OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

July 11, 2008

EPA-COUNCIL-08-002

The Honorable Stephen L. Johnson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Characterizing Uncertainty in Particulate Matter Benefits Using Expert Elicitation

Dear Administrator Johnson:

Prior to issuing the 2006 National Ambient Air Quality Standard (NAAQS) for particulate matter ($PM_{2.5}$), EPA's Office of Air Quality Planning and Standards (OAQPS) within the Office of Air and Radiation completed a multi-year effort to characterize the estimated benefits of reduced premature mortalities associated with exposures to $PM_{2.5}$. EPA used expert elicitation to quantitatively assess the relationship between exposures to $PM_{2.5}$ and the incidence of mortality, thus complementing and expanding the epidemiological literature on this subject and incorporating probabilistic uncertainty analysis. Completed in 2006, the $PM_{2.5}$-Mortality Expert Elicitation addressed the concentration-response function between $PM_{2.5}$ and mortality and provided probabilistic characterizations of uncertainty from 12 independent experts. The $PM_{2.5}$-Mortality Expert Elicitation received very favorable peer reviews (RTI International, *Peer Review of Expert Elicitation*, September 2006) and became the basis for assessing monetized benefits of the 2006 $PM_{2.5}$ NAAQS. EPA then asked the Science Advisory Board (SAB) Staff Office to convene an expert panel to review the application of the $PM_{2.5}$-Mortality Expert Elicitation results to the benefits assessment for $PM_{2.5}$. In particular, EPA asked for guidance on the interpretation of expert elicitation results and presentation in the Executive Summary, Press Release and Benefits Analysis chapter of EPA's Regulatory Impact Analysis associated with the 2006 $PM_{2.5}$ standard.

To conduct this review, the Advisory Council on Clean Air Compliance Analysis (Council) was augmented with noted experts in the health effects of air pollution and expert elicitation (see enclosed roster). The Council and invited experts met on May 8, 2008, to discuss charge questions from EPA. The Council's detailed advice and recommendations are provided in the enclosed Advisory with highlights below.

The Council endorses EPA's application of the expert elicitation results. The Council finds that EPA accurately characterized each expert's concentration-response function and expressed the uncertainty surrounding these functions in a technically sound manner. The probability distributions propagated for the experts' concentration-response functions were appropriately constructed and applied to estimate benefits. EPA thoroughly captured and expressed the breadth and diversity of opinion among experts and clearly differentiated between estimates based on empirical data (i.e., individual epidemiological studies) and those based on expert judgments (that are informed by epidemiological studies). The Council was asked whether EPA's benefits assessment responded to the National Research Council (NRC) recommendation to "move the assessment of uncertainties from its ancillary analyses into the primary analysis by conducting probabilistic, multiple-source uncertainty analyses." (NRC, *Estimating the Health-Risk-Reduction Benefits of Proposed Air Pollution Regulations*, 2002). Our answer is *yes*.

The Council was asked whether the Agency should move toward presenting a central estimate with uncertainty bounds or continue to provide separate estimates for each expert. The Council believes the answer to this question depends on the context of the expert elicitation and its results. On issues where experts have a wide range of opinions, it is important to provide separate estimates for each expert (or cluster of experts sharing similar views), thus emphasizing the uncertainty associated with the issue. But where experts largely agree, it would be appropriate to collapse the various estimates into a single distribution (or point estimate with uncertainty bounds) while still providing the individual estimates elsewhere, perhaps in an Appendix or website. In future analyses, the decision about aggregation must be made in the context of each analysis and its purpose.

On the critical side, the Council believes there is room for improvement in conveying the differences in assumptions (including the influence of key empirical studies) that drive the differences among experts' concentration-response functions. It would be useful to know why the experts agree on some things and disagree on others. The benefits chapter could be improved by devoting less space to the experts' quantitative judgments in exchange for more discussion to characterize their rationales. The text could better elucidate the relative importance of various sources of uncertainty: both those that were quantified and those that were not quantified. These issues could be addressed in the chapter and brought forward into the Executive Summary.

The Council has concerns about the Executive Summary and Press Release. The PM$_{2.5}$-Mortality Expert Elicitation showed a strong consensus among scientists, however the Executive Summary and Press Release failed to show this central mass of expert opinion. Instead, the Press Release presented the tails of the distribution, showing a range of $8 to $76 billion dollars in net benefits. Presented with this range, the casual reader could easily infer substantial differences in scientific opinion when, in fact, there was a pronounced central cluster of views on PM$_{2.5}$ mortality. To communicate with a wider audience, the Executive Summary and Press Release should have clearly stated that

scientific differences existed only with respect to the magnitude of the effect of $PM_{2.5}$ on mortality, not whether such an effect existed. The Executive Summary would have benefited from a short description of the $PM_{2.5}$-Mortality Expert Elicitation and the rationale for its use in the context of the $PM_{2.5}$ regulatory process. Additional efforts are needed to identify the most effective means of communicating both methods and results to different kinds of readers. This is of particular importance for the Executive Summary and Press Release, which are much more likely to be read in their entirety. The Council suggests that alternative and less complex graphics would provide much more useful information than the tables that are included in the Executive Summary.

Detailed recommendations are included in the enclosed Advisory. On behalf of the entire Council, we appreciate this opportunity to provide timely advice to the Agency. We hope these comments are helpful to EPA as it proceeds with this important work.

Sincerely,

/Signed/

James K. Hammitt, Chair
Advisory Council on Clean Air Compliance
Analysis

Enclosures

U.S. Environmental Protection Agency
Advisory Council on Clean Air Compliance Analysis Augmented for Benefits of Reduced PM-Mortality using Expert Elicitation

CHAIR

Dr. James K. Hammitt, Professor, Center for Risk Analysis, Harvard University, Boston, MA.

COUNCIL MEMBERS

Dr. David T. Allen, Gertz Regents Professor in Chemical Engineering, Department of Chemical Engineering, University of Texas, Austin, TX.

Dr. Dallas Burtraw, Senior Fellow, Resources for the Future, Washington, DC.

Dr. Shelby Gerking, Professor of Economics, Department of Economics, College of Business Administration, University of Central Florida, Orlando, FL.

Dr. Wayne Gray, Professor, Department of Economics, Clark University, Worcester, MA.

Dr. F. Reed Johnson, Senior Fellow and Principal Economist, RTI Health Solutions, Research Triangle Institute, NC.

Dr. Katherine Kiel, Associate Professor, Department of Economics, College of the Holy Cross, One College Street , Worcester, MA.

Dr. Virginia McConnell, Senior Fellow and Professor of Economics, Resources for the Future, Washington, DC.

Dr. David Popp, Associate Professor of Public Administration, Center for Policy Research, The Maxwell School, Syracuse University, Syracuse, NY.

Dr. Chris Walcek, Senior Research Scientist, Atmospheric Sciences Research Center, State University of New York, Albany, NY.

HEALTH EFFECTS SUBCOMMITTEE MEMBERS

Mr. J. Fintan Hurley, Scientific Director, Institute of Occupational Medicine (IOM),

Edinburgh, Scotland, United Kingdom.

Dr. Michael T. Kleinman, Professor, Department of Community & Environmental Medicine, University of California, Irvine, CA.

Dr. Rebecca Parkin, Professor of Environmental and Occupational Health and Associate Dean for Research and Public Health Practice, School of Public Health and Health Services, George Washington University Medical Center, Washington, DC.

INVITED EXPERTS

Dr. Aaron Cohen, Principal Scientist, Health Effects Institute, Charlestown Navy Yard, Boston, MA.

Dr. John Evans, Senior Lecturer on Environmental Science, Harvard University, Kuwait Public Health Project, Portsmouth, NH.

Dr. H. Christopher Frey, Professor, Civil, Construction and Environmental Engineering, College of Engineering, North Carolina State University, Raleigh, NC.

Dr. Ronald Wyzga, Technical Executive, Air Quality Health and Risk, Electric Power Research Institute, Palo Alto, CA.

SCIENCE ADVISORY BOARD STAFF

Dr. Holly Stallworth, Designated Federal Officer, Science Advisory Board Staff Office, Environmental Protection Agency, Washington, DC.

NOTICE

This report has been written as part of the activities of the U.S. Environmental Protection Agency's Advisory Council on Clean Air Compliance Analysis (Council), a federal advisory committee administratively located under the EPA Science Advisory Board (SAB) Staff Office. The Council is chartered to provide extramural scientific information and advice to the Administrator and other officials of the EPA. The Council is structured to provide balanced, expert assessment of scientific matters related to issue and problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the EPA, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use. Council reports are posted on the SAB Web site at: http://www.epa.gov/sab.

Advisory Council on Clean Air Compliance Analysis
Advisory on Characterizing Uncertainty in Particulate Matter Benefits Using Expert Elicitation

1. *In the PM NAAQS benefits chapter, has EPA accurately characterized each expert's concentration-response function as expressed in the PM-Mortality Expert Elicitation report and conveyed the differences in assumptions (including the influence of key empirical studies) that drive the differences among the concentration-response functions?*

In the benefits chapter (Chapter 5) of the particulate matter (PM) National Ambient Air Quality Standard (NAAQS) regulatory impact analysis (RIA), EPA has accurately described the experts' concentration-response (C-R) functions in general terms and has clearly summarized the implications of each expert's C-R function for the expected reduction in fatalities (and their monetary valuation) in Figs. 5-10 – 5-13. The benefits chapter does not report each expert's C-R function nor describe the factors (such as differences in assumptions and reliance on particular studies) that drive the differences among the C-R functions in the chapter. Extensive description of the individual experts' C-R functions as well as their perspectives, rationales, and reliance on empirical studies in formulating their judgments are described in the original reports of the expert elicitation study, including *The Expanded Expert Judgment Assessment of the Concentration-Response Relationship between PM$_{2.5}$ Exposure and Mortality* (Industrial Economics, 2006) and its technical support documents. The Council believes that it would be useful if Chapter 5 provided some discussion of the primary studies on which the experts relied and of the factors that drive differences among their responses, though we are sensitive to concerns that the regulatory analysis should not be overly long and complex.

2. *In applying the PM-Mortality Expert Elicitation results in EPA's benefit analysis, is our mathematical treatment of concepts such as the probability of causality, thresholds, and shape of the function technically sound, as well as transparent?*

The mathematical treatment of concepts such as the probability of causality, thresholds, and shape of the function is technically sound and transparent. For each expert, EPA combined the expert's quantitative assessment of his beliefs about annual average PM$_{2.5}$ and mortality hazard into an unconditional distribution of that expert's views of how mortality hazards change per unit change in annual average PM$_{2.5}$, at different baseline levels of that annual average in the range 4 – 30 µg/m^3. For experts who expressed a non-zero probability of a threshold, EPA made appropriate assumptions about how that probability was distributed within ranges of annual average PM$_{2.5}$ concentrations.

Some Council members expressed concern about how the derived unconditional distribution for each expert is used to produce estimates of mortality impacts. The PM NAAQS benefits chapter uses the C-R functions to estimate annual "attributable deaths,"

using an apparently simple static methodology. It is increasingly recognized (i) that there are difficulties underlying this concept, including that the estimated annual deaths do not reproduce year-on-year; and (ii) that these difficulties can be overcome by use of life tables, which also allow benefits to be expressed as gains in life expectancy (see, e.g., Rabl 2003, 2006). The Council encourages EPA to identify this conceptual issue in the use of both the expert elicitation and the cohort study results for benefits analysis. We believe both types of C-R functions could be used in life-table calculations. (Indeed, such calculations were apparently reported in an Appendix to the RIA, though not in Chapter 5). (We note that EPA suggests that measures of the gain in life expectancy may provide a "theoretically preferred" method to value changes in mortality risk (p. 5-56), but it does not discuss the assumptions used to estimate attributable deaths or attempt to quantify the uncertainties in impacts and valuation estimates that result.)

3. *Do the tables, text, conclusions, and Executive Summary adequately distinguish the benefit estimates based on data-derived components of the uncertainty assessment from those based on expert judgment? How should the mortality estimates based on the elicitation be compared to those derived from the empirical studies of the PM-mortality association?*
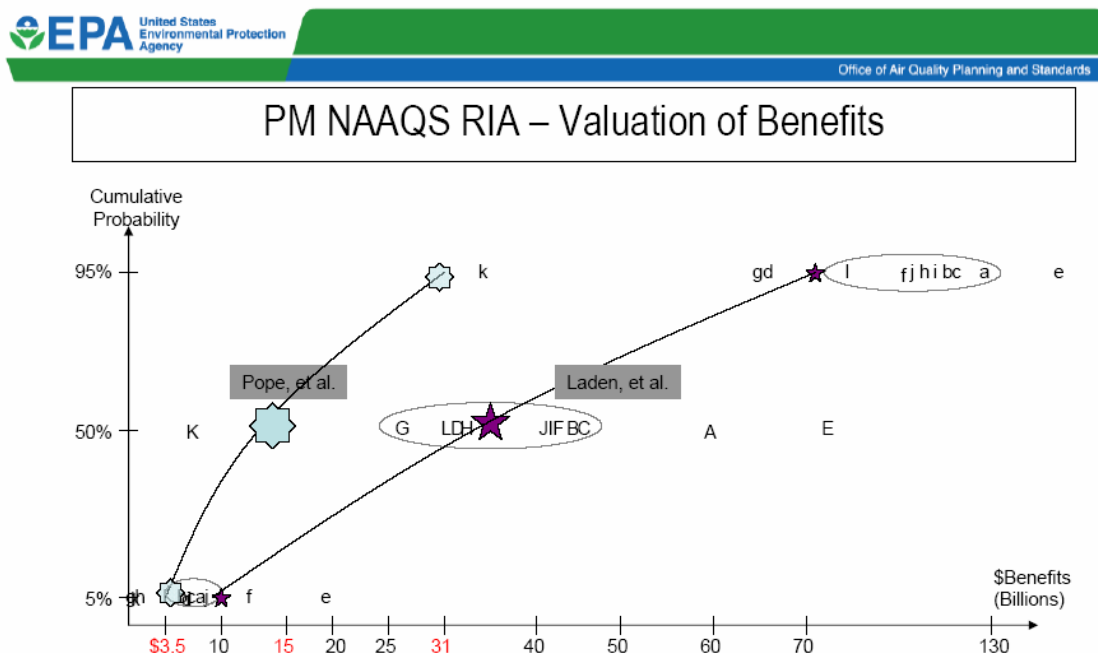
The tables, text, and conclusions of Chapter 5 clearly distinguish the benefit estimates based on direct application of C-R functions from epidemiological studies (so-called "data-derived components") from those based on the C-R functions elicited from the experts (that are, of course, informed by epidemiological studies and other data). Overall, the Council agrees that separately identifying the estimates based on their sources, and reporting estimates based on multiple relevant epidemiological studies and on each expert's C-R function, is a useful and appropriate method for accurately portraying the uncertainty about the effects of $PM_{2.5}$ on mortality (whether reported as number of premature deaths averted or as a monetary value).

It should be noted that the estimates derived from primary epidemiological studies and from expert elicitation are not fully comparable. The epidemiological studies cited are cohort studies used to estimate the longer-term influences of mortality; the expert elicitation addressed total mortality changes that could be associated "with a reduction in annual average PM2.5 including both changes in short–term (e.g., 24 hour) and long-term exposures to PM2.5."

In addition the two epidemiological studies cited are among several that were considered by the experts; it would be useful to present the range of estimates from several of the other epidemiological studies (see Exhibit 3-3 in the Report) considered to see whether a more comprehensive consideration of these studies yields as much variation in results as the expert opinions. This could provide greater insight into the variation in expert elicitation results.

The Council believes that graphical representations, such as the box-and-whiskers plots in Figs. 5-12 and 5-13 in Chapter 5, provide a clear and concise method to represent this

information. These figures convey information about the likelihood of different ranges of values as predicted from each C-R function (not simply a single range) and about the degree of clustering and overlap among the different C-R functions (i.e., from individual epidemiological studies and expert's judgments). The distribution functions presented in Figs. 5-14 and 5-15 provide slightly more information, but most Council members find them less informative, perhaps because the functions tend to stack on top of each other. The Council was enthusiastic about the graphic shown below that shows cumulative distributions of benefits calculated using two epidemiological studies and selected fractiles of each expert's C-R function, all clearly distinguished by using distinctive symbols. This graphic was presented to the Council but not included in the RIA benefits chapter itself. In contrast, the tables included in the chapter and Executive Summary permit only an impoverished representation of the degrees of certainty and uncertainty.

Graphic taken from "Characterizing the Uncertainty in Estimated Benefits of Reduced PM-Mortality Using Expert Elicitation," presentation by Lisa Conner, Bryan Hubbell and Harvey Richmond, Office of Air Quality Planning and Standards, Advisory Council on Clean Air Compliance Analysis meeting, May 8, 2008.

Recognizing that standard box-and-whiskers plots such as Figs. 5-12 and 5-13 are probably more complex than appropriate for the Executive Summary, the Council suggests that alternative and less complex graphics, such as the figure above, would still provide much more useful information than the tables that are included in the Executive Summary. Another option is a simplified box-and-whiskers plot including only the mean or median and the 5th and 95th percentile values for each epidemiological study and each expert.

In attempting to summarize the rich information about uncertainty it has developed, EPA evidently had difficulty choosing terminology. Table 5-1 and Table ES-3 present the same material, but the concepts that are labeled "Lower Bound Expert Result" and "Upper Bound Expert Result" in Table 5-1 are labeled "Low Mean" and "High Mean" in Table ES-3. These concepts are not defined or explained in either location. Moreover, the ranges for these concepts are not properly described. In Table ES-3, they are not even labeled. In Table 5-1, they are improperly labeled as "confidence intervals" rather than "credibility intervals." The concepts of confidence and credibility intervals are distinct and have different interpretations. A 90 percent confidence interval is a statistic (i.e., a random variable) constructed from data using a procedure such that the probability that the interval includes the true value is 90 percent (conditional on the model assumptions). A 90 percent credibility interval is an interval chosen by an expert who believes there is a 90 percent chance that the true value is in the interval (conditional on whatever assumptions he may specify).

4. *Does the EPA's present effort to incorporate uncertainty analyses and discussions into the primary analysis, as exemplified in the PM NAAQS RIA chapter, adequately address the NRC's request to move the assessment of uncertainties into its primary analyses? If not, what more could the EPA do to satisfy this request?*

The short answer to the first question is "yes." To understand why this is so, it is important to reflect on the way uncertainty was being addressed by the EPA at the time of 2002 NRC report. At that time benefits estimates for particulate air pollution would typically present a base-case analysis which relied on the Pope et al. American Cancer Society (ACS) study, a sensitivity analysis which provided upper estimates of effect drawing on the Dockery et al. Six Cities study, and lower estimates from the time-series literature, and these would be accompanied by a qualitative discussion of the issues related to drawing causal inferences from this literature, especially the cohort studies.

The NRC was unsatisfied with this form of presentation because it left unresolved the important question of determining how much weight to assign to these various alternative estimates. If the cohort studies did not reflect causation, the effect estimates would need to be based on the results of time series studies which are roughly a factor of 10 smaller than the ACS Study. And if they did reflect causal associations, then the relative plausibility of the coefficients from the Six Cities Study (that were about 3 times larger than those of the American Cancer Society Study) was left unspecified.

The NRC committee was of the view that most users of EPA regulatory analyses (i.e., regulators, Congressmen, the general public) were not in as good a position to evaluate these questions as scientific experts in epidemiology and toxicology would be and so it recommended that EPA explore the possibility of eliciting scientific opinion, using formal methods for probabilistic expert judgment, as a means of addressing this concern.

EPA's current effort reflects a careful attempt to do just that. EPA's analysis of uncertainty (in dose-response coefficients for $PM_{2.5}$) is integrative, quantitative, and

4

central. The analysis is integrative in the sense that it deals with all sources of uncertainty – both aleatory and epistemic – affecting PM dose-response functions. Aleatory uncertainty is the inherent variation associated with the physical system or the environment, sometimes referred to as stochastic uncertainty or irreducible uncertainty. Epistemic uncertainty stems from a lack of knowledge of quantities or processes of the system or the environment, sometimes referred to as model uncertainty or reducible uncertainty. In this way it differs from meta-analysis, which would be valuable if the only questions were about the magnitude of these relationships and not about the strength of evidence for causal interpretation of the epidemiologic studies. The analysis is quantitative in the sense that it provides probabilistic statements about the relative plausibility of alternative interpretations of the evidence – both about the relative strengths of various studies and also about the likelihood that these study results are artifacts of confounding or have little biological support. In addition, because the analysis presents separately the quantitative interpretations of 12 experts, it provides the user with a sense of the extent of scientific consensus among these experts. The analysis is central to the EPA's document in that these probabilistic characterizations of uncertainty are presented in the body of the RIA and not relegated to technical appendices or supporting documents.

While the Council commends EPA for this work that clearly responds to the NRC recommendations, we believe there are ways in which future efforts could be strengthened. The Council understands that there are limitations to any approach, including formal elicitation of expert judgment, to quantitatively characterize the nature and strength of scientific understanding of quantities (such as concentration-response slopes) relevant for environmental decision making. Among these are:

- Selection of experts – The first question in any effort to interpret ambiguous or conflicting scientific information is to determine which scientists to consult. The EPA's analysis relies on the views of 12 experts in epidemiology and toxicology. The fact that the EPA is open and transparent about who these experts are and how they were selected invites questions about whether the group was representative, whether the sample was a probability sample, whether the group was balanced (with regard to discipline, institutional affiliation, or other factors), and so on. These are certainly important questions. But it is necessary to recognize that any effort to resolve questions about the extent of epistemic uncertainty (which often is the dominant source of uncertainty) must rely on the interpretations of scientists and therefore involve these same issues of which scientists, how chosen, whether representative, how balanced, and so on. Thus, the question could have been asked of previous EPA regulatory analyses that also relied on professional judgment, but without the transparency of a formal expert elicitation.

- Aggregation of expert opinion – The EPA has chosen to first present separately the views of each of the 12 scientists who participated in their expert elicitation. This is entirely consistent with "best practices" in the field. But because of concerns about the scientific legitimacy of any approach for

aggregating expert opinion, the EPA has said that it declined to aggregate and it does not present any aggregate estimate of the central tendency of expert opinions. However, EPA does present a range bounded by the mean estimates of the experts with the lowest and highest mean estimates. The Council notes that this is, in fact, a form of aggregation that assigns positive weight to the most extreme judgments and zero weight to all the others (or perhaps suggests a uniform distribution between these extreme values). The Council feels this is not the best aggregation and recommends that the EPA consider other forms of weighting. e.g., assigning equal weight to each expert's distribution or assigning weights based on other approaches, such as peer weighting, self weighting, or performance on calibration questions. See the Council's discussion in Question 6b.

- Limits of rationality – All judgments are subject to well-known cognitive anomalies such as sensitivity to framing, anchoring, probability weighting, etc. Even trained experts such as scientists and physicians are not immune to such effects. To protect against these difficulties, this well-designed expert elicitation includes checks on consistency and logic and provide experts with an opportunity to reconsider and revise their evaluations.

- Costs – The EPA did not provide the Council with estimates of the costs of conducting this expert elicitation, but there is a sense that the costs were "high" (without giving a magnitude or any real comparison). This is a clear case, however, where the benefits of the expert elicitation for understanding and resolving the large differences in estimated PM regulatory benefits were even larger. There seems to be a consensus that the study provided benefits well beyond its costs for analysis of the effects of the PM regulation. However, it may be that similar expert elicitation efforts would not be appropriate for all RIAs. Estimates of the cost of this study, and any lessons learned about ways to reduce costs of future expert elicitation studies, would be useful for future regulatory analysis.

The Council recommends that EPA develop criteria for determining when systematic polling of scientific judgments would enhance the regulatory analysis, usefully inform decision making, and justify the associated analytical costs. The Council understands that the Agency is developing guidance on the use of expert judgment and encourages EPA to consider this topic if it is not already doing so.

The primary area in which the EPA's effort has not been responsive to the NRC report is in its explanation of the rationale behind the experts' judgments. Understanding why experts disagree about the implications of the available evidence may be as important as their specific judgments of the probability distribution for the $PM_{2.5}$ C-R function itself. Much information of this type was developed and reported in the final report (IEc, 2006) and other documents describing the expert judgment study, but it is not evident how, if at all, this information was reflected in the PM NAAQS benefits analysis.

Expert judgments involve intuitive weighting and interpretation of existing evidence. Thus in some ways expert elicitation is similar to meta-analysis, which also serves to integrate and synthesize a body of evidence. Identifying the specific data, decision weights, adjustments, and interpretations each expert employed could help inform discussions among a wider group of experts and stakeholders, foster explicit evaluation of the merits of experts' subjective criteria, and suggest opportunities for aggregating or clustering judgments across experts.

Some Council members believe it would be useful to analyze how, if at all, experts' judgments are correlated with factors such as: field of expertise (epidemiologists vs. toxicologists/clinicians); authorship of primary epidemiological studies (do authors put more emphasis on their own work?); and institution where the expert resides (the elicitation includes three experts from one institution and two from another).

The Council recommends that EPA evaluate the qualitative information collected during the elicitation, the post-elicitation workshop, and other interactions with the experts. This analysis could help decision makers understand the divergence of expert opinion, identify particular gaps or deficiencies in the evidence that experts believe contribute to uncertainty, and identify fruitful avenues for future research.

In summary, EPA's use of expert elicitation satisfies the NRC's request and represents a state-of-the-art example of expert elicitation methods. The benefits chapter serves as an excellent proof of concept for quantifying uncertainty in regulatory analysis. In this particular instance, the results serve to increase decision makers' and the public's confidence that the health benefits of $PM_{2.5}$ controls exceed costs by a comfortable margin. This is largely because, as a group, the experts have great confidence that the epidemiological studies upon which the EPA has relied reflect causal relationships between exposure and mortality (i.e., the experts place little weight on non-causal interpretations) and because they emphasize the relevance and validity of the cohort studies for answering the questions of interest to the EPA.

Despite our strong support of this analysis, the Council urges EPA to anticipate challenges to expert elicitation when it is used in more controversial applications. It is reasonable to expect that EPA will be required to defend the process used for expert selection. But as noted above, this challenge should apply to any effort to use expert opinion, whether through formal elicitation or informal consultation, in support of regulation.

5. *Has the EPA adequately communicated the uncertainty information associated with the PM premature mortality estimate to the audiences that the RIA addresses, including: scientists, policy analysts, decision makers, and the public?*

Not yet. The Council appreciates that the Executive Summary, and especially the benefits chapter, present the quantitative results in detail using diverse tabular and especially graphical approaches. However, we raise general concerns related to: 1) methods of

presentation most appropriate for the RIA's diverse readership; and 2) the proper metric(s) for characterizing the results of the elicitation with regard to the distribution of the experts' subjective probabilities and their effect on the health impact and economic valuation estimates.

The Council stresses that addressing these concerns will have important benefits for all forms of communication about the expert elicitation including the RIA, the Executive Summary, and the Press Release.

> **a.** *Considering the examples provided by the EPA, are there other methods the EPA should use, instead of or in addition to those employed, to summarize and communicate the results of the PM-Mortality Expert Elicitation in the benefits chapter and the Executive Summary for communication to technical and non-technical audiences?*
>
> Yes. The Council notes that the intended readership of the RIA is diverse, and we appreciate that EPA explored a range of approaches to presenting the results of the elicitation. Nonetheless, we believe that additional efforts are needed to identify the most effective means of communicating both methods and results to different kinds of readers. This is of particular importance for the Executive Summary and Press Release, which are much more likely to be read in their entirety by most readers. Specific suggestions include:
>
> - Provide in the Executive Summary a description of the elicitation with regard to the rationale for its use, what it comprised, and how it was conducted. Figure 5-1 could be useful in this regard.
>
> - Make more extensive use of graphical displays in the Executive Summary, rather than (or in addition to) tables.
>
> - Add some indication of the "bottom line" in the Press Release. For example, language could be added to state that disagreements among experts are limited to the <u>magnitude</u> of health benefits associated with $PM_{2.5}$ reductions, not whether those benefits exist. See also the Council's comments on reflecting central tendency in Questions 5b, 5c and 6b.
>
> **b.** *To what extent do the types of statements made in the Executive Summary of the PM NAAQS RIA successfully communicate the extent of uncertainty (and/or the certainty) in the estimate of PM premature mortality to those who are not familiar with the PM-Mortality Expert Elicitation?*
>
> As discussed above (in response to Question 3), the Council questions the use of ranges to characterize the uncertainty in impact and valuation estimates in the Executive Summary and its tables. Panelists note that the range appears to imply that any value within it enjoys an equal degree of support from the

experts while more detailed descriptions of the results show that the experts' judgments are more clustered. Some indication of the clustering in the elicitation results should be considered (see also Question 5c, point 1 below).

The Council notes two sources of uncertainty not explicitly addressed in either the chapter or the Executive Summary:

- The methods and criteria used to select the experts are neither presented in detail nor critically evaluated. Panelists note that this stage of the elicitation process is critical to ensuring that the panel of experts adequately represents expert opinion about the effect of PM on mortality. (As noted in Question 4, expert selection is also critical to alternative approaches such as consensus panels, but may be perceived as more salient for expert elicitation, perhaps because individual expert's distributions are reported.)

- As noted above (Question 2), estimation from cohort study data of annual numbers of attributable deaths, as opposed to measures of longevity (e.g., years of life lost), is problematic.

c. *Are there additional summary statements that are important to deduce from the results of the PM NAAQS benefits chapter to the Executive Summary?*

Yes. Panelists noted that the chapter lacks:

- A comprehensive statement of the "bottom line" with regard to the expert elicitation results, e.g., that it supports the conclusion that the benefits of $PM_{2.5}$ control are very likely to be substantial; and

- An integrated discussion of the relative importance of various sources of uncertainty: those that were not quantified (e.g., relative toxicity of PM sources/constituents) versus those that were, as well as the relative importance of the various uncertainties that were quantified (e.g., uncertainties in the C-R function vs. the valuation).

Table 5-5 identifies seven primary sources of uncertainty that are included in the RIA, however this recognition of multiple uncertainties does not permeate the rest of the chapter. It would be useful to acknowledge uncertainties at each stage of the analytic process and thus where the range of possible values increases. It might be helpful to have a chart that outlines each step, assesses the degree of uncertainty at that step, and reports how it is handled. This would allow the reader to understand how the final range of numbers reported depends on the various steps in the analysis, and to more easily see which uncertainties contribute most to the overall uncertainty. This information should also be summarized the Executive Summary.

6. *Has the EPA adequately summarized the results of the PM-Mortality Expert Elicitation across the experts in the PM NAAQS RIA benefits chapter and executive summary?*

The results are presented in summary form in terms of mean values and 90 percent confidence intervals. (As noted in response to Question 3, the latter should be referred to as 90 percent credibility intervals since they represent a judgment about uncertainty and not an inference from the sampling distribution of a statistic.)

The Executive Summary is adequate in terms of conveying the central tendency and range of estimates based on the experts' C-R functions. However, it could be made clearer that the results shown are not the actual judgments of the experts – i.e., the experts did not make judgments regarding the avoided premature mortality. Rather, the avoided premature mortality was estimated using the C-R function elicited from each expert. Similarly, the benefits assessment should not be attributed to the experts, but it should be made clear that the expert judgment was simply one of many inputs to the benefits assessment. This could also be made clearer in many of the tables (e.g., Table 5-32).

What is not apparent is why the expert judgments differ. For example, Figure 5-10 illustrates substantial inter-expert variability in results. Yet, the significance of this variability to the conclusions of the benefits assessment seems not to be addressed. Moreover, it appears that the experts could be grouped into clusters, such as a low cluster (Experts G, K), central cluster, and high cluster (Experts A, E, and perhaps B and C). It would be useful to know more about why the experts agree within clusters, and why they disagree between clusters. For example, do experts within a cluster tend to rely more heavily on a particular study than do experts in other clusters? If so, why do members of different clusters put more weight on different studies? Are there comments from the post-elicitation interviews that shed light on why the experts continue to disagree even after seeing each other's judgments? For example, do the experts differ with respect to which data sets they deem to be most representative or useful or regarding inference methods (e.g., biological plausibility, statistical power of empirically-based models)?

In the PM NAAQS benefits chapter, the EPA presents the mortality results based on each of the twelve individual expert's responses along with results based on concentration-response functions derived from empirical studies. The EPA has also considered employing methods to aggregate results based on the elicitation into a single combined estimate. In particular, the EPA considered calculating a simple average of estimates across experts after the concentration-response functions of each expert had been applied in the benefits model (i.e., the average of the resulting estimation of the change in mortality incidence). Other options for summarizing the results include: a weighted average of the resulting change in incidence, a trimmed

means approach, and a fitted distribution to the overall set of concentration-response functions.

    a. *Should the EPA continue to present the results of the individual experts in future benefits analyses as was done in the PM NAAQS RIA? Should the EPA develop metrics that aggregate across the individual experts? If aggregate measures are considered appropriate, should the EPA present these in addition to or instead of the individual estimates?*

The Council recommends that EPA continue to present the results of the individual experts in future benefits analyses, whether or not an aggregate or combined distribution is presented.

Results from an expert elicitation can be used for several purposes, and whether aggregation or combination across experts is useful depends on the intended purpose. If the goal is to characterize the uncertainty about an outcome (e.g., the benefits of controlling $PM_{2.5}$), a presentation that shows the distribution of estimated benefits conditional on each expert's C-R function (such as Figs. 5-10 – 5-11) is extremely useful, as it shows the implications of each expert's uncertainty and the diversity of judgments among experts. However, such a rich presentation of uncertainty is excessive for many purposes, and will inevitably be collapsed by some sort of aggregation method, whether by EPA or by others (e.g., news media).

The Council considered a number of approaches to aggregation but judges that none are ideal inasmuch as the appropriate form of aggregation may depend on the purposes of the elicitation and its results. For accurately portraying the range of expert opinion, reporting results using each expert's judgments may be most useful. For evaluating alternative regulatory options, a decision-analytic perspective suggests it may be most useful to combine the experts' distributions into a single distribution, though the best method for doing so is unclear. When the results suggest that many of the experts' judgments fall into one or a few clusters, it may be important to identify and describe those clusters and to separately describe any significant outliers. Alternatively, if the experts' judgments are approximately uniformly distributed across a range, a statement of this result may be most useful.

    b. *If a combination (aggregation) of results is considered appropriate, what technique for aggregation would you recommend?*

The basic goal is to explain both the range and distribution of judgments within it. There does not seem to be a perfect or single best technique, as this will depend on the purpose of the aggregation and the data. A virtue of formal decision analysis is that it provides a rigorous and theoretically justified method for mathematically combining information about uncertainty and

preferences. Alternative approaches that rely on a decision maker to holistically weigh multiple factors in his or her head are susceptible to cognitive limitations such as the heuristics and biases identified by Tversky and Kahneman (1974) and the tendency to overweight those factors that appear especially salient while neglecting others.

EPA could consider developing an operational approach to describing the distribution of experts' judgments, while acknowledging that there is no single best approach and that the choice of approach is a matter of judgment and context. The Council considered several examples of possible approaches:

- **Approach 1:** Present a range of experts' median or mean values. The range could be defined by fractiles (e.g., the interquartile range) or as a trimmed estimate by excluding judgments of the p highest and p lowest values. This has the advantage of being easy to explain and does not strongly imply a probabilistic interpretation, which is appropriate since the experts are not a probability sample. A limitation is that it does not take into account the range of uncertainty elicited from each expert or information from the experts with the highest and lowest central values.
- **Approach 2**: Similar to Approach 1 but instead of using a range of central values, use a range enclosed by ranking the 5th percentiles of the experts with the lowest such values and the 95th percentiles of the experts with the highest such values, and reporting some summary of these ranked extreme values (e.g., the mean or median 5th percentile; the smallest 5th percentile). This approach incorporates some information about the variation in both the locations and widths of the experts' distributions. It emphasizes the range of opinion but provides no information about any clustering within it.
- **Approach 3:** If there are multiple clusters, EPA could describe each cluster, perhaps using Approach 1 or 2. The identification of clusters may take into account qualitative information regarding similarities in the basis of the judgments of multiple experts in a cluster; it requires judgment.
- **Approach 4:** A combined distribution can be produced by aggregating the probability associated with each value across experts. This is equivalent to estimating a distribution by Monte Carlo simulation in which experts are sampled (with equal or unequal probability) and values drawn from the selected expert's distribution. Unlike the previously described approaches, this method has the advantage of using all of the information provided by all of the experts, is reasonably easy to explain, and has been shown to perform reasonably well in characterizing uncertainty (Clemen and Winkler, 1999). However, this method

has the disadvantage of suggesting that the experts constitute a probability sample of relevant opinion. If expert weights are used, these can be based on several sources, e.g., ranking of experts by each other or experts' performance in providing probability distributions in response to calibration questions. (Calibration questions ask about variables whose value is unknown to the experts at the time of elicitation but known to the analyst at the time of combining the experts' distributions; Cooke, 1991). Unequal weighting of experts' judgments can lead to superior performance if the weights (or calibration questions on which they are based) are well selected, but is often viewed skeptically because it appears to give inadequate respect to experts whose judgments are given little weight and/or may discourage experts from participating in a time-consuming elicitation process if their judgments may ultimately get little weight. Unequal weighting methods are more complicated to conduct and explain than some of the other approaches.

As is evident from these example approaches, none is perfect. Thus, EPA may wish to consider selecting a relatively simple approach that conveys the range and central tendency, while acknowledging that a variety of approaches are possible, explaining the advantages and disadvantages of candidate approaches, and explaining why the selected approach was chosen. It is possible to use multiple approaches in combination (e.g., Approaches 1 and 4), as the situation warrants.

c. *If a combined estimate is considered appropriate, what interpretation should be applied to the percentiles of the uncertainty distribution derived from the elicitation (e.g., the mean estimate of a combined elicitation function, or the 5th -95th percentiles)?*

The appropriate interpretation depends on the combination method, and is perhaps best characterized by explicit description of the method (e.g., for Approach 1 above the interquartile range of the experts' mean estimates). Because it is typically not appropriate to characterize the experts as a probability sample of some relevant population of expert opinion, it does not seem appropriate to characterize a combined distribution as a probabilistic distribution of expert opinion.

d. *If a combined distribution is not appropriate, how should the EPA characterize the estimates of the PM premature mortality effect? One option employed in the Executive Summary of the PM NAAQS RIA is to present the estimates as a range from the average value associated with the steepest concentration-response function to the average value associated with the flattest concentration-response function. Is this the best approach? What other options would you recommend?*

As described above, characterizing effects as a simple range provides a very limited summary of the rich information about uncertainty provided by the expert elicitation. The Council finds that graphical displays (such as Figs 5-10 – 5-11 and a graphic like that illustrated in response to Question 3) that portray the uncertainty about premature mortality conditional on each expert's judgment, and the variability in estimates among experts, provide a comprehensive summary. In attempting to compress this information into a simpler format, the Council suggests that the appropriate summary will depend on the data and encourages EPA to attempt to characterize the extent to which experts' judgments are congruent and overlapping or broadly distributed across an overall range.

References

Clemen, R.T., and Winkler, R.L., "Combining Probability Distributions from Experts in Risk Analysis," *Risk Analysis*, 19: 187-203, 1999.

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

Industrial Economics, Inc., *The Expanded Expert Judgment Assessment of the Concentration-Response Relationship between PM$_{2.5}$ Exposure and Mortality,* Final Report prepared for EPA Office of Air Quality Planning and Standards. Available at http://www.epa.gov/ttn/ecas/regdata/Uncertainty/pm_ee_report.pdf, 2006.

National Research Council, *Estimating The Public Health Benefits Of Proposed Air Pollution Regulations*, Washington, D.C.: National Academies Press, 2002.

Rabl, A., "Analysis of Air Pollution Mortality in Terms of Life Expectancy Changes: Relation between Time Series, Intervention, and Cohort Studies," *Environmental Health: A Global Access Science Source* 5: 1-11, 2006.

Rabl, A., "Interpretation of Air Pollution Mortality: Number of Deaths or Years of Life Lost?" *Journal of the Air and Waste Management Association* 53: 41-50, 2003.

Tversky, A., and Kahneman, D., "Judgment Under Uncertainty: Heuristics and Biases," *Science* 185(4157): 1124-1131, 1974.