



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

November 24, 2009

EPA-CASAC-10-003

The Honorable Lisa P. Jackson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Review of *Risk Assessment to Support the Review of the Particulate Matter (PM) Primary National Ambient Air Quality Standards – External Review Draft (September 2009)*

Dear Administrator Jackson:

The Clean Air Scientific Advisory Committee (CASAC) Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS) Review Panel met on October 5 - 6, 2009 to review the *Risk Assessment to Support the Review of the Particulate Matter (PM) Primary National Ambient Air Quality Standards – External Review Draft (September 2009)*. The Panel also held a public teleconference on November 12, 2009 to discuss and finalize its draft report. In this letter, CASAC offers general comments on the Risk Assessment, followed by our consensus responses to the Agency's charge questions. Comments from individual panelists are also attached.

Overall, CASAC found the Risk Assessment to be a clear and thoughtful approach to assessing health risks of airborne PM. It largely executes the path laid out in the *Scope and Methods Plan for Health Risk and Exposure Assessment* of February 2009. We are impressed with the progress that has been made in the development and application of risk assessment methodologies in the NAAQS process. The choices that EPA has made in terms of monitoring data, simulation of air quality, and selection of concentration-response functions are reasonable, appropriate, and justified. The estimates of health effects of various candidate levels for PM provide a reasonable basis for EPA's policy assessment. One weakness of the current Risk Assessment is that there is not a chapter that synthesizes the results, particularly across Chapters 4 and 5.

In this instance, EPA limited the Risk Assessment to only those outcomes judged to be “causal” or “likely causal” with the highest level of certainty and sufficient data. Thus only PM_{2.5} risks were assessed. We understand the practicality of focusing the Risk Assessment in this way, but we remain concerned about risks of coarse and ultrafine particles (UFP). Moreover, in focusing exclusively on PM_{2.5}, EPA may leave itself open to questions about regulatory decisions made with respect to PM₁₀ and PM_{10-2.5}. Rather than stopping short of addressing risks from the coarse fraction particles (PM_{10-2.5}), we suggest that EPA consider models that use the PM₁₀ data, along with PM_{2.5} data to improve understanding of risks associated with the coarse fraction component (PM_{10-2.5}). EPA could make use of PM₁₀ data and findings on health risks in western states where coarse particles are typically a substantial fraction of PM₁₀ and PM₁₀ would be a reasonable surrogate for PM_{10-2.5}. This approach could provide more information on rural and regional PM effects to complement EPA’s urban-focused assessment. Continued concern regarding risks associated with UFP is motivated by widespread proximal exposure to motor vehicle combustion exhaust. Although the association of UFP with health risk has been judged to be at the “suggestive” level in this review cycle, EPA should continue to track emerging health data in this developing research area. In addition, because the focus has been placed on those outcomes for which the evidence is at the highest levels of certainty, broad classes of other health outcomes (reproductive outcomes and lung cancer) do not become part of the quantitative risk assessment. EPA might face criticism because risk estimates were not made for conditions that may affect large segments of the population (>4.2 million births per year, and almost 200,000 deaths from lung cancer). Even considering the limited confidence that might be placed on such calculations, some semi-quantitative estimates of the range of risks for these conditions should be considered.

We concur with EPA’s structured approach for classifying uncertainty as well as the sensitivity analysis in the Risk Assessment. The sensitivity analysis would be strengthened by making a comparison of the selected concentration-response functions to those from studies that were not selected.

EPA chose not to carry out an exposure assessment as a component of this Risk Assessment even though exposure misclassification was characterized as the largest source of error. We understand the constraints faced by the Agency in carrying out a full exposure assessment. However, we recommend the use of quantitative (or semi-quantitative) exposure information and non-probabilistic based approaches to identify factors that contribute to observed variability in the concentration-response functions or to city-specific differences in risks. For example, EPA might provide some quantification of estimated inter-individual variability in daily average PM_{2.5} exposure/concentration ratios to demonstrate the extent of possible exposure misclassification. We urge the Agency to include an exposure assessment in the Risk Assessment and to further develop exposure modeling in order to be ready for the next NAAQS review cycle.

Looking beyond the current Risk Assessment, EPA should continue to carry out and encourage research on variation in risk of PM based on composition. Similarly, EPA should continue to broaden the monitoring approach for PM, measuring not just PM mass within particular size ranges, but continuing to expand collection of PM composition data across the size range of ambient exposures, from UFP through PM_{2.5} and PM_{10-2.5}. The resulting data should strengthen understanding of exposure to these PM size fractions and provide a broader platform for epidemiological research.

We thank the Agency for the opportunity to provide advice on the PM Risk Assessment and look forward to the review of EPA's second draft Risk Assessment and draft Policy Assessment early next year.

Sincerely,

/Signed/

Dr. Jonathan M. Samet, Chair
Clean Air Scientific Advisory Committee

Enclosures: A: CASAC Particulate Matter Review Panel Roster
B: CASAC Responses to Charge Questions
C: Individual Panelists' Responses to Charge Questions

Enclosure A

Clean Air Scientific Advisory Committee Particulate Matter Review Panel

CHAIR

Dr. Jonathan M. Samet, Professor and Chair, Department of Preventive Medicine, University of Southern California, Los Angeles, CA

CASAC MEMBERS

Dr. Joseph Brain, Cecil K. and Philip Drinker Professor of Environmental Physiology, Department of Environmental Health, Harvard School of Public Health, Boston, MA

Dr. Ellis B. Cowling, University Distinguished Professor At-Large Emeritus, Colleges of Natural Resources and Agriculture and Life Sciences, North Carolina State University, Raleigh, NC

Dr. James Crapo, Professor of Medicine, Department of Medicine, National Jewish Medical and Research Center, Denver, CO

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

Dr. Donna Kenski, Data Analysis Director, Lake Michigan Air Directors Consortium, Rosemont, IL

Dr. Armistead (Ted) Russell, Professor, Department of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA

CONSULTANTS

Dr. Lowell Ashbaugh, Associate Research Ecologist, Crocker Nuclear Lab, University of California, Davis, Davis, CA

Mr. Ed Avol, Professor, Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA

Dr. Wayne Cascio, Professor, Medicine, Cardiology, Brody School of Medicine at East Carolina University, Greenville, NC

Dr. David Grantz, Director, Botany and Plant Sciences and Air Pollution Research Center, Riverside Campus and Kearney Agricultural Center, University of California, Parlier, CA

Dr. Joseph Helble, Dean and Professor, Thayer School of Engineering, Dartmouth College, Hanover, NH

Dr. Rogene Henderson, Senior Scientist Emeritus, Lovelace Respiratory Research Institute, Albuquerque, NM

Dr. Philip Hopke, Bayard D. Clarkson Distinguished Professor, Department of Chemical Engineering, Clarkson University, Potsdam, NY

Dr. Morton Lippmann, Professor, Nelson Institute of Environmental Medicine, New York University School of Medicine, Tuxedo, NY

Dr. Helen Suh MacIntosh, Associate Professor, Environmental Health, School of Public Health, Harvard University, Boston, MA

Dr. William Malm, Research Physicist, National Park Service Air Resources Division, Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

Mr. Charles Thomas (Tom) Moore, Jr., Air Quality Program Manager, Western Governors' Association, Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

Dr. Robert F. Phalen, Professor, Department Medicine; Co-Director, Air Pollution Health Effects Laboratory; Professor of Occupational & Environmental Health, University of California Irvine, Irvine, CA

Dr. Kent Pinkerton, Professor, Regents of the University of California, Center for Health and the Environment, University of California, Davis, CA

Mr. Richard L. Poirot, Environmental Analyst, Air Pollution Control Division, Department of Environmental Conservation, Vermont Agency of Natural Resources, Waterbury, VT

Dr. Frank Speizer, Edward Kass Professor of Medicine, Channing Laboratory, Harvard Medical School, Boston, MA

Dr. Sverre Vedal, Professor, Department of Environmental and Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA

SCIENCE ADVISORY BOARD STAFF

Dr. Holly Stallworth, Designated Federal Officer, EPA Science Advisory Board Staff Office

NOTICE

This report has been written as part of the activities of the EPA's Clean Air Scientific Advisory Committee (CASAC), a federal advisory committee independently chartered to provide extramural scientific information and advice to the Administrator and other officials of the EPA. CASAC provides balanced, expert assessment of scientific matters related to issues 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 within the Executive Branch of the federal government. In addition, any mention of trade names or commercial products does not constitute a recommendation for use. CASAC reports are posted on the EPA Web site at: <http://www.epa.gov/casac>.

Enclosure B

Risk Assessment to Support the Review of the PM Primary National Ambient Air Quality Standards - External Review Draft (September 2009)

CASAC Responses to Agency Charge Questions

- 1) **After careful consideration of evidence provided in the second draft ISA and of the views expressed by the Panel in consulting on the Scope and Methods Plan, we have decided to quantitatively assess risk associated with both short- and long-term exposure to PM_{2.5} only. Based on our consideration of the evidence for health effects potentially associated with short-term exposure to PM_{10-2.5}, as well as to ultrafine particles and specific components, and in recognition of the limited available air quality data, we decided not to assess risk quantitatively for PM_{10-2.5}, ultrafine particles, or specific PM components as part of the current assessment. Is the Panel generally supportive of this scope? To what extent is the rationale for this decision clear and appropriate?**

Abundant epidemiological, clinical and animal toxicology studies support a causal relationship between exposure to PM_{2.5} and cardiovascular and respiratory disease. At the present time there is insufficient evidence of causal relationships between PM_{10-2.5}, ultrafine particles, or PM components. The rationale for this decision in the RA is both clear and appropriate. Based on these considerations, CASAC concludes that it is appropriate to focus the proposed quantitative risk assessment on PM_{2.5}. Our letter proposes that consideration be given to ways to carry out some exploratory analyses related to PM_{10-2.5}. Given the lack of sufficient information available for quantitative risk assessments on PM_{10-2.5}, UFP, and/or PM components, CASAC recommends that the Agency focus future research efforts on these topics. Moreover, this research should be supported with more comprehensive monitoring that is not limited to measuring only mass and particle size range.

- 2) **The final set of health effect categories included in the risk assessment for PM_{2.5} are consistent with those outlined in the Scope and Methods Plan (i.e., those classified as having a *causal or likely causal* association with PM_{2.5} exposure, as presented in the second draft ISA). We decided not to include any of the health effect categories classified as *suggestive* of a casual association in the second draft ISA, based on a number of considerations as described in section 3.3.1. Please comment on the approach taken and on the clarity of the rationale for selecting health effect categories for inclusion in the quantitative risk assessment.**

It is reasonable to exclude the “suggestive” category of endpoints from the quantitative Risk Assessment. CASAC certainly supports inclusion of the “causal” and “likely to be causal” categories in the Risk Assessment. The discussants unanimously appreciated the clarity in describing the approach and rationale for inclusion of the categories in the risk assessment.

The panelists, however, recognized that large segments of the potentially at risk populations for other outcomes (reproductive outcomes and lung cancer) would be left out of the Risk Assessment by this decision. One panelist suggests that a selective semi-quantitative risk assessment be carried out to provide the Administrator with a range of risks for these conditions to help in considering the margin of safety in selecting a standard. Another panelist raised

concern with the risk assessment for PM_{2.5}, which is a mass-based standard, as chemical components or particle number, rather than mass *per-se* may be driving the health effects.

3) Based on consideration of evidence presented in the second draft ISA, we have identified four combinations of 24-hour and annual alternative standard levels for analysis in the risk assessment. Please comment on the extent to which the rationale provided in section 2.5 appropriately supports these combinations of alternative standard levels for this assessment.

The discussion of combinations of alternative standard levels is focused on ambient concentrations of PM_{2.5} associated with adverse health outcomes in multiple large multi-city epidemiological studies. The advantages of utilizing these studies are clearly conveyed, and their use is justified.

Section 2.5 of the Risk Assessment focuses exclusively on fine particles, and thus is based entirely on PM_{2.5}. This scope is reasonable, since the most extensive epidemiologic evidence comes from studies on health outcomes in relation to PM_{2.5}. The Panel remains concerned about the need to consider coarse and ultrafine particles because evidence indicates that they are also associated with adverse health effects. New standards based on these other size fractions may be impractical now because of inadequate monitoring and because of limited epidemiologic data. However, we recommend that the Risk Assessment and Policy Assessment specifically address their potential risks and the need for further research.

The Risk Assessment adequately develops the rationale for alternative long-term PM standards. It was noted that the alternatives were close to each other, but CASAC thought that given the paucity of evidence at lower concentrations, a wider range of alternatives was not possible. Moreover, even a small increase in stringency of the NAAQS will increase the number of exceedances, and thus have implications for control.

Finally, there is a need for long-term monitoring and health effects research in order to better characterize the actual distribution of PM particles in the atmosphere and the sources, composition and effects of the different size fractions. CASAC encourages a long-term strategy to regulate coarse and ultrafine particles.

4) General approach:

- a) For this assessment, we have developed a primary set of risk results based on the application of modeling element choices (e.g., concentration-response (C-R) functions, lag periods) that we believe have the greatest overall support in the literature (referred to as the “core” results). As discussed in sections 2.4.1, 3.1 and 4.0, while it is not possible at this time to assign quantitative levels of confidence to these core risk estimates, staff believes that these estimates are generally based on inputs having higher overall levels of confidence, relative to risk estimates that could have been generated using other inputs identified in the literature. Consequently, the core risk estimates receive greater focus when we present, summarize and discuss risk estimates. What is the Panel’s view on the approach used and does the Panel consider it to be described appropriately and clearly?**

Overall, the Panel thought that a stronger justification was needed for the designation of the selected risk estimates as “core” and as being more certain than other potential inputs.

With respect to core concentration-response (C-R) functions, the methods used to select the C-R functions were clearly presented. The rationale for the focus on multi-city studies is clearly presented and appropriate, as is the basis for including select single city studies to assess PM_{2.5} mediated emergency department (ED) risks. However, the approach to selecting the particular multi-city studies is not sufficiently clear. There is no presentation of the rationale for selecting among several large multi-city time series studies. The same gap is evident with regard to the choice of long-term exposure studies, given the several available from which to choose.

CASAC found insufficient clarity on how the overall level of confidence or certainty was determined. The criteria used to determine staff decisions regarding “estimates...having higher overall levels of confidence” should be explicitly stated. Supporting analyses of the quantitative degree of uncertainty should be provided.

- b) Based on consideration of uncertainties associated with specifying C-R functions below the lowest measured level (LML) from a particular epidemiological study, we have decided to model risk for long-term PM_{2.5} exposures down to the LML, but not to extrapolate down to policy-relevant background (PRB). In contrast, when estimating risk associated with short-term PM_{2.5} exposures, because the LML is generally below the range of PRB values on some days during the study periods evaluated, we decided to model short-term risk down to PRB (see section 3.1). Is the Panel generally supportive of these approaches?**

There was support for modeling risk for long-term PM_{2.5} exposures to the lowest measured level (LML). For short-term risk estimates, EPA’s approach is appropriate, since as the document points out, the LMLs (which are daily) are below the PRB.

The results are rather dramatic and consistent. For the long-term effects, it appears that for substantial reduction of risk to be made, the alternative levels of PM_{2.5} for the NAAQS must be 12/25 µg/m³. In contrast, for short-term effects, little additional reduction is gained below a pairing of 13/35 µg/m³. The next draft will need to address this difference between the implications of the two sets of estimates.

5) Air quality inputs:

For this assessment, we have included an alternative approach for simulating air quality levels that just meet either current or alternative suites of standards in addition to the proportional analysis that has been used in previous analyses. Specifically, we have employed a hybrid (non-proportional) air quality adjustment procedure which simulates a combination of regional and local controls. The non-proportional rollback approach was used as part of a sensitivity analysis to examine uncertainty associated with this aspect of the risk assessment, while the historical proportional approach was used for the core analysis. To what extent does the Panel support the use of the alternative non-proportional rollback approach in the context of the sensitivity analyses? Please provide comments on the alternative approach as presented in section 3.2.3 and Appendix B.

The inclusion of a hybrid approach is appropriate given the various uncertainties. While the approach is *ad hoc*, any method would be at this point. A simplified approach allows an assessment of the sensitivity to a recognized concern that in some locations, a blend of local and

regional controls will be needed to reach attainment. The current approach likely overestimates this effect in some locations (e.g., those areas mainly affected by mobile source and secondary PM) and underestimates it in others (where specific industries and activities have a major local influence). The approach developed is reasonable, and of appropriate complexity, given the uncertainty as to the approach to lower emissions in specific geographic regions.

The explanation in Appendix B is clearer than the explanation in Section 3.2.3. As this section is relatively brief, it may be worth folding this appendix into the body of the report.

6) Selection of urban study areas:

We have included 15 urban study areas in the risk assessment, with the selection of these areas being based on a number of criteria as presented in section 3.3.2. To what extent does the Panel support the rationale provided for selection of the urban study areas and the specific locations considered?

The Risk Assessment understandably focuses on risk in the urban study areas, where the population is concentrated; and there appears to be an appropriate selection of cities, using defined criteria. However, little information is provided on rural PM effects. We are aware of regional data estimates that could have been used to justify inclusion (or exclusion) of these regions in carrying out the risk assessment. Some discussion of major traffic corridors in rural areas would seem warranted, if only to justify the exclusion of these data. The Risk Assessment should further consider exposures to rural populations and potential health effects. A discussion of east/west differences in the representativeness of PM_{2.5} as an indicator of the total PM effect by region would also be relevant. Although Appendix D provides comparisons of potential confounders of exposure within the 15 urban, these do not include evidence related to more rural regions. These considerations are relevant to the decision by EPA to carry out risk assessments only for PM_{2.5}. The regional differences in the proportion of total PM that is in the PM_{2.5} range varies widely. Thus risks associated with other components of PM may be selectively excluded by focusing only on urban areas.

7) Selection of epidemiological studies and C-R functions within those studies:

In estimating risks associated with PM_{2.5} exposures, we focused on selecting C-R functions from large multi-city studies based on staff's conclusion that these studies provided more defensible effect estimates (see section 3.3.1). Concentration-response functions from several single-city studies evaluating short-term PM_{2.5} exposures were also included to provide coverage for additional health effect endpoints (e.g., emergency department visits). To what extent is the Panel supportive of this approach for selecting C-R functions for modeling risk related to short-term and long-term PM_{2.5} exposures?

- a) **Specifically with regard to short-term exposure-related mortality, focusing on a study of 112 US cities by Zanobetti and Schwartz (2009), we obtained Empirical Bayes "shrunk" city-specific estimates from the study authors that provided a distinct C-R function for each urban study area location. For short-term exposure-related morbidity, focusing on a study of 202 U.S. counties by Bell et al. (2008), we used regionally-differentiated effect estimates provided by the study authors. Please comment on the selection of C-R functions for evaluating short-term morbidity and mortality effects. To what extent do the Panel members consider the rationales supporting the selection of C-R functions to be clearly and appropriately presented?**

The decision to emphasize multi-city studies and the reasons cited to support that choice are sound (p. 46). CASAC agrees with the specific choices of studies from which short-term C-R functions and with use of “shrunk” estimates for these functions. Because there is another large multi-city mortality study that utilizes PM_{2.5} as the exposure metric (Dominici et al. 2007), it is not absolutely clear why the Zanobetti et al. study was selected, although, as noted above (charge question 4a), it is a reasonable choice. That study also satisfies the selection criteria used (p. 46).

The selected single-city studies used in the Emergency Department (ED) risk assessments were also appropriate choices. Because the effect estimates from the single-city studies do not carry the same weight as those from the multi-city studies, additional care will be needed in interpreting the ED risk estimates. Also, these ED risk assessments are necessarily limited by being particularly relevant to the cities from which they were generated and of less certainty for others.

EPA should consider, as a sensitivity analysis, applying the large region-specific C-R functions from the Zanobetti et al. multi-city study (p. 47), instead of just the city-specific shrunk estimates, in order to allow the mortality risk estimates to more closely parallel the morbidity (hospitalization) risk estimates, which could only be based on region-specific (albeit different regions) effect estimates available from the Bell et al. study.

The choice to use the 2-day lag effect from the Bell et al. study for the respiratory hospitalizations C-R function, largely because the effect estimate was the largest, is less defensible – this is likely to be biased high. Support in the Risk Assessment for this choice is also based on a conclusion from the ISA (section 2.4.2.2) that respiratory health effects are strongest at a lag of 2 days. This conclusion is not supported by a review of the relevant tables in the ISA (Tables 6-10 – 6-14). Sensitivity analysis results with different lag estimates from the Bell et al. study (only lag 0 and lag 2 effects were published) would be informative. Consideration of lag choice should not be restricted to an evaluation of the Moolgavkar et al. (2003) studies.

- b) Specifically with regard to long-term exposure-related mortality, we identified a number of effect estimates using the extended follow-up of the American Cancer Society (ACS) study to use in the core analysis (Krewski et al., 2009). These effect estimates include standard Cox proportional hazards models, with 44 individual and 7 ecologic covariates, derived using two separate PM_{2.5} monitoring data sets (i.e., 1979-1983 and 1999-2000) (see section 3.3.3 of the RA). To what extent is the rationale for these choices clear and sufficiently justified as the basis for the core analysis involving long-term PM_{2.5}-related mortality?**

The choice of the specific study (the extended follow-up of the ACS cohort –Krewski et al. 2009) from which long-term C-R functions were derived is justified. However, there are now several large cohort studies that could potentially be used for this purpose. Specific justification for selecting the Krewski et al. study over these other cohort studies is not presented. A sense of the range of mortality effect estimates from the several cohort studies *vis-a-vis* the Krewski et al. study results, can be obtained by examining Figure 7-7 (p.7-124) of the ISA. The range of effect estimates is large, and indicates that the effect estimate chosen for use in the risk assessment would have a large impact on risk estimates. To avoid the need for an extensive sensitivity analysis that employs effect estimates from other cohort studies, a better justification for using

the Krewski et al. effect estimates is needed. Alternatively, some sensitivity analyses could be presented in the next draft.

Because of the myriad model estimates presented in the published findings, selection of “core” effect estimates was required for the risk assessment. The decision to select estimates from analyses that used the standard Cox proportional hazards model (i.e., the fixed effects model), that used a large set of individual-level covariates and a set of ecologic covariates, and that were based on two separate PM_{2.5} monitoring periods is reasonable. Justification of some of the features of the specific “core” model selected is relegated to a footnote (fixed effects vs. random effects, p.49). With regard to some model specifications, however, it is not immediately obvious which were used for the “core” risk estimates, and which were relegated to a sensitivity analysis. Without reviewing the original Krewski et al. report (only the log-linear specification was used for the fixed effects analysis), it only becomes clear that the log-linear specification of the model is being used for the core risk analysis by examining Table 3-8 (p. 61) in which it is noted that the other model specifications are examined in a sensitivity analysis. It is recommended that these model choices be made more obvious.

8) Addressing uncertainty and variability:

(a) Addressing uncertainty and variability -- The treatment of uncertainty and variability in the analysis is based on the multi-tiered approach presented in a recent WHO document (WHO, 2008). Specifically, as outlined in section 3.5, we have included qualitative analysis of both variability and uncertainty (WHO Tier 1), as well as single-factor and multi-factor sensitivity analyses aimed at identifying which potential sources of uncertainty have the greatest impact on the core risk estimates (WHO Tier 2). In addition, the sensitivity analyses have been designed to provide a reasonable set of alternate risk estimates to supplement the core risk estimates and inform consideration of uncertainty associated with the core analysis. To what extent does the Panel support the overall approach for addressing uncertainty and variability? To what extent does the Panel agree that the overall approach is appropriate and consistent with the goals of the risk assessment as outlined in chapter 1? Does the Panel have any recommendations for improving the characterization of variability and/or uncertainty?

The overall approach is reasonable, appropriate, and consistent with assessment goals.

As pointed out in Table 3-13, the largest source of uncertainty in the risk assessment is exposure misclassification, which may lead to bias and imprecision in risk estimates. Its potentially large contribution to uncertainty points to the importance and need for an exposure assessment to help interpret any estimates of risk. For example, information about inter-individual variability in exposures could be used to help assess potential bias in concentration-response functions, while information on factors that modify the exposure-ambient relationship could be used to interpret geographic or other heterogeneity in risk. We recommend that these exposure assessments be performed to clarify and highlight issues related to misclassification of exposures and heterogeneity in risks. While we understand that the Agency is not yet ready to apply probabilistic exposure assessment methods, we believe that simpler, non-probabilistic exposure assessment methods can be used for these purposes. We further urge the Agency to continue developing probabilistic exposure methods so that they will be ready for the next NAAQS review cycle.

- (b) The qualitative discussion of key sources of variability, and the degree to which the analysis design captures those sources of variability, are presented in section 3.5.2. Please provide comments on the approach used. Specifically, do the analyses sufficiently address the issue of variability? Are there key sources of variability that have not been addressed within the qualitative analysis but which could have an important impact on modeling population-level risk associated with PM_{2.5} exposure?**

Section 3.5.2 is generally very good. However, the use of the term “Key” in this section raises the question as to how the various potential sources of variability were compared and prioritized, and implies that there may be other sources of variability that are not “key.” This should be clarified and explained.

Six key sources of variability were identified and addressed qualitatively. These should either be modified or augmented to include differences in PM co-pollutant concentrations in the context of source variability, and to include land use, source locations, housing stock, and socio-economic factors in the context of demographics.

EPA should take credit for the sources of variability that are quantified in the assessment, such as spatial and temporal variability in ambient PM_{2.5} concentration. Thus, a list should be given of sources of variability that are quantified.

Although some of the factors discussed here are not quantified in terms of attempting to apportion exposure or risk by composition, demographics, activity patterns, and so on, EPA should provide insight into the variability of these factors and their implications for variability in the risk estimates. This examination could be analogous to the analyses done to examine the generalizability of the 15 cities to the rest of the US.

- (c) Table 3-13 provides a qualitative characterization of uncertainties including the potential direction, magnitude, and degree of confidence associated with our understanding of the sources of uncertainty. To what extent does the Panel support the characterizations of the key sources of uncertainties identified and the relative rankings of the importance of those sources of uncertainty? Are there additional uncertainties that should be considered?**

Table 3-13 is excellent. The panel supports the material contained in this table.

A source of uncertainty that was not included was the C-R function itself, which was developed from single studies. The source entitled “transferability of C-R functions from study locations to urban study area locations” should also take into account differences in C-R functional form associated with studies that addressed long-term or short-term effects for single or multi-city studies even if they were not the basis for the final set of C-R functions used in the RA.

Definitions should be given for the categories “low,” “medium,” and “high,” as a footnote in to the table with some discussion in the text. In the text, EPA states that “we” characterized uncertainty: it should be made clear as to whom is “we” and by what process and using what objective criteria were used for arriving at the uncertainty categories. The main point is to convey that these categories are not arbitrary, and were decisions made via a process.

Furthermore, it should be stated as to whether these categories are relative to each other, or based on some absolute scale of uncertainty.

EPA should comment on the extent to which there are dependencies among pairwise combinations of sources of uncertainty, and whether these dependencies would tend to offset or to increase the overall range and direction of uncertainty in the assessment results. For example, the statistical fit of the C-R functions, and the shape of the functions, are inter-related.

There should be a summary that describes implications of these uncertainties, including their relative importance, for interpreting results of the RA.

In the Results Section (Section 5.3), the results should be interpreted with respect to key sources of uncertainty – i.e. how robust are the results, and what are the likely biases. In particular, given exposure misclassification, it is likely that the estimates of Table 5-1 are biased low. This point should be conveyed consistently.

(d) The results of the sensitivity analyses have been used to gain insights into which sources of uncertainty significantly impact the core risk estimates and to provide a reasonable set of alternate risk estimates to supplement the core analysis. We are mindful that these estimates do not represent a true uncertainty distribution. With regard to the single- and multi-factor sensitivity analyses, to what extent is the Panel supportive of the approach used to conduct and characterize the results of the sensitivity analyses? Please provide comments on the extent to which the presentation of the results of the sensitivity analyses are clearly and reasonably described? Does the Panel have any recommendations for how the results of the sensitivity analyses could be used more effectively or appropriately in characterizing uncertainty associated with the core risk estimates?

The evaluation of alternative model structure is critically important, because model structure can potentially be a larger source of uncertainty than the range of values for an input to a given model. The range of uncertainty associated with confidence intervals for a given C-R function (an example of a Tier 3 assessment, which should be mentioned) should be compared to the range of estimates obtained by comparing alternative functional forms. This comparison would provide insight as to whether model structure or random error for a given model is a more important source of uncertainty.

EPA should indicate the direction of the percent changes in risk. In addition to the percent difference, the actual difference in risk should be reported to provide further context. This section should conclude with a brief but explicit summary of the decision to use the sensitivity results only from the long-term exposure mortality analysis, which is mentioned in Section 4.5.2.

The alternative model specifications appear to lead to larger values of adverse outcomes, indicating that uncertainties may be positively skewed. This is a notable finding and should be discussed. The implications could be that the point estimates of risk used for the assessment are under-estimating the true but unknown risk in part because of uncertainties in model structure, as well as because of the exposure misclassification issue.

A key question for the sensitivity analysis is whether the range of risk estimates is useful– i.e. do the lower and upper bounds from the results (as shown later in Figure 4-22) represent plausible lower and upper bounds on the true but unknown answer? For some readers and decision makers, a key question is whether the lower bound of the sensitivity analysis results (of 1.3% of total incidence of all cause mortality attributable to PM_{2.5}) is significantly greater than zero.

The results of the sensitivity analysis should be compared with the results from the qualitative assessment of uncertainty to offer judgments on the following: (a) how would the qualitatively characterized sources of uncertainty affect the quantitative answers (e.g., because of bias from exposure misclassification, the actual percent total incidence is expected to be higher than the numbers shown here); (b) what is the relative importance between the factors in the sensitivity analysis and the qualitatively assessed uncertainties; and (c) what is the bottom line in terms of a judgment regarding the robustness of the effects estimates?

9) A number of risk metrics as well as different approaches for presenting these metrics are included in tabular and graphical format for both the core analysis and sensitivity analyses. Please comment on the extent to which the risk estimates are clearly and appropriately characterized and presented.

A number of risk metrics as well as different approaches for presenting these metrics are included in tabular and graphical format for both the core analysis and sensitivity analyses.

Generally, the approaches and metrics for assessing various health risks are logically conceived, and the results of the “core” risk estimates, and sensitivity analyses are clearly presented in Chapter 4. The number of figures could be greatly reduced, since some appear almost identical and might better be moved to an appendix.

10) Evaluation of the representativeness of the urban study areas in the national context:

We completed a comparison of the 15 urban study areas against national distributions for key PM risk-related attributes. The goal of this analysis was to determine whether the urban study areas are more nationally-representative of these attributes, or are more focused on a particular portion of the distribution for a given parameter. In particular, given that one of the goals of the risk analysis is to provide estimates of risks for those areas likely to experience high levels of PM exposure and risk, this assessment provides insights as to the extent to which the assessment represents high PM_{2.5} risk locations. The results of this analysis were then used to evaluate, in part, whether the set of 15 urban study areas is likely to reflect the broader U.S. population with regard to PM_{2.5}-related risk, including coverage for those locations that represent specific at-risk populations. To what extent does the Panel support the approaches used? Please provide comments on the clarity with which the methods, results, and insights gained from this analysis are described.

We would characterize the question as referring to the representativeness of calculated health risks for the 15 selected urban sites. We have restated the question in terms of the calculated risks, rather than of the urban study areas, as the question implies that we accept the conclusions and figures in the preceding part of Chapter 4. There are some reasons why the premise may not be sound.

1) The results presented in Figures 4-1 through 4-13 are notable in several ways. They all look overly similar, with 13 of the 15 bunched closely together, and Tacoma being an outlier on the high end of risk, and Salt Lake City being an outlier on the low risk side. Does applying a national framework result in too similar risk estimates for 13 of the 15 cities? What accounts for the outliers? Are there differences in underlying disease risk, particularly comparing Salt Lake City and Tacoma? 2) Bringing up the outlier nature of the underlying mortality rates for New

York City as a footnote on page 99 is insufficient in terms of the adequacy of the models for predictive purposes. An effort should be made to understand what is driving the real, versus, the hypothetical risks, in our largest city, where there is a considerably higher IHD risk than in other cities, while at the same time, a considerably lower baseline all-cause mortality risk.

3) In more directly addressing Charge Question #10, the limitations of the modeling approach used, discussed above, warrant caution with regard to the extrapolation to risks in the total US population. In terms of applicability of urban (and suburban, i.e., the 31 county areas outside the 15 urban areas) to the population as a whole, including those living where there is essentially no PM or health risk data, the best approach may be to estimate exposures to regional levels of long-range transported PM_{2.5}, which vary greatly by geographic region and may have their own particular risks. The emphasis is on large cities where there is exposure to fresher traffic-related pollution that is not experienced so widely in other locations. For the populations living outside such areas where there are applicable data, it may be best to consider them not to differ substantially from the suburban populations in susceptibility determinants. The urban populations provide, as indicated in the RA, the best data for the upper end risks.

11) We completed a national scale assessment focused on long-term mortality associated with recent air quality conditions. To what extent does the Panel support the approach used? Please comment on the clarity and appropriateness with which the methods, results, and insights gained from this analysis are described.

The approach is reasonable based on the analysis presented in Section 4 and it is presented adequately. In the next draft, the national scale assessment should include an assessment associated with alternative standards. The results of the national scale and city-specific results should be synthesized, identifying considerations that might drive the revision of the NAAQS, and this synthesis should include how the alternative PM levels compare with risks recently calculated for the NAAQS for other criteria pollutants. (This should also be included in the Policy Assessment). The analysis indicates that the use of 2006-2008 baseline data has little effect on the results presented in section 5 as indicated by a sensitivity analysis. Results of this analysis should be in the text.

12) The national-scale long-term mortality risk assessment provides perspective for where the 31 counties associated with our 15 urban study area analysis fall along the national distribution of mortality risk? We note that this analysis is distinct from the representativeness analysis referenced above (and described in section 4.4) in that this analysis focuses on coverage of the 15 urban study areas for long-term mortality risk, while the earlier representativeness analysis focuses on coverage for PM risk-related factors. To what extent is the Panel supportive of this specific analysis and its intended use to provide insights into the extent to which the urban study area analysis broadly represents urban PM_{2.5}-related risk in the U.S?

The results in Section 5 are presented with sufficient clarity, and provide information as to how one might interpret the information provided by the detailed assessments presented in Chapter 4. Providing Figure 5-4 as an analysis of where the 31 counties included in the urban case study counties and showing that the chosen areas fall near the top of the CDF in the overall national risk distribution is helpful for putting the results in context. Instead of simply mentioning 2 representative counties in the lower end of the distribution and 2 in the upper end, a more

complete description of these specific areas should be included in the next draft, and there should be a reference to where a more complete description of all 31 is located elsewhere in the document. It would be useful to take the information in Chapter 4 and develop a national assessment for endpoints other than long-term mortality and to synthesize the results between the two chapters.

Enclosure C

Compendium of Comments
CASAC Particulate Matter Review Panel on
Risk Assessment to Support the Review of the
PM Primary National Ambient Air Quality Standards (September 2009)

Avol Comments (Professor Ed Avol)	2
Brain Comments (Dr. Joe Brain)	3
Cascio Comments (Dr. Wayne Cascio)	5
Henderson Comments (Dr. Rogene Henderson)	7
Helble Comments (Dr. Joe Helble)	8
Hopke Comments (Dr. Phil Hopke)	11
Lippmann Comments (Dr. Mort Lippmann)	12
Phalen Comments (Dr. Robert Phalen)	15
Pinkerton Comments (Dr. Kent Pinkerton)	16
Poirot Comments (Mr. Rich Poirot)	18
Russell Comments (Dr. Ted Russell)	20
Speizer Comments (Dr. Frank Speizer)	22
Suh Comments (Dr. Helen Suh)	24
Vedal Comments (Dr. Sverre Vedal)	28

Avol Comments (Professor Ed Avol)

I found the Risk Assessment document to be thoughtfully done, in great detail, and with many useful linkages to the previous review, CASAC comments, and agency decisions such that a demonstrably logical evolution to the current document was clear to the reader.

My specific charge was to consider the rationale and presentation for selection of the 15 urban study areas for subsequent risk assessment use. Here too, I thought the presentation was well-crafted, well-supported, and carefully linked to available data and design concerns.

The document understandably focuses on risk in the urban study areas (where major population concentrations are), but I was left wondering about rural regional PM effects. Do the considerations presented in the document and prioritized in the approaches utilized provide any substantive insights for rural population, rural exposures, and rural health effects? If any comments about these issues were presented in the main body of the document, I apologize for missing them, but they were not readily apparent. I raise this issue because recent studies have repeatedly emphasized the importance of near-road and proximity exposures, but there are reports of both local and regional effects on respiratory health (and possibly other health outcomes with which I am less familiar). In thinking about protecting the health of the public, shouldn't some comment regarding this segment of the public be included, or at least acknowledged?

Brain Comments (Dr. Joe Brain)

RA: (6) Selection of Urban Study Areas

Table 3-4 and the accompanying map, Figure 3-4, shows a reasonable distribution of the 15 urban study areas selected. They span the country, and they encompass varying mixes of pollutant sources and different meteorological conditions. In part, the rationale for the selection of these urban study areas reflects practical considerations, such as the availability of data and the relationship between these locations and the availability of appropriate epidemiologic studies. I also like the criterion of selecting locations that provide heterogeneity in regard to risk factor and demographics (for example, SES status, use of air conditioners, ethnicity, and PM sources).

Some concerns persist, such as the location of monitoring stations and their relationship to the most common human exposures. For example, this section and the current strategy do not deal adequately with the issues of heterogeneity of exposure. How do we include proximity to roadways or special sources like cement plants? *In toto*, however, section 3.3.2 seems well written and reasonable.

Response to Charge Question 3

Based on consideration of evidence presented in the second draft ISA, we have identified four combinations of 24-hour and annual alternative standard levels for analysis in the risk assessment. Please comment on the extent to which the rationale provided in section 2.5 appropriately supports these combinations of alternative standard levels for this assessment.

At the center of a discussion of relevant combinations of alternative standard levels is the range of ambient concentrations of PM_{2.5} associated with adverse health outcomes in multiple large multi-city epidemiological studies. The advantages of utilizing such studies are clearly conveyed, and their use appears to be justified.

There is an assumption early in section 2.5 that bears additional thought. The risk assessment focuses exclusively on fine particles, and thus is based entirely on PM_{2.5}. Perhaps this makes sense, since the most extensive epidemiologic data is health outcome in relation to PM_{2.5}. Do we have any reservations regarding ignoring coarse or ultrafine particles? A standard based on these other size fractions may be impractical, but can we indicate more clearly their presence and potential contribution.

Staff does a reasonable job of developing the rationale for the long term standard: 13 µg/m³ vs. 12 µg/m³. But they seem so close to each other. Do they represent significant alternatives? Can our relatively crude sampling strategies effectively distinguish 13 vs. 12? Either one is below the current annual standard. We are not sure there is any practical difference between the two. A more interesting alternative would be 11 µg/m³. Then we would have the current standard of 15

$\mu\text{g}/\text{m}^3$, which could be compared to $13 \mu\text{g}/\text{m}^3$ and $11 \mu\text{g}/\text{m}^3$. The 24-hour standard exhibits a greater range: 35 vs. 30 vs. $25 \mu\text{g}/\text{m}^3$.

Given the current alternatives, why not eliminate the third bullet from the bottom, “Alternative $\text{PM}_{2.5}$ standards: annual $12 \mu\text{g}/\text{m}^3$; 24-hours $35 \mu\text{g}/\text{m}^3$. Then we would be left with three alternative $\text{PM}_{2.5}$ standards, which would progressively be more conservative: 13 and 35, 13 and 30, 12 and $25 \mu\text{g}/\text{m}^3$. Currently, alternative 1 vs. alternative 2 offers too little choice, and the rationale for choosing between them seems unclear.

Cascio Comments (Dr. Wayne Cascio)

Charge Question 1

EPA staff has produced a very readable document and the decisions are clearly stated and largely justified. There is full agreement to exclude ultrafine PM and specific components from a risk assessment at this time. The question is whether the decision reached regarding the exclusion of PM_{10-2.5} is appropriate and can be supported by CASAC.

In the document, Section 2.2 describes the “Original Assessment Plan” for the risk assessment in which EPA staff proposed to: 1) a limited assessment of PM_{10-2.5}, and 2) to exclude ultrafine PM and specific components because of a lack of evidence to support a quantitative risk assessment. At that time a causal relationship of “suggestive” was felt sufficient to justify a limited risk assessment for PM_{10-2.5}. Subsequently, CASAC recommended in Dr. Samet’s letter to the Administrator that “priorities be established quickly in developing the health risk and exposure assessment, giving emphasis to those analyses that may be most informative for established PM standards”, and “provide a transparent algorithm for selecting endpoints based on the level of certainty and the relative and attributable risks.” Because PM_{10-2.5} was determined to be only “suggestive” of a causal relationship, and the health effects were less certain this previously planned analysis was dropped. This is a logical judgment given that the insufficient evidence and greater uncertainties attributed to PM_{10-2.5} are likely to result in more uncertainty in risk assessments which in turn and might be interpreted inappropriately without considering these uncertainties.

Nevertheless, several CASAC members view this decision to exclude a risk assessment of PM_{10-2.5} as too restrictive and does not place enough weight on the future needs of the Administrator to address the PM₁₀ standard or the unique constellation of health effects produced by coarse PM. Fig. 2.3 in the ISA clearly shows effect estimate data ordered by mean PM_{10-2.5} demonstrating a more consistent health impact at higher mean concentrations, and apparently independent of PM_{2.5}. While acknowledging the limited nature of the data and the greater uncertainty, there does appear to be a risk related to PM_{10-2.5} exposure. Providing at least a limited qualitative assessment of risk is recommended.

Is the Panel generally supportive of a quantitative risk assessment with both short- and long-term exposure to PM_{2.5} only.

Abundant epidemiological, clinical, and animal toxicology studies implicate a causal relationship between exposure to PM_{2.5} and adverse cardiovascular and respiratory outcomes, and mortality. Yet at the present time, and as summarized in the ISA there is inconclusive evidence of causal relationships between PM_{10-2.5}, ultrafine, or PM components and short-term and long-term cardiovascular and respiratory health endpoints. As such it is appropriate to focus the quantitative risk assessment solely on PM_{2.5} where there is convincing evidence of causality with cardiovascular effects and likely causal relationship with respiratory effects and mortality.

Likewise the long-term impact of PM_{2.5} exposure is well supported by the data and appears to be causal for cardiovascular effects, and likely causal for respiratory effects and mortality.

Charge Question 2

Comment on the approach taken and on the clarity of the rationale for selecting health effect categories for inclusion in the quantitative risk assessment.

For the purpose of the risk assessment an important issue is one of the certainty of the effect. Uncertainty in the level of association will be compounded by any risk assessment model and will yield predictions that will lack confidence. The present approach minimizes the uncertainty of the risk assessment by limiting the model development to only outcomes that are judged to highly and consistently associated to PM_{2.5} exposure, thereby judged causal or likely causal.

Charge Question 3

Comment on the extent to which the rationale provided in section 2.5 appropriately supports the four combinations of 24-hour and annual alternative standard levels for analysis in the risk assessment.

At the center of the determination of relevant combinations of alternative standard levels is the range of ambient concentrations of PM_{2.5} associated with adverse health outcomes and several large multi-city epidemiological studies. The advantages of utilizing such studies are clearly conveyed, and their use appears to be justified.

Henderson Comments (Dr. Rogene Henderson)

Charge question 1:

I am disappointed that the Agency still does not have enough information to evaluate the risk associated with exposure to coarse and ultrafine particles, but from a practical viewpoint, I think that is probably all you can do. The need to consider the composition of PM in relation to toxicity is major and should be addressed with some urgency. The NRC and BOSC have both urged the Agency to study this problem. The research required to address this issue is separate from what is normally done to set regulations. It is true that we do not now have the information to set a regulation based on PM composition, but I hope this will change in the future. Also,

I think we should point out that, in order to move in the direction of looking at PM composition, as the Agency has been urged to do, they are going to have to conduct more comprehensive monitoring and not just measure mass and size.

Charge question 2:

I very much agree to limiting the scope of the risk assessment to those health effects that fit in the causal or likely causal categories. The rationale for doing this was clearly presented. There is no indication that "suggestive" endpoints are more sensitive to PM exposure than the causal endpoints, so even if the "suggestive" endpoints are later found to be causal, the public should be protected by the standards set to protect against the causal endpoints.

Charge question 3:

I thought the rationale for the choice of possible short (25 or 30 ug/m³) and long term (12 or 13 ug/m³) PM standards to be considered was quite clear. The combinations shown were not as clear. For example, the short-term standards chosen were 25 or 30 ug/m³, but two of the combinations included 35 ug/m³. So I think the combinations chosen need a little more explanation

Charge Question 9:

A number of risk metrics as well as different approaches for presenting these metrics are included in tabular and graphical format for both the core analysis and sensitivity analyses.

Generally, the approaches and metrics for assessing various health risks are logically conceived, and the results of the "core" risk estimates, and sensitivity analyses are clearly presented in Chapter 4. The number of figures could be greatly reduced, since some appear almost identical and might better be moved to an appendix. While the PM Panel agrees that the risk assessment results based on a PM_{2.5} mass indicator are clearly presented here, we remain disappointed that no attempts were made to evaluate risks associated with the different PM components that are mixed in different proportions in the different urban areas."

Helble Comments (Dr. Joe Helble)

Chapter 3 – Scope , Charge Question 5

Air Quality Inputs: “Please provide comments on the alternative approach as presented in section 3.2.3 and Appendix B.”

The alternative, hybrid method used for simulating PM_{2.5} concentrations appears to be a reasonable approach to simulating the effects of local controls applied to point sources in combination with regional controls expected to achieve a proportional reduction in PM concentrations.

The explanation in Appendix B is clearer than the explanation in Section 3.2.3. Given that it is a relatively short amount of text, it may be worth folding this appendix into the body of the report.

Minor typographical errors in Section 3.2.3:

1. page 36 line 31, the \ should be deleted
2. page 37, line 20, missing left parenthesis in the denominator
3. page 38, line 6, missing left parenthesis in the denominator

Chapter 4 Results, Charge Question 9: “Please comment on the extent to which the risk estimates are clearly and appropriately characterized and presented.”

The risk estimates described in Chapter 4 of this document are presented at an appropriate level of detail. The initial sections of the chapter (4.1 and 4.2) are a bit tedious to read, but all of the necessary information is present. Later sections, particularly the sensitivity analysis, are well-written and clearly presented.

The tabular presentation of data in the Chapter and in the relevant appendices is generally clear. Tables are very detailed, and this is helpful when comparing risk assessment for different locations and different PM standards.

Table 4-1 is particularly clear despite the large amount of detail, and the individual references to specific tables in Appendix F are very helpful.

While the idea of including figures is sound, as they make it easy to compare the risk assessment resulting from the different NAAQS standards for a given location, overlapping city data make it difficult to follow trends. Many of the figures are nearly identical, and the data plotted are available in the accompanying tables. Given this, it might be better if only a representative year (2007) were shown. In addition, there is overlap between figures in the text and figures in Appendix E; it is not clear why both are needed. For example, Figure 4-1 is identical to Figure E-3, although the figure title is worded slightly differently. If this and similar figures are to be

included in the text, they should either be removed from Appendix E, or labeled identically in Appendix E with the name of the corresponding figure from the text indicated in a footnote.

Regarding Figure E-1 and related figures – since the independent variable here is in fact the current standard, recent [PM], or an alternate standard, it is misleading to label the axis “alternate standard.” “Current or Alternate Standard” or other terminology would be a more accurate descriptor of what is represented in the figures.

Additional editorial comments follow.

1. Page 92, line 19: define concentration-response here, rather than later (in line 23)
2. Page 92, line 22 – define PRB here (rather than later in line 25)
3. Page 93, line 8, 2nd word should be “effect,” not “affect”
4. Page 93 line 10, “are” should be changed to “is”
5. The entries in the tables (e.g, Table E-2) are the point estimate (absolute number) followed by the 95th percentile confidence intervals, as discussed on page 93, lines 16-20. This should also be noted in a footnote on the table.
6. A footnote regarding the significance of a negative value in the percent reduction tables (example, Table E-7) using language similar to that in the narrative (line 1, page 96) should be added to the relevant tables.
7. Page 99 line 2 – “Generally, results for the same are fairly similar... “ is followed by a discussion of one with 30% variability. It would be clearer if the text were more specific, e.g “...of the 15 cities considered, X were generally invariant (i.e. < y % year to year variation). Z of the cities showed greater variation, from *** up to 30%...”
8. p 100 line 30, “head” should be ‘had’
9. p. 101 line 15 – estimates, or estimate?
10. Discussion on p. 108 re number of monitors (line 4), would be helpful to reference Table 3-1 that lists the number of monitors at each study location
11. line 13, p. 123 – delete the right parenthesis)
12. p. 132 line 17, p. 134 line 34, p. 136 line 20 - shouldn’t this read Table 4-1?
13. p. 137 line 15, “compare” should read “comparison of”
14. p. 149 line 3 “shows” should be “show”
15. p. 149 line 19, “8oth” ‘should read “80th ”
16. explanatory notes written on Figure 4-14, 4-16 4-19 through 4-21 are helpful
17. p. 165 line 22, “see” should be “seen”

Chapter 5 – National Scale Assessment, Charge Question 11, Approach

The approach is reasonable based on the analysis presented in section 4. It is reasonably clearly presented.

p. 24 lines 24-25 indicate that the use of 2006-2008 baseline data have little effect on the results presented in section 5 due to a sensitivity analysis. Is this sensitivity analysis described elsewhere in the report (incl. Appendix G)?

Chapter 5 – National Scale Assessment , Charge Question 12, specific analysis

The results in Section 5 are fairly clearly presented. Providing Figure 5-4 as an analysis of where the 31 counties included in the urban case study counties lie on the overall national risk distribution is helpful for putting the results in context. Instead of simply mentioning 2 representative counties in the lower end of the distribution and 2 in the upper end, a more complete description of all 31 should either be included, or referenced here if it is located elsewhere in the document.

Minor comments on this section;

Page G-1, “Supplement” is spelled incorrectly in title

p. 172, line 3, pm_{2.5} should read PM_{2.5}

p. 172, footnote 48, last line, “is” should be “are”

Hopke Comments (Dr. Phil Hopke)

It is hard to provide any substantive comments on this document since they have followed the methodology that was laid out in the plan that had been previously reviewed by the Panel.

Air Quality Inputs

The air quality inputs are reasonable. Given what they want to explore, the rollback methodology that was used also seems reasonable.

Sources of Variability and Uncertainty

Again the approach is reasonable. Given the nature of the broader uncertainties in the whole RA process, I do not think there is much more that would be sensible to do.

Lippmann Comments (Dr. Mort Lippmann)

General Comments:

I commend OAQPS Staff for creating a straightforward text that clearly describes the objectives and methods used to develop risk assessments (RAs) for the PM_{2.5}-associated health effects judged in the PM ISA 2nd draft to be either causally or likely to be causally related to the exposures. As a long-term observer of the development of RAs for NAAQS, I am impressed with the progress that has been made in the development and applications of the methodology. I believe that the choices that were made in terms of monitoring data and concentration-response functions for this latest PM RA were reasonable and appropriate, and I therefore find the estimations of the health effects to be expected in meeting the current 15/35 $\mu\text{g}/\text{m}^3$ NAAQS, and of meeting the alternate NAAQS under consideration, provide a reasonable basis for the selection of the next suite of PM NAAQS.

I strongly recommend that the document be reduced in size to minimize the overlapping content with the PM ISA in terms of detailed descriptions of the key studies to be relied on to support the actual exposure and risk assessments. Readers should be referred to the details of the critical studies by citing the appropriate Section and/or page numbers in the ISA.

I did find a few nits to pick, and these are described below under Specific Comments.

Specific Comments:

Page(s)	Line(s)	Comment
18	1-4	Population exposure assessment is hardly a new issue in setting PM NAAQS. As long as it is focused on PM of outdoor origin, as it should be, I find little justification for putting it off until “the next PM NAAQS review.”
33	Table 3-1	There is only one New York City (NYC) risk assessment location. The Ito et al. (2007) was <u>not</u> restricted in New York County (Manhattan) but rather covered all five counties within NYC.
34	1 + 2	Delete
48	29	Delete “Manhattan”.
57	-	For the “Risk Assessment Location” entry for NYC, change the entry for “New York City (Manhattan)” to “New York City”, and indicate that the same five counties apply to the Ito et al (2007) paper.

66	Table 3-10	Correct the county listings for New York, NY as above.
72	11	Insert “be expected to” before “respond”.
75	5	Insert a comma after “variability”.
99	Footnote 38	<p>This footnote states: “Specifically, the baseline incidence rates for IHD mortality for New York City are 380 per 100,000 while national is 242 per 100,000 (See section 3.5, Table 3-9). This translates into New York City having approximately 1.5 times the rate of IHD deaths relative to the national average. All cause mortality baseline incidence also differs, although to a lesser extent, with New York City having 1,077 per 100,000 and the national average being 1,327 per 100,000. This translates into New York City having a baseline incidence rate for all-cause mortality that is 23% lower than the national average.”</p> <p>This cited quote is important information, which should be discussed in the final draft of PM ISA.</p>
109-116	Figures 4-1 through 4-8	<p>These figures indicate rather dramatic benefits from more stringent PM NAAQS in terms of reduced incidence of mortality (All cause, cardiopulmonary, and lung cancer), as well as substantial mortality benefits of achieving the pre-existing NAAQS for cities now in noncompliance. For 13 of the 15 cities, the 12/25 option would, if implemented, reduce PM_{2.5} –related mortality by 30-60% as compared to just meeting to current 15/35 NAAQS. It is important to note that the estimated benefits were based on neutral rather than conservative models of airborne PM_{2.5} concentrations and concentration-response relationships. While the estimations are subject to consideration of uncertainties in the data and models, it would be very hard to ascribe them to bias, and they do not rely on “margin of safety” considerations, which should, if anything, lead to even more stringent NAAQS.</p>
120-121	Figures 4-9 to 4-10	<p>Recognizing that the short-term mortality impacts of peak concentrations are considerably lower than those due to cumulative exposures, the similarities in the patterns lend credence to the validity of the benefits to be gained from more</p>

		stringent PM NAAQS.
125-126	Figures 4-12 to 4-13	The cardiovascular hospital admissions estimates are also supportive, especially in terms of new “coherence” with the mortality estimates.
130	Table 4-1	I suspect that the % differences have the wrong sign, since the alternate long-term exposure mortality study to the ACS cohort is the 6-cities cohort. The coefficient for 6-cities cohort is considerably greater than that for to ACS cohort.
133	24-40	I suspect that the % differences have the wrong sign, since the alternate long-term exposure mortality study to the ACS cohort is the 6-cities cohort. The coefficient for 6-cities cohort is considerably greater than that for the ACS cohort.

Phalen Comments (Dr. Robert Phalen)

Charge Question 2

The material is well written and the logic is clear with respect to the selection of health endpoint categories to include in the risk assessment. However, causality is a serious claim when it is applied to public health. Causality implies that the true culprit that is producing adverse health effects is known with sufficient certainty to both commit resources for control, and to disrupt people's lives (and productivity) in the process. Although the associations linking several health outcomes to PM_{2.5} exposures pass the criteria that EPA used to conclude causality, I don't believe that PM_{2.5} mass per se is responsible. Studies by Bell et al. (2008, 2009) convincingly indicate that of 20 components of PM_{2.5} only Vanadium, Nickel, and elemental carbon were statistically-significant with respect to cardiovascular respiratory hospital admissions in 65-plus year olds. The study included 106 continental U.S. counties with populations of 200,000 or more. The study represents a major advance in the process of uncovering valid specific causal factors. Having seen some unwelcome tradeoffs associated with mass-based PM NAAQS, I am concerned about accepting PM_{2.5} mass as a causal factor for adverse health outcomes. A formal risk assessment carries the assumption that the cause is clearly identified. As a result, I would drop the category "likely to be a causal relationship" from the health risk assessment. Also, the category "causal relationship" is questionable, and if it is included in the risk assessment, a discussion of the uncertainties related to a mass-based indicator should be added.

As an aside, now is not the right time to assume that the causal factors are known. True causal factors including specific components, component combinations, and exposure conditions must be identified in order to target efficient appropriate control actions. Inefficient control actions can do more harm to public health than good.

References cited.

Bell, M.L.; Ebisu, K.; Peng, R.D.; Walker, J.; Samet, J.M.; Zeger, S.L.; Dominici, F. (2008). Seasonal and regional short-term effects of fine particles on hospital admissions in 202 US counties, 1999-2005, *Am. J. Epidemiol.* 168(11):1301-1310

Bell, M.L.; Ebisu, K.; Peng, R.D.; Samet, J.M.; Dominici, F. (2009). Hospital admissions and chemical composition of fine particle air pollution. *Am. J. Respir. Crit. Care Med.* 179(12):1115-20.

Section 3.3, 2 reads well and I have no concerns regarding the rationale and study locations selected. Some minor suggestions are:

In Table 3-4: 1. Spell out "Los Angeles"; and 2. define "design value" in a table footnote, or in the main text.

Pinkerton Comments (Dr. Kent Pinkerton)

Comments to Charge question 1: The concept that PM_{2.5} drives health effects is correct, but the coarse and ultrafine size fractions of particles should not be excluded. The robustness of the available data is found in PM_{2.5}. The authors state the overview and discussion of key components of the quantitative risk assessment for PM_{2.5} are also applicable to the risk assessment conducted for PM_{10-2.5}. However, the scope of the risk assessment for PM_{10-2.5} is much more limited, reflecting more limited epidemiological data and air quality information available for PM_{10-2.5}.

Risk Assessment: key design elements (10 listed – 9 will be used)

- 1) PM size fractions
- 2) Selection of health effects categories (PM_{2.5})
- 3) Selection of health effects categories (PM_{10-2.5}) – not considered
- 4) Selection of urban study areas
- 5) Simulation of air quality levels that just meet either current or alternative standard levels
- 6) Characterization of PRB
- 7) Selection of epidemiological studies to provide C-R functions
- 8) Characterization of uncertainty and variability
- 9) Representativeness analysis for the urban study areas
- 10) National-scale health impact analysis

The above design elements emphasize the need to clearly outline the purpose of the analysis, with specific ways in which the results would be used to interpret the estimates generated from the risk assessment. The EPA authors have done an excellent job to communicate each of these points in Chapter 2 of the Risk Assessment document. It will be interesting to see the outcome of EPA's further development of population exposure analysis methodology.

In general, I would agree with the decision to only use PM_{2.5} for risk assessment for short-term and long-term exposure. The EPA does acknowledge that evidence for health effects associated with thoracic coarse particles, UFPs and PM components will be addressed as part of the evidence-based analysis to be presented in the forthcoming draft Policy Assessment (PA).

Comments to Charge question 2RA, with emphasis on section 3.3.1.

For PM_{2.5} risk assessment, I favor both causal and likely causal relationships for quantitative risk assessment. It is my opinion the exclusion of health categories as suggestive of a causal association does not seem to greatly lessen the power of the risk assessment outcomes presented in Chapter 3 of the Risk Assessment document. The evidence presented in the ISA provides numerous examples, along with strong evidence of significant health endpoints for exposure to PM_{2.5}. However, we should not be too quick to dismiss similar, although less well documented health endpoints for PM_{10-2.5}.

The health effect categories and assessment of perceived risk level associated with short-term and long-term PM_{2.5} exposure appear to be highly appropriate, based on epidemiological, controlled human

exposure, as well as toxicological studies. These include a causal relationship for cardiovascular effects and likely causal relationship for respiratory effects and mortality for both short-term and long-term PM_{2.5} exposure.

Selection of the 15 urban study areas is reasonable with adequate breadth and distribution of these cities as representative of the country along with adequately conducted epidemiological studies in each of these sites have relatively elevated 24-hour and/or annual PM_{2.5} monitored levels. Such conditions allow for assessment that will provide potential insights into the degree of risk reduction associated with alternative standard levels.

Comments relevant to charge question 2, but throughout Chapters 2 and 3:

The urban areas selected for study reflect sites with adequate variable ambient PM_{2.5} levels that will allow for the evaluation of potential health impacts across a wide range of diverse locations flush with a adequate data and higher statistical power.

I am highly supportive and enthusiastic of the scope and methods plan presented. The rationale provided is adequate and presented in a clear and appropriate fashion.

Comments to Question 3.

The alternative standard levels for assessment, based on four alternative sets of standards for the annual and 24-hour PM_{2.5} standards are reasonable to evaluate quantitative risk assessment. Again, concern is expressed to exclude ultrafine particles, chemical speciation and PM_{10-2.5} from consideration of alternative standard levels for assessment. Something needs to be stated beyond the declaration that inadequate data exists for these diverse size fractions and/or chemical speciation.

Will the closeness of the annual alternative standard concentrations of 12 µg/m³ versus 13 µg/m³ be sufficient to determine differences or to make an impact on determining health consequences? I realize this is the precise purpose for studying such differences, albeit small. I am more confident of the 24-hour alternative standards of 35, 30, and 25 µg/m³.

Given the current alternatives, it would be fascinating to use the current standard of 15 µg/m³ 24-hour and annual standard of 35 µg/m³ to these proposed alternate standard combinations.

Poirot Comments (Mr. Rich Poirot)

Charge Question 9:

A number of risk metrics as well as different approaches for presenting these metrics are included in tabular and graphical format for both the core analysis and sensitivity analyses. Please comment on the extent to which the risk estimates are clearly and appropriately characterized and presented.

Generally, the approaches and metrics for assessing various health risks are logically conceived, and the results of the “core” risk estimates, sensitivity analyses and national representativeness are all clearly presented in Chapter 4. The section 4.5 “summary and key observations” is especially well-written, and in some cases easier to understand than the more detailed presentation of the same information earlier in the chapter. It might be helpful to move this summary to the beginning of the chapter, or at least adopt some of its clear wording in earlier sections. The section 4.4 evaluation of the “representativeness” of the 15 urban study areas in the larger national context is also clearly written, and the presentation of results in both tabular and graphical form is excellent! I also thought the sensitivity analyses in section 4.3 was well conceived and clearly presented, with an informative and concise summary in Table 4-1.

Section 4.1, which basically describes the contents of the tables in Appendix E, is tedious to read and/or requires frequent referrals to the appendices to see the results (or to maintain interest). It might be more effective to just include some of the referenced tables in the chapter rather than only as appendices.

The section 4.2 assessment of risks associated with just meeting current and alternative standards is clearly written. However, I’m not sure the graphical presentation of results in Figures 4-1 through 4-13 is all that effective. For one thing, I note that (except for the different legends which convey details that could be presented as well or better in tabular form), Figures 4-1 through 4-8 appear to me to be exactly (or very, very nearly) the same figure. In a similar way, I can’t see any differences in Figures 4-9 through 4-11, and 4-12 and 4-13 look the same to me (and are darned similar to 4-9 through 4-11). I wonder if there’s possibly a mistake here, and if there isn’t, what point is being illustrated by so many figures that can’t be discerned from each other? I don’t suppose it’s possible that the Y axes are actually showing the % reductions in concentrations (rather than in the indicated effects) as this might account for why all the figures look the same. Possibly some of these apparently redundant figures could be moved to an appendix, or possibly similar graphs which show differences (assuming there are some) among the different effects endpoints for each individual city might convey more useful or interesting information.

Specific Comments on PM Health Risk Assessment, Chapter 4

p. 92, line 26: It seems counterintuitive that the lowest measured levels were lower than the policy-relevant background across all studies. Could a brief explanation for this be provided here?

pp. 98, lines 10 & 12: Change “Seep” to “See” and add “ly” to “significant”.

p. 100, line 19 (& elsewhere): Is “reflecting use of” the right phrase here? I would think the negative lower bound estimates of incidence are the indicators of insignificant effects estimates rather than reflections of the use of insignificant effects estimates.

p. 100, line 30: Change “head” to “had”.

p. 107, line 13: The word “conditions” seems out of place and could be deleted.

p. 118, line 32: “long-term morality” is indeed a noble aspiration, and a quick word search shows similar references to “exposure morality” (p. 135, line 40), “non-accidental morality” (p 134, line 12) and (my favorite) “premature morality” (p. 45, line 22).

p. 140 or elsewhere in this section: A majority of the sensitivity analysis results summarized in Table 4-1 (or Table 4-5 and Figure 4-22) seem to show a positive bias (i.e. a larger degree of risk than indicated by the core analyses. Is there any implication to this apparent “directionality”? If so (or if not), should it be discussed here?

p. 148, line 18: Change “hear” to “heart”.

pp. 151-158, Figures 4-14 through 4-21: I really like these figures. Might it be possible to use different colored vertical lines and a legend to indicate which of the study cities is which?

Russell Comments (Dr. Ted Russell)

The first draft of the Risk Assessment to Support the Review of the PM Primary NAAQS (hereafter, RA) represents a significant amount of work, and provides a good deal of information to inform the Administrator, as well as other stakeholders, as to issues relevant to the review of the Primary PM NAAQS. It largely executes the path laid out by the Scope and Methods document and provides quantitative information as to the various health risks related to PM exposure and how those risks may respond to revised PM NAAQS (primarily only PM_{2.5} with some discussion relating to PM_{10-2.5}). The Sensitivity Analysis section was probably the best I have seen in any of the RAs and was informative. I believe that Chapter 4 is well set up to provide the location specific analyses of the range of health endpoints of interest and how the chosen locations are representative of the broader national conditions.

While the RA has largely followed the SM, and accounted for CASAC comments, there is significant room for improvement.

First, I note that it would be very desirable to have an upfront (Chapter 2?) summary as to the approach and results. This should build upon the policy-relevant questions identified in the PM ISA, and provide answers as appropriate (or identify why such answers are not provided).

Second, while the RA does represent a tremendous amount of work, it is not as effectively presented as it could be. First, it is rather repetitive in places, particularly the transition from Chapter 3 to Chapter 4. It seems as though Chapter 3 tells us what is going to be done in Chapter 4, and then Chapter 4 goes back over the same. Given the detail in Chapter 3, it is possible to just jump in to results in Chapter 4. Next, Chapter 4 is a bit laborious to get through, and after reading the material, it is a struggle to keep it all sorted out in one's mind. The various paragraphs in Sections 4.1 and 4.2 providing numerical results of the various risk assessments for different endpoints (there are 27 such paragraphs) lose punch as one goes through them. It is recognized that the results are somewhat condensed from what is more thoroughly presented in over 100 tables in the appendices, but a few summary tables or graphs would go a long way here. Choose the most influential endpoint(s), and summarize across cases and alternatives. Further, the figures in this chapter are not overly effectively presented as Figs. 4-1 through 4-8 and Figs. 4-9 through 4-11 are rather indistinguishable and the text does not identify what is really different and important between them.

I was hoping to see more from Chapter 5 as I was looking for a national scale assessment of a broader set of endpoints. While I recognize the detail and care that went in to matching of studies in Chapter 4, something should be done to take the city-specific results from Chapter 4 to provide the national scale assessment of more health end-points than just long-term mortality. Further, Chapter 5 did not assess the change in risk at different alternative standards.

Given that the SM planned to do the national scale assessment only on long-term mortality, what really may be missing is a synthesis as part of Chapter 4 or 5, or a new Chapter 6 that really synthesizes the results from both Chapters 4 and 5. Looking back at the chapter, this synthesis is started by the section discussing how the chosen areas are representative of the nation, but more is needed to interpret how those results reflect likely risks to the population, and this is the point where absolute numbers of individuals likely to be impacted (and which endpoints) would be valuable. As part of a new Chapter, or somewhere else, how the risks associated with the alternative PM standards compare to the other

NAAQS. I why one might stop with just investigating down to $12 \mu\text{g m}^{-3}$ (annual)/ $25 \mu\text{g m}^{-3}$ (24-hour) when it appears substantial risks are still found at the levels currently in the RA, and the choice of alternatives might consider how those alternatives compare to the NAAQS (proposed or effective) for other pollutants.

Response to Charge Questions:

5. The inclusion of a hybrid approach is appropriate and a nice extension to the analysis. While the approach is rather ad hoc, any method would be at this point, and a simplified approach allows assessing the sensitivity to a recognized concern that in some locations, there will be a blend of local and regional controls to reach attainment. The approach currently likely overestimates this effect in some locations (e.g., those areas mainly affected by mobile source and secondary PM) and underestimates in others (where specific industries and activities have a major local influence). The approach developed is reasonable and of appropriate complexity given the vast unknowns as to how specific areas would choose to control emissions.
10. I was generally pleased with the approaches used to demonstrate how (or how not) the chosen urban areas represent the nation as a whole. The discussion of how one should interpret the cases where specific risk attributes in the chosen areas are/are not similar to the nation as a whole, and the use of CDFs, was informative and at a good level.
11. As noted above, I was disappointed with the national scale assessment as it is limited to one endpoint and did not include an assessment associated with alternative standards or a synthesis of the results of the city-specific analyses. Consider a final chapter synthesizing the results from Chapters 4 and 5, identifying the key considerations that might drive the revision of the NAAQS, and how the alternatives compare with other NAAQS.
12. Showing that the chosen areas fall near the top of the CDF is a good start, and does provide information as to how one might interpret the information provided by the detailed assessments presented in Chapter 4. Again, what is missing is going the other direction, that being taking the information in Chapter 4 and developing a national assessment for endpoints other than long-term mortality.

An overall concern for both the Primary NAAQS and Urban Visibility Reviews: The way the current RA and Visibility documents are currently formulated, the potential importance of controlling sources of elemental and primary organic carbon are understated versus other components. Health studies are suggesting that EC/OC are more associated with cardiovascular disease issues than many other components (e.g., ionic inorganic species making up much of the mass of $\text{PM}_{2.5}$ in much of the US), and EC absorbs sunlight and can exacerbate climate warming. On the other hand, inorganic ionic species likely lead to a net cooling. While a visibility assessment can be more confidently done (or the results would be subject to less uncertainty), climate impacts are a much greater concern (at least to me and I think a great fraction of others). Information that is transmitted to decision makers should more fully express the importance of controlling sources of EC and primary OC.

Speizer Comments (Dr. Frank Speizer)

Scope of the Assessment and Methods used for the urban case studies

1) Choice of assessing PM_{2.5} only

Page 15, section 2.4.1, first and third bullet:

At this point I would argue that insufficient detail is provided to justify dropping doing a quantitative risk assessment for PM_{10-2.5}. Unless more detail is provided Staff is making the same decision make in previous PM assessment. Since in Chapter 6 (and chapter 7) of the ISA conclusions that PM₁₀ are likely causal, we asked for more modeling of how PM₁₀ along with PM_{2.5} data might be used to improve understanding of the course fraction component. In the ISA effort in this direction is taken, and seemingly some quantification is reported. Why, having gone to the trouble there not use it here? Reference is made to section 3.3.1. The relevant section is at the end of the section, just before the start of 3.3.2. I do not believe the argument is sufficient to drop the course fraction and this will need to be debated at the CASAC meeting.

2) Selection of causal or likely causal associations with PM_{2.5} only.

Again, this will need to be discussed. It is not clear that there was or is a consensus as claimed that CASAC was not interested in looking at the suggestive category. In fact, there might be a consensus that for different disease outcomes different levels of certainty of causality, with appropriate considerations of the size of the margin of safety might be considered. (For example, a risk affecting 1% of newborns with a lower level of certainty might have a substantially greater impact than a risk affecting 5% of emergency room admissions of elderly patients with respiratory disease with a greater degree of certainty. Simply stating that the degree of certainty would make quantitative estimates less useful without providing some calculations seems inappropriate.

3) Rational provided in section 2.5 for alternative standard levels for assessment.

Logically and well described in section.

4. General approach

The considerations discussed in the remainder of Chapter 3 are logical and clear. However, I cannot accept that they would only apply to PM_{2.5}. If the same arguments were made for the other components of PM one would expect that the analysis itself would be able to show where the uncertainty becomes so great as to make the calculations useless and where some additional information was obtained that would help the us and the Administrator make informed judgments about the risks. By simply dismissing doing such calculations we simply don't have the data with which to make a decision.

The choice of LML to assess long-term PM exposures, but not extrapolating to PRB (if I understand what was done seems appropriate. Similarly, for short-term PM exposures going to PRB rather than LML, assuming continuous exposure-response function is reasonable. Thus, I would be generally supported of these approaches.

The results are rather dramatic and consistent. For the long term effects it appears that for substantial reduction of risk to be made that the alternative levels of PM_{2.5} must be

12/25. In contrast for short term effects it looks as though anything below 13/35 is relatively flat. How this will be used in reporting will need to be discussed.

7. Selection of epi studies and C-R functions within these studies.

For what was done, the descriptions are fine and the choices made seem justified in the text. My argument remains that the same criteria could have been applied to other PM components, or at least modeled with PM₁₀ and PM_{2.5} to get estimates of PM_{10-2.5} and similarly put to the test as to whether the data were adequate. Staff may have done this, but the data are either buried in an appendix and not referenced to or wasn't done, thus making the decision not to do it less defensible. Similarly with regard to end point decisions, by simply choosing not to explore in more detail the suggestive categories, whole disease categories are being left out. (eg reproductive outcomes and lung cancer). This latter omission is particularly troublesome since the exposure characteristics from the ACS studies are being used for the long term exposure mortality and it would have been a simple matter to do the same analysis with lung cancer. We would then be able to see how the uncertainty would play out at least for one disease category.

Suh Comments (Dr. Helen Suh)

The RA set forth a clear and thoughtful approach to assess particle-mediated health risks. The goals of the RA are well-defined and for the most part the RA does an admirable job of achieving these goals. As an overall comment, the decision to forego a risk assessment for $PM_{10-2.5}$ should be discussed further, as a limited risk assessment for $PM_{10-2.5}$ would provide information helpful to the standard setting process in a manner consistent with the document's stated goals. This limited risk assessment could be based on the same multi-city studies as used in the risk assessment for $PM_{2.5}$ (Zanobetti and Schwartz, 2009 and Peng et al., 2008). Correspondingly, the decision to forego a population exposure assessment should also be revisited. Although the previously proposed approach requires further development before its application to the RA, a simpler, more targeted exposure assessment approach could be used to help identify factors that contribute to observed variability in the C-R functions or to city-specific differences in risks.

The Summary and Key Observations section at the end of the document provided a very nice summary of the key findings. Coming at the end of the document, this section was buried and it would be helpful if it was moved forward, perhaps before or after Chapter 2. [If before, a small paragraph on the scope should probably be added.]

Charge Question 7: Selection of epidemiological studies and C-R functions within those studies:

In estimating risks associated with $PM_{2.5}$ exposures, we focused on selecting C-R functions from large multi-city studies based on staff's conclusion that these studies provided more defensible effect estimates (see section 3.3.1). Concentration-response functions from several single-city studies evaluating short-term $PM_{2.5}$ exposures were also included to provide coverage for additional health effect endpoints (e.g., emergency department visits). To what extent is the Panel supportive of this approach for selecting C-R functions for modeling risk related to short-term and long-term $PM_{2.5}$ exposures?

a) Specifically with regard to short-term exposure-related mortality, focusing on a study of 112 US cities by Zanobetti and Schwartz (2009), we obtained Empirical Bayes "shrunk" city-specific estimates from the study authors that provided a distinct C-R function for each urban study area location. For short-term exposure-related morbidity, focusing on a study of 202 U.S. counties by Bell et al. (2008), we used regionally-differentiated effect estimates provided by the study authors. Please comment on the selection of C-R functions for evaluating short-term morbidity and mortality effects. To what extent do the Panel members consider the rationales supporting the selection of C-R functions to be clearly and appropriately presented?

The methods used to select the C-R functions were clearly presented. The rationale for the focus on multi-city studies is clearly presented and is appropriate, as is the reason for the inclusion of select single city studies to assess $PM_{2.5}$ -mediated ED risks. While the Zanobetti and Schwartz (2009) and the Bell et al. (2008) studies are excellent studies to select, it is not clear why the analysis is limited to just these studies. The specific reasons for their selection and for the omission of other multi-city studies should be provided for clarity. Currently, other multi-city studies meet the three criteria set forth in the document, namely that they be:

- published, peer-reviewed study that was evaluated in the PM ISA and judged to be adequate by EPA staff
- direct measurements of PM_{2.5} had to be used on reasonable proportion of the days
- could not rely on GAMs using S-Plus software

In this effort, the definitions of “adequate” and “more defensible estimates”, which were used to describe the selection process, would be important.

The use of “shrunk” estimates to obtain the C-R function for mortality is reasonable, especially for the small cities. It may also make sense to use regional specific C-R functions as well, perhaps as a sensitivity analysis, since these regional specific estimates would correspond to and help interpret the appropriateness of using regional C-R functions to assess risks for hospital admissions.

b) Specifically with regard to long-term exposure-related mortality, we identified a number of effect estimates using the extended follow-up of the American Cancer Society (ACS) study to use in the core analysis (Krewski et al., 2009). These effect estimates include standard Cox proportional hazards models, with 44 individual and 7 ecologic covariates, derived using two separate PM_{2.5} monitoring data sets (i.e., 1979-1983 and 1999-2000) (see section 3.3.3 of the RA). To what extent is the rationale for these choices clear and sufficiently justified as the basis for the core analysis involving long-term PM_{2.5}-related mortality?

The rationale for choosing effect estimates from the extended follow-up of the ACS cohort was clearly and logically presented. Given the size of the ACS cohort, it makes sense to select effect estimates from this study for the core analysis. However, given that there are only a relatively few number of chronic cohort studies have been conducted to date, it would be interesting and useful to see how the selected C-R functions compare to those from other studies. This could be done as part of a sensitivity analysis.

Charge Question 8: Addressing uncertainty and variability

- a) *The treatment of uncertainty and variability in the analysis is based on the multi-tiered approach presented in a recent WHO document (WHO, 2008). Specifically, as outlined in section 3.5, we have included qualitative analysis of both variability and uncertainty (WHO Tier 1), as well as single-factor and multi-factor sensitivity analyses aimed at identifying which potential sources of uncertainty have the greatest impact on the core risk estimates (WHO Tier 2). In addition, the sensitivity analyses have been designed to provide a reasonable set of alternate risk estimates to supplement the core risk estimates and inform consideration of uncertainty associated with the core analysis. To what extent does the Panel support the overall approach for addressing uncertainty and variability? To what extent does the Panel agree that the overall approach is appropriate and consistent with the goals of the risk assessment as outlined in chapter 1? Does the Panel have any recommendations for improving the characterization of variability and/or uncertainty?*

The approach used to examine variability and uncertainty in the risk estimates is generally well described and consistent with the stated goals of the risk assessment. The WHO framework is an appropriate and well established approach to assess uncertainty in risk estimates. The decision to forego a probabilistic (or WHO Tier 3) analysis to examine uncertainty and variability in risk estimates seems appropriate given the resource- and time-constraints. The two additional analyses intended to place the risk results for the 15 study areas in a broader national context is an important addition to the document.

- b) *The qualitative discussion of key sources of variability, and the degree to which the analysis design captures those sources of variability, are presented in section 3.5.2. Please provide comments on the approach used. Specifically, do the analyses sufficiently address the issue of variability? Are there key sources of variability that have not been addressed within the qualitative analysis but which could have an important impact on modeling population-level risk associated with PM_{2.5} exposure?*

For the assessment of variability, six key sources of variability were identified. The identified six sources are appropriate; however, their definitions should be broadened or additional categories should be included. For example, differences in PM co-pollutant concentrations (e.g., overall pollutant mixture) should be included as a source of potential variability in PM-associated risks. This factor could be included by broadening the PM_{2.5} composition category to include all pollutants. [While co-pollutants are a source of uncertainty, they may also be a source of variability if there are synergistic or multiple pollutant effects.] Correspondingly, demographics could be broadened to also include land use, source locations, housing stock, and SES.

Beyond source identification, additional work should be performed to assess the potential impact of the variability sources on the risk estimates. While it is true that it is not possible to separate their contribution to risk estimates completely, single- or multiple-factor, WHO Tier 2 analyses could be conducted, with results used to examine the impacts of these variability sources on the C-R function or risk estimates. This examination could correspond to the analyses done to examine the generalizability of the 15 cities to the rest of the US, possibly through a simple weighted-regression of the city-specific risk estimates on the variability source.

- c) *Table 3-13 provides a qualitative characterization of uncertainties including the potential direction, magnitude, and degree of confidence associated with our understanding of the sources of uncertainty. To what extent does the Panel support the characterizations of the key sources of uncertainties identified and the relative rankings of the importance of those sources of uncertainty? Are there additional uncertainties that should be considered?*

Table 3-13 is relatively complete and provides a good overview of the sources of uncertainty and their potential impacts. A source of uncertainty that was not included was the C-R function itself, which was developed from single studies. In this regard, it would be helpful to broaden or alter Source J. (Transferability of C-R functions from study locations to urban study area locations) to include examination of long-term risks using C-R functions from different long-term mortality studies (WHI, NHS, ASHMOG, etc.) or of short-term risks using C-R functions from other cities included in the Zanobetti and Schwartz or Bell et al. studies or from other multi-city studies that include at least one of the target cities. If possible, it would also be helpful to define, perhaps as a Table footnote and even if only vaguely, what is meant by the categories “low”, “medium”, and “high”.

- d) *The results of the sensitivity analyses have been used to gain insights into which sources of uncertainty significantly impact the core risk estimates and to provide a reasonable set of alternate risk estimates to supplement the core analysis. We are mindful that these estimates do not represent a true uncertainty distribution. With regard to the single- and multi-factor sensitivity analyses, to what extent is the Panel supportive of the approach used to conduct and characterize the results of the sensitivity analyses? Please provide comments on the extent to which the presentation of the results of the sensitivity analyses are clearly and reasonably described? Does the Panel have any recommendations for how the results of the sensitivity analyses could be used more effectively or appropriately in characterizing uncertainty associated with the core risk estimates?*

The approach to the sensitivity analysis is reasonable. It would be worthwhile in the text to indicate the direction of the percent changes in risk. Further, the large and variable percent changes by city for some analyses (such as that for seasons-specific C-R) raises concerns over the use of the percent difference to characterize the findings. These findings suggest that in addition to the percent difference, the actual difference in risk should be reported to provide further context. This section should conclude with a brief but explicit summary of the decision to use the sensitivity results only from the long-term exposure mortality analysis, as I think that it now only appears in the summary of results (Section 4.5.2).

Vedal Comments (Dr. Sverre Vedal)

Question 3. Standard levels for risk assessment.

The difficulty here is that epidemiological studies do not provide much help in deciding on the level of the standards. They certainly do not provide much information on levels below which no effects are seen. So, attempting to use them, as is valiantly done here, to identify standards of interest, is not easily justifiable.

Long-term concentration levels. Based on the observations made about mean study concentrations and confidence in effect estimates, it is difficult to understand how the judgments as to the concentrations to be used for this purpose were actually made. The focus on means seems reasonable in light of the form of the annual standard. Using long-term mean concentrations from short-term studies (line 12, p. 19), however, seems a strange approach, and harkens back to the time when the long-term standard was used to attempt to reduce short-term exposure effects.

Instead of using a number of unconvincing approaches to justifying selection of alternatives, why not just take the simple approach of going below the current standard in increments of 1 $\mu\text{g}/\text{m}^3$, say 14, 13 and 12 $\mu\text{g}/\text{m}^3$? That should pretty much cover it for our purposes. I would not want to exclude 14 $\mu\text{g}/\text{m}^3$ because that was clearly a level of interest at the last round and remains of interest.

Short-term concentration levels. I would make the same point here about the exercise of wrestling with the short-term study concentrations to try to arrive at some justifiable levels being ultimately unsatisfactory. Here the simple approach of going down below the current standard in increments of 5 $\mu\text{g}/\text{m}^3$ to, say 25, would have been an equally defensible one, and interestingly, would have resulted in the same concentrations that were in fact selected.

In short, then, I would suggest a 3x2 matrix of standards for use in making risk estimates: three long-term levels of 12, 13 and 14 $\mu\text{g}/\text{m}^3$ and two short-term levels of 25 and 30 $\mu\text{g}/\text{m}^3$.

4.a. Core CRFs: selection approach and description.

A core set of CRFs with identified lag periods were selected. Sensitivity analyses assess the importance of these selections in affecting/influencing risk estimates.

Based on the arguments provided on selection of endpoints for which CRFs will be chosen and risk estimates made, which I agree with, the choice of lung cancer mortality as an endpoint (line 21, p. 40) is inconsistent.

The rationale for choosing to emphasize multi-city study estimates for short-term and long-term CRFs is sound. However, the approach to selecting which multi-city studies on which to focus is not particularly clear. Although I could probably supply some arguments, readers of the RA might wonder why the Zanobetti and Schwartz study rather than the Dominici 2007 study was being used for short-term mortality CRFs, for example. This is important because effect estimates differ between the two studies. That is, there is no presentation of the rationale for deciding alternative large multi-city time series studies. The same applies to the choice of long-term exposure studies, given the several there are to choose from currently, and here the effect estimates differ dramatically.

I agree with the choice and rationale for not proceeding with estimating risks of coarse PM exposures. First, no causal assessment for any effect of coarse PM rises above the grade of “suggestive,” and second, selection of adequate CRFs for coarse PM would be problematic at this time.

4.b. Short-term and long-term lowest modeled levels.

I agree with the basic argument that we should only be concerned with estimating risk above the so-called policy relevant background (PRB) as in previous risk assessments, in spite of there being controversy as to how PRB is calculated. I do not see the point of estimating risks associated with PM that cannot be influenced by human activities and estimating risks down to a zero concentration.

In the absence of equally compelling alternatives, I agree with the choice to only estimate risk down to the lowest measured level (LML) in the core study used for long-term CRFs. However, it would be valuable to see the impact even here of estimating risks down to PRB, even though extrapolation is needed; however, I see that is done in a sensitivity analysis, and that is sufficient. For short-term risk estimates, the choice is easy because the LMLs (which are daily) are below the PRB.