution mix could be different. The people could be different. Smaller sample size." His final estimate was taken from Six Cities extended reanalysis (Laden et al., 2006 (Table 3, Model 1: 0.1484 (0.0417)). "Advantages: biggest one is the monitor placement and the city size that's used. The city sizes are small, so the monitors are probably most representative. And the sample is more random and representative of those areas. Third, it's been replicated by Krewski. The downsides are … small sample [size that only includes the] East Coast [and the] Midwest."

Expert I based his maximum value on adding two standard deviations to the mean effect estimates from the Laden et al., 2006 study, arguing that it would be "hard to believe" it could be higher.

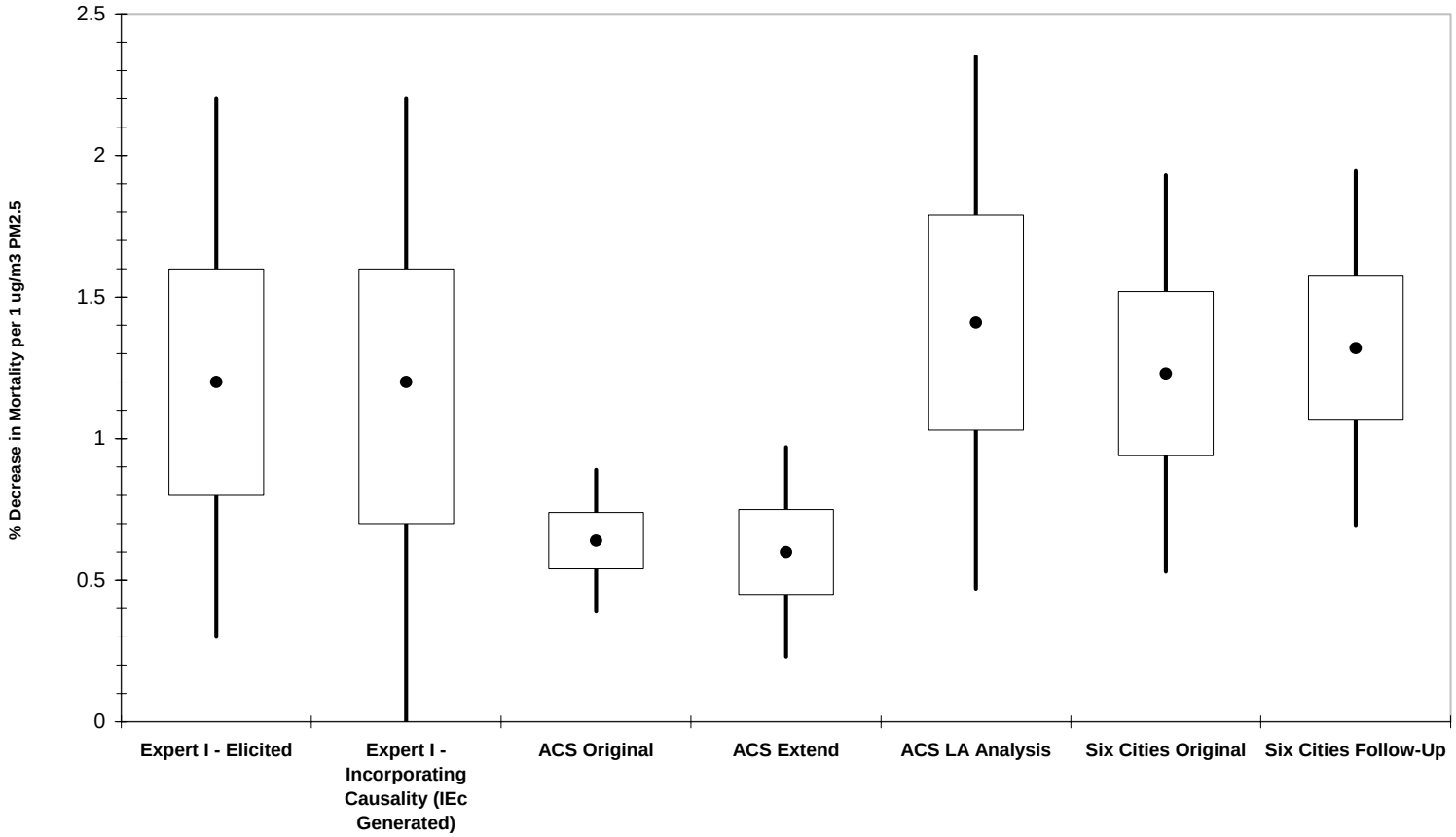
Expert I ran through the combination of these three studies with three different weighting schemes. The first run included equal weights for each study. For the second run, he placed a weight of 0.25 on the Pope et al., 2002 estimate and a weight of 0.375 on both the Jerrett et al., 2005 and the Laden et al., 2006 estimates. His final run included an estimate from the Pope et al., 2002 study adjusted for educational attainment as an alternative to the original Pope estimate; this run also assigned equal weights to each study. After examining the resulting distributions from each run, Expert I ultimately decided to use the results from the second run, which did not weight Pope et al., 2002 as highly as the other two studies, for the following two reasons: "One is that the measurement error might be a little bit more [in the Pope et al., 2002 study] … And then two, because of the [lack of] national representation."

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Expert I

Percentile	Percent Change in Mortality Elicited Distribution**	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	0	0
5th	0.30	0
25th	0.80	0.70
50th	1.2	1.2
75th	1.6	1.6
95th	2.2	2.2
Maximum	2.3	2.3

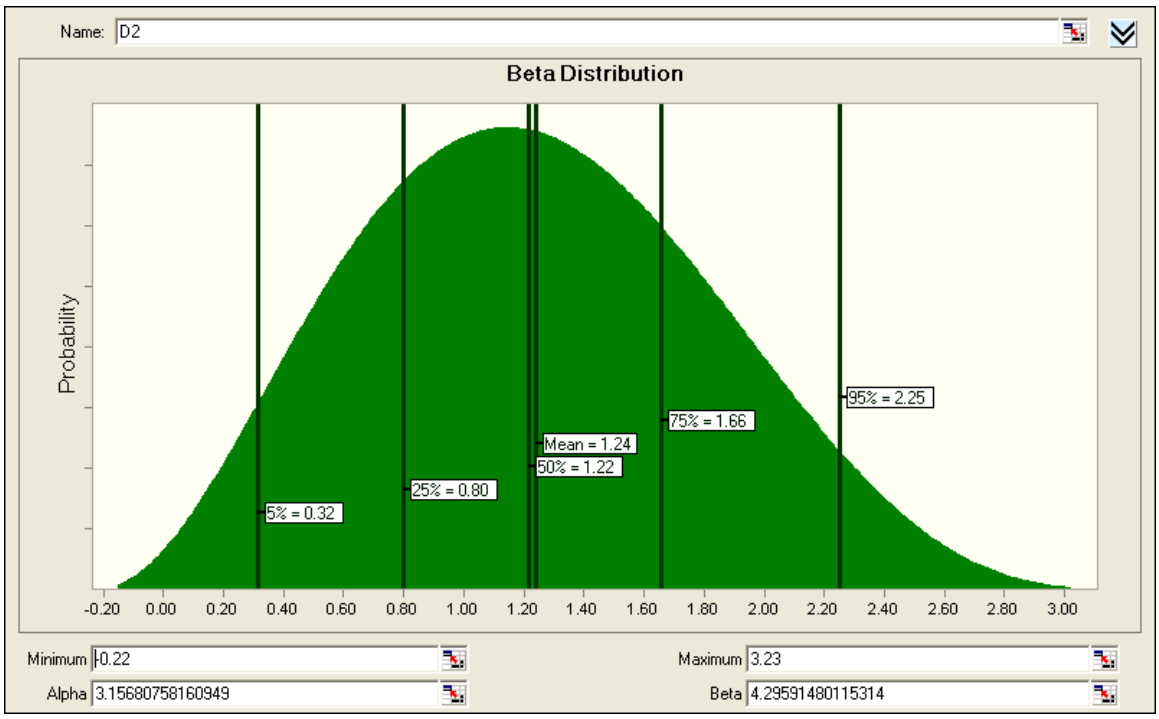
** Parameters of fitted beta distribution: $\alpha = 2.9$, $\beta = 3.7$, min = -0.16, max = 3.0

Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distributions from Expert I

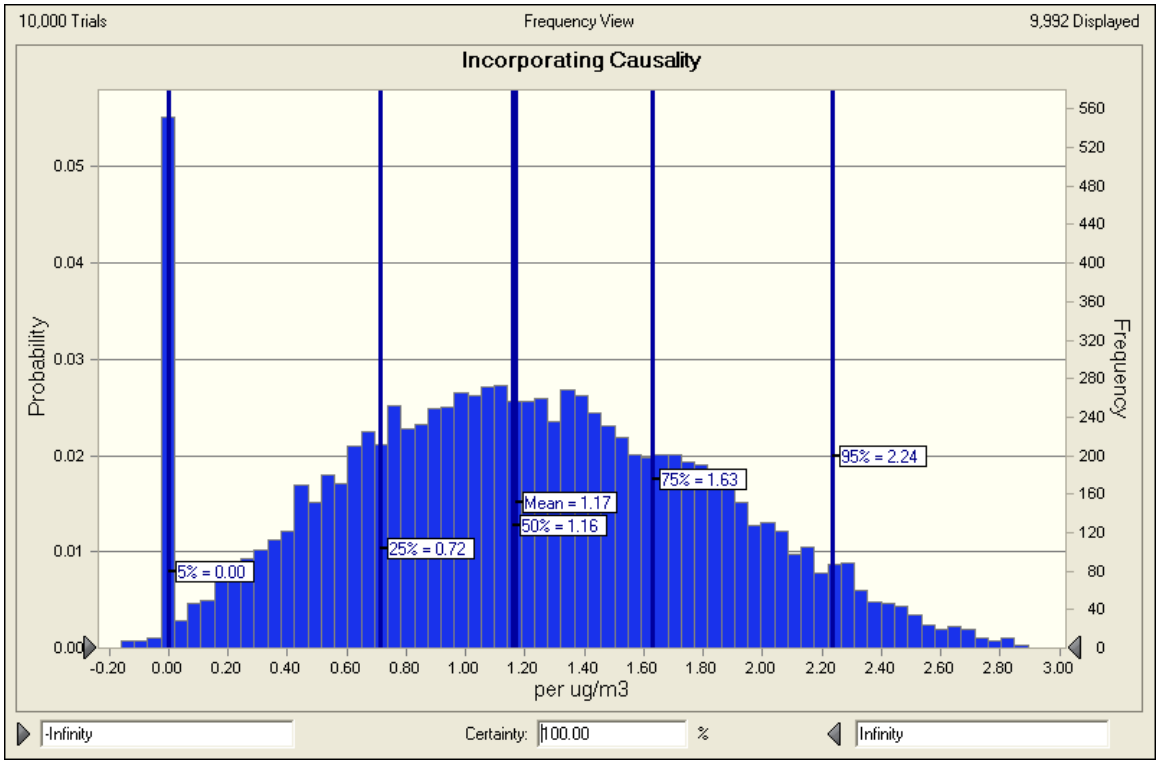


• = median □ = interquartile range | = 90 percent confidence interval

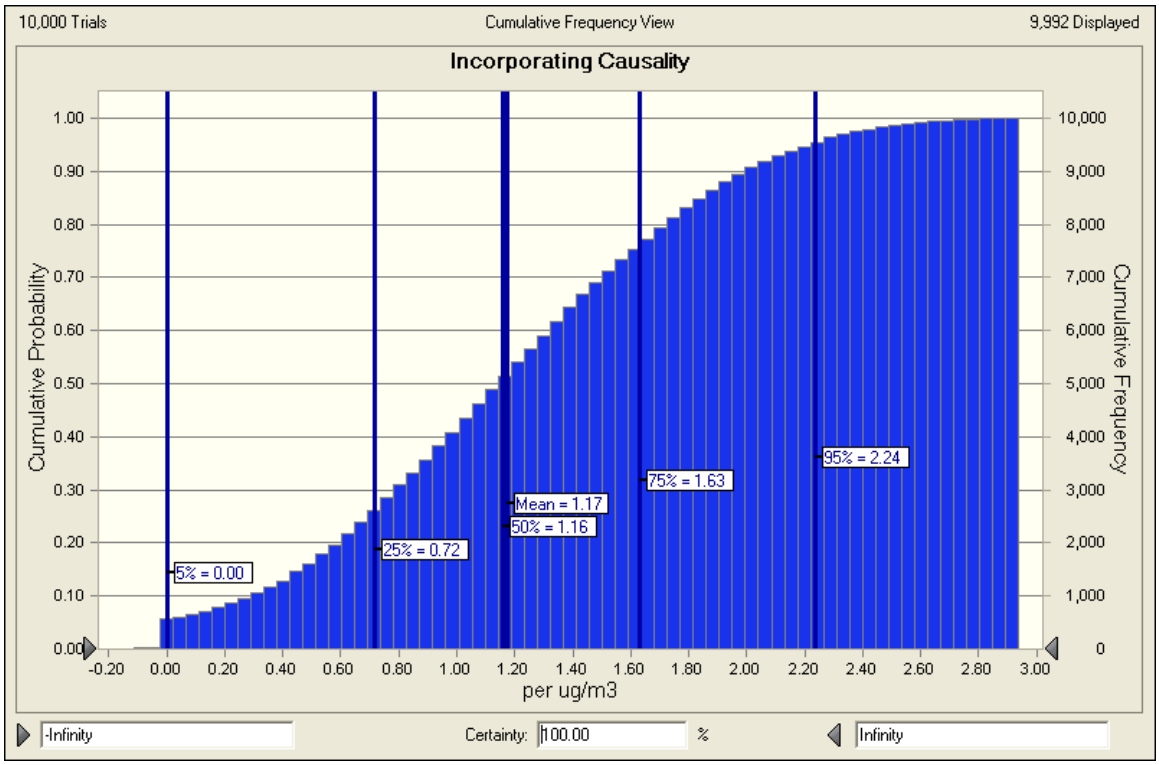
Elicited Distribution



Distribution Incorporating Causality - Probability Density Function (IEc Generated)



Distribution Incorporating Causality - Cumulative Density Function (IEc Generated)



U.S. EPA EXPERT ELICITATION STUDY OF THE CONCENTRATION-RESPONSE RELATIONSHIP BETWEEN ANNUAL AVERAGE $PM_{2.5}$ EXPOSURE AND MORTALITY

Modification to Expert Judgments

Expert I

Date: 7/7/06

Section of Protocol Affected (Section Number and/or Title):

Part 4 – Elicitation of Quantitative Judgments

Description of Change (e.g. to a specific percentile, or to a qualitative opinion or statement of belief):

1. Expert I chose to use Run 3 of the meta-analysis for his final distribution, rather than Run 2. This entailed using estimates for the beta (slope) and standard error, from Pope et al., 2002 that he adjusted upward for educational attainment. He then applied equal weights to the three studies used in the meta-analysis (Pope et al., 2002; Jerrett et al., 2005; and Laden et al., 2006).

2. Expert I decided to fit the weighted average data to a normal distribution, rather than a beta distribution, as in the original interview.

3. Expert I selected a minimum value of 0.2 based on the lower 95 percent confidence limit on the original Pope et al., 2002 all-cause mortality effect estimate for the average of the two exposure periods (1979-83 & 1999-2000).

Rationale for Change:

1. He thought that first adjusting the Pope et al. upwards for education and then weighing the three studies equally was a better approach then weighting the original Pope estimate lower than the other two studies. He also wanted to clarify that although the Jerrett et al. study is a subset of the ACS study, the exposure assessment was so different than the original study, that he considered it to be an independent analysis. Therefore, he chose to weight it equally along with the other two studies.

2. His rationale was that the weighted average data fit to the normal distribution was almost equally as good (the normal distribution was the second most highly ranked distribution in terms of fit according to the Anderson-Darling test), the values of his final uncertainty distribution differed very little from those resulting

from the fit of the beta distribution, and that the normal distribution was more intuitive and straightforward to explain.

3. He selected this minimum value based on the lower 95 percent confidence interval bound of the Pope et al. (2002) all-cause estimate for the average of the two exposure periods (1979-83 & 1999-2000).

Expert J
Interview Summary

Interview Summary
Expert J

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Relying on a schematic diagram, Expert J outlined the general pathophysiologic mechanisms and pathways that he thought could explain the mortality effects of $PM_{2.5}$ exposure. In order to be plausible, Expert J noted, mechanisms should be consistent with what is observed in the epidemiologic studies (respiratory and cardiovascular mortality for short-term exposures and cardiovascular and to a lesser extent, respiratory mortality, in the studies of long-term exposures). He also thought the mechanisms should explain the findings of morbidity studies (primarily respiratory and hospital admissions associated with short-term exposures). His views have been shaped particularly by the emergence over the last six years of numerous studies from different disciplines that seem to converge on this set of pathways. Expert J chose to discuss mechanisms for short-term and long-term exposures together. "I think that it's hard to talk about [short-term and long-term exposures] individually. There's no question that some of the pathophysiologic links are different whether you're talking short-term exposure versus long-term exposure. But there's a lot of it that's very much the same."

The primary health effects that Expert J thought might be linked by the same pathophysiologic mechanisms included chronic obstructive pulmonary disease (COPD) and ischemic heart disease. He discussed data showing both acute and chronic impacts on lung function (Harvard 24-city study (Raizenne et al., 1996), Adventist Health and Smog (AHSMOG) study, Children's Health Study (Gaudermann et al., 2004)). In particular, he thought there was "substantial evidence that long-term exposure to air pollution increases the progression of chronic obstructive pulmonary disease…[and] that short-term elevated exposure exacerbates existing COPD. [T]here is also substantial literature, more recently, that suggests that the existence of COPD substantially increases [the] risk of heart disease and complications related to heart disease, especially … ischemic heart disease and death." He thought the mechanism for exacerbation of COPD was likely to involve pulmonary inflammation.

The second key hypothesis for the impact of PM on mortality that Expert J discussed was that "long-term, repeated exposure to fine particulate matter may help initiate and accelerate the progression of atherosclerosis and that short-term elevated PM exposures may also contribute to the acute thrombotic complications of atherosclerosis." He noted that this hypothesis is not independent of the COPD hypothesis, but rather that the process begins in the lung with pulmonary inflammation leading to systemic inflammation and oxidative stress that contribute to development of atherosclerosis, plaque instability, and increasing vulnerability to ischemic events. He cited Godleski et al (2004), Brook et al. (2004), Kunzli's atherosclerosis work in LA (2005), Sun et al.'s (2005) work at New York University (NYU) in mice, and work done in Hogg's lab work looking at bone marrow responses (van Eeden et al., 2001 & 2002; Terashima et al.,

1997; Mukae et al., 2000 & 2001; Fujii et al., 2002; Goto et al., 2004) and in hyperlipidemic rabbits (Suwa et al., 2002) as some of the evidence supporting this hypothesis.

Expert J thought that another, related set of intermediate effects on the vascular system that several studies have associated with exposures to PM contribute to the risk of ischemic events (i.e., endothelial dysfunction, vasoconstriction, and hypertension). He noted that the PM findings are "at least qualitatively consistent with" the smoking literature in which one sees that "smoking accelerates COPD ... increases respiratory symptoms, [reduces] lung function. You see systemic inflammation, … oxidative stress, … ischemic heart disease, … ischemic cerebral vascular disease, … changes in the vasculature."

He discussed an additional hypothesis for cardiovascular mortality related to impacts on the autonomic nervous system, which has arisen out of the work of Annette Peters et al. and others. He noted there are now about a dozen studies showing a relatively robust association between short-term exposures to PM (mostly $PM_{2.5}$ but sometimes PM_{10}) and short-term changes in heart rate variability (HRV). However, Expert J indicated that it is not yet clear to him how these observations relate to the other pathophysiologic mechanisms discussed above.

Two other hypotheses for which he found the evidence to be less compelling included those involving systemic translocation of ultrafines, altered immune responses, and hypoxemia (resulting from declines in lung function and oxygen saturation of the blood).

3.2. Conceptual Framework for Mortality Effects of Short-Term and Long-Term $PM_{2.5}$ Exposures

While he thought Kunzli's paper on this topic was an important contribution, he thought that, in practice, allocation of health effects would be difficult to achieve using this framework. He preferred to view the effects of short- and long-term exposures in terms of time scale. Referring to the work by Schwartz et al. (2000 & 2001), and others (Zeger et al., 1999; Kelsall et al., 1999; Dominici et al., 2003) exploring the question of harvesting, 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. Because the adverse health effects of particulate air pollution are dependent on both exposure concentrations and the length of exposure, it's fully expected that long-term repeated exposures would have larger, more persistent cumulative effects than short-term transient exposures." He thought that the data are consistent with the view that long-term chronic exposure contributes to the progression of disease and that short-term exposure exacerbates underlying disease and contributes to short-term changes in mortality. He argued that the exposure-response relationships are essentially linear and that if the right time scale and weighting were known, one could estimate the long-term effects exposure from the time-series studies; if

it were not linear, and the effects of short-term exposures only occurred above some level, the results of two types of studies would be additive. He concluded that the time-series studies are only giving a "glimpse" of the total health effects of PM and that the long-term cohort studies are capturing nearly all of the effects.

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

The following table summarizes the discussion of the types of impacts that Expert J thought were observable using different study designs:

Study Design	Type of Exposure Effects Captured
Prospective Cohort	• Long-term • Cumulative short-term
Intervention Studies (Intermediate term length)	• Cumulative short-term • Intermediate long-term • Misses effects of long-term exposures
Ecological Studies	• Cumulative short-term • Long-term
Multi-city, intermediate time-series studies (distributed lag, unconstrained distributed lag ~ 5-40 days)	• Cumulative short-term up to length of lag • Not capturing effects of long-term exposure to chronic disease and mortality

Overall, he thought the American Cancer Society (ACS) cohort studies (Pope et al., 1995, 2002 & 2004; Krewski et al., 2000; Jerrett et al., 2005) and the Six Cities cohort studies (Dockery et al., 1993; Krewski et al., 2000; Laden et al., 2006) were the most suitable for estimating the total health impacts of a 1 $\mu g/m^3$ reduction in annual average $PM_{2.5}$. He thought they were likely to capture both the mortality effects of long-term exposures but also the cumulative effects of short-term exposures (distributed lag).

He thought the intervention studies (Pope et al., 1996 ("Utah Valley" study); Clancy et al., 2002 ("Dublin" study)) were primarily indicative of the effects of cumulative short-term exposures and intermediate exposures. He did not think they captured the full effects of chronic exposure (contribution to progression of chronic disease and mortality).

Expert J thought that the "ecological studies" (Lave and Seskin; Thurston and Ozkaynak; Evans, Tosteson and Kinney) have come to represent much the same findings as the prospective cohort studies. Their results were dismissed early on because they did not control for smoking or other individual level risk factors (several contextual ecologic variables were examined). He thought it was ironic that subsequent studies including Krewski's re-analysis of the ACS study (2000) and the ACS LA analysis Jerrett et al. (2005) have suggested that these individual factors are not likely to have had a huge impact on the results.

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

Expert J noted that none of the studies currently published really represent the ideal study one would want for predicting the impact of $PM_{2.5}$ on U.S. mortality. He thought an ideal study should have a larger sample size, include more cities, include cities where air pollution differs across communities and over time, and should collect information on clinical and subclinical measures of disease (e.g., atherosclerosis) or markers of inflammation. He thought the ideal study would also have not just ambient monitors but monitors in homes, in workplaces, and some personal monitoring.

He noted that the Six Cites study had many characteristics of an ideal study in that, unlike most epidemiologic studies that tend to be opportunistic, it was specifically designed to look at the relationship between air pollution exposure and health. He indicated that it included cities with a range of exposures, and with the extended follow-up, a few of the cities have shown substantial reductions in exposure. He felt that the study's limitations are relatively small study size, and that it did not anticipate the link with cardiovascular disease, and therefore did not include some of the key clinical and subclinical measures of disease.

Expert J thought that the ACS study and its reanalysis, although more opportunistic that the Six Cities study, had the strong advantage of representing a much larger number cities, included a large population, and collected very good information on individual risk factors. The ACS study, like the Six Cities study, has also undergone extensive re-analysis and has been corroborated by independent investigators.

When asked which studies he would most like to focus on during our later discussions of confounding, effect modification, and exposure issues, he indicated that he would most like to rely on the ACS study and then the Six Cities studies. For the ACS study, he included the original study its reanalysis (Pope et al., 1995; Krewski et al., 2000) its extended analysis (Pope et al., 2002), the analysis of cardiovascular disease (Pope et al., 2004) as well as the analysis of spatial resolution of exposure in LA by Jerrett et al. (2005). For the Six Cities study, he relied upon the original analysis and reanalysis (Dockery et al., 1993; Krewski et al. 2000) and the extended analysis (Laden et al., 2006).

Expert J discussed more briefly a second set of studies: the AHSMOG cohort studies (Abbey et al., 1991 & 1999 and Chen et al., 2005); Hoek et al.'s work in the Netherlands looking at proximity to roadways (2002); the Finklestein et al. (2004) study in Ontario, Canada also looking at proximity to roadways and finding similar results as Hoek et al.; and the series of work staring with Woodruff (1997) looking at infant mortality. We then discussed the series of papers based on the Veterans Administration (VA) cohort by Lipfert et al., 2000, 2003 & 2006. He expressed concern about the lack of robustness of the results, the convoluted nature of the analysis and the write-up, and their publication in journals where they are less likely to have received peer review by epidemiologists. He

had not had an opportunity to evaluate fully the Enstrom et al. (2005) study in California, but expressed some concerns similar to those he had for the VA studies.

3.5 Confounding

We conducted a lengthy discussion of confounding in both time-series analyses and in cohort studies. Expert J generally felt that most confounders had been adequately controlled in the studies upon-which he would rely for his quantitative estimates of the PM mortality effect.

For time-series studies, the potential confounder that remained a concern for Expert J was the issue of co-pollutants (SO_2, NO_2, CO, and ozone). Expert J thought that co-pollutants are positively correlated with PM and inadequately controlling for co-pollutants will tend to overstate the importance of PM alone. However, he argued that if one is using PM as an indicator for air pollution in general, excluding the other co-pollutants might underestimate the effects of air pollution. He wanted to clarify that this argument about co-pollutants can only be carefully evaluated in the context of large, multi-city studies like the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) or Air Pollution and Health – A European Approach (APHEA). He thought analyses of co-pollutant effects in single cities (e.g., by Moolgavkar et al.) are problematic given that individual cities are likely to differ in the relative levels of co-pollutants. In general, he argued that the literature suggests that PM is driving more of the mortality effect seen than the co-pollutants.

For the cohort studies, Expert J discussed smoking, occupation, weather, diet, socioeconomic variables, contextual variables, and differential migration. He thought smoking had been well-controlled in the ACS and Six Cities studies, and the AHSMOG study included only non-smokers. He thought occupation, weather, and diet were well-controlled for in both ACS and Six Cities studies (score = 1). Neither study controlled for exercise directly, though each assumed body mass index (BMI) was at least a partial measure for exercise.

He had some concern about general socioeconomic status (SES) variables and the fact that the Jerrett et al. (2005) study found lower effect estimates when it controlled for more contextual SES variables. He thought it was unclear whether these analyses were evidence of residual confounding or over-fitting the models. “Probably some of both.”

Expert J thought that of all the confounders discussed, co-pollutants and differential migration had been addressed the least in his key studies. He thought co-pollutants had the potential to be a major factor (score ≈ 3) in overstating the impact of PM alone (as discussed above). Differential migration he thought had the potential to be a major negative confounder (score ≈ 3) if for example, sick people are more likely to move to the cleaner areas of the country (e.g., to Arizona) whereas the healthy people do not move.

3.6 Effect Modification

After an extensive discussion, Expert J summarized his views on effect modification by saying that the only effect modifier that has emerged from the analysis and re-analysis of the Six Cities and ACS studies has been educational attainment. He thought the relative risks associated with PM were clearly higher in the population with less than high school education. He explained that educational attainment per se is not a risk factor but that it correlates with various socioeconomic variables that may more directly bear on exposure or health outcomes (e.g., housing, income, access to medical care, or use of air conditioning).

Expert J thought that the higher level of educational attainment in the ACS cohort relative to the U.S. population could lead to the conclusion that the overall ACS risk estimate likely understates the true relative risks in the broader U.S. population. He cited a presentation at the Pre-elicitation Workshop in which the Pope et al., 2002 effect estimates were adjusted to account for the distribution of educational attainment in the U.S. population. Depending on the data used for the adjustment, the original estimates rose by about 30-50 percent or from "six, seven percent per 10 $\mu g/m^3$ estimate to about a nine or 10."

Expert J also discussed whether race is another potential effect modifier given that neither the Six Cities or the ACS study were representative of the racial mix in the U.S. On the one hand, he noted that work by Joel Schwartz's group has shown a role for genetic differences in response to particulate pollution and changes in HRV. On the other hand, he did not think that the epidemiologic evidence in studies in populations around the world suggests that different nationalities respond differently to air pollution. He did not think sufficient evidence exists to speculate whether and/or how differences in racial composition of the cohorts might affect the generalizability of the results to the US population.

3.7 Exposure Issues

Expert J raised five potential issues with regard to how exposure is measured in the cohort studies:

- Central versus personal exposure monitors;
- Spatial resolution of exposure monitoring;
- Migration of the population;
- Co-pollutants; and
- Temporal changes in exposure.

Expert J thought that the first three issues were most important as each could contribute to a substantial underestimate of the PM mortality effect (score = 3). "To the extent that variability in central site monitoring only gives you part of the variability in exposure, we have substantial exposure measurement error that likely is biasing our effect estimates downward." He thought a related issue is the degree of spatial resolution of the exposure measurements. He referred to an expert's presentation at the Pre-elicitation Workshop,

which suggested that finer spatial resolution of the Six Cities study relative to the ACS study might account in part for the higher relative risks observed in that study. He noted that the spatial resolution in the L.A. sub-cohort of the ACS study (Jerrett et al., 2005) was finer than in the original ACS study and found results that are more comparable to the Six Cities study. He pointed out, however, that it is difficult to know what the "right" spatial resolution should be in these studies; if it's too big, there is exposure measurement error, but it it's too small, there may be exposure measurement error because individuals are not staying in the vicinity of the monitor to which they're assigned (e.g., they go to work elsewhere). He also thought migration of individuals in the cohort over the course of the study could lead to exposure error and underestimation of the "true" relative risk. He indicated that if, on average, some individuals in high pollution areas move to lower pollution areas and some move from low to high pollution areas, it is another form of random exposure misclassification. He noted another expert's presentation on differential migration patterns according to socioeconomic status at the Pre-elicitation Workshop, which offers a "reasonable hypothesis" for the differences in relative risk estimates by education. This expert's analysis of U.S. census data indicated that individuals with higher SES were more likely to move outside than within Metropolitan Statistical Areas (MSAs) as compared with individuals with lower SES, the implication being that their exposure might be subject to more misclassification. Expert J did not think the data were available to know how to estimate the magnitude of this issue's effects on relative risks.

In his discussion of co-pollutants, Expert J raised the same issues as he had under confounding. In essence, the question is whether one is focusing on a pure PM effect or an air pollution effect for which PM is an indicator. Expert J argued that a weakness of the elicitation question is that it is hard to know which of these two is truly the focus of the study. He noted that, in reality, the evidence does not allow us to understand what happens if PM changes without any other change in co-pollutants or components; when the sources are reduced, the mix of co-pollutants changes. If one wants the marginal effect of PM alone, which is the actual focus of the elicitation, the PM effect estimates in the studies could be moderate over-estimates (score = 2) of the true PM effect for the US population.

Finally, we raised the question of the time course of historical exposures that were relevant to the health effects observed in the follow-up period. Expert J did not think this issue would matter much in terms of hypothesis testing, though it might matter some in the magnitude of the relative risks. Theoretically, "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." However, he believed that the health effects observed in the cohort studies are primarily due to exposures over the previous 5 to 10 years (based on looking at the progressive size of mortality effects over studies looking at the past few days (time series), to few years (intervention studies), to eight years (the Six Cities) where the largest effects are seen). He therefore did not think he would want to adjust his estimates to account for this issue (score = 1).

3.8 Causality

Expert J described how the development of the body of evidence over the past several years has convinced him that it is highly likely that the relationship between exposure to fine particles and mortality is a causal one. The literature has gone from the early, naïve epidemiological studies to a variety of study designs showing a consistent and coherent relationship between particulate air pollution and respiratory and cardiovascular disease as measured by various health endpoints ("changes in lung function, changes in cardiac autonomic function, changes in hospitalization for both respiratory and cardiovascular disease, changes in mortality for respiratory hospitalization and mortality"). He indicated that there are no studies showing associations with health endpoints that one would consider *a priori* to be unrelated. In addition, he stated that there are "semi-controlled [studies] … looking for sub-clinical measures of disease" such as measures of pulmonary and systemic inflammation, C-reactive protein, and cardiac autonomic function. He thought that the toxicological studies by Hogg et al and the NYU group looking at such things as "changes in inflammation, changes in bone marrow responses, changes in atherosclerosis, etcetera" lend further credibility to the whole hypothesis. Ultimately, Expert J noted that he is not able to define an alternative explanation for the phenomena observed across these studies; nor has he heard one put forward by others. For example, the early criticism of publication or analytical bias in the individual time series studies has been muted by the large multi-city studies (e.g., NMMAPS, APHEA).

Expert J did not draw a distinction between short-term and long-term exposures and mortality. While he expressed the view that, at some level, it has been more difficult to imagine that a short-term exposures to PM could lead to a cardiac death than to believe that chronic exposure to PM could, like exposure to environmental tobacco smoke (ETS), lead to an increased risk of cardiovascular disease. However, he felt that the empirical evidence is too strong to ignore.

Expert J's quantitative characterizations of the causal relationship reflect his inherent conceptual difficulty with this question. In the initial discussions on causality, Expert J expressed the view that the overall likelihood of a causal relationship between annual average $PM_{2.5}$ exposures and mortality ranged between 75 percent and 98 percent with a best estimate of 90-95 percent. However, in the development of his quantitative estimates, Expert J revised these views, ultimately stating that he could not argue for a 5 percent likelihood of a non-causal relationship. He then provided an estimate of no more than a 1 percent likelihood of a non-causal relationship (i.e., his final estimate of the percent likelihood of a causal relationship was 99 percent).

3.9 Thresholds

Expert J's views on thresholds were driven by the empirical data. While conceptually he might have difficulty believing that low levels of $PM_{2.5}$ could cause damage, he stated that the empirical evidence does not point to an observable threshold for the effects of either short-term or long-term exposures. Piecing together the literature on smoking, occupational exposures to particulates (e.g., coke ovens), ETS, and ambient air pollution,

there's a trend that "basically looks linear." He thought the epidemiological studies with sufficient power to make statistical inferences about population thresholds (e.g., multi-city time series studies, the ACS study) have not been able to detect them even with using more sophisticated statistical techniques (e.g., meta-smoothing in multi-city time series studies, Krewski et al., 2000 reanalysis of the ACS study). He indicated that these studies consistently support a generally linear relationship. He did not think that toxicological studies or clinical studies were useful for finding population thresholds because they lack sufficient statistical power and are conducted at the high end of, or well beyond the range of environmentally relevant exposures.

3.10 Other Influential Factors

Expert J discussed additional sources of uncertainty that were not raised as part of the protocol. He expressed concern about whether the PM index most commonly used in the literature ($PM_{2.5}$) is the best measure in terms of both understanding potential health comes and directing regulatory efforts. For example, he expressed the opinion that if there are characteristics of combustion or non-combustion related PM more toxicologically relevant, and PM is not a good index for them, we may be underestimating the real effects of exposure. He thought that an example of this was with historical measurements of PM_{10} or total suspended particles (TSP). "When we were using PM_{10}, we were underestimating the effect estimates, especially in those studies where PM_{10} was not highly correlated with $PM_{2.5}$ … and with TSP, there [were] no effects." Furthermore, he thought we might be regulating the wrong sources.

Expert J also discussed the fundamental issues raised by the 2002 National Research Council (NRC) study about application of probability models that assume random behavior or sampling to observational data in epidemiologic studies and whether the standard error is therefore an appropriate expression of uncertainty. He noted that analysts are aware that this approach is not "exactly right" but that in the "real world, the actual data never matches the ideal." He argued that none of the people who do these studies "thinks that any one of the studies precisely defines not only the point estimate [or] the uncertainty around that estimate."

He also expressed a lingering concern about scientific and human fallibility and that the current scientific consensus might eventually be proven wrong.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert J thought that the concentration-response (C-R) function was log-linear and consistent over the entire range of annual average $PM_{2.5}$ concentrations that were the focus of the study (4-30 $\mu g/m^3$).

As discussed in Section 3.8, Expert J ultimately chose to provide a C-R function that directly incorporated his assumptions about the likelihood of a causal relationship.

Expert J specified his maximum, minimum, 5th, 50th, and 95th percentiles and then used the Weibull distribution in Crystal Ball™ to generate the interquartile range (25th and 75th percentiles).

Expert J placed his minimum at zero in order to be consistent with his views that the likelihood of a causal relationship was less than 100 percent. He loosely based his estimate of the 5th percentile on published estimates from time-series studies with distributed lags (meta-analyses from Levy et al., 2000; Steib et al., 2002; Anderson et al., 2005; Ostro et al., 2005; Schwartz et al., 1996 (Six cities), Klemm et al., 2000; Burnett et al., 2003) He thought NMMAPS value (0.04 percent per $\mu g/m^3$) represented somewhat of an outlier in this group and more of a minimum value.

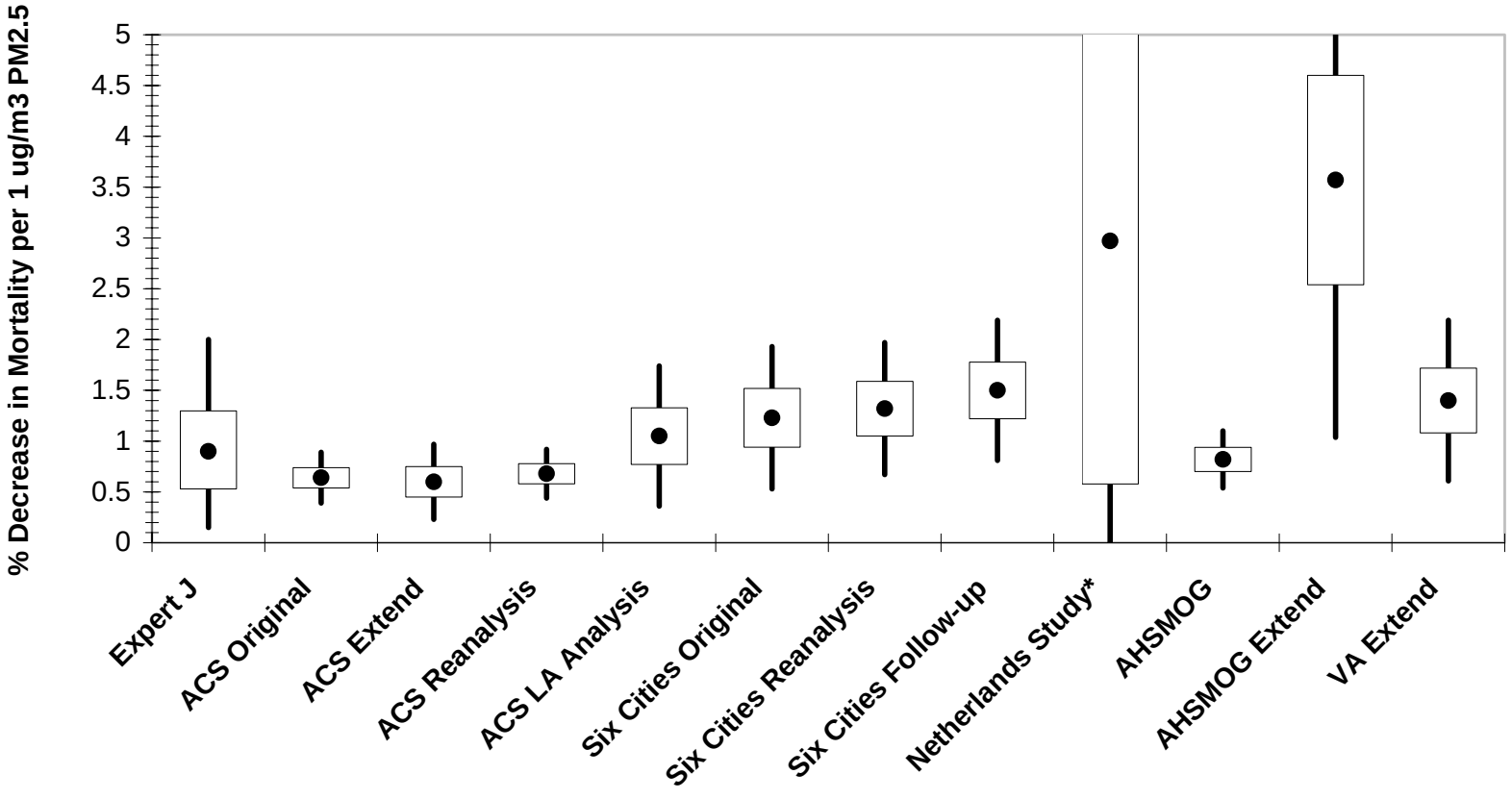
For his 95th percentile value, he first discussed results from the Hoek et al. (2002) work in the Netherlands which, though based on measurements of black smoke, suggested effects on the order of 3 percent per $\mu g/m^3$ which he thought was on the high end. He next considered results from the Six Cities cohort studies and the Jerrett et al. (2005) paper (1.7 per $\mu g/m^3$) and observed that one standard deviation or so above the primary estimates lead to values of roughly 2-2.5 percent. The maximum value (3 percent) was based on estimates from Hoek et al., (2002) as well as unpublished findings from the Women's Health Initiative study.

His 50th percentile was initially based on adjusting the ACS cohort estimate (0.62 percent per $\mu g/m^3$ from Pope et al., 2002 (average of exposure periods, all-cause mortality)) upward to account for exposure misclassification due to central site monitoring and migration, and effect modification by education and also giving consideration to the Harvard Six Cities and ACS-L.A. studies. He then down-weighted that estimate to take into account some of the more negative studies (AHSMOG, VA, Enstrom). When elicitors pointed out that his estimate was only slightly higher than that in the ACS, he indicated that he was relying on the fact that the ACS study was larger than the other published studies.

Exhibit 1: Subjective Estimate of the Percent Change in Annual Mortality Associated with a 1$\mu g/m^3$ Change in Ambient Annual Average $PM_{2.5}$ Concentrations

Percentile	Percent Change in Mortality
Minimum	0
5th	0.15
25th	0.53
50th	0.90
75th	1.3
95th	2.0
Maximum	3.0

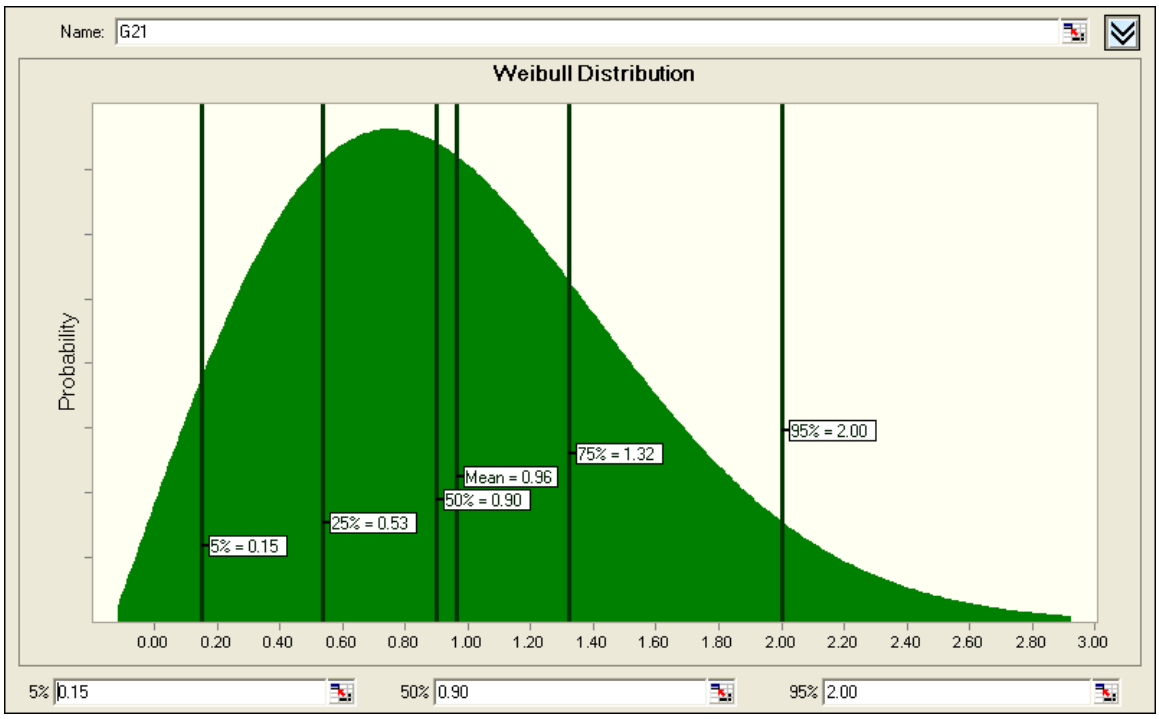
Exhibit 2: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distribution from Expert J



• = median □ = interquartile range | = 90 percent confidence intervals

* Note that this estimate is for a 1 μg/m^3 increase in black smoke for cardiopulmonary mortality.

Expert J Distribution



Expert K
Interview Summary

Interview Summary
Expert K

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert K discussed the biological mechanisms for short-term and long-term exposures separately. "I think the mechanisms probably have some relationship between acute and chronic. But [since] we're talking about mortality, then I think there are some differences." He began with a discussion about short-term mechanisms.

Short-Term Exposures

Expert K thought that short-term mortality from PM is caused by "an exacerbation … of an underlying disease process."

He discussed cardiovascular mortality, specifying two plausible underlying mechanisms for acute events: arrhythmias and endothelial dysfunction. He discussed studies looking at PM concentrations and ventricular arrhythmias in patients with defibrillators (Peters et al., 2000; Dockery et al., 2005). He thought that this could be related to effects on the autonomic nervous system but "how that happens exactly I don't think we know at this point." He thought another plausible mechanism for cardiovascular death was endothelial dysfunction. He cited a study by Brooks et al. from Michigan (2002) that measured brachial artery diameter and found that this vessel constricted following exposure to ozone and particles. He thought that "[i]f it happens in a peripheral or a central artery, there's some evidence that these kinds of things can happen in coronary arteries as well." He thought that this "vasoconstriction or absence of normal dilation" in an individual with significant vascular disease is a possible explanatory mechanism for an acute event. Expert K thought that the "attractive basic science mechanism right now" for these responses is that particles could induce inflammation, which could stimulate production of reactive oxygen species (ROS). He indicated that *in vitro* studies have shown production of ROS with cellular exposure to PM (Nel et al., 2001). The ROS could trigger a series of events in someone with underlying endothelial dysfunction, or in the case of arrhythmias, could an effect on the autonomic nervous system.

Expert K indicated that it was conceivable that respiratory mortality may be related to particles if one argued that particles cause inflammation in the lung, resulting in broncho-constriction and possibly mortality. However, he had less confidence in this mechanism, pointing out that asthmatics, although they could have particle-induced exacerbations of their disease, do not usually die from asthma.

Expert K discussed the limitations and/or gaps in the mechanistic literature. His major concern was about exposure particularly in epidemiological studies, which he thought were uncertain because they are measures of ambient, rather than personal exposures. He thought it was possible that those with defibrillators could be spending their time indoors,

for example, and would not be exposed to ambient pollution. He thought the *in vitro* studies that provide evidence on the "more basic science or the cell biology question, the production of reactive oxygen species" could not be directly extrapolated to humans. He indicated that although the exposures are more controlled, these studies rely on isolated cell systems whereas "the way most things happen in biology is that … there's a signal from one cell to the next … [This gap] creates great uncertainty."

When asked about data suggesting C-reactive proteins (CRP) or other factors as a markers of inflammation related to air pollution exposures, he explained that CRP levels are non-specific; they may be elevated in response to an infection, to chronic artery disease, or to other chronic inflammatory diseases like lupus. Similarly, there is "some very interesting data showing that when you breathe ultrafine particles and you don't vasodilate, it affects your nitric oxide production [Shah et al., submitted for publication]. Nitric oxide can certainly be related to reactive oxygen products. But it isn't a specific marker."

Long-Term Exposures

Expert K indicated that he thought the mortality effects from PM are "largely cardiac" although he expressed uncertainty about whether there was a true independent chronic effect of long-term exposure and of acute exposures, or whether the long-term effects published were more the result of acute exacerbation of existing chronic disease. He cited the Sun et al. (2005) study looking at the impact of six-month exposures to concentrated ambient particles (CAPs) in mice in which the authors found evidence for the "development of vascular response and atherosclerosis." While he thought this study had the advantage of controlled exposure and a genetically modified mouse model, he also thought it still raised questions. "Is it something specific about the particles or … are there lots of other things that would just do this because of [the animals'] genetic susceptibility? … [I]f we gave [them] salt every day or something else, would we see plaque formation just because most any inflammatory event might initiate this kind of change?" He thought the Kunzli et al. (2005) epidemiologic study of intima-media thickness in carotid arteries, as a measure of atherosclerosis progression, was intriguing but that exposure uncertainties limit its usefulness.

He thought the series of studies by Godleski et al. looking at myocardial infarction (MI) in animals with underlying chronic disease after exposure to particles were indicative that "in fact you'd need the chronic disease to lead on to the acute." Expert K thought Devlin et al.'s study in Baltimore that found acute changes in heart rate variability (HRV) and some arrhythmia in an elderly population exposed to CAPs provided evidence of "the potential that particles would put you at risk, [but] is it acute or chronic? I don't know."

Expert K also discussed possible evidence for the effect of long-term exposures on respiratory mortality. He cited a study by Diaz-Sanchez et al. (1999) that he felt provided some evidence not just for exacerbation but possibly development of asthma. In this study, the investigators showed IgE production, a marker of allergy/immunological response, following nasal exposures to diesel; the response was enhanced with joint

exposure to ragweed and diesel. He thought Gauderman's paper in the *New England Journal of Medicine* on the effect of air pollution on lung function were not specific to particles but indicated a NO_2 effect as well.

When asked about lung cancer, Expert I answered that he did not think there was a strong link with particle exposures. He thought the signal from smoking is so strong, and self-reporting on smoking histories are so often inaccurate (i.e., self identification as a non-smoker is unreliable) that he thought associations found in cohort studies that did not involve a formal medical history taking could be the result of residual confounding by smoking. He noted, for example that the ACS study relied on a questionnaire filled out at one point in time (1982).

3.2. Conceptual Framework for Mortality Effects of Short-term and Long-Term $PM_{2.5}$ Exposures

Expert K thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures. "I don't know if it's valid or not, but it's a good way to say, if indeed short-term or long-term or perhaps both can affect mortality, how can we envision that that can happen? I think this is a very reasonable sort of framework for it."

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

Expert K thought that cohort, time-series, and intervention studies each provided somewhat different information about the potential mortality effects of changes in annual average $PM_{2.5}$ concentrations, as indicated in the table below:

Study Design	**Type of Mortality Effects Captured (e.g., short-term, long-term, or both)**
Cohort Studies	Long-term
Time-Series Studies	Short-term
Intervention/ "Accountability" Studies	Short-term, Long-term, if studied long enough

Expert K thought the intervention studies were certainly capturing effects of short-term reductions in $PM_{2.5}$ on both morbidity and mortality and that they offered the hope of understanding the effects of long-term reductions. However, he would want to see longer-term follow-up, use of clear markers of cardiac disease (e.g., cardiac catheterization) and comparison to a community without the same intervention.

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert K thought that the following characteristics would be part of an ideal epidemiologic study to characterize the $PM_{2.5}$-mortality relationship in the U.S. population:

- Good exposure data documenting the link between ambient exposures and the specific outcome;
- Prospective study design (similar to the Six Cities);
- An intervention in exposure;
- Good characterization of underlying cause of death (e.g., details on myocardial infarction or ventricular arrhythmia, not just cardiac arrest); and
- Broad geographic coverage of the U.S.

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert K mentioned the American Cancer Society (ACS) study, the Six Cities study, the Clancy et al. (2002) study in Dublin, and the Jerrett et al. studies in Los Angeles (L.A.) and New York. He felt the Six Cities study was strengthened by being specifically designed to test the hypothesis of a relationship between air pollution levels and health, although he would have liked to have seen more personal monitoring. Among the analyses of the Six Cities studies, he focused on the original study (Dockery et al., 1993) and its reanalysis by Krewski et al. (2000). He was not as familiar with the Laden et al. (2006) follow-up to the Six Cities study, but felt that it found similar results as the original and reanalysis.

He preferred the prospective design to what he characterized as "data mining," which he defined as the use of large data sets to look retrospectively for associations. He expressed a general discomfort with the fact that investigators often find associations with PM and several outcomes. "Every time I see an association between breathing and an adverse outcome [e.g., reproduction, central nervous system, birth outcomes], it makes me a little uncomfortable. I could be wrong, but I worry that it's in the way the data are analyzed, and that there are so many unknowns in terms of the actual exposures. And yet every time the analysis is done, it shows an effect." In general, he thought it would be beneficial if the clinical endpoints of these studies were more fully understood by the investigators.

He acknowledged that his concern about the role of the statistical analysis in somehow generating the associations primarily applied to time-series studies. He thought the Krewski re-analysis of the Six Cities and ACS studies was very credible.

He thought the Dublin intervention study (Clancy et al., 2002) was important because of the potential advantages of intervention studies discussed previously.

He thought the Jerrett et al. analyses of the ACS data in L.A. and New York were intriguing because they had "a better handle on the exposure." However, he wasn't sure what the implications of the improved exposure were given that the two studies appeared to find very different relationships.

3.5 Confounding

In his discussion of confounding, Expert K focused on the ACS and Six Cities studies. He initially first discussed several factors that he thought could still be potential confounders of the mortality effect (smoking, socioeconomic status, pre-existing health status, indoor air exposures, co-pollutants, and diet). He ultimately reasoned that smoking, co-pollutants, and diet were the most important of these for each study and assigned them scores according to the table below.

Study	Potential Confounder	Direction of impact	Score
Pope et al., 2002	Smoking	Overestimate	1-2
	Diet	Overestimate	1
	Co-pollutants	Uncertain	2
	Indoor Air Exposures	Uncertain	2
Dockery et al., 1993	Smoking	Overestimate	1
	Diet	Overestimate	1
	Co-pollutant	Uncertain	1

Expert K first discussed smoking. He acknowledged that the ACS study had a variable for smoking in the model but thought that there could be some residual confounding (see discussion under mechanisms) that could lead to inflation of the effect estimate (score of 1 or 2, smaller than the effect of co-pollutants).

He also thought that cardiovascular mortality from PM could be overestimated due to confounding by diet; he thought poverty, smoking, poor diet are probably highly correlated. He indicated that the effects of diet on the estimation of the PM/mortality relationship are likely to be a major focus of future research.

Expert K raised co-pollutants as an issue of concern. "You have a very large population and there you could have fairly different exposures in terms of the co-pollutants." He discussed the Pope et al. (2002) and the Krewski et al. (2000) reanalysis, which indicated that SO_2 was the only co-pollutant that could account for a lot of the $PM_{2.5}$ effect, but ultimately Expert K did not think there was a sound biological explanation for an independent SO_2 role in mortality at ambient concentrations. Expert K then indicated that he did not think that the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) had not eliminated the possible role of co-pollutants. He particularly thought that ozone has not been ruled out as a factor. In summarizing his views, Expert K indicated that the issue of co-pollutants is not straightforward. He thought there are recent data relating elevated ozone levels to increased mortality rates. But more importantly, there may be unregulated co-pollutants that are neither regulated nor measured which could important confounders. He thought that these unmeasured co-pollutants could have a moderate effect (score = 2) on the mortality effect estimate for the ACS study estimates.

Finally, he thought differential exposures in indoor air across study areas could cause the effect estimates to be confounded. He was uncertain about either the direction or the magnitude of the effect since indoor air was not monitored in either the ACS or Six Cities study. However, he argued that it was more likely to be an issue in the ACS study because it spanned a broader geographic area than the Six Cities study.

Expert K thought the Six Cities study could be affected in a similar way by the confounders discussed for the ACS study. He thought that this study controlled for smoking and diet a little bit better than the ACS study because they interviewed the cohort participants rather than just had them fill out questionnaires, but still thought that there could be some residual confounding. He did not think indoor air affected the Six Cities results because he did not think there was as much potential for differential exposures across the cities given their geographic distribution largely in the East.

3.6 Effect Modification

Expert K discussed three issues that he thought could be effect modifiers in specific studies: educational attainment, co-pollutants, and weather (as it affects prevalence of seasonal viruses and building effects).

Expert K thought that the ACS study results could be modified by educational attainment, since the ACS cohort was a more highly educated population than the general U.S. population. He thought this differential in educational attainment could cause the mortality effects to be underestimated and, given the Krewski analysis, assigned it a score of 2.

Expert K thought that differences in the co-pollutant/component mix in the Six Cities study relative to the rest of the country could result in potential effect modification of their results. Specifically, he thought that the SO_2 and acid particles could be higher in the six cities selected for the study than in the rest of the U.S., but thought that the ozone levels could be lower. Therefore, he was uncertain of the direction of bias, and assigned it a score of 2.

He thought differences in viral epidemics (e.g., adeno-viruses (not influenza)) could lead to an overestimate because of "certain kind[s] of winter and seasonal viruses which may be different in this temperate zone that in the southern temperate zone ... It might be a problem when you're extrapolating outside of those cities."

He also thought that temperature could be an effect modifier, but was uncertain of the direction. "And weather … the ability of outdoor air to penetrate indoors, depending on the amount of insulation and how tight your windows are and so on. I don't know if we're stretching or those are real issues, but there are differences that exist."

3.7. Exposure Issues

Expert K thought a major exposure issue in both the ACS and Six Cities studies was central site versus individual exposures. "Is there any confidence that what we say we're

monitoring is actually what people are exposed to? ... The other exposure issue … that's covered in this is going to be … where you live [with respect to a roadside]." He discussed Jerrett et al. work with the ACS data that attempted to improve spatial resolution of exposure. He indicated that the L.A. analysis (2005) seemed to show that improved exposure measurement leads to higher effect estimates. However, the analysis in New York (unpublished) did not find the same relationship. He did not necessarily think that the New York results refuted the L.A. results, because of city-specific characteristics such as the existence of tall vertical buildings that could cause differences in the exposures. He thought that exposure misclassification was probably more of a problem in the ACS study, and assigned it a score of 2 to 3. He thought the Six Cities study had better exposure measurement and placement of central site monitors, and assigned a score of 2. There was further discussion of the statistical arguments about the influence of exposure misclassification on effect estimates. He was uncertain about the direction of influence central site versus individual monitors might have on the effect estimates.

3.8. Causality

Expert K expressed the view that "causality doesn't simply arrive from statistical correlations." He indicated that there are several time-series and cohort studies that "suggest or indicate an association between PM concentrations and … mortality," but he thought of them as associations rather than evidence for causality. He thought that there are several mechanistic explanations for a PM/mortality effect, but he did not feel that there was a single, well-established pathway. "Reactive oxygen species, as we talked about this morning, [is] a nice model. There are some very elegant *in vitro* work that show these kinds of effects. There are whole animal studies [where] one can find lots of different mediators, lots of different electrophysiological responses, inflammation. But I don't think we have a clear understanding of a mechanism that goes from exposure to whatever it is in the ambient air to an effect that we want to establish as mortality or certainly morbidity. I think those are the two … major reasons that I'm still not certain about a causal relationship."

He thought that there were some differences in causality between short-term and long-term exposures, in that he thought the evidence was stronger for short-term. He cited the defibrillator studies indicating that if these studies had personal exposures measurements so that, "if I knew that the people who develop the arrhythmia, that they were the people that we actually knew that their exposure went up, not at a central site … that to me would get very close to establishing causality." As for the long-term studies, he found the kind of questionnaires and correlations not as powerful, particularly in the ACS study, for establishing a causal link. Furthermore, the exposure issues (discussed in Section 3.7) also undermined his belief in a causal relationship. He indicated that, "I don't know that there is a clear model that's going to get us there on the long-term studies until we actually are able to more carefully define the exposure."

Expert K estimated the likelihood of a causal relationship between PM and mortality to be about 20 percent. His overall estimate the range was rooted in his low confidence in the completeness of the existing science. In terms of cardiovascular mortality, he felt

there was uncertainty in both the link between exposure to particles and atherosclerotic plaque and in the progression from that stage to death. He observed that, "the only observations that really tie the long-term exposure with the plaque are the NYU studies." Although he thought these studies represent "an initially powerful set of studies," he thought that there were problems with extrapolating from studies performed on mice genetically altered to be at high risk for atherosclerosis to humans. In particular, he did not think that it was possible to tell yet whether it was just the particles or whether there were a number of other factors that could equally well trigger atherosclerosis in this sensitive mouse model.

He thought the likelihood of a causal relationship could be as high as 50 percent. To be more convinced, he would want a study to find similar results using a more normal animal model.

After quantifying his uncertainty in the C-R function, Expert K revised his estimates for the likelihood of a causal relationship. He thought that his original estimate of 20 percent was too low. He then specified a range of 20 to 50 percent likelihood with a most likely value of 35 percent.

3.9 Thresholds

Expert K thought that it is possible to make a conceptual argument that a population threshold exists. He drew an analogy with smoking, indicating that among heavy smokers, only a proportion of them gets lung cancer or demonstrates an accelerated decline in lung function. "I think that the idea that somehow there's no level that biologically is safe just doesn't fit with toxicology, where in fact we usually think about the dose making the toxicity of the material, and normal host defenses, normal responses, I think that there is going to be a level below which the population is fine."

When asked about what types of studies would be appropriate for determining values for a potential population threshold, Expert K answered that short-term studies with personal exposure monitoring data and clear outcome (e.g., arrhythmia) that could be associated with mortality would be helpful. He also thought that an animal model that did not use animals that were genetically altered to be highly susceptible to disease and that used exposures that were progressively closer to ambient levels would be necessary. However, he thought the likelihood that such a model would find no threshold would be small.

He did not think that a population threshold was detectable in the currently available epidemiologic studies. He indicated that in 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 falls between 0 and 5, and a 20 percent chance that it

falls between >5 and 10. The elicitation team took this information and created a probabilistic distribution in Crystal Ball™ with 50 percent of the weight at zero, 40 percent of the weight between >0 and 5, and 10 percent of the weight between >5 and 10.

3.10 Other Influential Factors

Expert K discussed additional sources of uncertainty that were not part of the protocol. Expert K indicated that he was unsure if publication bias existed, but noted that there were a large number of studies published with positive results on this topic (primarily time-series studies). In addition, Expert K expressed the view that the fact that investigators are finding associations with PM and "different, totally unrelated outcomes" such as reproductive outcomes, effects on fertility, central nervous system, stroke, and Alzheimer's disease made him uncomfortable. He thought that there was a possibility that, "there is a systematic problem that results in these kinds of associations."

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert K chose to characterize his C-R function within the specified range of annual average $PM_{2.5}$ concentrations (4-30 $\mu g/m^3$) in a piece-wise linear fashion. He specified one set of numbers that applied to concentrations of 4-16 $\mu g/m^3$ (hereafter, "Range 1"), and another set that applied to concentrations of >16-30 $\mu g/m^3$ (hereafter, "Range 2"). He chose the split point, 16 $\mu g/m^3$, because it was at about the mid-point of the range.

Expert K elected to provide his initial distributions conditional on the existence of a causal relationship and on being above a threshold. The elicitation team then combined his conditional distributions with his percent likelihood of causality specified in Section 3.8 and his probability distribution for a threshold specified in Section 3.9. In order to create a single distribution (hereafter, "Example Applied Distribution") representing the incorporation of all these concepts, the elicitation team probabilistically combined the distributions for threshold, causal likelihood, and concentration response with a distribution of population-weighted annual average $PM_{2.5}$ concentrations in the U.S. from EPA's BenMap model using Monte Carlo simulation.

Expert K began by specifying the percentiles for the slope of Range 2. He indicated that he put more weight on the ACS study than the Six Cities study because it was more geographically representative. He thought that the maximum value would be 1.5, slightly above the mortality estimates for the original Six Cities study. In order to specify his median, he indicated that he would start with an estimate of 0.6 from the ACS study and 1.25 for the Six Cities. We discussed again his view that smoking at diet might not have been fully controlled for in both studies and his uncertainty about the role of co-pollutants. He chose to weight the ACS study roughly twice as much as the Six Cities study and arrived at a median of 0.7. He then selected a 5^{th} percentile of 0.1 after considering a range of confidence intervals from the cohort studies. He then had the elicitation team fit a normal distribution to his estimated 5^{th} and 50^{th} percentile values. The resulting 25^{th} percentile value was 0.45, his 75^{th} was 0.95, and his 95^{th} percentile was 1.3.

He then specified the percentiles for Range 1, which overall he thought should reflect his view that at "these lower levels, there should be less effect for a small change." He provided a theoretical maximum of 0.8. He specified a 5th percentile of 0.1 and a 50th of 0.4**.** The elicitation team again fit a normal distribution to the 5th and 50th percentile values. His resulting 25th percentile was 0.28, his 75th was 0.52, and his 95th was 0.7.

Viewing his distribution with causality incorporated prompted further discussion of his assumption about the likelihood of causality. He reconsidered his views thinking his initial estimate of 20 percent was perhaps an "overreaction." He still thought he couldn't place more than 50 percent likelihood of a causal relationship but decided to change his most likely estimate from 20 percent to 35 percent, with a range of 20-50 percent.

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Range 1 (4 – 16 μg/m^3)

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)	Percent Change in Mortality Distribution Incorporating Causal Likelihood and Threshold (IEc Generated)
Minimum	-	0	0
5th	0.10	0	0
25th	0.28	0	0
50th	0.40	0	0
75th	0.52	0.29	0.28
95th	0.70	0.59	0.58
Maximum	0.80	0.80	0.80

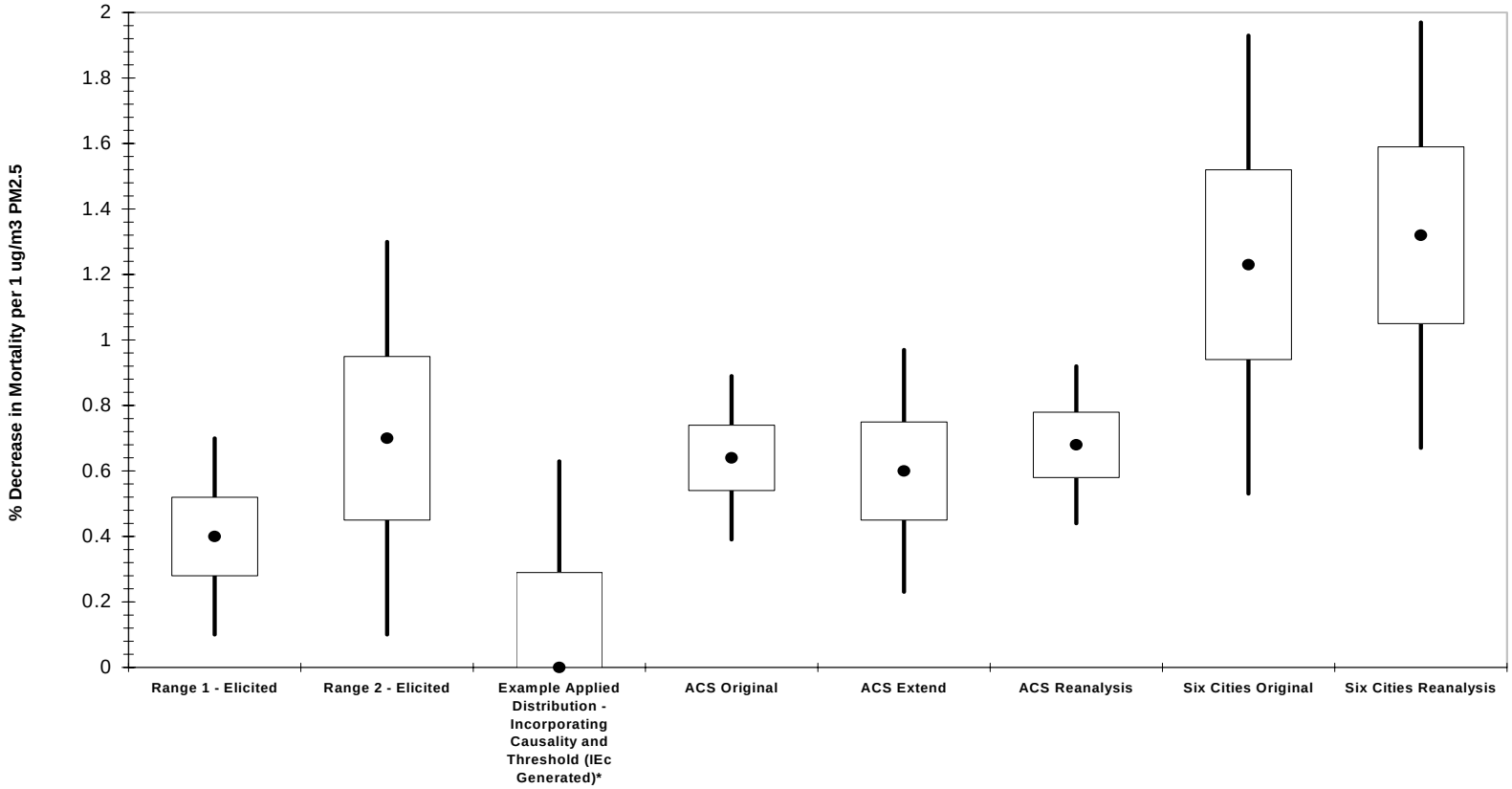
Exhibit 2: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1µg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Range 2 (>16 – 30 µg/m^3)

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)	Percent Change in Mortality Distribution Incorporating Causal Likelihood and Threshold (IEc Generated)
Minimum	-	0	0
5th	0.10	0	0
25th	0.45	0	0
50th	0.70	0	0
75th	0.95	0.48	0.48
95th	1.3	1.1	1.1
Maximum	1.5	1.5	1.5

Exhibit 3: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1µg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations After Incorporating Causality and Threshold and Applying the C-R functions to the Population-Weighted Annual Average $PM_{2.5}$ Concentration Distribution in the U.S. from BenMap - Example Applied Distribution (IEc Generated)

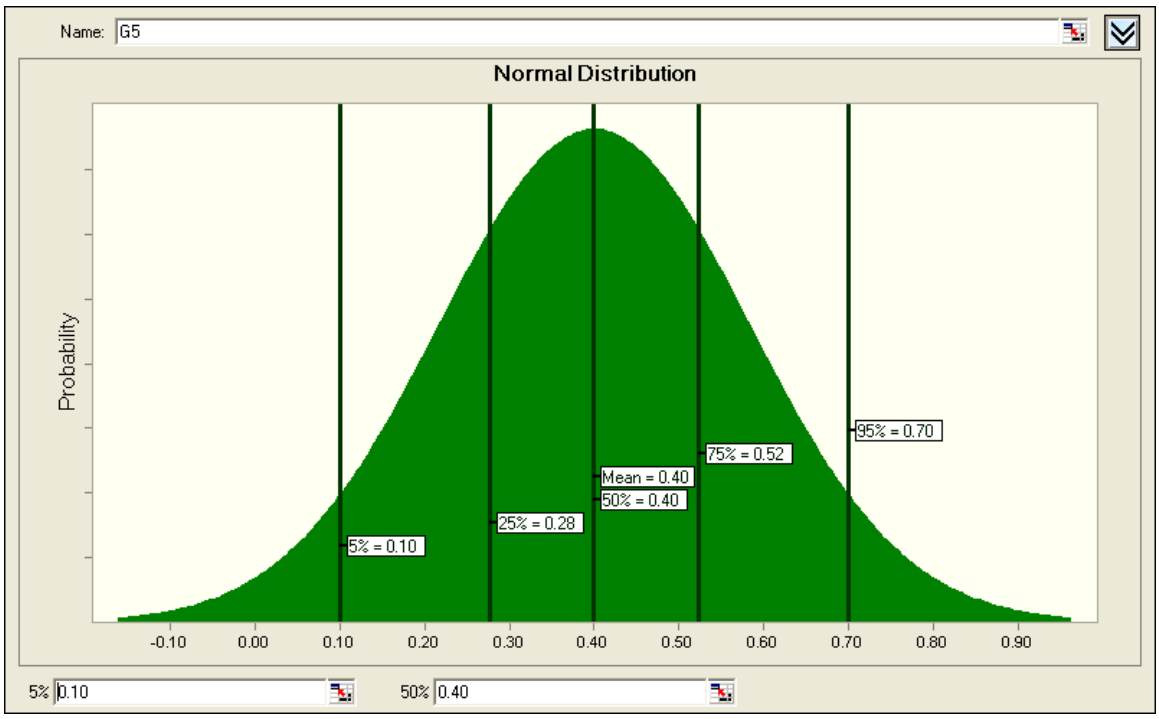
Percentile	Percent Change in Mortality
Minimum	0
5th	0
25th	0
50th	0
75th	0.29
95th	0.63
Maximum	1.5

Exhibit 4: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distributions from Expert K

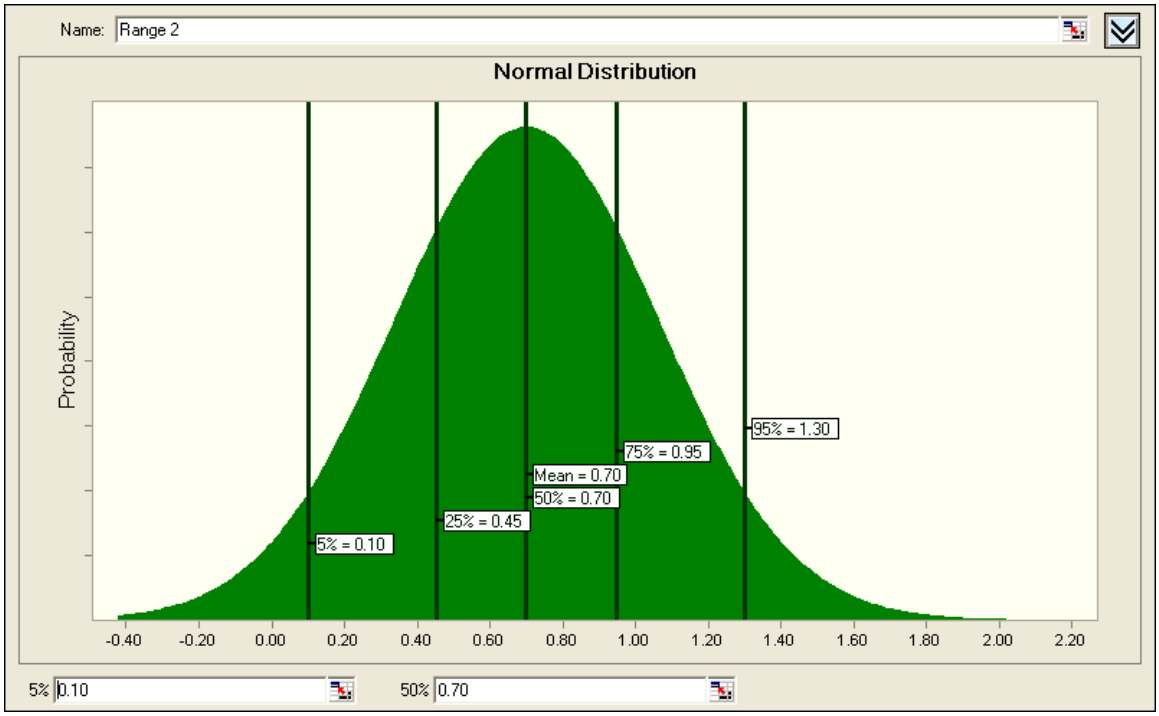


* = Distribution incorporating causality and applying the C-R functions from Ranges 1 and 2 to a 2002 population-weighted annual average $PM_{2.5}$ concentration distribution in the U.S. from BenMap.
• = median □ = interquartile range | = 90 percent confidence interval

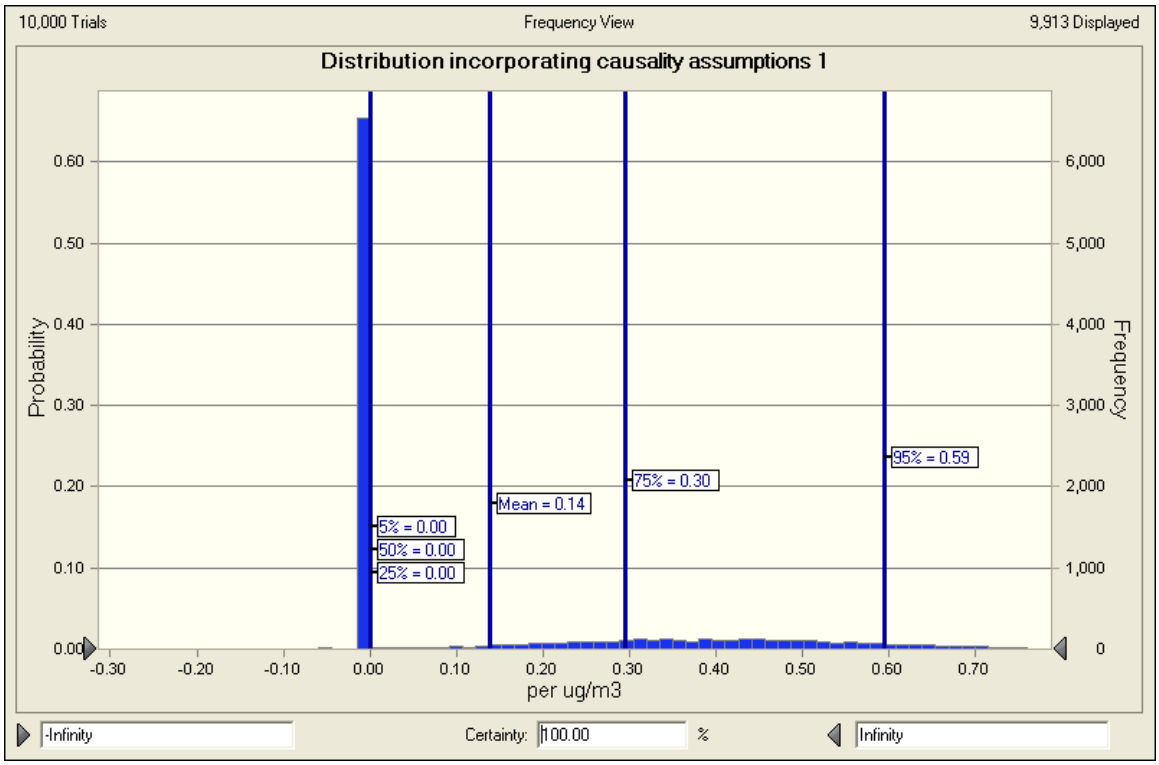
Elicited Distribution – Range 1



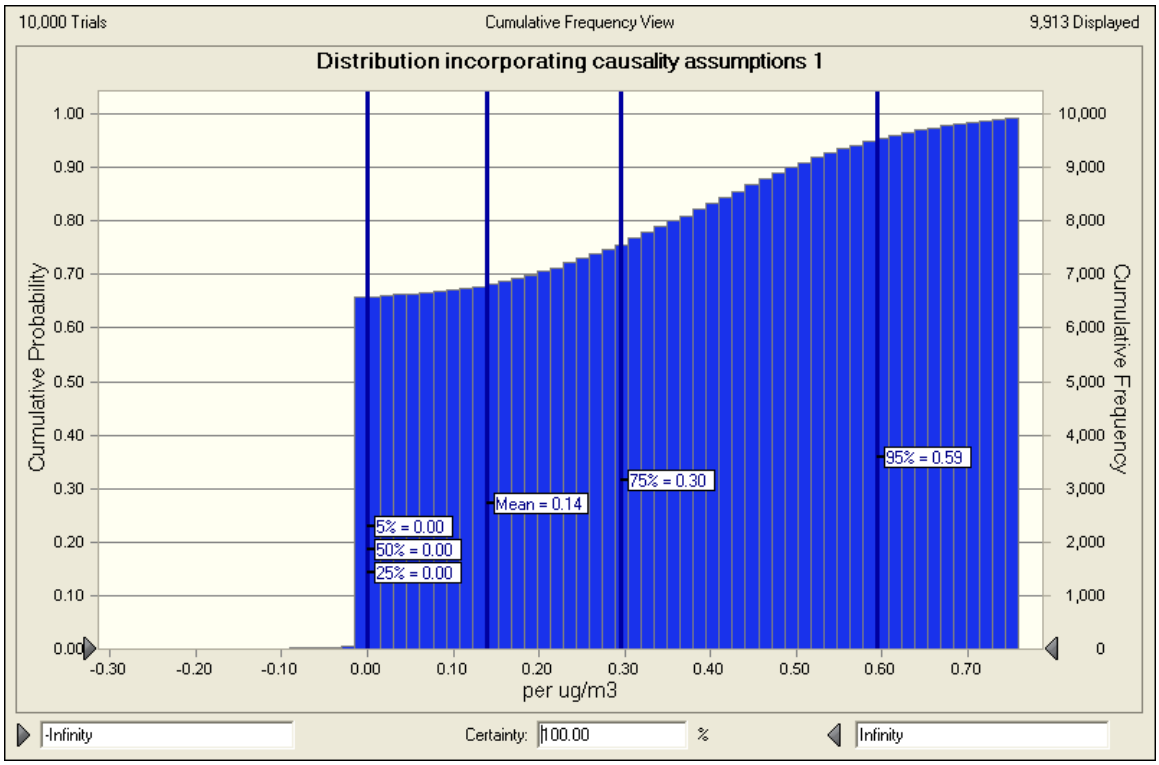
Elicited Distribution – Range 2



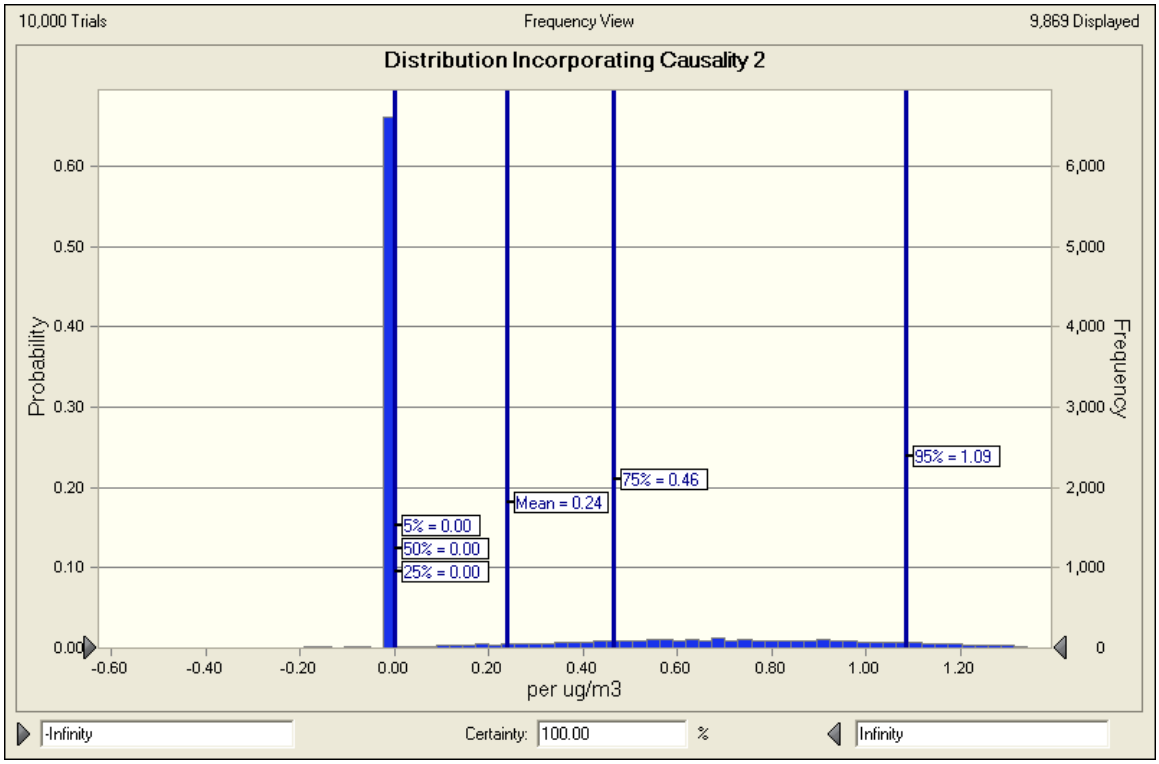
Range 1 Incorporating Causality - Probability Density Function (PDF) (IEc Generated)



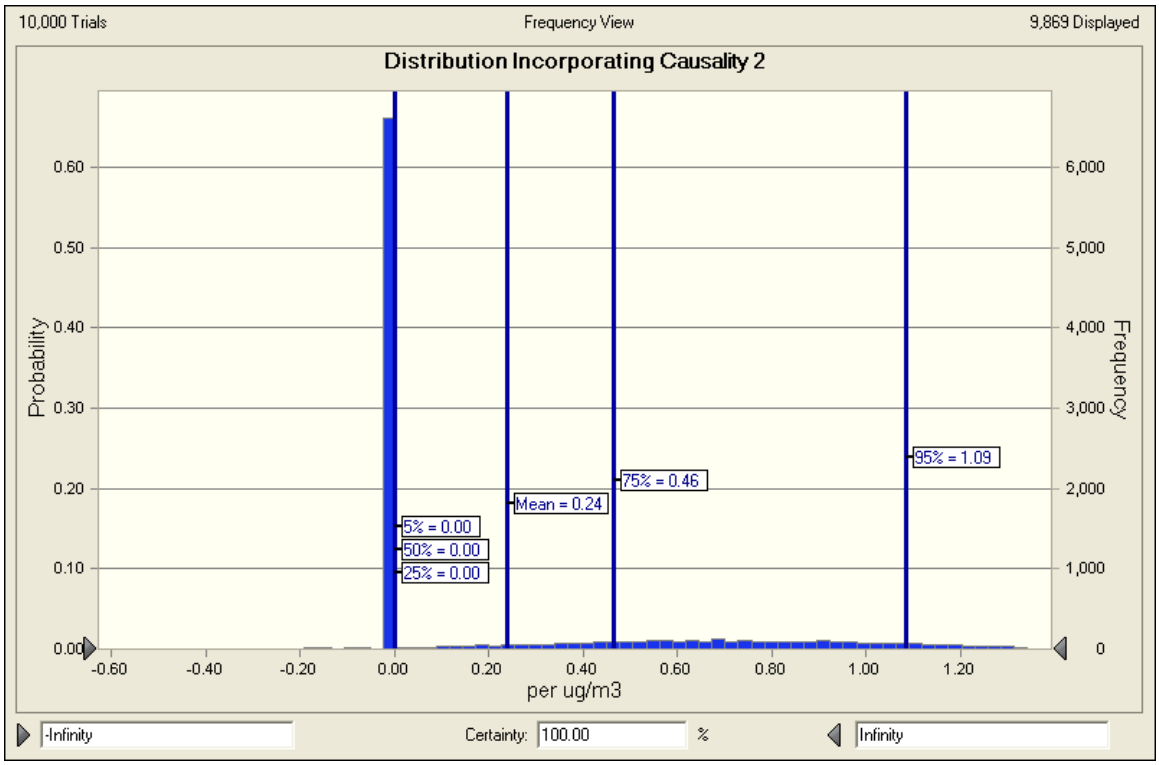
Range 1 Incorporating Causality - Cumulative Density Function (CDF) (IEc Generated)



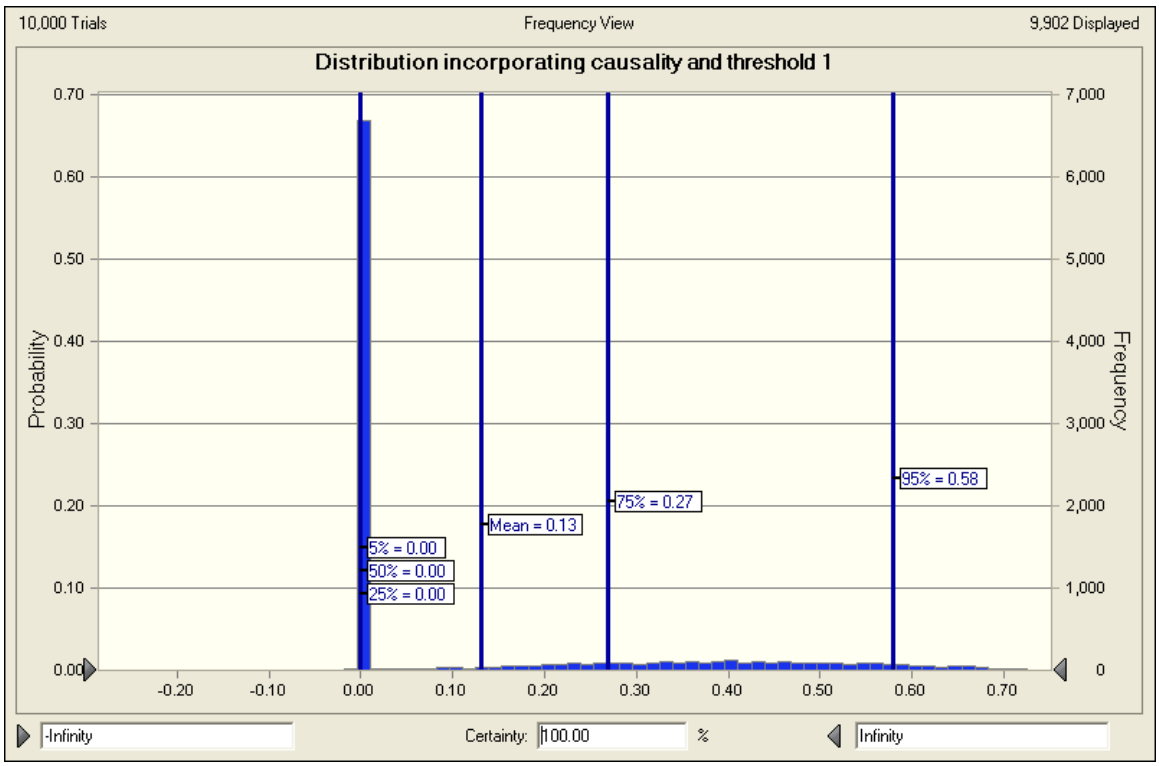
Range 2 Incorporating Causality – PDF (IEc Generated)



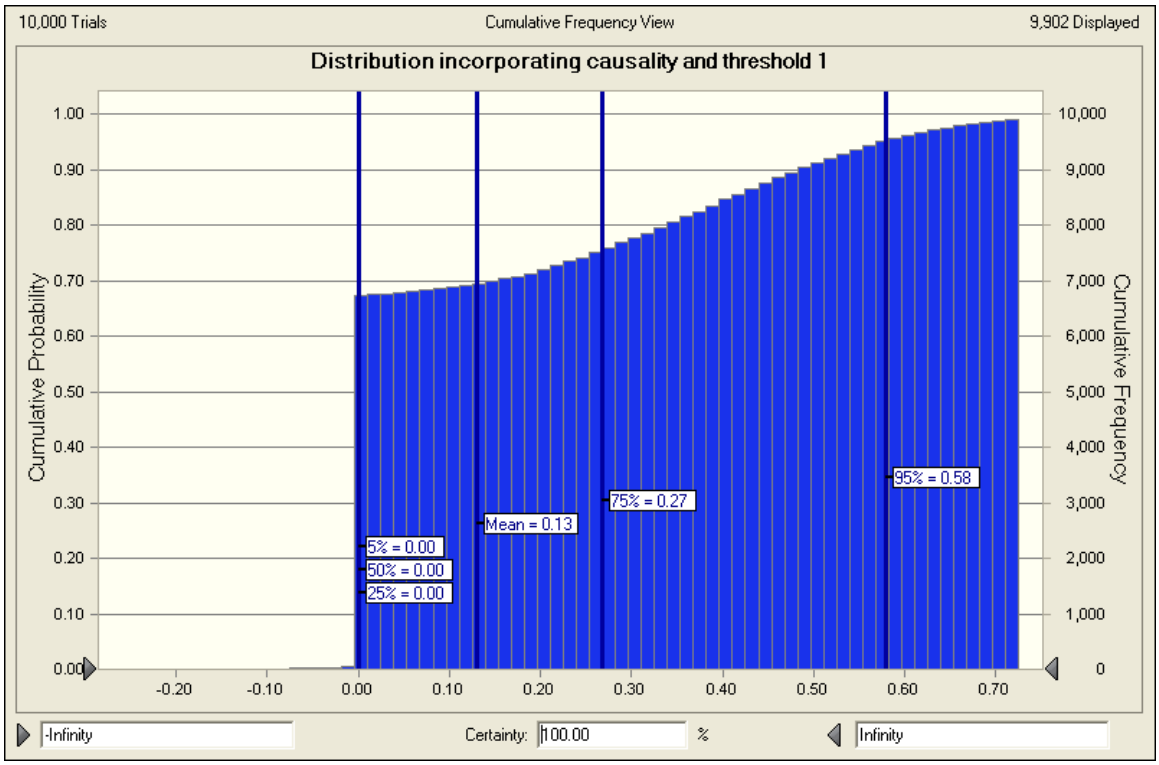
Range 2 Incorporating Causality – CDF (IEc Generated)



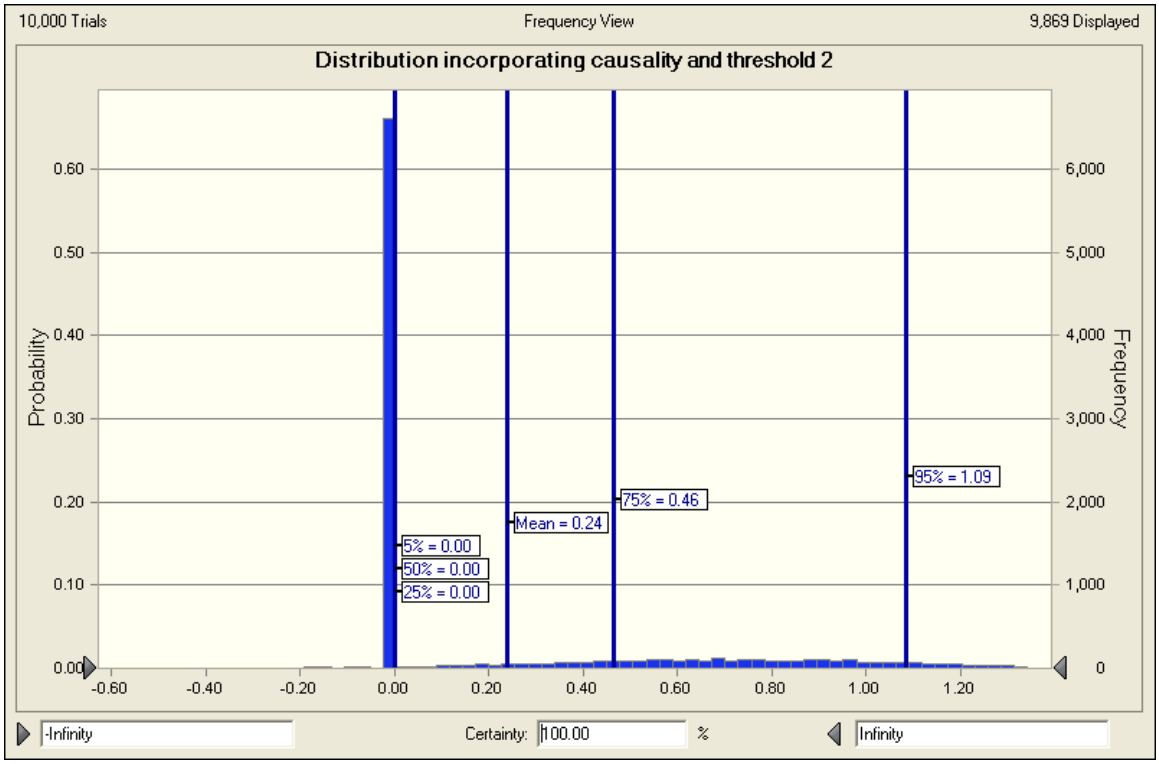
Range 1 Incorporating Causality and Threshold – PDF (IEc Generated)



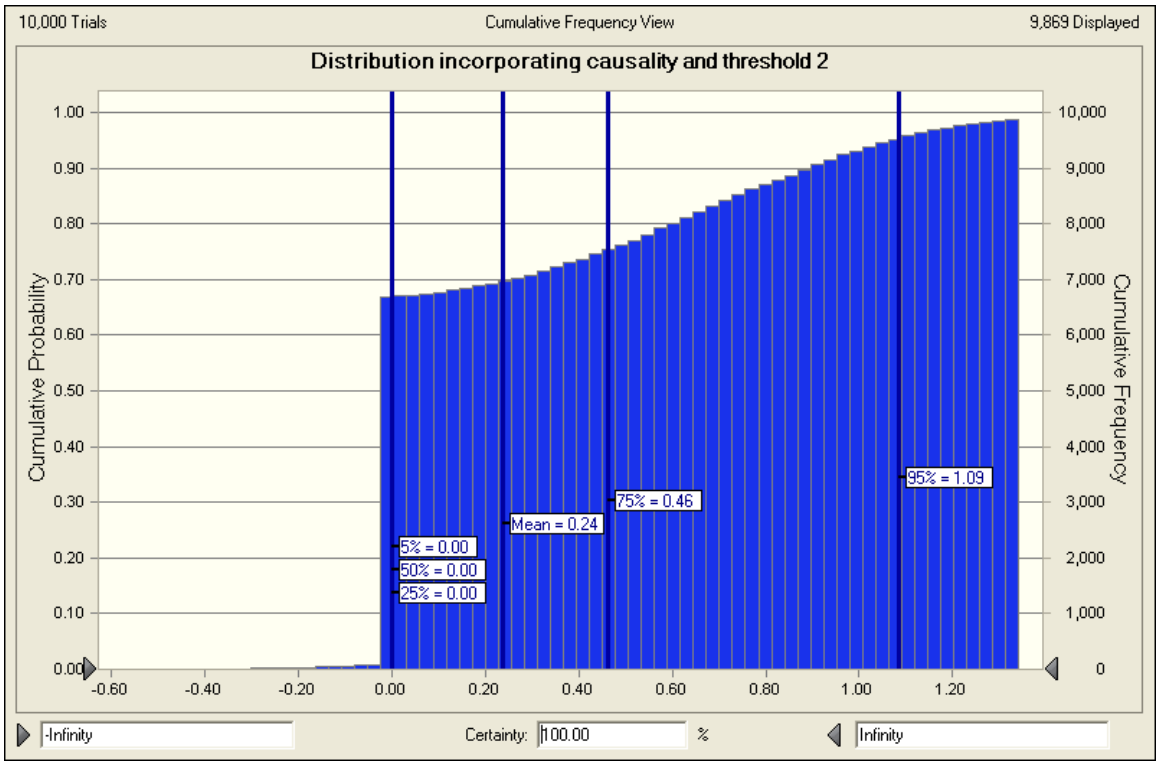
Range 1 Incorporating Causality and Threshold – CDF (IEc Generated)



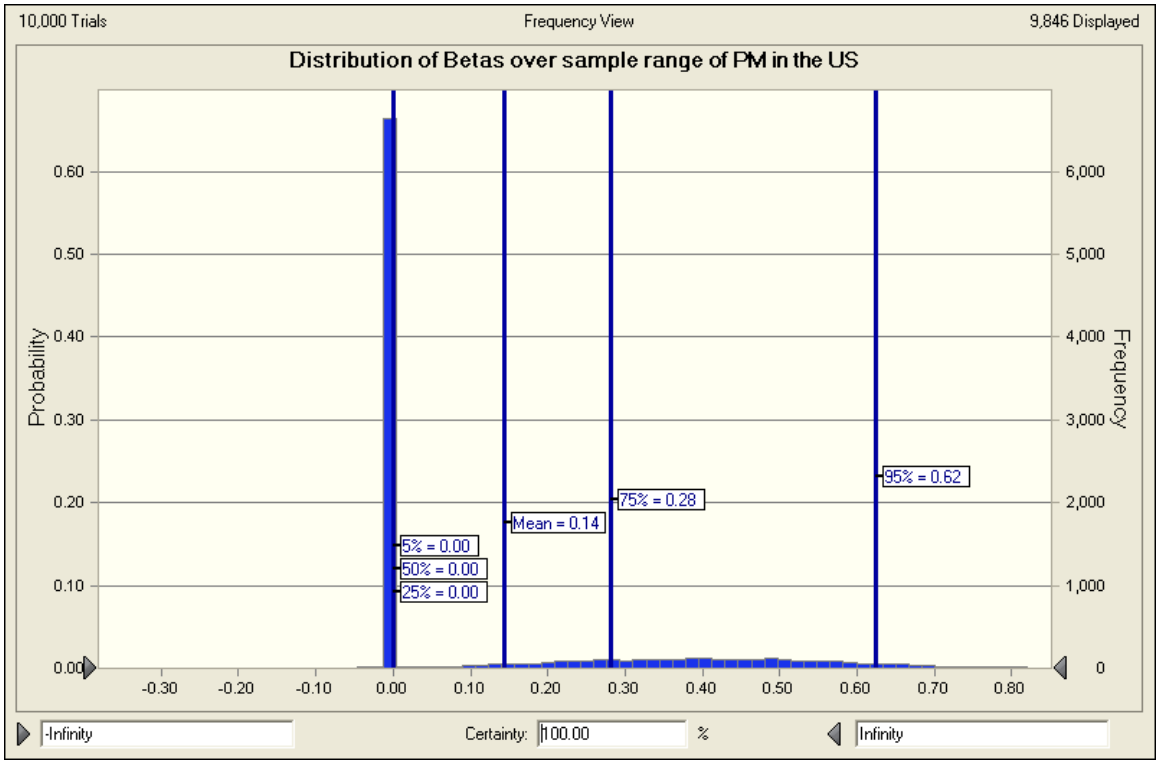
Range 2 Incorporating Causality and Threshold – PDF (IEc Generated)



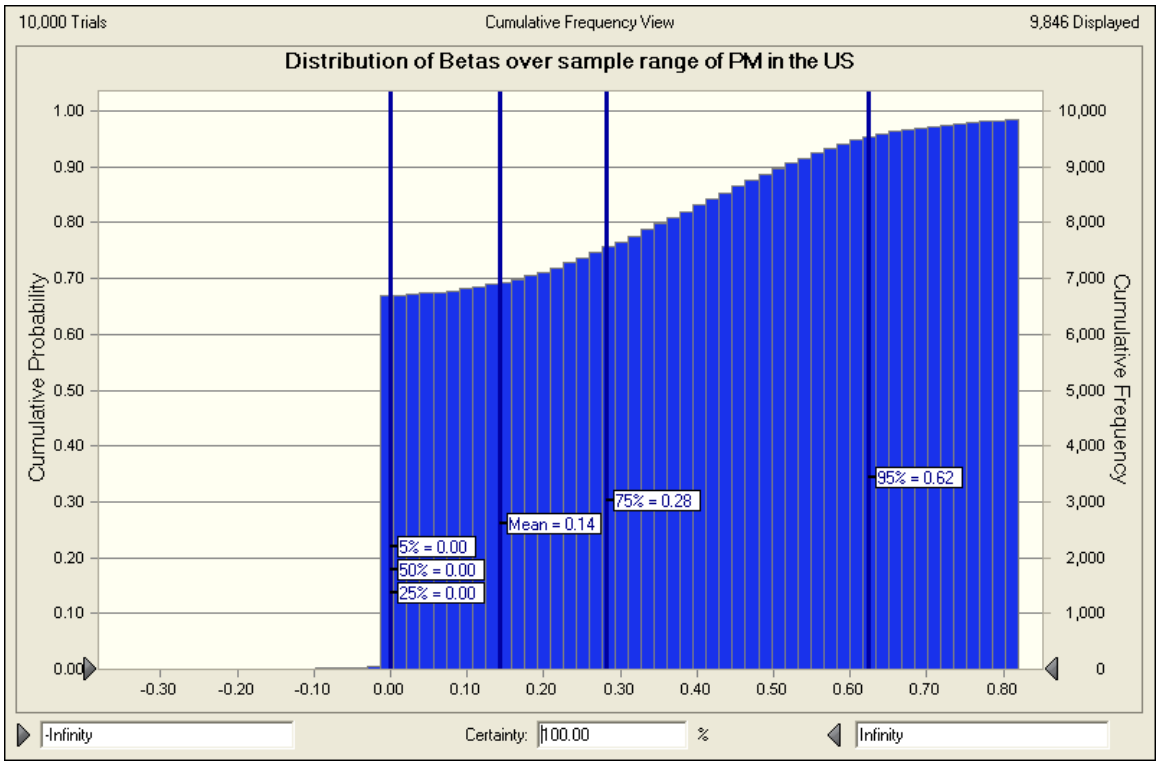
Range 2 Incorporating Causality and Threshold – CDF (IEc Generated)



Example Applied Distribution – PDF (IEc Generated)



Example Applied Distribution – CDF (IEc Generated)



U.S. EPA EXPERT ELICITATION STUDY OF THE CONCENTRATION-RESPONSE RELATIONSHIP BETWEEN ANNUAL AVERAGE $PM_{2.5}$ EXPOSURE AND MORTALITY

Modification to Expert Judgments

Expert K

Date: 7/10/06

Section of Protocol Affected (Section Number and/or Title):

3.8 Causality

Description of Change (e.g. to a specific percentile, or to a qualitative opinion or statement of belief):

Expert K changed his range of values for the likelihood of a causal relationship. The new range is from 5 to 50 percent. The most likely value of 35 percent remains the same.

Rationale for Change:

My confidence (or lack thereof) in the data showing that reductions in $PM_{2.5}$ levels translate into changes in mortality (e.g., annual averages of 4-30 $\mu g/m^3$) goes back to my concerns about both exposure measurements (errors) as well as all of the other unaccounted for variables that are changing in the environment. Unfortunately, your question is somewhat confusing to me: Are you asking about a 1 $\mu g/m^3$ reduction or as much as a 25 $\mu g/m^3$ reduction. I would certainly have more confidence that a large reduction could impact mortality and virtually no confidence that we could meaningfully measure the impact of a reduction of 1 $\mu g/m^3$. It would simply be an exercise in more modeling and extrapolation at such low level reductions. My estimates will certainly increase for larger decreases in $PM_{2.5}$ levels.

Expert L
Interview Summary

Interview Summary
Expert L

PART 3. CONDITIONING STEP: PRELIMINARY QUESTIONS

3.1 Mechanisms for Effects of Exposure to $PM_{2.5}$

Expert L discussed the biological mechanisms for short-term and long-term exposures separately because he thought, "different mechanisms come into play." He preferred to begin with a discussion of the mechanisms of mortality related to short-term exposures because "long-term effects are to some extent also the result of a mechanism that happened in the acute domain."

Short-Term Exposures

Expert L's views on causes of death related to fine particle exposure were informed by a broad overview of the scientific evidence, which he thought provides support to the idea that "seemingly different pathways are interconnected." He thought smoking is a good parallel for fine particles because both are complex exposures; particles are "markers for something for that is much more complex than just the mass … [which makes it] very possible to assume that [a] whole range of pathways are initiated."

Expert L thought that a major cause of death from short-term exposures was cardiovascular disease, specifically myocardial infarction (MI) and stroke. He thought work by Annette Peters (2001), Joel Schwartz, and others on "a whole range of … acute effect studies" was informative for these outcomes (Wellenius et al., 2005; Tsai et al., 2003; Hong et al., 2002; Zanobetti et al., 2000; Schwartz et al., 2003). He also thought the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) was very important because of the wide geographic coverage and standardization of methods. He felt that the U.S. findings are strengthened by the European studies, particularly the large multi-center studies like Air Pollution and Health – A European Approach (APHEA), where publication bias is not an issue.

He thought one pathway for cardiovascular disease included particle-induced oxidative stress leading to systemic inflammatory responses and release of fibrinogen, which are associated with outcomes like MI, and to a lesser degree, with stroke. He also thought cardiovascular outcomes could be caused by particle-induced arrhythmias or changes in autonomic function that might also be linked to inflammatory pathways. To support this, he cited work by Joel Schwartz, which found that changes in HRV are modified by factors on the oxidative stress or inflammatory pathway (Schwartz et al., 2005). He indicated that there are fewer studies on stroke, but they are of a high quality. Expert L thought that the mechanistic evidence is strengthened by the fact that "where we expect MIs [to] happen, strokes should happen as well, and the fact that [published] studies show this, I think this very much [strengthens] the evidence."

Expert L thought that another cause of death related to short-term exposures was respiratory disease and that oxidative stress and inflammation were again part of the pathway. "The changes in [neuro] reflexes in the lung ... structur[al] changes [like] narrowing of the airways." He thought these impacts create particular challenges for asthmatics. Expert L also indicated that PM could cause weakening of clearance and the immune system, lowering an individual's ability to fight infection.

In addition, he thought there were subpopulations that are particularly susceptible to the effects of PM, including people who are in "terminal health states," such as those with lung cancer or diabetes. He noted Mark Goldberg's time-series studies showing that lung cancer patients die earlier when air pollution is high. He thought diabetics were more likely to have underlying cardiovascular disease or atherosclerosis, in addition to reduced defense mechanisms against infections and therefore would be vulnerable to acute exposures to PM.

Long-Term Exposures

Expert L thought that overall, "long-term effects … are really changes in the underlying pathophysiology that ultimately leads to ailments, to chronic conditions which we know lead to premature death."

He first discussed lung cancer as a cause of death related to long-term exposures. He cited a Swedish study by Nyberg et al. (*Epidemiology*, 2000: Vol. 487) that found lung cancer associated with air pollution exposures 20-30 years in the past. He also thought that there was "toxicological evidence for some of the relevant pathways that are well-described for carcinogenesis, that they also relate to toxicity of particles, leading to chronic persistent inflammation in the lung and airways, with genotoxic and proliferative effects in the lung tissue and remodeling." He indicated carcinogens can adsorb to the particles and individuals can be exposed in that way as well.

Expert L next discussed cardiovascular disease as a cause of death related to PM. He cited four groups of researchers that published recent animal studies providing evidence that particles contribute to the progression of atherosclerosis (a study in rabbits by Suwa et al., 2002; a particle instillation study in rabbits showing inflammation by Goto et al., 2004; a study from Mort Lippmann's group (Sun et al., 2005) showing progression of atherosclerosis in rats; and a rat model by Lemos et al. 2006). He also mentioned a Japanese study by Yamawaki et al. (2006) that found endothelial dysfunction with exposures to carbon black, which he thought was an important part of developing atherosclerosis. He thought additional support of this mechanism was provided by an epidemiologic study by Kunzli et al. (2005) in Los Angeles (L.A.) showing increasing carotid intima-media thickness (CIMT) associated with outdoor particles.

He thought another mechanism related to long-term exposures was chronic lung inflammation leading to decreased lung function, although he acknowledged that lung pathologists would agree that it is difficult to line up all the pathways that lead to reduced lung function as it may involve air ways, lung parenchyma, and the pulmonary

vasculature. He indicated that, "[l]ower lung function is one of the strongest predictors of survival, of life expectancy, of mortality." He thought one of the most important pieces of evidence for this mechanism comes from the Children's Health Study, which found that children's lung function develops more slowly if they reside in areas with poor air quality. It also showed that when children moved into higher pollution areas, the development of their lung function decreased. He thought in general, respiratory deaths were less correlated with air pollution in the major cohort studies than other causes. For instance, the Six Cities follow-up (Laden et al., 2006) had found a positive signal with respiratory deaths that was non-significant. He indicated that some studies have found signals for chronic obstructive pulmonary disease (COPD), but in general the signals are weak. He thought an explanation for this could be misclassification of cause of death (e.g., older people are more likely to be classified as dying from cardiovascular disease) or a lack of clarity in the definition of COPD itself. He indicated that there was one German study that observed an association between air pollution and lung function markers of COPD (Schikowski et al., 2006). In addition, he said that the Adventist Health and Smog (AHSMOG) study shows incidence of COPD associated with PM. "From the mechanistic perspective … it is very important to use lung function studies in assessing the plausibility … of deaths being related to particles or to air pollution."

Finally, he discussed the development of asthma – a story he thought was not as clear. He indicated that several recent studies (Children's Health Study (McConnell et al., 2006); a European study (Venn et al., 2001); a study in Alaska (Gordian et al., 2006)) are suggesting a connection between traffic related exposures and development of asthma. He thought these studies were indicative air pollution's role in the development of chronic disease

3.2. Conceptual Framework for Mortality Effects of Short-term and Long-Term $PM_{2.5}$ Exposures

Expert L thought the Künzli diagram was a good conceptualization of the relationship between long- and short-term exposures. "I think it is … a framework that is useful … and it allows us to think, for each death … about what story might be important." He thought the difficulty is how the literature relates to the diagram and he indicated that one has to be careful in discussing this relationship.

3.3. Role of Epidemiologic Study Design in Characterizing the Total Impacts of $PM_{2.5}$ Exposures on Mortality

Expert L thought that none of the study designs alone were sufficient for capturing "total" mortality from PM exposures. He thought time-series studies with a 2-3 day exposure window are useful measures of the most acute effects. He indicated that these estimates are certainly not a measure of "total" mortality, nor even of total acute if one imagines that "air pollution … triggers an MI, many MI [patient]s do not die within 3 days but end up in a phase of treatment … and some of them die within six weeks, or during their convalescence." He thought Ari Rabel's work arguing that we don't observe the true total acute effect was a useful theoretical contribution.

The next class of studies he discussed was the cohort studies which he thought provided some measure of the effects of long-term exposures, but which were unlikely to capture the cases where life is shortened by less than five to ten days up to two or three weeks, for example. He noted that we have to be "clear and honest that these are very, very crude approaches in terms of characterizing exposure, and also in terms of measuring time lost … [loss of] life expectancy. When asked how much mortality the cohorts might miss, he responded, "I don't know the signal exactly, but maybe it's the NMMAPS type of signal, which is …[about] a half percent [per 10 $\mu g/m^3$] … It seems to me this is [part of the] noise of what we discuss for the cohort [studies]."

He thought case-control studies had value for estimating the risks for specific health outcomes like lung cancer, or possibly COPD. He did not think they were appropriate for assessing the effects of exposure on total mortality.

The final type of studies he discussed was the intervention study, which he thought provided strong evidence of a benefit of reducing air pollution, particularly in the short-term. However, he thought they might pose a challenge for assessing the long-term effects of reduction. For example, he thought the Utah Valley study (Pope et al., 1996) provides "particularly strong evidence for the acute, sub-acute domain." He thought the Clancy et al. (2002) study in Dublin was the appropriate design for assessing the effects of long-term exposures but that the longer the follow-up time, the more difficult it becomes to isolate the effects of PM changes alone. It might be argued that changes in other factors, for example, economic development, diet, smoking rates, or ETS could also contribute to improvement, or declines (in the case of obesity) in health status and susceptibility. Thus, the intervention studies could not necessarily be a "gold standard" for chronic effect studies in his view.

The effects captured by each study design are shown in the table below:

Study Design	**Type of Effects Captured (e.g., short-term, long-term, or both)**
Cohort Studies	Long-term, excludes very short-term effects
Case-Control Studies	Long-term for specific health outcomes only
Intervention Studies	Acute, sub-acute, intermediate acute (depending on study)
Time-Series Studies with a 2-3 day lag; distributed lag (up to 3 months)	Short-term; subacute

3.4. Epidemiologic Evidence for the Impact of Exposures to PM on Mortality

Expert L thought that the following characteristics would be part of an ideal epidemiologic study to characterize the $PM_{2.5}$-mortality relationship in the U.S. population:

- Based in the U.S.;
- Large sample size covering the entire U.S.;

- Random population sample, including susceptible individuals;
- Spatially resolved exposure data at the neighborhood level;
- Lifetime, or at least 10, 20, or 30 years of individually assigned exposure;
- Recent cohort (with PM exposures at current levels);
- Control of potential confounders and susceptibility factors (e.g., smoking, physical activity, diet, obesity, body mass index (BMI), family history of disease);

When asked to review the epidemiologic studies that have been most informative about the percent change in all-cause mortality related to a reduction in annual average ambient $PM_{2.5}$ concentrations, Expert L first discussed Pope et al's work with the American Cancer Society (ACS) cohort (Pope et al., 1995 & 2002). He thought that strengths of these studies included large population, geographic representation of the U.S., long-term follow-up, control of confounding, and the degree of scrutiny and re-analysis that they have undergone. He thought that limitations of the studies included the very large areas that were assigned exposure from a single monitor and the fact that the population was not a random sample. He also discussed the ACS reanalysis in L.A. He thought that this study had very strong exposure assessment and control of confounders. However, he thought that the fact that it only included one city was a limitation. "[I]t opens the question of whether these effects are larger because of the better exposure, or whether it's a population … with different types of more toxic exposures from traffic, more toxic PM due to the ozone environment or the susceptibility [in] the population."

He next discussed the Six Cities study, both the original (Dockery et al., 1993) and the extended analysis (Laden et al., 2006). He thought a strength of these studies were that the cohort was a more random sample than the ACS study, with better exposure assignments (smaller areas assigned to each monitor). However, he thought small sample size and geographic focus on the eastern U.S. were limitations.

Expert L discussed the AHSMOG, the Enstrom et al., 2005, and the Veteran's cohort studies by Lipfert et al. Although he considered these studies for this exercise, he did not end up relying on the estimates when quantifying his C-R function. However, he did discuss the following strengths and limitations:

- **AHSMOG**: Expert L thought that this study was overall supportive of a PM effect but less clear. He indicated that the population was not a random sample and had restricted geography, small sample size, and was non-representative in terms of susceptibility factors. He did think the study had full control of smoking, good exposure assignment, making "attempts to assign exposure on an individual level, more than any other [current] study."
- **Enstrom et al., 2005**: He thought that this study was limited geographically, in that it only included 11 counties in California. In addition, he indicated that they used a single monitor for each county, which could have introduced exposure error. It is also limited by having a very elderly population so the relevant periods of exposure may extend back a long time.
- **Veteran's cohort:** He said that this study used traffic as a surrogate for PM, which was consistent with a Canadian study by Finkelstein et al. (2004) and the Netherlands cohort study by Hoek et al. (2002). He thought that the statistical

analysis was a limitation, indicating that the authors used models that included both traffic and $PM_{2.5}$, which would yield uncertain results since the two factors are correlated. He also indicated that the study population was not representative of the general population because it included a large percentage of smokers and included only hypertensive individuals, which could cause uncertain results due to potential treatment interaction effects.

- Other work that he thought was supportive of a quantitative relationship included the Finkelstein et al. study and the Hoek et al. (2002) study, which are more "traffic-driven" and therefore not directly applicable given the relationship to PM.

3.5 Confounding

Although Expert L discussed potential confounding by smoking and SES, he thought that confounding did not affect the published estimates from the ACS studies (Pope et al., 1995 & 2002), the ACS L.A. reanalysis (Jerrett et al., 2005), and Six Cities Studies (Dockery et al., 1993; Laden et al., 2006) in a significant way. "It's my impression that confounding has been addressed … I don't see much strong evidence for an uncontrolled confounder that one could claim, and convince me that [it is] really confounding. [That it is both] associated with air pollution and these outcomes." In particular, he felt that the Krewski et al. (2000) reanalyses of the Six Cities and the ACS studies, as well as the ACS L.A. reanalysis (Jerrett et al., 2005) included extensive sensitivity analyses with several different variables without finding a large effects on the estimates. For example, Jerrett et al. included 44 individual covariates in the analysis. Expert L thought that it was possible that the Jerrett analysis could have overadjusted. In particular, he was concerned about the models with 44 individual covariates plus contextual covariates because they include variables that are indicators of exposure, such as air conditioning other pollutants. The contextual covariates include things like SES (a complex covariate), or urban land use that could also be indicators for exposure. He thought SES could also be a surrogate for susceptibility (e.g., correlated with obesity, diet, antioxidants) but was likely not a confounder per se. In general he argued the quantitative effects of possible confounding were within the range of effect estimates presented in the key studies. He thought the model from the Jerrett study that included 44 individual covariates and parsimonious contextual covariates, with relative risk of 1.11, represented the "lowest end" of the PM effect estimates.

3.6 Effect Modification

Expert L framed his discussion of effect modification by stating that he sees two different aspects: 1) real biological modifiers of the PM effect (i.e., that influence susceptibility to PM) and 2) the presence of co-pollutants in the ambient environment that may modify the effect of particles. For this discussion, he focused on the first issue. The second issue, which he felt related more to regulatory approaches, was left for discussion of exposure issues.

Expert L's discussion focused on the ACS study. Expert L thought that the ACS study population was not reflective of the actual U.S. distribution of SES; that is that it

underrepresented individuals with lower SES as indicated by educational attainment (those with less than a high school education). He thought this could be one of the reasons that the effect estimates in this study are lower than those found in the Six Cities study. He reiterated that SES could be a marker for several factors, such as obesity, diet, or underlying diseases. He thought it would be useful to adjust the ACS study for educational attainment, which might get at some of the underlying SES effect modification. He would want to include it in his list of effect estimates as part of characterizing his uncertainty. He thought the sample adjustments presented at the Pre-elicitation Workshop, showing 30-50 percent increases in the ACS mortality effect estimate when adjusted for educational attainment, were consistent with what he would have expected.

Effect modification in the Jerrett et al. (2005) and the Six Cities studies were not discussed, and he thought that the paper did not disclose much about heterogeneities.

Expert L expressed the general concern that there are important heterogeneities (patterns in susceptibility) in populations that are likely to change over time. "This type of risk assessment [the expert elicitation] would be much less uncertain if our data would show no heterogeneity. And that's not true. The evidence is really increasing that we have heterogeneity."

3.7. Exposure Issues

Expert L thought that exposure misclassification due to central monitors versus individual exposures caused the ACS study results to be biased downwards. He discussed the ACS L.A. reanalysis by Jerrett et al. (2005), indicating that he thought that it was "strong evidence that improvement of exposure assignment matters [quite a bit]." However, he thought that "there is a limitation [in the Jerrett paper] … we are not sure to what extent [the] much bigger signal [is] a result of being in California rather than in the rest of the states, or in another susceptibility domain than the rest."

He thought another exposure issue was changes in the characteristics of ambient air pollution over time. He thought that "exposures, emissions change with technology. And [it] is … not … possible to take into [this into] account with the current type of data, which is just mass-based."

Expert L discussed his theoretical concern that "in heavily polluted centers, over the last 30 years, improvements in air quality have been much … bigger than in the cleaner areas, where it might even have deteriorated or [become] more or less stable. This has implications on the estimate, and not knowing exactly which time window is the relevant one for the mortality, this becomes an unresolvable challenge. Exposure gradients and errors in assignment also change with the time window of exposure." He thought this probably could be more of a problem in the Six Cities study because it was smaller, so "the large change in one city has a big impact." He thought that this was the only argument he could find that "could lead to an overestimation of the main effects." However, he noted that the Laden et al. analysis, by looking into how changes in air

quality affected the outcomes, controlled for the problem to some extent. He thought this issue was less of a problem in the ACS study, pointing to the sensitivity analysis in the Krewski reanalysis and a comparison of effect estimates from different time periods in a paper by Jerrett et al. (2003).

Expert L discussed regional differences in the concentration-response (C-R) function. He indicated there are not enough data in the studies of long term exposures to stratify by region so "one can go to the acute effect studies ... for evidence of … geographic heterogeneity." He noted that NMMAPS (Samet et al., 2000) found evidence that "there are geographic patterns" although the reason for these differences was not easily explained. He thought they could be due to differences in PM toxicity, or susceptibilities of the population. He thought that the ACS study was "a decent estimate for the [entire] U.S.," and although he thought the Six Cities study was limited by the inclusion of only six cities and was probably an estimate of east coast exposures, he thought "these differences [in PM composition] are not sufficiently big to expect too much of a difference in this estimate." He went on to say that, "I'm convinced [the constituents of PM] does matter, but it's very difficult to say how. And the answer might be different for the different outcomes. Might be different in acute and in the chronic domain. And given all these uncertainties, it still seems that measuring the mass concentration in the real world of real exposures that happen today, real emissions, it still seems to be a pretty useful marker for a much more complex story."

3.8. Causality

Expert L indicated that he relied on a variety of evidence to evaluate the strength of the causal relationship between PM and mortality:

- Cohort and time-series mortality studies;
- Intervention studies (epidemiological);
- Studies examining the underlying mechanisms, "in terms of morbidity or intermediate outcomes, [it] is extremely important to support the evidence"; and
- Toxicological studies, especially those using concentrated ambient particle (CAP) exposures. In particular, he mentioned an animal study by McDonald et al. (2004) that involved an experimental intervention "changing the diesel emissions with a filter [or] sulfur content of the gasoline and showing that all these acute effects on the molecular and cellular level are almost disappearing or very … strongly reduced, all the inflammatory responses, oxidative stress-related responses."

His overall conclusion was that he does not "see any coherent way to argue that this is a fluke that has nothing to do with air pollution and particle pollution, but would rather be explained by SES factors or the weather or smoking."

He thought that there was strong evidence of a causal relationship with mortality for both long-term and short-term exposures. He thought that "there is a continuum, and given that we have this everyday exposure, we repeat the short-term effects every day in the

long run. [B]ased on the mechanism discussion we had yesterday, I think evidence is really increasing almost every month in the direction of leading to the conclusion that these short-term acute mechanistic pathways that are initiated, they have longer-term effects."

Expert L thought that the likelihood of a causal relationship between PM and mortality fell between 90 and 100 percent, with a most likely value of 99 percent. He based these estimates on the fact that he did not think that there was an alternate explanation for the effects seen, particularly in light of the more recent evidence. He cited studies showing atherosclerosis ("this piece was entirely missing and now it's here"), oxidative stress and DNA changes due to diesel particles (Knaapen et al., 2004; Borm et al., 2004) as well as new epidemiologic studies looking at cardiovascular disease and other outcomes. In fact, he suggested that he would probably have provided different numbers three years ago.

3.9 Thresholds

Expert L thought that in order to determine whether there is a threshold in the PM/mortality relationship, he would want to rely on the published epidemiological studies that include scatter plots of their data, that are inclusive of a wide range of PM, and those that try more formally to investigate the existence of a threshold. He thought that a formal investigation of threshold would be difficult because of the small effect sizes seen. He thought that this had been done to some extent in the time-series studies, with authors concluding that, "there is no evidence for any well-defined threshold." However, he thought that this type of assessment was more difficult in the long-term studies, although some had attempted it and had also come to the same conclusion as in the time series studies. "And if there is one, it must be at the very lower end of this range … based on the lack of evidence for a clear threshold, and the positive findings in [relatively] clean air studies." He thought it might be possible to assess the presence of a threshold if one were to focus on a susceptible population and follow them for long periods of time. Or he thought the threshold question might be addressed by repeated acute effect time-series approaches. "Just to repeat … every five years under the new pollution conditions, to see whether one still sees the signal and whether one sees it down to … very low exposures." He did not think toxicological or clinical study designs were likely to resolve the question given concerns about interspecies comparisons, high to low dose extrapolations, differences in susceptibilities, and interaction of effects of different pollutants.

When asked if his views on threshold differed for long-term and short-term exposures, Expert L thought that they were related, in that if "we are clear about the effects and shape of the response function in the very short-term domain, this answers part of the long-term question … if it is true that in the very short-term, there is no threshold, this would be true for the long-term public health impact to the extent that the latter is in part also the accumulation of acute effects."

He indicated that overall, he did not think there was evidence for a threshold, indicating that the scatter plots in the ACS study do not show a threshold across the range of exposures (about 6 up to 23 $\mu g/m^3$ for the later exposures (1999-2000) and about 10 to 30

μg/m^3 (1979-1983) for the earlier exposures). Nor did he think they are statistically detectable by current methods. Expert L did not elect to incorporate a threshold into his C-R function.

3.10 Other Influential Factors

Expert L discussed additional sources of uncertainty that were not part of the protocol. When asked whether he felt that there had been adequate exploration of alternative ways to model PM/mortality effects, he answered that the ACS and Six Cities studies had been analyzed and scrutinized extensively and that the overall conclusions of the main published estimates remained the same. Expert L did not think there were any other outstanding issues not already covered by the protocol.

PART 4. ELICITATION OF QUANTITATIVE JUDGMENTS

Expert L chose to characterize his C-R function within the specified range of annual average $PM_{2.5}$ concentrations (4-30 μg/m^3) in a piece-wise linear fashion. He specified one set of numbers that applied to concentrations of 4-10 μg/m^3 (hereafter, "Range 1"), and another set that applied to concentrations of >10-30 μg/m^3 (hereafter, "Range 2"). He did so to reflect his greater uncertainty about the shape of the C-R function in the lower concentration range.

Expert L elected to provide C-R functions conditional on the existence of a causal relationship. He thought that the discussion of causality should be separate from that of the C-R function itself. "I don't see how to integrate the probability of the causality with the probability of these estimates, because I honestly do not believe that this is a multiplication of probabilities. [There is] [t]he causality, and then these estimates." The elicitation team then combined his conditional distributions with his percent likelihood of causality specified in Section 3.8 (99 percent likelihood of a causal relationship). In addition, the two distributions were applied to a distribution of population-weighted annual average $PM_{2.5}$ concentrations in the U.S. from EPA's BenMap model to create a combined distribution (hereafter, "Example Applied Distribution").

His general approach was to characterize his uncertainty using the "best estimates" from a variety of studies, reflecting different strengths and weaknesses. He did not think the published statistical confidence intervals were as relevant to answering the elicitation question, given that they reflect precision of the data, choice of model, and covariates.

He began by specifying the percentiles for Range 2. He started with a minimum value of 0.2 percent per 10 μg/m^3, basing it on the NMMAPS paper (Samet et al., 2000) because "biologically, mechanistically, and epidemiologically, it is really clear that there must be more than just these immediate acute effects that you observe in two days … evidence is very strong that things happen in a longer-term way." His 5th percentile value, 2 percent per 10 μg/m^3, was based on a study by Schwartz et al. (2000), which used a distributed lag model up to 3 months**.** To determine his 25th percentile, he began with estimates from the cohort studies. He selected a 25th percentile value of 4 percent per 10 μg/m^3 based on the lower estimate from the Pope et al., 2002 ACS paper. He noted that the maximum is

a particularly difficult value to estimate. He ultimately chose a maximum of 30 percent per 10 μg/m^3 based on a rough average of the upper 95 percent confidence limits from several models published in the Jerrett et al. (2005) paper. For the 95th percentile value, he used the individual covariate-adjusted estimate from the Jerrett et al. (2005) paper, 17 percent per 10 μg/m^3. For the 75th percentile value, Expert L elected to use 15 percent per 10 μg/m^3 based on the estimate adjusted for social factors from the Jerrett et al. (2005) study as well as the Laden et al. (2006) extension of the Six Cities study. He then chose his 50th percentile value based on the fully adjusted estimate from Jerrett et al. (2005) as well as the fact that it fell between the other Jerrett et al. estimates, the Laden et al. estimates, and the Pope et al. estimates. For Range 1, Expert L specified the same percentiles, with the exception of the minimum, which he placed at zero to account for the uncertainty in the threshold at low levels. We discussed how published error bars tend to be greater around the lower and upper ends of a range, but Expert L felt that those were statistically driven and did not make sense for these estimates.

After viewing his distributions for Ranges 1 and 2 incorporating causality in Crystal Ball, he felt that the they did not accurately reflect his views about where the probability mass should be centered. For example, his interquartile range was between 4 and 15 percent per 10 μg/m^3, which would indicate that 50 percent of the time, the true mortality effect estimate should fall within that range. However, he felt that the probability weight assigned to that range should be greater than 50 percent, and opted to change the weight to 70 percent. Therefore, his original 25th and 75th percentiles became his 15th and 85th percentiles. The elicitation team then built a custom distribution in Crystal Ball in order to calculate his new interquartile range, displayed below.

Exhibit 1: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Range 1 (4 – 10 μg/m^3)

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	0	0
5th	0.20	0.16
25th	0.60	0.54
50th	1.0	0.98
75th	1.4	1.4
95th	1.6	1.6
Maximum	2.7	2.7

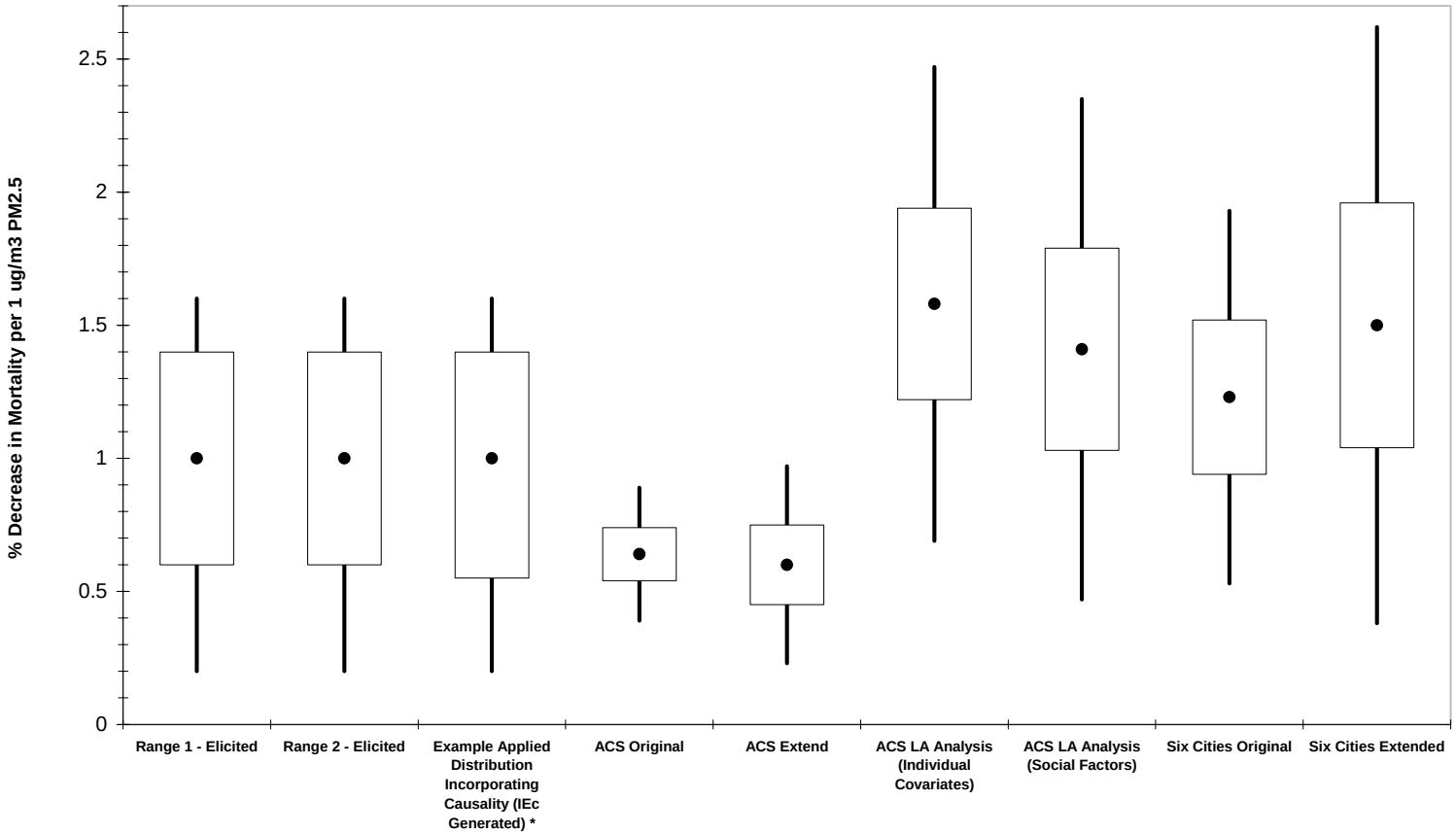
Exhibit 2: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1µg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations for Range 2 (>10 – 30 µg/m^3)

Percentile	Percent Change in Mortality Elicited Distribution	Percent Change in Mortality Distribution Incorporating Causal Likelihood (IEc Generated)
Minimum	0.02	0
5th	0.20	0.17
25th	0.60	0.56
50th	1.0	0.99
75th	1.4	1.4
95th	1.6	1.6
Maximum	2.7	2.7

Exhibit 3: Subjective Estimates of the Percent Change in Annual Mortality Associated with a 1µg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations After Incorporating Causality and Applying the C-R functions to the Population-Weighted Annual Average $PM_{2.5}$ Concentration Distribution in the U.S. from BenMap - Example Applied Distribution (IEc Generated)

Percentile	Percent Change in Mortality
Minimum	0
5th	0.20
25th	0.55
50th	1.0
75th	1.4
95th	1.6
Maximum	2.7

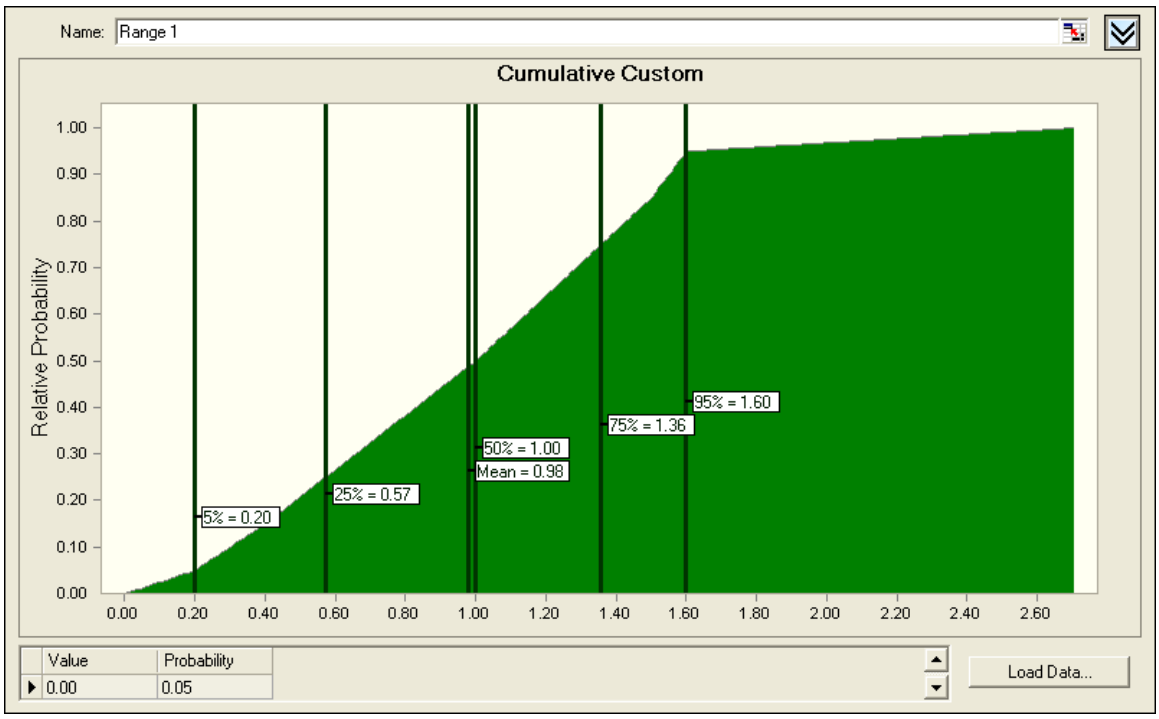
Exhibit 4: Percent Change in Annual Mortality Associated with a 1μg/m^3 Change in Ambient Annual Average $PM_{2.5}$ Concentrations: Comparison of 90 Percent Confidence Intervals for Various Studies to Distributions from Expert L



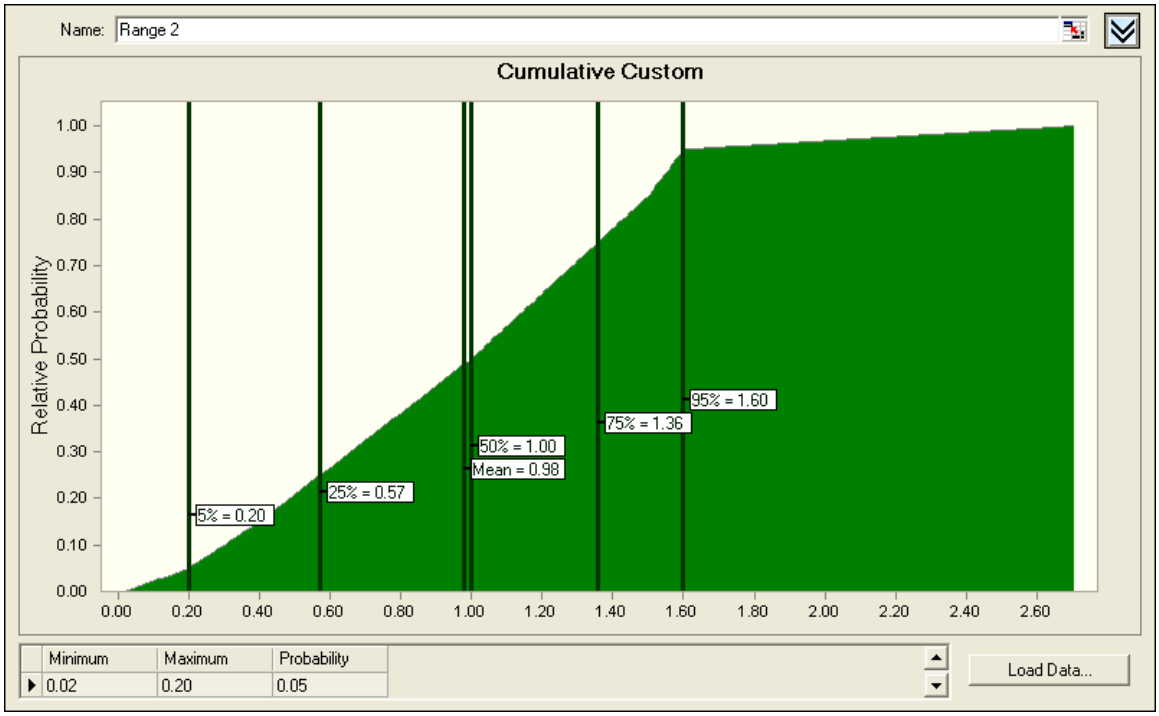
* = Distribution incorporating causality and applying the C-R functions from Ranges 1 and 2 to a 2002 population-weighted annual average $PM_{2.5}$ concentration distribution in the U.S. from BenMap.

• = median □ = interquartile range | = 90 percent confidence interval

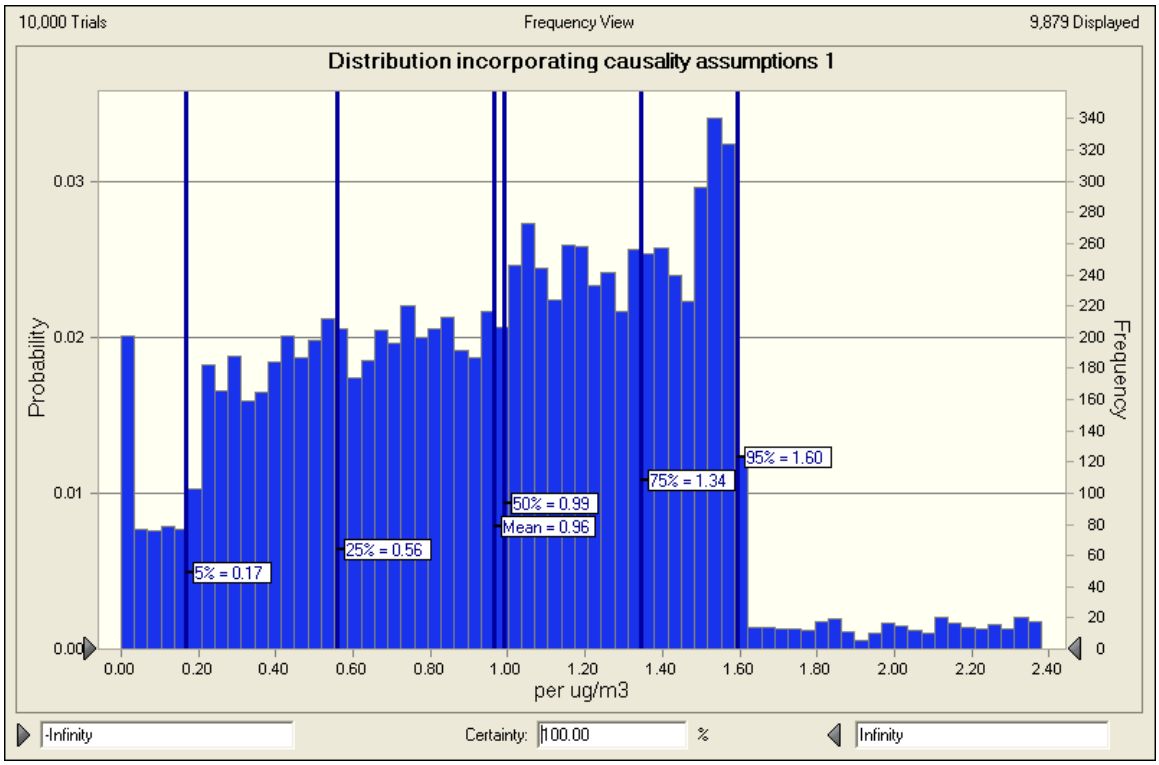
Elicited Distribution – Range 1



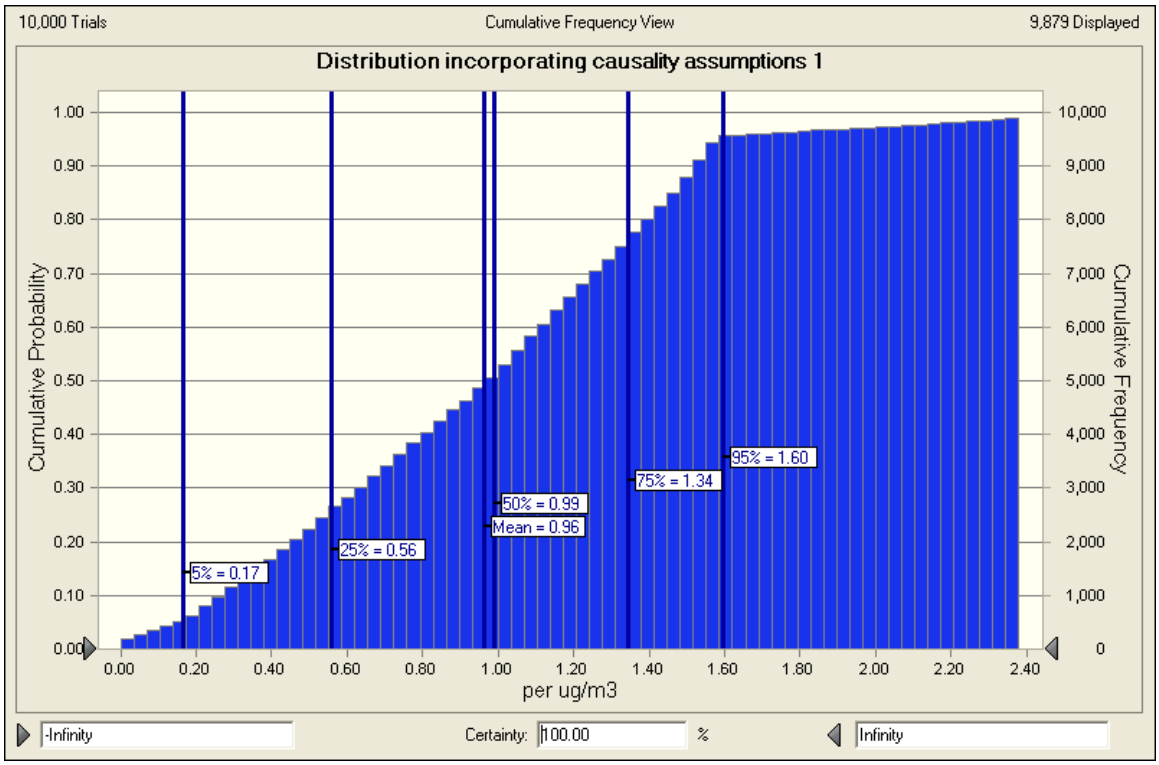
Elicited Distribution – Range 2



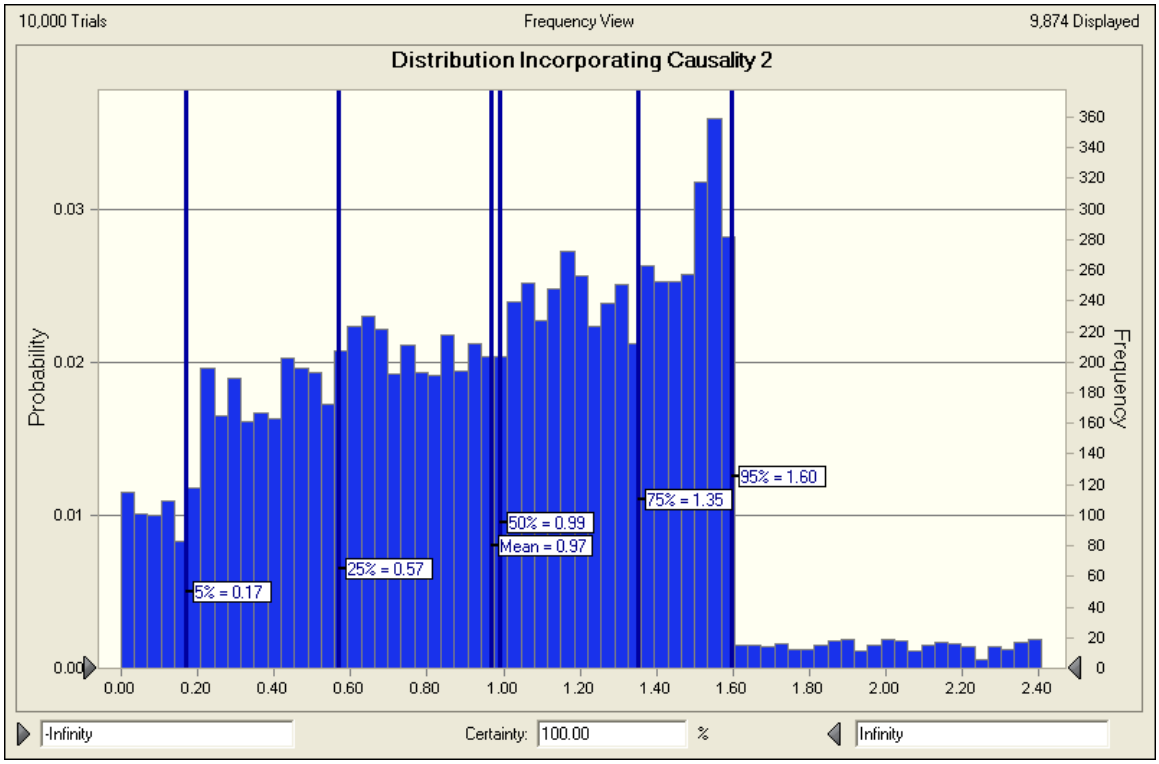
Range 1 Incorporating Causality - Probability Density Function (PDF) (IEc Generated)



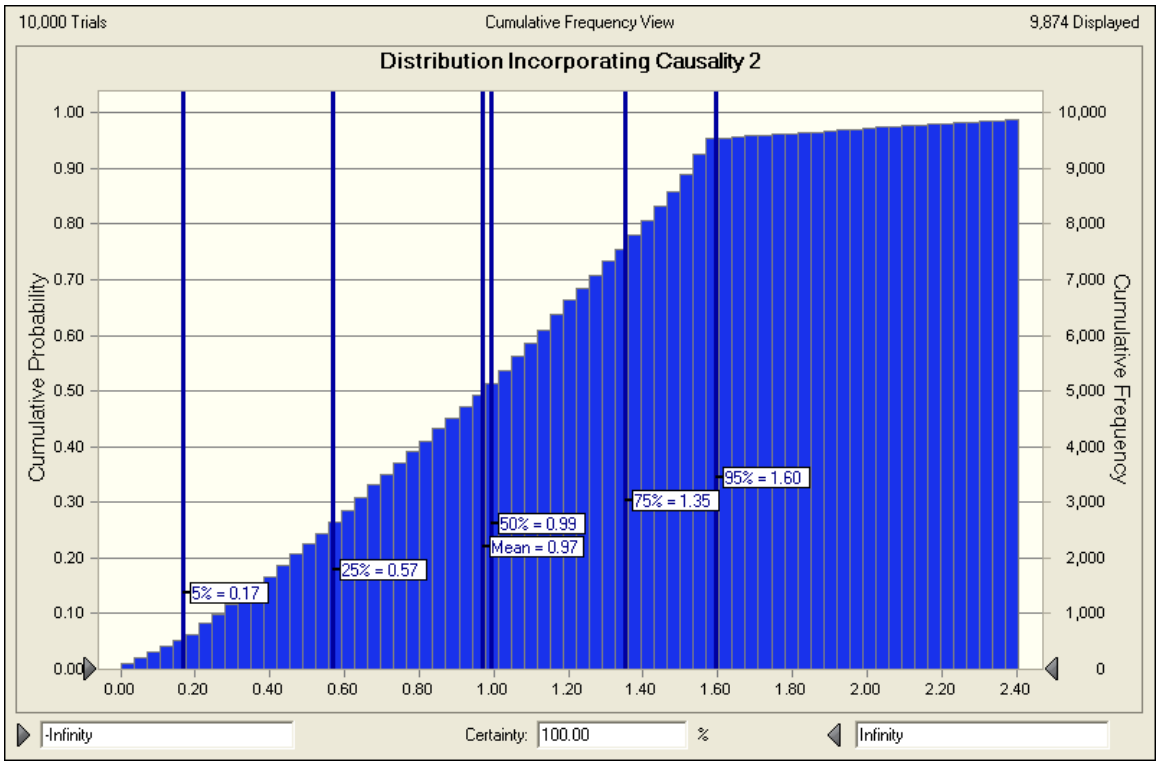
Range 1 Incorporating Causality - Cumulative Density Function (CDF) (IEc Generated)



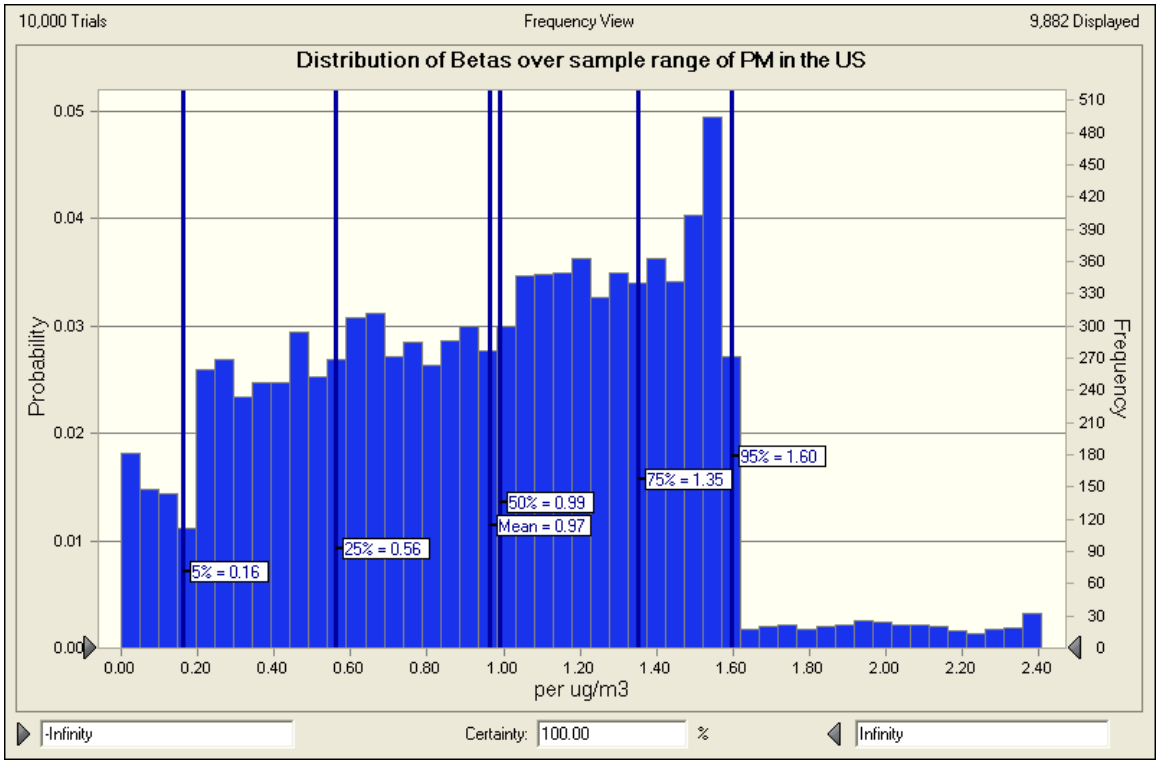
Range 2 Incorporating Causality – PDF (IEc Generated)



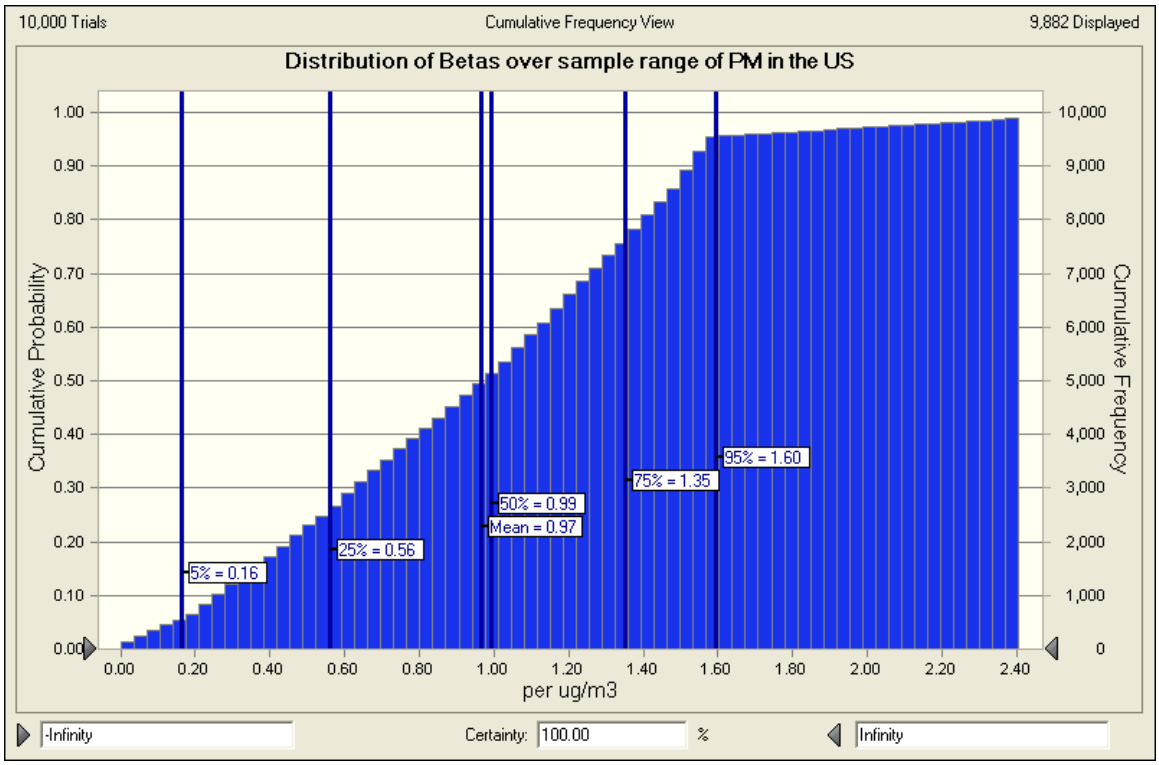
Range 2 Incorporating Causality – CDF (IEc Generated)



Example Applied Distribution – PDF (IEc Generated)



Example Applied Distribution – CDF (IEc Generated)



U.S. EPA EXPERT ELICITATION STUDY OF THE CONCENTRATION-RESPONSE RELATIONSHIP BETWEEN ANNUAL AVERAGE $PM_{2.5}$ EXPOSURE AND MORTALITY

Modification to Expert Judgments

Expert L

Date: July 7th, 2006

Section of Protocol Affected (Section Number and/or Title):

Causality

Description of Change (e.g. to a specific percentile, or to a qualitative opinion or statement of belief) (Causality versus Change of Mortality):

1. Regarding the range of probabilities of a CAUSAL RELATIONSHIP I would like to express two issues:

   In the range of >10 μg/m^3, my Min/Max values are 90 and 100 with the 'more likely' being 99 because we can never be "100 percent sure."

   In the range of 0-10 μg/m^3 my Min/Max range estimates are 0 and 95 with the 'most likely value' being 75.

   ARGUMENTS:

   It is very hard to me to find a plausible coherent alternative explanation for all the findings – at least in the acute effect domain – showing effects of PM in the range >10 μg/m^3. It is inherently more difficult to know what is going on in the lowest ranges. If the theory of 'distributions of susceptibilities' does apply (which I endorse), it means to have no threshold on the population level, thus effects among susceptible individuals are likely to occur in the 0-10 μg/m^3 range, too. Empirical evidence is just not abandoned to disregard the minute probability of 'no effect' in this lowest range (or maybe in the lowest range of this low range, e.g., 0-5 μg/m^3.

2. I ask for strict separation of the "causality distribution" and of the response function distribution (in the two concentration strata). It is important to not provide any figures or tables that combine "Percent Change in Mortality" with "Causality." What might be done are combinations of the overall distributions (percent change, number of cases etc.) with causality in the estimation/presentations of costs.

ARGUMENTS:

As discussed at the Post-elicitation Workshop, the science to establish causality has its own research domains, which may lead to a steady increase (as in case of air pollution) in favor of 'causality,' or the opposite could happen as well. The question of the quantitative association between an agent and an outcome needs an independent assessment. Evidence for causality may, e.g., increase while the updated literature may give no indication at all that concentration response functions changed – or if they do change in future studies it may be due to constituents or susceptibility changes rather than 'causality.'