DRAFT Preliminary Individual Comments on the Characterization of Uncertainty in the Estimated Benefits of Reduced PM-Mortality using Expert Elicitation

DO NOT CITE OR QUOTE. These Preliminary Comments Are Not Approved by the Advisory Council on Clean Air Compliance Analysis and do not represent consensus or final advice to EPA.

5/7/2008

Dr. Ronald E. Wyzga
Electric Power Research Institute

Initial Pre-Meeting Comments

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

Overall comments:

In addressing this question there are also elements of other charge questions present, given the close relationship among the various charge questions. I have tried, however, to focus upon the first question.

The method used to develop concentration-response functions is reasonably well-described although the details of the elicitation process are not presented. This is not necessary if the Final Report describing the process is widely available. I would urge that efforts be made to incorporate key portions of this report into articles suitable for publication in the peer-reviewed literature thus making this material even more widely available.

The influence of differences in assumptions among experts is not presented in much detail at all in Chapter 5. There is considerable discussion of this in the Final Report. Again this detail is probably not necessary in Chapter 5 if it is accessible elsewhere.

It is apparent that the cohort studies influenced expert elicitation response. The estimates derived from 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." Hence in the paradigm of Exhibit 3-1 of the report, the cohort studies address areas A and B; the experts were including portions of C as well. This leads to several points that could be considered in Chapter 5:

1.) The differences in the objectives of the epidemiological studies and the expert elicitation are not articulated; Figures 5-10, 5-11, 5-12, and 5-13 are misleading to the extent they imply and compare the results of the epidemiological studies with the expert elicitations.

2.) The two epidemiological studies cited are among several that were considered by the experts; it would be interesting 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.
3.) Since several experts were co-authors and investigators in some of the epidemiological studies, it would be of interest to see whether their responses were more heavily influenced by their own studies than the responses of non-authors. (The Report does not provide this information, and provision of this information could compromise the anonymity of experts.)

Related comments:

Figures 5-10 and 5-11 are accurate; at issue is whether additional information should be presented or should the estimates be presented or described differently. For each expert the distribution of estimates is presented. Should this information be condensed, and if so how? I struggled with this issue and concluded that the presentation is probably correct. The elicitation results are a function of the chosen subjects. Another set of subjects could have yielded different results. Since there is no well-defined sample space from which the subjects were chosen and they were not randomly chosen, their responses are much more difficult to interpret, and I believe it is better to present them individually.

I find it curious that there is substantial uncertainty presented for the concentration-response functions, but that there is no uncertainty represented for the economic values associated with this response.

Detailed comments:

The relationship between the description of the methods and results (e.g., Figures 5-10; 5-11) would be improved if these tables were referenced in the methods description.

Figures 5-10, 5-11, 5-12, 5-13 refer to dotted red lines which are not evidenced in black and white copies. The legends could drop the word “red”; it should be stated that the numbers in these figures refer to the mean values.

The material in Figures 5-10, 5-11, 5-12, 5-13 is used in Tables 5-16, 5-18, 5-20 and 5-21, 5-23, and 5-24 yet the latter are presented first. It might be better to reorganize Chapter 5 to present the expert elicitation results before they are utilized.

Dr. Dallas Burtraw
Resources for the Future

(Assigned question #1): Characterization of expert's C-R function

The benefits chapter devotes considerable length to the characterization of the expert's assessments about mortality effects. This is integrated thoroughly in the chapter. This information is presented in a manner that makes it available to the reader in parallel fashion to information drawn from epidemiology studies.

There may be room for improvement in trading off less quantity of space to the expert's and more quality in characterizing their input. The expert assessment is portrayed in several different types of figures. In my opinion, the box and whiskers plot conveys all the information that appears in the other formats including the 3-d bar graphs and the cumulative distribution functions. Since an improved uncertainty analysis in the future may represent uncertainties in other parts of the assessment, an effort to be parsimonious in the presentation of information will be important so that the report does not grow in length unnecessarily.

The information summarizing the expert elicitation in the Executive Summary does not fully convey what the reader needs to know. An explanation of the labels "low mean" and "high mean" should be included. I believe the sentence(s) at the end of the first paragraph on page 5-33 should appear in the executive summary. This should be added to a brief discussion of other factors such as consideration of thresholds and causality so the reader understands a bit more about how the expert assessment was conducted. Also, throughout the report or at least earlier in the report the term "credible interval" should be introduced representing the expert's assessment and contrasted with "confidence interval." Also, the description in footnote b on page 5-3 should be brought up to the Executive Summary.

The charge question explicitly asks whether the report "conveyed the differences in assumptions (including the influence of key empirical studies) that drive the differences among the c-r functions." The report does not present rationale of the experts or information about what studies they embrace in order to arrive at their own assessments. This would be valuable information to convey because it would provide an implicit, transparent review of the literature by the experts.

Note in passing, I especially liked the narrative on page 5-90.

Overall comments on other charge questions

We are asked in general whether the EPA has addressed the NRC's recommendation to make uncertainty an integral part of the assessment. Yes they have, in the case of mortality, although as noted there opportunity for improvement. The assessment presents the same information in a variety of formats that contributes little and adds to length.

(Q #6)Another question is whether the expert judgments should be consolidated into a single uncertainty measure. My opinion they should not. I feel it would convey a false sense of the information.

In addition, I feel there should be a method for qualitatively indicating the differences in the epidemiology and expert assessments. It is fine for them to appear side-by-side, but not implicitly given equal weight to each. Perhaps one part of the table could be shaded in some manner.

(Q #4) Finally, the agency needs to consider other sources of uncertainty and how they would be addressed. It is clear given the magnitude of benefits from mortality that it is most important. However, one might ask what is the value of additional information regarding uncertainty in other parts of the assessment. I wish there was more discussion of uncertainty elsewhere in the damage function pathway. For example, variation in valuation would have comparable effect on the range of estimated benefits. This assessment does not adequately incorporate this source of uncertainty.

<u>Other comments</u>

It is unusual for an Executive Summary to discuss items that do not appear in the body text. I do not remember a discussion of the last bullet point on page ES-14 regarding units of measure for direct PM emissions elsewhere in chapter 5.

Another note, p. 5-96. I appreciate the discussion of benefits by major PM component but it should go a bit further to repeat for the reader that these do not account for the limited speciation information. I understand these estimates apply the same C-R coefficients and simply look at the overall constituents of PM.

Dr. Aaron Cohen
Health Effects Institute

Preliminary responses to Charge Question 5

Has 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?

For the most part, the current Executive Summary and draft chapter present the quantitative results at appropriate levels of detail and comprehensiveness. The major weaknesses that I see in one or both documents are: 1) inadequate description of the methods, especially as regards the Expert Elicitation; 2) discussion of the likely relative importance of different sources of uncertainty: those that were quantified vs. those that were not; and 3) failure to consider uncertainties related to the estimability of numbers attributable deaths using cohort study data.

a. *Considering the examples provided by the EPA, are there other methods that 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?*

I cannot suggest additional specific methods beyond the tabular and graphical displays that EPA has used in the current draft. That said, I found the overall description of the process (especially the material on Page 5-58) difficult to follow. Perhaps, putting more detailed methodologic material in text boxes would improve the narrative flow.

In contrast to the relatively large amount of technical information, I was unable to find in the chapter a clear explanation of how the experts were asked to consider particular current sources of uncertainty, such as the possibility of different exposure-response functions at different spatial scales or for different social classes (as per M Jerrett's 2005 analysis of the ACS LA sub-cohort, G Hoek's 2002 Dutch Dietary Cohort Study, and the 2000 HEI Reanalysis of the ACS and 6-Cities studies),or the induction time for a change in pollution levels on mortality (as per D Dockery's 2002 analysis of the Dublin Coal Ban). Presumably, it was the experts' views on these issues that determined the direction and magnitude of their estimates.

The chapter should also discuss the justification for affording the estimates of all experts equal weight given that they had quite different levels of expertise with regard to the evaluation of the largely epidemiologic evidence.

The Executive Summary currently lacks an adequate description of the Elicitation with regard to the rationale for its use (i.e., what it is intended to add above and beyond the results of the 2 studies), what it comprised, how it was conducted. A brief description should be added, perhaps using the language in Section 5.1.5.2 (Page 5-23, last para.) accompanied by Figure 5-1 (Page 5-24), and perhaps one representative box plot.

b. *To what extent do the types of statements made in the Executive Summary of the PM NAAQS RIA successfully communicate the extent of the uncertainty (and/or the uncertainty) in the estimate of PM premature mortality to those who are not familiar with the PM-Mortality Expert Elicitation?*

The Executive Summary currently lacks an overall description of the various sources of uncertainty and how they were addressed (or not). Perhaps Table 5-5 (Page 5-17) could be modified for this purpose.

Also, see below**.**

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

The draft chapter lacks 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, or could not be, quantified, as well as the relative importance of the various uncertainties that were quantified (uncertainties in the impact function vs. the valuation). This discussion should be provided and then be brought forward into the Executive Summary.

At least one important source of uncertainty was alluded to but not discussed as such. Recent work suggests that estimation from cohort study data of annual numbers of attributable deaths, as opposed to measures of longevity (e.g., Years-of-Life-Lost or Life Expectancy), is problematic at best, and that such analyses may overstate the annual attributable numbers (and by extension the valuation of those deaths). (See Brunekreef B et al. The brave new world of lives sacrificed and saved, deaths attributed and avoided. *Epidemiology.* 2007 Nov;18(6):785-8; Rabl A. Air Pollution Mortality: Harvesting and Loss of Life

Expectancy. *J Toxicol Environ Health.* 2005; Part A, 68:1175-1180; Rabl A. Analysis of air pollution mortality in terms of life expectancy changes: relation between time series, intervention, and cohort studies. *Environ Health.* 2006 Feb 1;5:1). EPA discusses this issue on Page 5-56, and acknowledges that measures of longevity would be "theoretically preferred" over numbers of deaths. They do not, however, discuss, let alone attempt to quantify, the real uncertainties that this theoretical issue poses for their actual estimates.

Dr. H.C. Frey
North Carolina State University

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. The latter should be referred to as 90 percent probability ranges or 90 percent credibility ranges (we might discuss the appropriate terminology at our meeting), since they represent a judgment about uncertainty and not an inference from a sampling distribution of a statistic.

The summary is adequate in terms of conveying the central tendency and range of the estimate based on the Concentration-Response (C-R) function of each expert. However, it could be more clear 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 based on C-R functions elicited from the experts. Similar, the benefits assessment should not be attributed to the experts, but it should be clear that the expert judgment was simply one of many inputs to the benefits assessment. This could be more clear in many of the tables, such as Table 5-32 and others.

What I did not readily find was an explanation of why the expert judgments differed, nor much if anything in the way of evaluation or assessment/interpretation of the judgments. For example, in Figure 5-10, it is clear that there is substantial inter-expert variability in results. Yet, the significance of this variability to the bottom line 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 why the experts agree within the clusters, and why they disagree between clusters. For example, why does Expert E have a more sensitive C-R function than Expert K? Also, are there comments from the post-elicitation interviews/verification step that would shed light on why the experts continue to disagree even after seeing each other's judgments?

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?

I recommend that the judgments of multiple experts be kept disaggregated as has been done here. However, EPA could infer whether there are subgroups or clusters of experts in which there is agreement, and also identify as to whether the judgments of one set of experts leads to a different bottom line conclusion than that of another set of experts. In this case, the bottom line conclusion may be whether net benefits are positive.

Overall, the approach used here to keep each expert separate and to evaluate the implications of the results based on different experts is useful. However, the results could be summarized a bit more, perhaps (can discuss more at the meeting).

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

There does not seem to be a perfect or best technique, and thus this becomes a matter of preference or style. Thus, I am not a proponent of combining or aggregating results. However, if an approach like this is to be used, a key point is to develop a weighted mixture distribution of multiple experts, not to average the judgments. This typically requires some kind of judgment as to the validity or relative merit of each expert, which adds an additional layer of subjectivity to the analysis that may be difficult to justify or controversial.

Alternatively, one could just assign equal weight to each expert. This may or may not be valid in the sense that the experts may not be truly independent of each other (i.e. they may share similar viewpoints on theory or interpretation of data). Thus, if several experts "agree," equal weighting among the experts will tend to give emphasis to the experts with similar views, even though collectively they may represent only one school of thought or point of view.

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)?

I do not recommend that a combined estimate be used unless the results from each expert are first shown separately and it is clear that the estimate is based on something like the mean of the experts and the range implied by the experts. It may not be appropriate to attempt to assign a probabilistic interpretation to such a range unless there is some reason to believe that the experts are a probability sample of some kind. Rather, the range could be treated as a sensitivity analysis and used to assess the robustness of the key findings of the assessment.

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?

The approach used here, perhaps with some additional post-analysis summarization, seems appropriate.

Dr. Wayne Gray
Clark University

Assorted Comments

I read over the PM NAAQS RIA materials (Executive Summary and Chapter 5), and looked over the NOAA report on dealing with uncertainty in the climate change area. I didn't look at the underlying Expert Elicitation report, so can't comment on whether EPA represented the original EE results properly.

Overall, the RIA materials are doing a good job at dealing with very complicated issues. Some of my initial reactions to the RIA report centered on ways that things could get more complicated (e.g. is "causality" really "all-or-nothing", or could the observed results reflect some "true" causality and some bias from confounding factors?) - but those thoughts are probably not helpful, or at least wouldn't lead to a more useful document, in terms of better conveying results about uncertainty.

A few suggestions that might be more helpful (I'm not always sure exactly which charge question(s) they'd fall under, so didn't bother trying to characterize them):

The Executive Summary could have included a short description about the connections between the uncertainty numbers in Tables ES-1, ES-2, and ES-3. I think that Table ES-2 shows the "fundamental" uncertainty about benefits in terms of mortality which the EE analysis was trying to measure, while Tables ES-1 and ES-3 multiply out the mortality numbers from ES-2 into dollars using some value of a statistical life, and add in the morbidity outcomes from ES-2, again multiplied by some dollar values. The discussion of uncertainty and limitations in Section ES-3 is quite good, but a brief summary around the earlier Tables would help (e.g. showing the "multiplication chain" from regulatory stringency to emissions to exposure to health effects to dollar values, but indicating that many/most of the steps in are being considered as fixed numbers, while the expert elicitation is concentrated on the uncertainty about concentration-response functions - at least that's what I think is happening).

As a broader note (perhaps related to question 4), there is clearly uncertainty on all of the steps in the "chain", although some steps (like who lives where) might be relatively carefully measured. I'm not sure how much more we can expect EPA to do in terms of dealing with all of the uncertainties in the chain. Working on the concentration-response side of things seems like a good place to start, and I'm not sure how much progress is available on the other fronts (it seems especially difficult to deal with uncertainty in connecting emissions to ambient concentrations, at least as I understand available air transport models). I suppose we could decide that a few steps in the chain are well

enough known (who lives where, from Census data), or serve the role of "constant terms" by being used in all these analyses (dollars per statistical life saved), so that EPA could just plug in the appropriate disclaimer about those sorts of uncertainties each time (rather than re-visiting these uncertainty issues over again in every rule-making).

In Table ES-2, the distinction between EE and ACS results might be highlighted even further, by having the main categories in the "Estimate" column being Expert Elicitation results (mortality only) and American Cancer Society study results (mortality and morbidity) - as a way to focus on the two different methods being used to capture uncertainty.

I'd like to suggest an alternative (or additional) way of capturing the uncertainty across the experts in the EE approach. This would be a "quartile of quartiles" approach (I just made up the name and the method, and I'm not sure if either or both have already been used in this area). The idea would be to pool the quartile-extreme values (.25 and .75) for each of the experts, then drop the top and bottom quarter of the values, and report the remaining range. This is relatively easy to explain (?), or at least easy to compute, and with an even number of reviewers it gives a clear outcome. The method gives much more of a central tendency than the low mean/high mean comparison, which is driven by the extremes of the researcher values. It's true that low/high mean collapses the distribution within each researcher, but it doesn't get any sort of "consensus" across the reviewers. For example, in the data in Figure 5-10 we would get a range of 4000-7500 for mortality, while in Figure 5-11 we would get a range of 4400-17000. I think this gives a better indication of how the overall expert distribution changes between these figures, as compared to the low/high mean ranges (which are 1200-13000 and 2200-24000, respectively). The quartile of quartiles range captures not only that the expected benefits have grown or that the uncertainty has grown, but also that there has been a much bigger shift on the upper-end of the benefits distribution.

It might be argued (as is mentioned in the NOAA report) that people, including experts, tend to put too little weight on "surprises" when calculating probability distributions, so that we want to include the extreme expert opinions to counteract this tendency. That approach seems to me to be misguided (since we have no way to judge whether the two "errors" are of comparable magnitudes in particular applications), and less useful when we're trying to come up with a tool to use in lots of different rule-making applications for different policy interventions, where we might want to compare the uncertainty associated with the different interventions in a way that provided a better sense of the overall views of the experts, rather than emphasizing the extremes.

Dr. Fintan Hurley
Institute of Occupational Medicine
United Kingdom

Charge Question 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?

Summary

Yes. The individuals who took part in the elicitation exercise gave their assessments in quite a wide variety of ways. Relative to its aim of estimating attributable deaths, the benefits analysis dealt well with this diversity, by having an overall approach that was common to all the individual assessments, and by taking due account of differences between individuals in how that approach was implemented. Along the way I have a small number of specific questions and comments.

Question 2 in context of the overall strategy in assessing benefits

Question 2 deals with technical aspects. These technical aspects have meaning in relation to the overall framework of assessing, separately for each individual who took part in the elicitation exercise, the mortality benefits among adults in the United States of reducing annual average $PM_{2.5}$.

This answer does not address the question of whether individual results only, or some 'average' result, should be reported.

For each individual, three main sources of uncertainty have been identified and quantified.

a. The probability that the association between annual average $PM_{2.5}$ and mortality hazard is causal.
b. Assuming causality, the probability that there is a population threshold for the relationship between annual average $PM_{2.5}$ and mortality hazard.
c. Assuming causality, the shape of the CR relationship (at annual average $PM_{2.5}$ above threshold).

At issue is how these were integrated in order to estimate benefits.

The description given is consistent with methods to estimate annual 'attributable deaths'. While seemingly simple, it is increasingly recognised that there are difficulties underlying this concept. It is however a widely used approach; and I think the spirit of Question 2 is not to question it, but to see whether, within this framework, the various probabilities have been dealt with well.

My main response is within this framework; I add some comments about annual attributable deaths.

EPA's approach to combining probabilities information from the three kinds of issues above

Given these data, EPA's seems to have made a clear decision, that what is appropriate for a benefits analysis is, for any annual average $PM_{2.5}$ in the range 4-30 µg/m^3, to use the unconditional distribution (across the three aspects of probability of causality, threshold and shape of CR function) of an individual expert's assessments of these issues.

I think that is the correct approach.
My only question concerns the probability of causation, how the experts understood this in the elicitation exercise, and how this has affects understanding of the derived benefits. It is clear from the Extended Elicitation Report (Section 3.1.7) that this was a difficult aspect of the underlying elicitation exercise, in that (even after discussion) the question may have been understood differently by different participants.

I think the benefits assessment differs according to how this question is interpreted.

i. A 'hard' version of the question is: does exposure to $PM_{2.5}$ *per se* cause mortality?
ii. A 'soft' version is: Assuming that factors other than air pollution remain unchanged, do changes in annual average $PM_{2.5}$ lead to changes in risk of mortality among adults in the US?

'Yes' to (i) implies 'Yes' to (ii). However, 'Yes' to (ii) does not necessarily imply 'Yes' to one if, for example, one understands the associations with $PM_{2.5}$ as expressing the effect of a mixture.

I think version (ii) is the relevant one for benefits analysis.

If experts interpreted the causality question more restrictively that implied by (ii), then in principle the unconditional distribution as assessed may underestimate the benefits.

In practice, for most experts the probability of causation was (very) high; and expert K, where it matters most, interpreted causation in the spirit of (ii). So in this analysis the issue is not practically important.

In using this benefits analysis to learn and improve, there may be something here to consider. I think the issue is more one for the design of the elicitation exercise, to eliminate ambiguity here, rather than for how EPA has used the results of the elicitation.

Putting the approach into practice

I think this is good – for any individual, disaggregate to the various options and sample between them in proportion to the probabilities.

The approach can look complicated when the expert's underlying judgments (e.g. K) are different at different annual average $PM_{2.5}$; but it's really a question of being sensible (when added assumptions are needed) and careful, and this benefits analysis has been both.

There is an added assumption, that where an expert gave a probability of threshold applied to a range of annual average $PM_{2.5}$, then this probability can be apportioned within that range. I think that's reasonable.

My one doubt about it concerned the probability of threshold at less than 5 $\mu g/m^3$, when the analysis is restricted to 4 $\mu g/m^3$ or higher. However, the assessment of relevant expert (K) seems to refer to all concentrations below 5 $\mu g/m^3$, so I think the EPA's use of this information is good.

Communicating the approach and the results

The approach looks complicated, because some of the underlying expert judgments are complicated. But it really is quite straightforward.

It may be possible to improve communication by focusing on the underlying simplicity, and giving examples of this; and of having the detail on complexity – which is essential to have – in and Appendix. This may be easier for the reader who does not want to deal with the mathematics, while giving the necessary information to those who do want that detail.

Annual attributable deaths and life tables

It is increasingly recognised that there is a benefit in using life tables for assessing mortality impacts of changes in annual average $PM_{2.5}$ (whether these benefits are expressed in terms of deaths or life years). This is because the approach incorporates aspects of population dynamics that are not included in a 'static' approach, as seems to have been used in the present benefits analysis.

I think the EPA's approach to the issues of Charge Question 2 carry over to life table applications also; i.e. while there are some differences in application (e.g. in not using impact functions in the same way), I do not see any real new issues in how the experts' judgments are used.

Dr. Virginia McConnell
Resources for the Future and UMBC

General comments:

Why are all benefits shown in 1999$ dollars? I think it would be easier for the public to understand estimates that are in dollars of the date of release of the RIA.

I would like to see Chapter 5 have some discussion of how well the experts understood their task of providing the probabilistic distribution function described on page 5-33 (top paragraph, line 7). Even for someone who is doing research in this area, this seems like this could be quite difficult. There is some indication in the discussion of the experts' responses that some of them may not have been entirely clear on what they were doing. Was there some attempt to address this….I am sure there was, but a brief discussion here might be important.

I am not clear from chapter 5 about how exactly the different threshold level of benefits were estimated. I am sure there will be a chance to ask about this on Thrusday. But it may need to be more clearly explained in the chapter itself.

The results shown in section 5.2 starting on page 5.65 do not explain that the dollar benefit estimates in tables 5.20 – 5.26 reflect uncertainty in both the C-R functions *and* the economic valuation functions. The general approach is suggested on page 5-14 but it is not discussed or mentioned again in the text when the dollar estimates are presented, or in the section on valuation (previous to 5.2). In fact, on page 5-55, it sounds like a point estimate of $5.5 million is used.

Around page 5-44 of Chapter 5, the section headings and labels change. Section labeling should be checked to make sure it is consistent.

Response to Charge Question #5.

Executive Summary.

The maps are very helpful. What about a map or set of maps that shows the reductions needed to attain the annual 15ug/m^3 and 14 ug/m^3 standards? Those are the major sets of comparisons in the analysis and summarized in ES-1.

Title of Table ES-1 should include the fact that these are annual benefits and costs. It should read "…Annual Benefits with Annual Social Costs….."

A problem with Table ES-1 is that the benefits based on the ACS study (top part of table) are presented as point estimates, and the benefits based on expert elicitation survey reflect uncertainty. Ideally, full uncertainty analysis would be reflected consistently in both estimates of costs and benefits. Since that is not possible at this point, I think there should be a brief footnote explaining why uncertainty is addressed for certain benefits and not costs. I like the division between the ACS study and the expert elicitation study results, but there should be a range of benefits shown for each. I do think the expert elicitation study results can summarized in a table like this, but I don't see exactly where the low mean and the high mean were defined. This definition is not in Chapter 5. Footnote e as it stands is not clear.

A bar chart summarizing Table ES-1 might make the results easier to understand. There could be two separate charts, each comparing across studies and across standards, one at the 3% interest rate and one at the 7% interest rate.

Table ES-2. What the actual numbers are could be made more clear: e.g. expected fewer deaths; or reduced cases of ……

Chapter 5.

Table 5-16. Title should say that these are annual estimates (right?). Is the threshold of 10 assumed in this table? (if so, it should be in the footnotes). Also, I think the headings "Mortality Impact Functions Derived from….." are confusing. They seem to imply the numbers are the impact functions. Those headings could be changed to : " Mortality reductions based on functions derived from epidemiology literature" etc.

Also for table 5-16. Expert K from the discussion on page 5-38 and 5-39 doesn't quite match with the results in Table 5-16….200 vs. 210? And, can there be a short footnote explaining what the "modeled partial attainment" and "residual attainment"are. It is mentioned on the page before and that discussion refers the reader to Chapter 4. But the difference in these seems important for the Table, so I think at least a short footnote would help.

I like the box and whiskers plots in Figures 5-10 and 5-11. They clearly show the distributions across the experts. The headings across the top are not clear however.

I am not sure that 5-12 and 5-13 are useful. Some aggregation of the experts would be better before estimating dollar benefits. This is also true of Table 5-27. This table would be more effective if it included a summary measure from the expert elicitation results.

The results could be compared across the different types of studies, and across the different standards.

Why are the benefits for the East so high in Figure 5-9? It is not clear to me why this would be the case, and I didn't see it explained in the text.

I think 5-32 and 5-33 are not useful. The pie charts in figure 5-16 are good, but let the reader partition the benefits among pollutants, if there is interest in that. 5-32 and 5-33 are very busy, but they don't tell us very much.

The study that Jim Hammett sent out the link for on 5/5/08 suggests that in terms of presentation:

- it is very important to get feedback from target audiences (through focus groups or other ways) on how clear different ways of presenting the uncertainty information are, and then iterate to get an improved presentation.
- It is always useful to show the uncertainty in a number of different ways. I think the current version of Chapter 5 does this well.

I will think more about more of the specifics of question 5 for the discussion on Thursday.

Dr. David Popp
Syracuse University

General

There is a lot of variation in the expert responses. More importantly, these generally seem to be above the results from the epidemiology literature. The report doesn't make clear why this is the case. Is there something that epidemiology studies are missing that expert elicitation can pick up? Should we consider the expert elicitations as biased upwards? It would help to include more discussion of this discrepancy. I think that expert elicitation can be a useful tool, but I need some context to understand why the results are so different, and if this difference is "noise" or "value added".

Comments on charge question #6: Summary of Expert Elicitation

The EPA should continue to present results from individual experts. The fact that these vary widely is important information. At the same time, it should present aggregate estimates in addition to the individual responses. Given the wide range of values, it is difficult to know what to do with any one value. A combined distribution would be of use. I would use a fitted distribution to the overall set of concentration-response functions. As for interpreting the percentiles, it seems straightforward to describe them as the percentiles from a select group of experts. Unfortunately, it is difficult to generalize from this, as the experts are not a randomly chosen sample (nor should they be).

Dr. Rebecca Parkin
George Washington University

Q1. Has EPA accurately characterized each expert's…function…and conveyed the differences in assumptions...that drive the differences in the functions?

The accuracy of the characterization cannot be evaluated without access to the raw data and software used by the analysts. However, this reviewer is not aware of any significant problems with the methods used to characterize the experts' functions.

Differences in assumptions that drive the differences in the experts' functions were not easy to identify in the chapter or Executive Summary (ES). This reviewer suggests enhancement of this information in the text, perhaps including a table of such assumptions.

Q2. …is our mathematical treatment of concepts...sound, as well as transparent?

This reviewer is not well qualified to address this question. Although mathematical approaches are presented and discussed, the text would not be easy for non-technical readers to understand.

Q3. Do the tables…adequately distinguish the benefit estimates based on data-derived components…from those based on expert judgment? How should the mortality estimates based on the elicitation be compared…?

See Q5 for discussion of "adequacy."

This reviewer found the tables and figures valuable supplements to the text; they increase the reader's ability to find and understand the distinctions. In this reviewer's opinion, the conclusions and ES, however, did not distinguish as clearly as the rest of the materials.

Q4. Does the EPA present effort to incorporate uncertainty analysis...adequately address the NRC's request…?

This effort addresses some but not all of the NRC's request.

The NRC recommended the following (quoted or paraphrased from NRC (2002), pp. 9-12), and the EPA response in this chapter is noted.

- Reflect the plausibility and uncertainty of the concentration-response function, such as imprecision of exposure and response measures, potential confounding

factors, and extrapolation from the study population to the target population in the benefit analysis. *Some but not all of these elements are addressed in the chapter.*

- Decision-makers may want to know the effects of a regulation on different subgroups of a population… *The chapter includes subgroup estimates as the peer-reviewed literature permits.*
- Move the assessment of uncertainty…into the primary analyses. *This chapter accomplishes this change.*
- Use sensitivity analyses but…include more than one source of uncertainty at a time. *Sensitivity analyses are used in the chapter.*
- Methods, rationales, and results should be clearly described…and important factors should be highlighted. *Methods and rationales are not always described thoroughly (see comments below), although results seem to be well characterized in this chapter.*
- Elements for a detailed summary (on p. 12) are listed. *These are included in this chapter, but not all to the same level of description.*
- Highlight all assumptions that have a substantial impact on the results of the analysis. *It is not clear whether the Agency determined thoroughly which assumptions had greater impact on the results. No summary comparison of the impacts is provided. A table of such assumptions would aid the reader's comprehension. The comments in the Discussion section are quite general and do not give the reader a sense of which assumptions were the most crucial to the estimates.*
- EPA should estimate benefits at reasonable intervals, such as every 5 years… *The analysis meets this recommendation.*

<u>Q5. Has the EPA adequately communicated the uncertainty information …to the audiences that the RIA addresses…?</u>

Several key issues are raised by this overarching question.

- What is meant by "adequately communicated?" What are the Agency's communication goals for the ES and chapter? Is that set of goals the target for judging "adequacy" or is some other standard of interest? Perhaps adequacy is in the eyes of the "audiences."
- What are the "audiences'" information needs and priorities? What are their pre-existing knowledge and views of the topic? While "audiences" and "communicated…to" imply passive receivers of information, it is unlikely that some of these "audiences" will be passive parties. Would "decision-makers and stakeholders" be a better way of characterizing the readers of the report?

Without empirical evidence that indicates what the "audiences'" information needs and preferences are, it is difficult to determine whether this chapter overall "adequately communicates."

Therefore, the following questions focus on whether the uncertainty information is communicated in alignment with "best practices" (e.g., synthesis of peer-reviewed literature).

- What types of uncertainty are included in the report? (Some are noted on pages 5-8 and 5-15 through –17.) Are appropriate methods of analysis used to address these types? Are these methods and the Agency's process of method selection and implementation (e.g., criteria, judgments, assumptions, etc.) clearly described?
- Does the report discuss the quality of data and/or criteria for assessing data quality used to characterize the uncertainties?
- Is there empirical evidence indicating what uncertainty information is crucial to meeting the "audiences'" information and decision-making needs?
- Are the uncertainties characterized in quantitative and/or qualitative terms? Are these terms ones that are familiar to and understandable by the various "audiences?"
- Are both text and graphics used to characterize the uncertainties? If so, are effective forms used? (e.g., discussion and presentation of widely differing individual expert elicitation results)
- If graphics are used, are they clearly labeled and readily interpreted?
- For each uncertainty, are both the nature and extent of the uncertainty described?
- Are the uncertainties presented in a context so that they can be interpreted for decision-making purposes?
- How are the uncertainties framed? Are odds and/or probabilities shown?

a. …are there other methods the EPA should use…to summarize and communicate the results…?

Overall, the methods to summarize the results seem appropriate. What cannot be determined on face value is whether these methods will "communicate" effectively to the intended readers and users. Some authors who present expert elicitation have used to color to focus the readers' attention on key points. Color may have been used on Figures 5-14 and 5-15, but this reviewer's copy was in gray tones. If color was not used but is an option, this reviewer suggests that EPA consider this additional element for complex figures.

This reviewer suggests that EPA consider using additional graphics in the Executive Summary.

b. To what extent do the types of statement made in the Executive Summary…successfully communicate the extent of uncertainty…in the estimate…to those who are not familiar with the PM-Mortality Expert Elicitation?

Judging success in communication depends, in part, on the perspective taken. If we take the viewpoint of the Agency, we need to know the Agency's communication goals; this reviewer did not find such goals in this draft. If we take the viewpoint of the readers, it is not possible, without empirical testing and refinement of the statements among representatives of the "audiences," to determine whether the statements will be successful in meeting their needs

Comments regarding the ES "Uncertainties and Limitations" section follow:

- The section is not clearly aligned with the organization of Chapter 5.
- There is no opening statement to this section to indicate the nature of the bullets and/or which relate to "uncertainty" and which relate to "limitations." Further, the four bullets included under "Benefits" are identified as "key assumptions" in the text (pp. 5-6 and 5-7). The non-technical reader may not know how "assumptions" relate to "uncertainty" or "limitations." How are the other bullets noted in the text? There appears to be a mixture of types of bullets and level of importance in their impacts on the uncertainty characterizations.
- The sources and nature of uncertainties are not described in the ES, nor are they quantified. These aspects of uncertainty are presented in the text.
- There is no figure or graphic to support the uncertainty summary. Using both text and graphics typically strengthens reader comprehension. In this case, one of the box plots of the expert elicitation results may be an appropriate example.
- While limitations of the air quality modeling and emissions are discussed, there is no mention of the limitations of the data used in the model or of the expert elicitation method.

The primary problem with formulating an answer to this question is the lack of empirical evidence indicating what types of statements would successfully communicate to/with the "audiences" about this particular topic.

c. Are there additional summary statements that are important…?

While general recommendations are suggested above, there is no replacement for empirical testing and refinement of statements to ensure that the uncertainty information is received and understood as intended by the Agency.

There are many points made in the chapter about uncertainty and assumptions, but a fairly small set of these appear in the ES. It is likely that one or more of the chapter-bound statements should be lifted into the ES. This reviewer found statements to consider, for example, on the following pages:

- Uncertainty: pp. 5-6, 8, 14-23, 25, 27, 55-57, 85, 90, 96, and 101.
- Assumptions: pp. 5-6 through 9, 13-15, 19-21, 44, 46, 48, and 54.

Two important statements regarding the estimates are near the top of pg. 5-85, EPA should consider adding the following to the ES:

- Last sentence of the first paragraph
- Entire or shortened form of the third paragraph

Both of these statements provide crucial contextual information to assist readers in understanding and interpreting the results.

The criteria used to compile an appropriate set of experts are not mentioned in the chapter, nor is a reference given (e.g., Roman et al, 2008) where such information can be found. Given the importance of the process used to identify the set of experts interviewed, this reviewer strongly suggests that the process and criteria noted in Roman et al be included in the chapter.

One comment that could be considered for the ES is in the last paragraph on p. 5-22, regarding the incomplete capture of health effects results in "downward bias" of the benefits presented.

Additionally, statements about the impact of assumptions on the experts' input (e.g., on p. 5-44) could be included in the ES.

The culminating sentence found on p. 5-101 (last sentence of the third full paragraph) could be added to the ES.

<u>Q6. Has the EPA adequately summarized the results…across the experts…in the benefits chapter and ES?</u>

Note: The same issue of "adequacy" as in Q5 applies here.

Presentation of the individual experts' responses is appropriate, because there is wide divergence in their judgments. An aggregated approach runs the risk of masking information that may be crucial to readers and decision-makers. Clarifying the upper and lower bounds is important, as shown on Table 5-1 (although the same data is marked differently in Table ES-1). This reviewer agrees that not aggregating the results is the better choice in this case.

In future analyses the decision about aggregation must be made in the context of each analysis and its purposes. It would be appropriate for the Agency to explore the impacts of the experts' differences on the regulatory decision that is being made.

Dr. Chris Walcek
State University of New York, Albany

Preliminary Comments Charge Question #3

Charge Question 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 with those derived from the empirical studies of the PM-Mortality association?

The tables, text, conclusions and Executive Summary definitely distinguish the benefit estimates based on data-derived components from those based on judgments elicited from experts. My only advice would be to take the additional information about uncertainty one step further and actually place reasonable upper and lower bounds to the 5th and 95th percentile uncertainty limits by aggregating and merging the data-derived and expert judgments. For example, if you want the 95th percentile confidence limit, simply average the 95th confidence limits suggested by BOTH the peer-reviewed literature estimates and the 12 elicited experts. In this manner, the overall presentation would be considerably simplified: There would be only one set of uncertainty ranges presented. The information breaking down this uncertainty among the 12 experts and two data-derived published studies could be retained as an appendix, along with an explanation of why the uncertainty ranges are much wider than published estimates.

The underlying technical differences between these two methods of quantifying uncertainty are ambiguous and somewhat arbitrary. These two distinct methods of quantifying uncertainty are probably more overlapping that acknowledged. Presumably the only difference between these two methods is the reliance on **published** peer-reviewed literature vs. the estimates of experts empanelled directly in response to this study. However, the authors of the ACS and Harvard six-cities studies are clearly "experts", probably looking at largely the same published observations as the 12 experts elicited for this study. Were experts A-L all independent of the published literature authors?

It appears that in nearly all cases presented that the ranges defined by the "expert elicitations" are considerably wider than the ranges derived from the epidemiological literature. Presumably, the "experts" are all deriving their judgments from largely the same published observations, and the EPA must assume that the much larger uncertainty encompassed by the expert-elicited estimates more accurately quantifies the actual uncertainty in the "state of the science" regarding these matters. The only "weakness" of this assumption would be the nature of the qualifications of the "experts" that were elicited for this survey. Someone (maybe one of the other experts) might complain that

one or more of the expert's estimates of uncertainty are "way out of the range" and therefore might recommend ignoring some of the ranges noted. However, if the professional qualifications of the elicited experts are reasonably quality-assured, then I see no defensible reason to question any of the estimates presented.

Some minor technical suggestions

1) Throughout the report, reference is made to "90th percentile confidence intervals". To avoid any uncertainty in the meaning of this description, I would suggest calling this "5th-95th percentile confidence interval"

2) Costs are presented in dollar amounts, which can lead to ambiguity in whether these are one-time costs or spending rates. Only after more careful reading are these costs revealed as "annual" costs. Therefore, I would suggest that all costs presented here be DEFINED as $ per year ($ yr^{-1} or $/yr) explicitly in the title. Furthermore, to more clearly note the inflation effect, I would suggest that all dollar amount quoted be denoted with an explicit subscript defining the 1999 year. Therefore, all costs quoted in tables and figures should be clearly displayed as "$\$_{1999}$/yr". Therefore in Table 5-1 the title to the table should read "billions $\$_{1999}$ per year" rather than
"billions 1999$"). Even individual entries in the table should read $\$_{1999}$ 17/yr rather than "$17"

Additional questions that I have:

a) Costs are "monetized" at 3% and 7% discount rates. I would assume that another key factor affecting these methods of "spreading costs" would be the assumed term of amortization. What is the term? Is it a fixed 20 years or something? Shouldn't there be a similar sensitivity presented for the amortization term (e. g. 10 year vs. 30 year amortization period)? If not, why not?

b) In table 5-3, I'm surprised that the "elasticity estimates" are all less than one. I would expect that this number <u>could be</u> greater than one under some conditions. E. g., a poor person is willing to pay only an extremely small amount (WTP=$1) to improve visibility, whereas a rich person might be willing to pay considerably more (WTP=$1000).

c) Table 5-17: Why are some of the morbidity changes NEGATIVE? How is it possible that there is LESS morbidity with poorer air quality?

d) Table 5-18: Are premature mortality numbers "deaths per year" or are they accumulated deaths over some time period. This information should be clearly defined in the table column headings or the table title.

e) Fig. 5-8 is really confusing. I don't even know what is plotted here. Is this same information in some other table?

f) Table 5-28: Maybe I don't understand the purpose of this table. I'm surprised to see apparently significant health impacts at any threshold. I thought thresholds were by definition adopted such that there should be minimal impacts at the legal threshold.

Dr. Michael Kleinman
University of California, Irvine

Charge Question 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 for developing a probabilistic representation of causality could be more explicitly developed in the document. The strength of association between PM2.5 exposure and health effects is appropriately assessed via the weight of evidence in the selected referenced material. However, the extension of assumption 2 (pg 5-6) that "all fine particles, regardless of their chemical composition, are equally potent…" becomes severely stretched when making the logical jump that one can apportion effects among constituents to evaluate modeled attainment strategies (pg 5-20, Fig 5-16, pg 5-97).

Recently published results suggest increased potency of ultrafine particles compared to fine particles and that the carbon-containing components (EC and OC) may be more toxic in some non-mortality endpoints. These components are not sufficiently characterized in most mortality studies to examine their contribution to the observed heath effects. Perhaps it might be suggested that a future expert panel could try to extrapolate observed data into mortality estimates.

There are significant differences between the benefits computed assuming a non-threshold distribution vs. an assumed 10 μg/m3 threshold under some situations. This appears to increase the level of uncertainty in the calculations of benefits. The impact of threshold assumption is large in Western US states (excepting CA) because the annual ambient concentrations are low but is less important in Eastern and states and CA where concentrations are higher. It might have been useful to take into account population as well since the states with low PM are also low in population. Would the threshold impact be more appropriately examined using some population-weighted construction?

Dr. Shelby Gerking
University of Central Florida

Benefits Assessment of the $PM_{2.5}$ Standard—Charge Question #3

A. 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? Total health benefit estimates obtained from data-derived components differ from total health benefit estimates obtained by eliciting expert judgments because of differences in the mortality benefit component. Data-derived total benefit estimates relied on the American Cancer Society Study and the Harvard Six-City Study to obtain epidemiological relationships between mortality and exposure. These relationships were statistically estimated, so error margins for coefficients of interest are available. Expert judgments were obtained from a panel of 12 experts who were asked for information useful in constructing a probabilistic relationship between mortality and exposure. Monetary benefits of reduced mortality were obtained in both cases by applying VSL estimates from the same probability distribution of values. Total data-derived health benefits and total expert judgment health benefits then were obtained by adding in morbidity values from the published scientific literature. Monte Carlo simulations based on available point estimates and error margins were used to construct confidence bounds for total health benefit estimates obtained from both the data-derived and expert judgment components.

   The text, tables, and conclusions in Chapter 5 of the report clearly describe these procedures and adequately distinguish between data-derived components of the uncertainty assessment and those obtained from the expert judgments. The executive summary, on the other hand, would benefit from a small amount of additional narrative to indicate more precisely both the common and different features of the two types of benefit estimates. In making this comparison, it might be helpful to present Table ES-2 first, followed by Table ES-3. If Table ES-3 could be revised to include the cost figures from ES-1, then Table ES-1 may not be needed. Also: (1) what is now Table ES-2 might be slightly revised so that it is clear that the low and high means in the "range based on expert elicitation results" are the means obtained from the low expert and the high expert and (2) what is now Table ES-3 might be revised to compare mortality benefit estimates from expert judgment with those obtained from the mortality function estimates in the ACS study. As it stands, total health benefits are compared even though the methodology used to obtain the morbidity estimates is the same in each case. The added detail requested here (a comparison of data-derived mortality benefit estimates with expert judgment mortality benefit estimates in addition to a comparison of the two types of total health benefit estimates) would have the

advantage of indicating the relative magnitudes of mortality and morbidity health benefits in computing total health benefits.

B. How should the mortality estimates based on the elicitation be compared to those derived from the empirical studies of the PM-mortality association? In, for example, Tables 5-16 and 5-21, the 95% "credible" intervals presented for each of the 12 experts are interesting and respond to guidance provided in the NRC report on estimating health benefits of air regulations. The NRC report recommended greater emphasis on uncertainties in health benefits analysis. The report further suggested that implementation of this recommendation will require specification of a probability distribution for each uncertainty source using both data and expert judgment.

In these tables, as well as in related tables, it might be useful to present information for the "average" expert as well as for each of the 12 individual experts. Inclusion of the "average" expert would give some sense for the center of the distribution of expert judgment estimates. A point estimate of benefits and an accompanying interval for the "average" expert could be obtained using Monte Carlo simulation by choosing values at random from the probability distributions of the 12 experts. While details about the variability of the individual expert estimates is clearly valuable in addressing uncertainty, the "average" expert might be more directly compared to the data-derived estimates from the ACS and Harvard Six-City studies. Information about the "average" expert then would become an additional line item in Tables 5-16, 5-21, ES-3 (and possibly other tables) and also would become an additional discussion point in the narrative.