Dear Administrator Jackson,

SAMET:
This letter provides our principal comments concerning the Second External Review Drafts of the *Quantitative Health Risk Assessment for Particulate Matter* and the *Particulate Matter Urban-Focused Visibility Assessment.* The responses to the charge questions, along with comments from individual Panel members follow.

In general, CASAC considered that the Second External Review Drafts had been substantially improved and that the changes in the documents were responsive to comments provided by CASAC following its initial review. With the additional changes recommended following review of the Second External Review Drafts, the documents will be a solid foundation for the Policy Assessment.

While CASAC found the *Quantitative Health Risk Assessment for Particulate Matter* to be greatly improved over the first draft, several major issues remain to be addressed as the final draft is prepared:

- **The urban case study analyses, which are central to the risk assessment, use three different approaches for simulating just meeting the current and alternative suites of $PM_{2.5}$ standards: the proportional approach, the hybrid approach, and the peak-shaving approach. CASAC found that these three alternatives were not defined with sufficient clarity, and recommended a graphical approach, along with the inclusion of examples based on the 15 study areas. CASAC also suggests additional discussion of the three approaches in relation to potential control scenarios. The hybrid approach merits the greatest emphasis, while the proportional and peak-shaving approaches represent bounding scenarios. CASAC recommends an alternative label for the "peak shaving" approach, since this term does not describe the actual method.**
- **Chapter 5, providing national estimates of $PM_{2.5}$ attributable total mortality, has been added since the first draft. It is useful for assessing the generalizability of the findings in the 15 urban study areas, but its findings are otherwise not central to the risk assessment. Consequently, CASAC recommends that the chapter be moved to an appendix and that discussion of the key findings, summarized in Figure 5-4, be placed at the appropriate point in the current Chapter 6.**
- **The current Risk Assessment explores the burden of disease that can be avoided under different scenarios of reduction of $PM_{2.5}$. The scenarios**

involve specified values of 24-hr and annual standards. However, the Risk Assessment would more appropriately explore avoidable disease burden under a range of alternatives that are not defined by potential regulatory options. For example, the current document includes scenarios of $PM_{2.5}$ concentration as low as 12 µg/m^3; the risk assessment might be reasonably extended to lower values. We recommend that EPA develop and apply specific criteria for determining the exposure concentration values to be considered in the risk assessment. Mounting uncertainty at lower concentrations would be a reasonable basis for doing so.

- **Chapter 6 is arguably the most critical chapter of the document. We commend EPA for being responsive to our prior comments and adding this chapter. Given its importance, we recommend a careful rewriting and editing to assure that its findings are clearly presented.**

CASAC Responses to Charge Questions

Chapter 3 – Urban Case Study Analysis Methods TED

1) Air quality inputs (section 3.2): We have expanded the consideration of alternative approaches to simulating just meeting the current and alternative suites of PM2.5 standards (i.e., rollback approaches) to include a peak shaving approach, in addition to the hybrid and proportional approaches considered in the first draft assessment. This peak shaving approach is intended to represent more localized, rather than regional, patterns of PM2.5 reductions (discussed in section 3.2.3.3).

a) To what extent does the Panel believe that the use of the peak shaving approach provides useful additional exploration of variability associated with how ambient PM2.5 concentrations are simulated to change upon just meeting the current and alternative suites of standards?

RUSSELL:
The "peak shaving" approach does provide useful information, but should be better explained and motivated, both in section 3.2 and Appendix B, preferably with a specific example of what situation(s) and potential control approach(es) it is meant to simulate. Appendix B should also provide the specific mathematical formulation. It would be instructive to provide an example of the three rollback approaches being applied to one or two cities in the section, or in Appendix B.

b) We have used comparisons of composite monitor annual averages generated using the different rollback approaches as a surrogate for differences in long-term exposure-related mortality in looking across all three rollback approaches. To what extent does the Panel believe that this is a reasonable approach for assessing the impact of variability associated with simulating changes in air quality patterns on estimates of long-term exposure-related mortality?

RUSSELL:
A very brief description of how annual design values are currently calculated should be provided to help motivate the procedure developed here. The Panel is uncomfortable with the approach used for imputing missing values, and has discussed alternatives in their individual comments. Whatever method is ultimately chosen, the impact of the approach taken can be well tested using data from stations where there is little missing data, then removing data, applying the approach, and testing to see how closely the method reproduces the original annual average. It would make the approaches to compositing and imputing missing values easier to follow if equations were provided.

2) Selection of model inputs (section 3.3): We have expanded and clarified the discussion of our rationale for identifying modeling choices comprising the core risk model, focusing in particular on selection of C-R functions (section 3.3.3). To what extent does the Panel consider this discussion to be clear and the model selections appropriate?

LIPPMANN:
The Panel commends the authors for expanding and clarifying their rationale for identifying modeling choices comprising the core risk model in a logical and satisfactory manner. Their model selections were appropriate for this review cycle, which is focused on PM2.5 exposure and cardiovascular responses. A suitable rationale provided a good foundation for the selection of the epidemiological studies that were utilized to establish C-R functions. The expansion of the discussion and integration of the ISA was useful in that it provided an opportunity to reinforce the gaps in knowledge, as on p. 3-20, where it was stated that there were no multi-city studies for the category of short-term exposure to PM2.5 and emergency department visits for cardiovascular and/or respiratory illnesses. The summary tables (Tables 3-5 through 3-8) provide a useful synopsis of the model inputs for the core risk models and sensitivity analyses.

Charge Question 3: Addressing uncertainty and variability (section 3.5): We have clarified the process used to evaluate sources of variability and added coverage for specific sources of variability (section 3.5.2); expanded our discussion of the qualitative analysis of uncertainty (section 3.5.3); and included analyses of pair-wise interactions of sources of uncertainty (section 3.5.4). To what extent does the Panel consider these discussions to be clear and appropriate?

FREY:
In general, the second draft REA appropriately identifies and discusses key sources of variability and uncertainty, and includes sensitivity analysis that provides insight regarding the impact of some sources of uncertainty on the core risk estimates.

EPA provided a footnote explaining the rationale for identifying "key" sources of variability. The document should indicate if the same process was used to identify "key" sources of uncertainty.

In the discussion of key sources of variability, EPA added material regarding copollutant concentrations and on demographic and socioeconomic status, as requested in CASAC comments on the first draft of the REA. As a minor comment, it is not entirely clear that age of housing only affects infiltration rate because of air conditioning use. Newer homes are typically "tighter" than older homes, and thus have lower infiltration rates. Climate zones are another factor in infiltration. For example, northeastern homes do not have as high a proportion of central air conditioning as southeastern homes. Given its effect on particle composition, concentrations, ventilation and activity patterns, it would make sense to also include seasonality in the list of variability sources.

We note that EPA did not address a comment on the first draft REA to the effect that "exposure modeling should be included in the REA. A probabilistic Tier 3 approach should be used for the exposure assessment." While we understanding that timing may have precluded adequate treatment of this topic, we expect that

EPA will develop this capacity for future revisions of the standard. CASAC asked for this five years ago, and would like to see this in the next revision.

With regard to uncertainties, in response to CASAC comments on the first draft of the REA, EPA has included uncertainty in the C-R function itself, which was developed from single studies. EPA has appropriately taken 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 REA.

EPA has provide some explanation of the meaning of uncertainty categories of "low", "medium", and "high." This discussion is adequate. However, there is a confusing statement to the effect that "high" sources of uncertainty "are likely to influence the interpretation of risk…" "if those sources of uncertainty are reduced or more fully characterized." The parenthetical "if" clause seems to confuse the issue, and should be deleted. A reader of the paragraph on lines 6-20 of p. 3-63 might wonder how "staff consensus" was achieved, and whether consensus is an appropriate goal when characterizing uncertainty. A potential concern is that achievement of "consensus" might mean that some opinions over-ride others and that the resulting characterization of uncertainty might be biased. There are short-comings of group-based elicitation processes, such as dominance by strong personalities or a tendency to provide opinions about goals rather than state of knowledge. It would be useful to explain the process by which "consensus" was achieved.

EPA has done a nice job on commenting 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. EPA has provided a nice treatment of this on page 3-71.

Based on quantifiable sensitivity analysis, the report generally clearly conveys that the "core" estimates appear to be at the low end of alternative "plausible" estimates. However, particularly in Chapter 6, the role of sources of uncertainty treated qualitatively should also be addressed. In particular, given exposure misclassification, it is likely that the core estimates are biased low. This is an important point to convey consistently. The core estimates seem to be conservative in the sense of being underestimated, which is not typical practice for public health endpoints, given that virtually all of the sensitivity analyses result in higher risk estimates as compared to the core.]

Charge Question 4: Sensitivity analysis results (section 4.3): We have included a discussion of how the results of the sensitivity analysis can be used as an additional set of reasonable risk estimates to inform consideration of uncertainty in the core risk estimates

(see section 4.3.2). What are the Panel's views on how we have used the sensitivity analysis results to support consideration of uncertainty in the core risk estimates?

FREY:
Overall, the sensitivity analysis section 4.3 is very good and nicely covers a complex topic. Table 4-3 is a useful summary. The classification of descriptive categories for small, moderate, moderate-large, and large contributions is useful. However, it would be more appropriate to refer to these as contributions to "sensitivity" rather than "uncertainty." As noted in several place, the sensitivity analysis represent plausible alternatives to the core estimate, but are not a probability sample. Thus, there is not a probabilistic interpretation to the sensitivity analysis results. EPA has appropriately addressed this point and has clearly articulated, quite reasonably, that the sensitivity analysis results represent plausible and scientifically defensible estimates. The range of these estimates provides an indication of the implications of uncertainty.

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 results in Table 4-3 indicate, for example, that the random effects log-log model provides larger risk estimates than the fixed effects log-liner model used for the core estimates. This information is very useful and is an excellent addition to the REA. The more thorough treatment of model choices and alternative C-R functions provides plausible alternative estimates to the core estimate.

It was not apparent that EPA responded to this comment on the first draft REA: "The range of uncertainty associated with confidence intervals for a given C-R function (which is 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 would provide insight as to whether model structure, or random error for a given model, is a more important source of uncertainty." It would be useful to make this comparison, which can be discussed qualitatively.

Per CASAC's comments on the first draft REA, EPA indicates 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. The second draft REA seems to put emphasis on relative changes in risk. However, the NAAQS are intended to be protective of public health, and therefore the magnitude of the risk estimates is ultimately a more useful policy-relevant metric.

The sensitivity analysis related to peak shaving and "peakiness" was not very clear. What are the main points to take away from these analysis could be clarified. Is there an implication of some sort of risk trade-off between the cities with and without "peakiness", as shown in the comparison of results for the different roll-back approaches?]

What is the premise of the overarching conclusions – e.g., that there are strong regional or inter-city effects? Would this carry forth to the policy analysis in some way? This might be clarified.

5) Consideration of design values and patterns of PM2.5 monitoring data in interpreting core risk estimates (section 4.5): To enhance our interpretation of the patterns of core risk estimates generated for both the current and alternative suites of standards, we have included analyses of 24-hour and annual design values together with patterns of PM2.5 monitoring data for the 15 urban study areas. This reflects the fact that these two factors play a key role in determining the degree of risk reduction estimated upon just meeting the current and alternative suites of standards under alternative rollback approaches. As part of the consideration of design values, we have also contrasted the 15 urban study areas with patterns of design values seen for the broader set of urban areas in the U.S. in order to help place the urban study area in a broader national context.

a) To what extent is the Panel supportive of these additional assessments?

b) Does the Panel have any recommendations for additional insights based on consideration of patterns in design values and PM2.5 monitoring data across the 15 urban study areas and at the national level?

VEDAL:
The graphical presentations depicting the 24-hour and annual average design values for US urban areas and the 15 urban areas used in the risk assessment were very useful for understanding the concept of the controlling standard, the implications of reducing either the annual or the 24-hour standard, or both, and the representativeness of the 15 urban areas. To further enhance the value of these graphical presentations, more complex color coding could be used to provide information on US region of each urban area. In these plots, unwarranted conclusions were drawn for cities lying on the border between zones. Graphical presentation of the design values by monitoring site for the 15 risk assessment urban areas provided valuable insights into the role of patterns of PM monitoring data in different cities in determining effects of various control strategies. It is recommended that the use of "peaky" to describe PM patterns be applied consistently, and that the main observations from both sets of plots be summarized, especially as regards impacts of the alternative rollback procedures.

SAMET/SPEIZER Comment on Ch. 5:
Chapter 5 of the revised RA provides estimates of the numbers of deaths attributable to long-term PM2.5 exposure, based on air quality estimates from the Community Model for Air Quality (CMAQ) and the environmental Benefits Mapping and Analysis Program (BenMap), and uses the risk estimates derived for the Krewski, 2009 assessment of the ACS data with a LML of 5.8ug/m3. A principal purpose for inclusion of this chapter is to place the PM2.5-associated risks for the 15 urban study areas within the distribution oif risks nationally. Figure 5-4 provides

the key findings in regard to this purpose and indicates that the selected urban study areas in large part fall in the highest 20% of the distribution of sites. . We recommend that Chapter 5 be moved to an Appendix with inclusion of Figure 5-4 within the current Chapter 6. The figure provides information relevant to the generalizability of findings from the 15 areas to the entire United States. However, the estimates themselves are not directly relevant to the overall purpose of the RA; the estimation approach differs from that used for the 15 areas; and the chapter is brief and does not adequately set out sources of uncertainty and variability. By placing the chapter's contents into an appendix and specifically acknowledging its purpose, it will not distract from the flow of the RA and the major objective of the analysis will be met by inclusion of Figure 5-4 in the revised Chapter 5.

Chapter 6 – Integrative Discussion of PM2.5-related Risks

6) We have developed an integrated discussion of the PM2.5-related risk estimates which considers the results of the qualitative and quantitative treatment of uncertainty and variability together with the various national-scale assessments completed for the analysis to support interpretation of the core risk estimates. As part of the integrative discussion, we also provide key observations that bear on policy-relevant risk-based questions.

SPEIZER:

The CASAC panel was unanimously pleased to see this chapter added to the document. The chapter is the culmination of a rational approach to summarize the many detailed analyses carried out in Chapter 4 and extended in the Appendices. However, the results presented could be summarized more effectively and rather than being presented in detail should have warranted a more integrative discussion. The choice of the 15 urban study areas is previously discussed in Chapter 4. Similarly, the choices of endpoints were previously documented. Clearly, the use of IHD mortality (as opposed to all cause cardiovascular mortality) represents an upper bound of effects and appears to be the direct effect of the results available from Krewski's 2009 analysis, which presumably was chosen as the most defensible effect estimate.

There was general agreement that the current standards are not adequate and CASAC spent some time discussing just how the selection of alternative standards, given a no threshold linear effect, should be made. In the document as presented the lowest levels assessed were 12/25, however preliminary analyses at 10/25 were offered indicating a variable effect across the 15 urban study sites, resulting from the very different characteristics of some of the areas. We believe it is important for Staff to come up with some criteria by which they will choose a lower bound to assess and provide a discussion as the justification for choosing this level. It is clear that staff must face the dilemma of choosing criteria that will minimize the potential for leaving a susceptible group unprotected while providing some estimate of the certainty that effects estimated are based on reasonable expectations that they are something more than chance. One criterion suggested was to choose the lowest levels at which reported results are in the ISA. Another is to choose a level that would provide "margin of safety" where by the level chosen represents half or two-

thirds of the lowest level of reported results that have been confirmed in multiple studies (rather than the lowest reported value). We believe carrying out such a procedure and documenting the criteria used would be consistent with the assessment of uncertainty and variability well discussed in the chapter.

a) To what extent does the Panel believe that we have captured the key policy-relevant questions that can be addressed by this risk assessment?

SPEIZER:
Staff has done an excellent job in presenting and capturing the key-policy relevant questions. However, as indicated early on in this document there was to be a qualitative discussion on PM10-2.5 and on those effects that were deemed only "suggestive" but might have important public health implications (e.g. lung cancer, reproductive effects), but for which quantitative risk assessment was not thought warranted but that would appear in the PA (page 2.6). I would have thought that some remarks in this chapter would be necessary to assure that the PA would discuss the issues.

b) We provide a set of key observations related to estimates of risk associated with simulations of just meeting the current and alternative suites of standards. These observations are based not only on consideration of trends in risk reduction across alternative suites of standards and residual risk remaining after simulation of just meeting specific suites of standards, but also on additional factors that can impact risk (e.g., the role of annual and 24-hour design values, the peakiness of PM2.5 distributions within a study area, and application of different rollback approaches). To what extent do the Panel members believe that the observations presented in section 6.2 are well supported by the results of the analyses? Are there other observations that might be made that would help to address the policy-relevant questions identified at the beginning of the chapter?

SPEIZER:
Key observations are presented and discussed with adequate discussion of the relevant contribution of the role of annual and 24 hour design values and the role of "peakiness" of distributions. One observation that appears to be focused upon and may be a driving force is the uneven distribution among the 15 urban sites on the impact of the various scenarios and whether this fact is sufficiently taken into account in scaling up for the national estimates. More discussion and or analyses on this point may be warranted. For example what role does the actual estimates from these 15 sites play is coming to the estimates of 3-9% excess mortality? It may be too much to expect (in spite of the statistics) that 63,000-88,000 premature deaths would be prevented. Part of the country is already well below the proposed alternative levels and thus would not contribute to lives saved. Are there additional alternative sensitivity analyses that would provide either alternative estimates or put more confidence in these estimates by taking into account better population weighted C-R analyses?

c) Part of our interpretation of the core risk estimates presented in section 6.2 is our characterization of confidence in the core risk estimates and in observations made based on those estimates. These assessments of confidence are based on consideration of the results of the sensitivity analysis as well as on the qualitative assessment of uncertainty and variability. To what extent does the Panel believe that the characterizations of confidence in the core risk estimates and associated policy-related observations are reasonable given available information?

SPEIZER:
See above. In spite of the last comment, the uncertainties and variability of the core assessments seems to be as good as it can be.

d) As part of the integrative discussion, we use the results of several national-scale analyses (i.e., the national scale PM2.5 mortality analysis, the representativeness analysis, and the new exploration of design values and patterns of PM2.5 monitoring data presented in section 4.5) to place the results of the risk assessment in a broader national-context. What are the Panel members' views on appropriateness of this effort to place results of the analysis in a national context?

SPEIZER:
Evaluation of the several national scale analyses, as indicated above is of some concern. If I read the Tables in Appendix E correctly, the effect of moving to the lowest alternative (25/12) in some cases within the 15 urban sites produces a range of 32-67% (with one outlier at 11% and one at 100%) reduction in the IHD compared to the current standard. The question is, is this the best baseline for the comparison or should it be the current recent measurements, which would drop the percent changes considerably (and perhaps provide a more realistic estimated of the potential benefits from implementing changes). Obviously, the proportional ranking and changes would be the same, but the impact on "lives saved" on a national scale might be considerably less and more realistic.

e) We conclude chapter 6 with a list of key observations. Does the Panel believe that we have appropriately highlighted key findings of the risk assessment in these observations? Of particular note is the observation that, while alternative 24-hour standard levels can be used to reduce annual-average PM2.5 concentrations and thus to reduce estimated risk, the results are likely to be highly variable across urban areas. More consistent lowering of annual-average PM2.5 concentrations across study areas, and thus more consistent reductions in estimated risk, may result from application of alternative annual standard levels. We also note that simulation of the alternative 24-hour standard level of 25 μg/m3 resulted in reductions in annual-average PM2.5 levels for some study areas that were well below the lowest annual standard level assessed (i.e., below 12 μg/m3). As a consequence, we observed risk reductions reflecting these changes in annual-average PM2.5 levels below 12 μg/m3. Given these results, does the Panel believe that there is utility in estimating risks for alternative annual standard levels below 12 μg/m3?

SPEIZER:
Key observations seem to be presented in a balanced and fair way. Although the national assessment suggests a range of 63,000-88,000 premature deaths per year attributable to PM2.5 does not jive with a fairly often quoted figure from 2006 that moving the annual standard from 15 to 14 ppm would result is "more lives than perished in 9/11". (That figure translated into about 3000 lives.) Staff acknowledges that the range of effects are in two categories: 3-9% and 0-3% in two halves of the country. This may be less precise than what the data indicate in that it would appear from their own estimates that the bulk of the effect comes from the upper end of the exposure in the counties (pg 5.8, line 12-15).