UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

May 18, 2009

EPA-CASAC-09-007

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

Subject: Clean Air Scientific Advisory Committee's (CASAC) Review of *EPA's Risk and Exposure Assessment (REA) to Support the Review of the SO_2 Primary National Ambient Air Quality Standards: Second Draft*

Dear Administrator Jackson:

The CASAC Sulfur Oxides Primary National Ambient Air Quality Standards (NAAQS) Review Panel (see Enclosure A for Panel Roster) is providing review comments on the Environmental Protection Agency's (EPA) second draft *EPA's Risk and Exposure Assessment (REA) to Support the Review of the SO_2 Primary National Ambient Air Quality Standards: Second Draft.* This letter provides CASAC's overall comments, highlighting the most important issues to be addressed in revising the second draft REA. Responses to the specific charge questions (Enclosure B) and comments from individual Panel members (Enclosure C) follow.

The second draft REA was greatly improved and CASAC found that its comments on the first draft had largely been addressed. The REA builds successfully on the Integrated Science Assessment (ISA) and generally describes the REA's methodology well. CASAC supports the approach taken and concludes that the REA offers the analyses and findings needed for determining the four elements of the NAAQS for SO_2. Chapter 10, specifically relevant to that purpose, sets out a framework of evidence for decision making and identifies key uncertainties. CASAC is in agreement with having a short-term standard and finds that the REA supports a one-hour standard as protective of public health. It is also in agreement with the proposed range for a one-hour standard of 50-150 ppb. The REA gives preference to the 99th percentile for the form of the standard, but this preference needs better justification. The panel recommends further consideration of analyses already conducted, perhaps supported with additional analysis if found warranted, to better characterize the implications of selecting the 98th or 99th percentile for the form. As discussed below, promulgation of a one-hour standard has implications for 24-hour and annual standards.

The principal comments to be addressed as the document is revised primarily relate to organization and the need for greater clarity in communicating key aspects of the REA's methods and findings, particularly those related to uncertainty and variability. The panel strongly recommends the following:

- Every chapter in this or any REA (as well as in the ISAs) should end with a summary section of findings relevant to setting the NAAQS, as presented in Chapter 10 in this REA. Each chapter's summary section should state the key findings/conclusions in the chapter and specifically address:

  > *What scientific evidence and scientific insights have been developed since the last review that either support or call into question the current public-health-based and/or current public-welfare-based NAAQS, or indicate that alternative levels, indicators, statistical forms, or averaging times of the standards are needed to protect public health with an adequate margin of safety and to protect public welfare?*

- CASAC found the discussions of uncertainty in individual REA chapters to be lacking in clarity, with incomplete descriptions of methods and findings. The panel recommends rewriting with more complete description of methods and highlighting of key findings, perhaps with bullets, rather than in lengthy text. Sensitivity analyses need to be distinguished from those addressing uncertainty. More explicit chapter-by-chapter discussions of uncertainty characterization will inform the summary discussions in Chapter 10 about the NAAQS.

- The health endpoints in the clinical studies, increase in airway resistance (sRaw) and decrement in forced expiratory volume in one second (FEV_1), need to be better framed as indicative of an adverse consequence of SO_2 exposure. There needs to be expanded discussion of the clinical implications of these endpoints and why these endpoints are considered informative measures for setting the NAAQS.

- Chapter 3 needs extensive revision. It reads poorly and does not satisfactorily define or address the key concepts of susceptibility and vulnerability. The EPA should carefully compare the content of this chapter, and particularly the definitions of these concepts, to that of similar chapters in other ISAs and REAs, and even to other EPA documents using these concepts. CASAC found the discussion of vulnerability and susceptibility in the ISA and REA for particulate matter to be better developed and more informative.

- To the extent possible, the REA should better address the representativeness of the locations with SO_2 monitors considered in the REA, as well as the representativeness of Greene and St. Louis Counties, where the risk analysis was carried out.

- The REA should explain what considerations and analyses will be needed to inform a decision with regard to changing or revoking the 24-hour and annual average standards, if a one-hour standard is implemented.

We now point out two aspects of this review that also apply to the other criteria pollutants and to analyses and documents that will be developed by the Agency about them.

1. CASAC received this second draft REA without a specific separate document that provided a formal, easily accessible summary of Agency responses to previous CASAC comments on the first draft REA and also without any indication of changes made since the prior draft. CASAC reiterates its expectation that all revised drafts will be accompanied by such materials, both to enhance the efficiency and targeting of its review and to provide a transparent record of the basis for Agency changes in these important science assessments.

2. With reviews in progress for the gaseous criteria pollutants as well as for particulate matter (PM), CASAC notes the inherent oversimplification of handling these components of the ambient air pollution mixture on an individual basis. Consideration needs to be given to how the existence of the criteria pollutants in mixtures can be better acknowledged and to approaches for moving towards regulatory strategies that are built on understanding of health risks of ambient pollution mixtures.

In closing, we hope these comments will help the Agency revise the second draft REA for SO_2 and that the comments will provide useful advice as EPA considers revisions of the standard for this criteria pollutant.

Sincerely,

/Signed/

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

Enclosures

Enclosure A

ROSTER
U.S. Environmental Protection Agency
Clean Air Scientific Advisory Committee
Sulfur Oxides Primary NAAQS 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, Philip Drinker Professor of Environmental Physiology, Department of Environmental Health, Harvard School of Public Health, Harvard University, 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

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

Dr. John R. Balmes, Professor, Department of Medicine, Division of Occupational and Environmental Medicine, University of California, San Francisco, CA

Dr. Douglas Crawford-Brown, Professor Emeritus, Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC

Dr. Terry Gordon, Professor, Environmental Medicine, New York University School of Medicine, Tuxedo, NY

Dr. Dale Hattis, Research Professor, Center for Technology, Environment, and Development, George Perkins Marsh Institute, Clark University, Worcester, MA

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

Dr. Patrick Kinney, Associate Professor, Department of Environmental Health Sciences, Mailman School of Public Health , Columbia University, New York, NY

Dr. Steven Kleeberger, Professor and Lab Chief, Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC

Dr. Timothy V. Larson, Professor, Department of Civil and Environmental Engineering, University of Washington, Seattle, WA

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

Dr. Richard Schlesinger, Associate Dean, Department of Biology, Dyson College, Pace University, New York, NY

Dr. Christian Seigneur, Director, Centre d'enseignement et de recherche en environnement atmosphérique (CEREA), Ecole nationale des ponts et chaussées (ENPC), Université Paris-Est, CEREA - ENPC, Marne la Vallée, , France

Dr. Elizabeth A. (Lianne) Sheppard, Professor, Biostatistics and Environmental & Occupational Health Sciences, School of Public Health, University of Washington, Seattle, WA

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

Dr. George Thurston, Professor, Environmental Medicine, NYU School of Medicine, New York University, Tuxedo, NY

Dr. James Ultman, Professor, Chemical Engineering, Bioengineering Program, Pennsylvania State University, University Park, PA

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

SCIENCE ADVISORY BOARD STAFF

Dr. Angela Nugent, Designated Federal Officer, 1200 Pennsylvania Avenue, NW 1400F, Washington, DC, Phone: 202-343-9981, Fax: 202-233-0643, (nugent.angela@epa.gov)

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. The 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
CASAC Responses to Agency Charge questions

Discussion and response to Agency charge questions relating to characterization of air quality (chapters 2, 5, 6, and 7)1:

1. Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?

There were substantial improvements in this work since the last version. There has been a good effort to incorporate more information about sites but concern remains that siting features are not well understood and that the monitor location selection may be key to the inferences that have been drawn from the monitoring data. SO_2 concentrations are highly influenced by local sources. The air quality analysis assumes that the universe of monitoring data represents a reasonable sample for analysis, yet it is not known whether the available monitors are representative of the concentrations at which exposures are received by the community. Furthermore, there is an assumption that site-years are exchangeable. There should be a deeper description of the monitoring network design in Chapter 2 and further characterization of the network features in Chapter 7.

A second concern is that only one 5-minute exceedance per day was counted. While from some policy perspectives this choice is reasonable, this approach needs to be described and justified early in the document. Consider new wording "number of days with at least one 5-minute concentration above potential health effect benchmark levels" instead of "numbers of daily maximum five-minute concentration exceedances" or similar language.

2. In order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO_2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

The panel recognizes that proportionally increasing the concentrations up to just meeting the standard does not fully account for all the reasons why current SO_2 levels are under the current NAAQS, but agrees with the decision to use a simple, transparent approach that does not involve making a number of additional assumptions.

3. In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?

Choosing urban areas with multiple monitors for inclusion in the analysis is certainly reasonable. The increase to 40 counties in the exposure assessment was seen as a substantial

improvement over the prior document. Including an additional metric targeting those urban areas with levels relatively close to the current standard is also reasonable. It would be interesting to know how many of these counties are classified "c" with respect to their coefficient of variation (potential for relatively high peak to mean ratios), and alternately, how many were not included. This information is in the Appendix and could easily be extracted in a few sentences.

The panel generally supports use of both of the criteria that EPA chose to use for selection of the 40 locations. Selection by the lowest mean adjustment factor means selecting by relatively high SO_2 levels and thus offers EPA the opportunity to evaluate the effectiveness of candidate alternative standards in places where standards may produce the greatest benefits. It is also defensible to select counties with at least two working monitors so that the analysis will be based on a more robust data set than would be the case if only a single monitor were used to characterize the whole county. One panelist offered another possible selection criterion-that is based on relatively high values of the Coefficient of Variability statistic for the temporal variation in SO_2. Some panelists expressed disappointment the data from these 40 counties were not explored in greater depth.

4. *What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?*

The panel commends the EPA for progress in this arena. The panel has expressed an over-arching concern about the communication of uncertainty in the REA and provide general advice relating to uncertainty on pp. 11-12 of these consensus comments.

Specific to air quality, the REA should be revised in several ways. In the table summarizing air quality uncertainties, the REA should replace the general term "uncertainty" with the more specific term "imprecision," as appropriate, and add an assessment of impact of each source. A few key uncertainties are omitted: representativeness of the monitoring network (both the full network and the two subsets with 5-minute data), and the assumption that site-years are exchangeable. The revised REA should also re-evaluate the uncertainty characterization of the spatial representation.

Discussion and response to Agency charge questions relating to characterization of health effects evidence and selection of potential alternative standards for analysis (chapters 3, 4, 5)

1. *The presentation of the SO_2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO_2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?*

The panel finds the presentation of SO_2 health effects evidence in the draft REA to accurately reflect the overall characterization contained in the final ISA for SO_2. However, the

panel advises EPA to revise Chapter 4 extensively. The REA should not simply list relevant studies; the REA should instead present an integrated discussion of health effects.

Another concern relates to the discussion of the concepts of susceptibility and vulnerability in Chapter 3. In particular, while the panel appreciates the addition of Table 3-1, it advises EPA to correct the list of specific susceptibility and vulnerability factors. CASAC found the discussion of vulnerability and susceptibility in the ISA for particulate matter to be better developed and more informative and the panel suggest revising Chapter 3 along the lines of this discussion. The discussion about the choice of endpoints and their adversity needs to be expanded

One major issue that is not dealt with directly in the draft REA is the different time frames in which responses are observed between the results of controlled human exposure and epidemiological studies. The former show brief exposures to SO_2 cause transient bronchoconstriction and respiratory symptoms and the latter observed associations between exposures to SO_2 and respiratory symptoms in asthmatic children after multi-day lags. The draft REA would be strengthened by discussion of potential mechanisms by which brief exposures to SO_2 might lead to exacerbations of asthma a few days later. The REA should also address the different ages of subjects in the clinical and epidemiological studies and how this difference relates to the characterization of health effects evidence.

2. *The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?*

The presentation was generally sufficient and appropriate. A substantive case was made for consideration of a shorter-term standard that might obviate the need for the existing forms of the standard. EPA should seek continuity in approach and presentation across pollutants and documents (e.g., the susceptibility/vulnerability presentation in this document should incorporate recommendations made in a similar section in the PM document).

The alternatives focus on the clinical studies carried out over many years and, particularly for this pollutant, form a justified basis for the selection of range of exposure for which susceptible individuals (asthmatics) are consistently responsive. The choice of the endpoints needs to be rationalized in terms of their adversity and clinical significance. Some discussion is needed that indicates why the risk assessment for this pollutant, in contrast to others, is limited only to health effects that are classified as sufficient to infer causality. Additional discussion indicating that a small fraction of potential susceptible subjects might remain even at the lowest levels of assessment is needed to inform the question of uncertainty as it applies to considering an adequate margin of safety.

Discussion and response to Agency charge questions relating to characterization of exposure (Chapters 6 and 8):

1. Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?

The approach to estimating 5-minute peak SO_2 levels is reasonable and clearly communicated, as is how air quality is adjusted to meet various benchmark levels. The use of APEX and AERMOD are appropriate for conducting the exposure analysis.

A key weakness is the failure to acknowledge the lack of evaluation of APEX for SO_2. APEX is a complicated model, and not being able to evaluate the results in this application should be discussed. The number of exceedances predicted by APEX above the benchmarks analyzed is a small fraction of the total number possible, and they represent the exposures at the high end of the distribution. As such, they will be very sensitive to biases in the input air quality fields, but this sensitivity has not been explored. EPA should also consider previous efforts to evaluate or partially evaluate APEX for estimating exposure to other pollutants, and their applicability to evaluating the use of APEX for estimating exposure to SO_2, as well as key differences.

2. The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

The committee endorses the inclusion of St. Louis in this analysis, thereby capturing an urban area with a relatively high population density and moderately high SO_2 emissions compared with other urban areas.

3. What are the views of the Panel regarding the approaches taken to model SO_2 emission sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?

CASAC encourages EPA to include a more extensive discussion of the agreement between the model and measurements in these areas. There is insufficient attention paid to characterizing potential biases in AERMOD results at the higher levels (95th percentile and above), and how those impact the predicted exceedances of benchmark levels from APEX. Further, the ability to simulate the 5-minute peaks should be assessed for Greene County where such observational data are available. Concern is expressed above as to the adjustment of non-point source emissions.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

The panel provides general advice relating to uncertainty on pp. 11-12 of these consensus comments.

In characterizing exposure, the assessment of uncertainty and variability is extensive and the REA does a good job in suggesting potential biases in the results due to the uncertainties discussed. uncertainty that is missing, and might be large, involves APEX results As noted above, the APEX exposure results are not evaluated for this application, and while it may be viewed that the uncertainty in any one model component may be small or medium, the overall combined uncertainty of the model results may be large, and significant biases may exist. Further, the model is being used to predict extreme events which further challenge the system's (AERMOD + APEX) capabilities. Not being able to evaluate the system has implications as to how one might perceive the risk characterization results. The exposure assessment need investigate further the validity and implications of the assumption that activity patterns are similar for asthmatics and healthy individuals.

5. *What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?*

The characterization focused on time spent outdoors and distribution of asthma prevalence. These were reasonably characterized although the higher prevalence of asthma in the northeast suggests future analyses should focus on that region. The discussion of representativeness of these two areas should also consider other spatial locations in the U.S., regardless of presence of SO_2 monitoring data. An assessment of the key features that distinguish St. Louis vs. Greene County may lead to insights about other U.S. locations. The committee found the staff analysis presented at the meeting comparing these two locations with the 40 county-based monitors to be useful and informative and recommend that this analysis be included in the document.

Discussion and response to Agency charge questions relating to characterization of health risks (chapters 7, 8, 9):

1. *Based on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO_2 exposures, we have adjusted our range of 5-minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO_2 exposures evaluated in the ISA?*

The authors have conscientiously used conclusions from the ISA to appropriately adjust the range of five-minute potential health effect benchmark values. Potential health effect benchmark values from 100 to 400 ppb have been carefully characterized in this REA by using clearly appropriate parameters and clearly detailed applications of models. The discussion was generally clear and compelling, but clarity of specific terminology usage (such as "benchmark") would be aided by a glossary for ready reference.

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?

The draft REA has done a comprehensive job of characterizing the health risks of SO_2. However, some minor additions would improve the presentation of the work. Choosing FEV_1 and sRaw as measures of health effect in the quantitative assessment while deciding against the use of respiratory symptoms should be better rationalized. Why the focus of the risk characterization was on a single hourly peak concentration at the exclusion of possible health effects caused by multiple peaks within an hour needs to be better explained. It would also be helpful to include data supporting the use of a concentration benchmark that is independent of the particular level of physical activity (once ventilation per unit body surface area is above a threshold level).

3. *A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?*

The EPA staff has done an excellent job of conducting the selected quantitative risk assessment for the two chosen indicators in these specific two counties. While the rationales for using St. Louis and Greene counties for the analyses was deemed reasonable, the fact that these two counties appear to be in the upper half of 40 US counties (with respect to emissions, exposure, proximity to population centers, etc) rather than in the extremes, raised some concern as to whether the full range of the situations has been as fully characterized as possible. The "Additional Representativeness Evaluation of St. Louis and Green County Air Quality" table presented to the CASAC panel at the April 16, 2009 public meeting was found to be helpful in addressing this concern. While some members raised concerns regarding the usefulness of applying the available SO_2 epidemiological studies to the risk assessment, in agreement with the EPA's choice not to do so, two panel members expressed in their comments the opinion that the application of the SO_2 epidemiological study concentration-response coefficients to EPA's BENMAP model for the full 40 counties would provide another useful perspective that would strengthen the risk characterization overall.

4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

In general, the panel supports EPA's efforts to characterize uncertainty. The comments here are intended to guide EPA in more fully interpreting the uncertainty characterization that has been conducted. The panel recommends that EPA revise the material on uncertainty analysis taking into account the following points:

- A clear purpose should be stated for the assessment of uncertainty.
- A primary purpose of the uncertainty characterization is to make a judgment of the weight of evidence and degree of confidence supporting the assessment endpoint. EPA

has handled the weight of evidence aspects well, and has appropriately identified causality between SO_2 exposure and respiratory morbidity. EPA should further expand its assessment of the degree of confidence or certainty with which a particular alternative level, form, and averaging time is protective of public health.

- EPA should provide comment on the degree to which characterization of uncertainty can be used to inform issues of margin of safety. This is done to some extent in the current draft REA, such as in the discussion at the end of chapter 10 of the implications of considering that the clinical data are not likely to include the most sensitive subgroups among asthmatics. However, the implications of uncertainty for margin of safety should be explored systematically for all key sources of uncertainty.
- The panel is careful to point out that uncertainty is not the same as doubt. One can have adequate weight of evidence to support a determination of causality, and have uncertainty regarding the exact relationship between exposure and health effects, and nonetheless have an adequate basis for regulatory decision making.
- Another purpose is to compare, on a relative basis, the uncertainty in the assessment endpoint attributable to specific sources for the purpose of identification and prioritization of data collection and research needs. In this regard, the qualitative uncertainty characterization can be used to infer a research agenda that could be implemented to improve the state of knowledge for the next revision of the standard five years from now.

The REA should discuss the robustness of the technical analyses regarding air quality, exposure, and health effects assessment to aid the Administrator in interpreting the assessment results. It is reasonable to point out that uncertainty typically exists in complex scientific assessments such as this. Quantification of uncertainty is good scientific practice, and robust inferences are possible even in the face of uncertainty, and that there is a long track record of Agency decision making in the face of uncertainty (a recent example being the April 12, 2009 decision by the Administrator regarding the "Proposed Endangerment and Cause or Contribute Findings for Greenhouse Gases under the Clean Air Act.) There are ongoing efforts within EPA, such as by the Probabilistic Risk Assessment (PRA) working group of the Risk Assessment Forum, to address how decisions regarding risk management have been, can or should be made taking uncertainty into account.

There are various specific comments on the uncertainty assessments of Chapters 7, 8, and 9:

- EPA has adapted a qualitative assessment methodology based on the World Health Organisation, Harmonization Project Document No. 6, Part 1: *Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment* (2008). EPA should provide an explanation of why a qualitative approach was selected, the specific adaptations made, and justification of the adaptations.
- EPA must carefully define terms such as "bias" and "uncertainty." "Uncertainty" is often interpreted to include components of bias and imprecision, also referred to as (lack of) accuracy and (lack of) precision, or systematic and random error, respectively. EPA should clarify how it is addressing random error or imprecision, as distinct from bias.

- For Tables 7-14, 8-16, and 9-10, a clear statement is needed of what is the assessment endpoint for which uncertainties are being characterized.
- Some of the material, especially in Section 7.4, is difficult to follow and should be rewritten.
- In general, Chapters 7, 8, and 9 include significant discussion of uncertainty. There is, however, little discussion of uncertainty, but little discussion of variability. EPA should briefly summarize how variability is characterized, its implications, and how variability is distinct from uncertainty.
- Another issue for clarification is the choice to base predictions of sRaw responses only on a fit to the logistic model. A probit model fit is considered but the discussion is based only on goodness-of-fit criteria which are reportedly equivalent between the two models. This neglects the *a priori* reason to prefer the probit model (based on a hypothesized lognormal distribution of individual thresholds for response). The discussion could be strengthened and uncertainty more fully communicated by noting that there is an approximately 5-6 fold difference between probit and logistic model predictions for the aggregate number of sRaw responses for St. Louis as shown in Table 9-2.

Overall, the panel commends EPA for undertaking a systematic assessment of uncertainties.

Policy Assessment (Chapter 10):

1. *The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?*

Overall, Chapter 10 was well written and the integration was clearly communicated and appropriately characterized. Staff did due diligence in consideration of the available evidence for consideration of current and potential alternative standards. However, the suggested decision process associated with considering an alternative shorter-term (e.g., one-hour average) standard, and the implications of a shorter-term standard on compliance with longer-term (e.g., 24 hour or annual average) standards, could be made clearer. The document seems to convey that a starting point for the decision is to determine the need for a one-hour average standard, and set its form, level, and indicator. For indicator, SO_2 is clearly the preferred choice. The document implies that there may be a sequential decision process for short and long-term standards, given that compliance with a possible one-hour SO2 standard might imply 24-hour and annual averages below the current standards.

The document implies that if a 1-hour standard is to be developed, that the choice of level should be informed by keeping in mind that health effects are associated with 5-10 minute exposures. Hence, the analysis supporting a one-hour average level that offers protection from peak five-minute average concentrations should be explained and interpreted more thoroughly, also taking into account implications for margin of safety.

The final "Conclusions regarding level" (section 10.5.4.3) was internally inconsistent. The beginning statement in the section "provisionally concludes that the evidence and exposure

and risk information reasonably support a 1-hour daily maximum standard within a range of 50-150 ppb" and concludes "if the alternative standard selected is not expected to prevent ambient SO_2 concentrations from exceeding the levels of the current standards, it would be appropriate to consider retaining the current NAAQS." The evidence presented throughout the "Potential Alternative Standards" (section 10.5) and the language used clearly fall in support of a 1-hour daily maximum standard and the conclusion should reflect this.

.

Chapter 10 should better address uncertainty in identifying alternative NAAQS for SO_2. In particular, the uncertainties discussed in the health risk characterization should be considered in specifying a NAAQS that provides adequate margin of safety. One particular source of uncertainty needing acknowledgment is the characteristics of persons included in the clinical studies. The draft REA acknowledges that clinical studies are unlikely to have included severe asthmatics that are likely to be potentially at greater risk than those persons included in the clinical studies.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?

Assuming that EPA adopts a one hour standard in the range suggested, and if there is evidence showing that the short-term standard provides equivalent protection of public health in the long-term as the annual standard, the panel is supportive of the REA discussion of discontinuing the annual standard. Chapter 10 does a good job showing that an annual standard is not justified; Tables 10.3 and 10.4 are very useful in this regard.

Despite much discussion demonstrating the inadequacy of the current 24-hr standard, however, the text did not make a strong statement about whether the 24-hour standard should be retained, although the evidence presented (Table 10.3) was convincing that some of the alternative one-hour standards could also adequately protect against exceedances of the current 24-hour standard. The panel agrees that a one-hour standard is the preferred averaging time and the text should clearly justify why a one-hour standard would be preferred over a five-minute standard. The chapter, while a very good synthesis of the rest of the REA document, should expand its major conclusions as a final synthesis of conclusions in the REA.

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?

The authors of Chapter 10 have done an excellent job in distilling the information in the ISA and the REA. They show that the proposed alternative 1-hour daily maximum SO_2 standard is predicated upon the intersection of airway hyperresponsiveness [asthma] combined with exercise. The conclusions are presented in a systematic fashion and are coherent and compelling.

The panel supports serious consideration of a one-hour standard. The panel agrees that the current 24-hour and annual standards are not adequate to protect public health, especially in relation to short term exposures to SO_2 (5-10 minutes) by exercising asthmatics. However, there

is ambiguity as to whether the one-hour daily maximum SO_2 standard should replace the 24-hour and annual standards. The REA should explain how analyses will inform a decision with regard to changing or revoking the 24-hour and annual average standards, if a one-hour standard is implemented. The merits of a single versus two or three-standard approach should be presented. Recommendations and the supporting rationale should be clear.

The form of the standard was also discussed. There is adequate information to justify the use of a concentration-based form averaged over 3 years. There is also a provisional suggestion to use the 99th percentile to reduce the number of days allowed to exceed the selected level. We recommend that the REA better discuss the rationale for selecting the 99th rather than 98th percentile The comparison of the 98^{th} versus 99^{th} percentile form in Figure 7-18 may provide a useful starting point for more complete discussion.

In conclusion, the panel finds the rationale for a 1-hour daily maximum standard is convincing. The panel believes that more effective protection against short term health effects is critical.

4. *Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50- 150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?*

Information regarding weight of evidence and uncertainty can be used to inform choices of the margin of safety (a policy choice) with which to develop a standard that protects public health. Chapter 10 clearly provides sufficient rationale for the range of levels beginning at a lower limit of 50 ppb. An upper limit of 150 ppb posited in Chapter 10 could be justified under some interpretations of weight of evidence, uncertainties, and policy choices regarding margin of safety. The draft REA appropriately implies that levels greater than 150 ppb are not adequately supported. The panel agrees that the posited range of 50 to 150 ppb and the exposition of factors to consider when comparing values within the range are appropriately conveyed. However, the REA should more thoroughly explore the implications of the characterization of uncertainties with respect to interpretation of the degree of confidence regarding key metrics by which potential levels could be evaluated, such as: (a) the estimated number of days per year where five-minute daily maximum SO_2 concentrations exceed selected benchmarks among 40 monitoring sites; and (b) the number and percentage of asthmatics at elevated ventilation rates who experience one or more exposure events above selected benchmarks in the St. Louis area. In particular, the implications for bias and imprecision in the estimates and for margin of safety in comparing possible levels should be further discussed.

REFERENCE

World Health Organisation, Harmonization Project Document No. 6, Part 1: Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment, International Program on Chemical Safety, World Health Organization, and Co-sponsored by International Labour Organization, and the United Nations Environmental Programme, WHO Geneva, Switzerland, 2008. (http://www.who.int/ipcs/publications/methods/harmonization/exposure_assessment.pdf)

Enclosure C:
Compilation Comments from Individual Panel Members

Compilation of Individual Panel Member Comments on EPA's second draft *Risk and Exposure Assessment (REA) to Support the Review of the SO_2 Primary National Ambient Air Quality Standard*

This enclosure contains post-meeting comments from individual members of the Clean Air Scientific Advisory Committee (CASAC) Sulfur Oxides of Nitrogen Primary National Ambient Air Quality Standards (NAAQS) Review Panel. The comments are included here to provide both a full perspective and a range of individual views expressed by panel members during the review process. These comments do not represent the views of the CASAC or the CASAC Panel.

Comments Received:

Comments from Prof. Ed. Avol

Charge Question Responses:

Characterization of Air Quality:

1. Yes, the document presents the steps taken and results found in a generally logical and understandable manner.
2. The approach seemed was understandable and a reasonable one. However, there is one technical concern – if Staff lack confidence in the robustness of the national 5-minute SO2 data, but seem prepared to accept that there may be measurable and significant health effects from 5-minute exposures, then it would be logical to have a recommendation forthcoming for an expanded network of five-minute reporting data sites.
3. Expansion of the number of counties for evaluation and use of the 2001-2006 time frame seemed justified in the document.
4. The uncertainty/variability presentation was useful, and I especially appreciated the clarity and utility of Table 7-14 (summarizing the qualitative uncertainties). One outstanding aspect of the presentation is that, regardless of whether on agrees or disagrees with the merits of the presentation, the basis for the determinations are clearly presented and generally transparent to the reader.

Characterization of Health Effects Evidence…

1. The discussion and presentation seems consistent with the findings of the ISA. However, the sections and discussions presented regarding susceptibility and vulnerability are incomplete and in some cases, inconsistent and in need of revision (see specific comments on Chapter 3 below). In some sections (see specific Chapter 3 comments below, it seemed that the REA was reproducing sections of the ISA, rather than drawing from it in summary fashion.
2. The rationale for potential alternative standards selection was generally clear and sufficient. I found the discussion to be useful and appropriate.

Characterization of Exposure

1. The exposure analyses seemed sound and well-communicated.
2. It was insightful to follow the presentations for St. Louis and Greene counties; the presentation was informative; I don't have any specific concerns to voice at this time.
3. I will defer to the modeling experts for definitive guidance on the approaches taken. APEX seemed an appropriate choice. Selections for AERMOD and decisions in the course of model settings seemed clearly presented for the reader to follow. The model runs seemed to capture the general shape of ambient levels well, if not the absolute magnitude of them. There seemed to be ample description and explanation of what was being done, and the choices being made.
4. The uncertainty and variability discussions were helpful and added to the credibility of the document.
5. The Staff argument for the representativeness of St Louis and Greene counties in representing the entire country seemed a little thin. The conclusion that "…some were smaller, some were larger…" seemed vague. Should a "high" and a "low" county have been chosen to demonstrate more of the possible range, instead of two counties somewhere in the range?

Characterization of Health Risks

1. The rationale and decision process to adjust the range of five-minute potential health effect benchmark values to 100-400ppb SO2 is well-described and supported by the references studies.
2. The risk characterization results seem to be aprporiately presented, explained, and documented.
3. The rationale for using St. Louis and Greene counties for the analyses seem reasonable, but the fact that these counties appear to be in the upper half of US counties (with respect to emissions, exposure, proximity to population centers, etc) rather than in the extremes, leaves me wondering how or if the "bottom-line message" might have changed if more extreme edges of the county distribution (perhaps a 5^{th} percentile and 95^{th} percentile, or a 10^{th} and 90^{th}) had been used instead.
4. The use of a tabular summary to codify the magnitude and direction of various uncertainties (Table 9.10) is very helpful. The text discussion of uncertainty seemed appropriate and sufficient, but the variability discussion seemed minimal. However, since the tenor of the variability discussion seemed to be "we don't know", perhaps not much more needs to be said).

Policy Assessment

1. In my reading, the policy chapter did integrate the risk and exposure information in an understandable manner.
2. The discussion of considerations related to adequacy of the current standards was appropriate and sufficient.
3. The policy chapter presented the implications of the alternate 1hr standards in an understandable manner.
4. The rationale for a 1hr standard seemed understandable and well-presented. The tradeoffs and implications as to how a 1hr standard in the range of 50-150ppb SO2 would compare to the current NAAQS was also well-presented.

<u>General Comments on REA 2^{nd} Draft</u>

The document reads well, is generally easy to follow and understand, and usually clearly makes its summary points. In that context, the summary sections ("Key Observations") at chapters' end, with bullet summaries of the key points, is especially useful and should serve as a prototype for all similar future documents. Lessons learned in other criteria pollutant reviews ought to transcend specific pollutants whenever possible. For example, comments and concerns regarding susceptibility and vulnerability, presented in the context of the PM review, should be carried over to the SOx documents. Treatment of and decisions about the five-tier causality scheme should be applied across ALL pollutants consistently (or the reasoning as to why this is not consistent across pollutants should be presented).

<u>Specific Comments on REA 2^{nd} Draft Sections</u>

Chapter 1:

1. In the 2^{nd} draft REA for SOx (P11, Section 1.2.2 Species of Sulfur Oxides Included in Analyses), it is explained that only gaseous components of sulfur oxides are considered under the SOx review, because sulfates will be considered under the PM review.

However, in the PM review, the decision is made to consider PM on the basis of size-fractionation, rather than chemical composition. This underscores the continued difficulty of dealing with pollutants in ambient air as if they were single-entity exposures (which they clearly are not), rather than the complex mixtures of gases AND particles (which they clearly are). Looking towards the future, Staff needs to consider how to deal with multi-pollutant exposure scenarios.

Chapter 3:

1. Table 3-1, P 18 – The "Vulnerability Factors" portion of this table needs some re-examination, as several of the listed factors are sub-sets of other factors (for example, increased exertion levels are a component of increased activity patterns; geographic location is not clearly a vulnerability factor but is a part of geographic location; lower education level is often considered a part of lower SES), and other listed factors (such as limited air conditioner use) seem a part of something else (microenvironmental location?). The delineation between susceptibility and vulnerability may be a useful distinction to make, but the current presentation does an ineffective job of making it.
2. P19, Susceptibility discussions – These sectional discussions could be made more focused and useful if they concluded with a summary statement about the subject of the section. For example, the section summarizing what is known about susceptibility of pre-existing disease could conclude that evidence exists for concern about subjects with pre-existing respiratory disease, but that the implications of pre-existing cardio-vascular disease are inconclusive at this time.
3. P20, lines 5-8 – The summary nature of this REA is being violated here, by a reporting/review of what was found in a specific study (which would seem more appropriate for the ISA or annex materials). It would be sufficient to reference the study as having demonstrated a genetic association, but that the overall body of evidence was still too limited to reach broader conclusions.
4. P20, Susceptibility discussions – Summary judgments are provided on the strength of evidence for age, genetics, and pre-existing disease, but the other listed susceptibility factors in Table 3-1 (gender, race, ethnicity, obesity, adverse birth outcomes) are not mentioned. Are these not important? Is nothing known about these other "factors"? A comment about them would seem appropriate, or else their inclusion into the table seems odd and possibly unsupported.
5. P21, Section 3.5 Vulnerability – As with the preceding section on Susceptibility, this section rightfully sets out (I think) to summarize the strength of evidence about vulnerable populations, but only mentions three of many factors (microenvironmental location, increased exertion levels, SES). Moreover, the section's conclusion is about the limited information about SES, and does not say anything about the larger topic of vulnerability and whether such a state has been adequately demonstrated for a subset of the population.
6. P21, Section 3.6 Number of Susceptible or Vulnerable Individuals – The conclusion of this section, that there are substantial numbers of people potentially at risk, seems appropriate, but also seems inconsistent with the tenor of the previous paragraphs leading up to it. This may be an example of the appropriate conclusion being reached, without showing the appropriate reasoning. I recommend this section on susceptibility and

vulnerability – which is entirely appropriate and valuable – be reviewed and modified to reflect a more complete and logical path to conclusions.

Chapter 4

1. P23, lines 2-4 – It is stated that, for this document, the threshold used for characterizing health risks associated with SO2 exposure is evidence sufficient to infer a causal relationship (the uppermost level of the five-tier causal weights of evidence being applied. However, for the PM review, the first *two* levels (causal and likely causal) were proposed for use. This raises the question of consistency between criteria pollutant reviews; why is the threshold of causal evidence higher for SO2 than for PM?
2. P28, lines 13 forward to the chapter's end – The detailed discussion of specific studies seems more appropriate in the ISA or in an annex document. It is my understanding that the detailed discussion of specific studies is *not* the function of the REA. The summary determinations are useful and build upon the ISA, with appropriate references to supporting articles and data, but these final chapter sections seem to slide into a review of several studies (which has already been done in the ISA).

Chapter 5

1. P34, line 12 – "Indicator" is spelled incorrectly.
2. P35, line 19 – If it is indeed the case that Staff lack confidence in the robustness of the national 5-minute SO2 data, yet seem prepared to accept that there may be measurable and significant health effects from 5-minute exposures, then it would be useful to have a recommendation for an expanded network of five-minute reporting data sites.
3. Figures 5-4 on P41, and Figure 5-5 on P42 – The cited study author is incorrect in both of these figures – the author is "Lin", not "Linn".

Chapter 6

1. P51, line 10 – PMR used here without definition...until subsequent equation appears (define abbreviations the first time they appear in the text).

Chapter 7

1. P66, lines 5-20 – This is a somewhat convoluted and confusing discussion, making it difficult for the reader to follow. After re-reading it several times, some of the points began to come through, but a clearer presentation here would be a dramatic improvement.

Chapter 8

1. P199, lines 6-12 – The discussion regarding air conditioning prevalence rates raises a small question: does the 95.5% value used refer to presence of an air conditioning unit or the actual usage rate of such units (in other words, were usage rates assumed based on the presence of the unit at the home, or was some determination made regarding presence of units and electrical consumption in light of exceeded some temperature degree day threshold)?

Chapter 9

1. P248, line 4 – "Introduction" is mis-spelled.

Comments from Dr. John Balmes

My comments will be focused on the charge questions and confined to the areas of my expertise.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)
1. The presentation of the SO_2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO_2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?

The draft accurately reflects the characterization of the evidence regarding health effects of SO_2 in the ISA. The presentation is clear and appropriately balanced.

2. The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?

The rationale for the selection of potential alternative standards is clear and sufficient to justify their use in the air quality, exposure and risk analyses.

Characterization of Health Risks (Chapters 7, 8, 9):
1. Based on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO_2 exposures, we have adjusted our range of 5-minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO_2 exposures evaluated in the ISA?

As the draft points out, clinically relevant bronchoconstriction has been demonstrated in a substantial proportion of asthmatic subjects exposed to 200 ppb (the lowest concentration of SO_2 used) for 5 minutes (Linn et al., 1987). Given that only mild-moderate asthmatic individuals participated in this study, it is reasonable to infer, as does the draft, that exposure to lower concentrations for 5 minutes would cause some asthmatic individuals, especially those with more severe disease, to experience bronchoconstriction. The 100-400 ppb range for potential benchmark values adequately reflects the evidence from controlled human exposure studies presented in the ISA. However, there is epidemiological evidence that short-term exposure to levels below 100 ppb increases the risk of respiratory morbidity. Because the risk estimates presented in Chapter 9 included those for a 50 ppb alternative standard, there is some inconsistency across chapters of the draft.

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?

The risk characterization and lung function quantitative risk assessment appear to be technically sound and appropriately characterized. Communication of the results could be crisper. For example, Chapter 8 would benefit by a concluding "Key Observations" section that both Chapters 7 and 9 have.

3. A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?

While the risk assessment was limited to just two areas in Missouri and thus the generalizability of the results is an appropriate issue, the risk estimates do provide a useful perspective on the magnitude and distribution of bronchoconstrictor responses of asthmatic individuals for the alternative standards considered.

4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

The discussion of uncertainty and variability is improved in this draft, especially for the quantitative risk assessment in Chapter 9. While Chapter 9 has a text discussion of uncertainty and variability, a table listing the key uncertainties and a summary bullet, Chapter 8 only has a text discussion and Chapter 7 has no explicit discussion of uncertainty and variability. Chapter 10 again has a nice discussion of the implications of the key uncertainties for decision-making about the SO_2 air quality standard.

Policy Assessment (Chapter 10):
1. The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?

The integration of health evidence in Chapter 10 is technically sound, clearly communicated, and appropriately characterized.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?

The draft of Chapter 10 adequately characterizes the public health implications of the current SO_2 standards.

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?

The draft of Chapter 10 adequately characterizes the public health implications of the potential 1-hour daily maximum SO_2 standards.

4. Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50- 150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?

The draft policy chapter provides sufficient rationale for consideration of the proposed range of 1-hour daily maximum SO_2 standards.
.

Comments from Dr. Joseph Brain

Charge Question: Policy Assessment (Chapter 10)

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?

In general, the authors of Chapter 10 have done an excellent job in distilling the information present in the ISA and the RSA. They point to critical studies and discuss uncertainty, especially as it relates to sensitive individuals. The alternative 1-hour daily maximum SO_2 standard is predicated upon the intersection of airway hyperresponsiveness in asthmatics combined with exercise.

The chapter is clear and compelling, and succinctly summarizes the evidence for a standard that would better protect exercising asthmatics. I believe it makes sense to suggest a range of 50-150 ppb. The panel also concurs that an appropriate indicator for ambient SO_x is the continued use of SO_2. The summary supporting that conclusion is well documented.

The panel supports serious consideration of a 1-hour standard. Although there is little epidemiologic evidence to characterize brief exposures to SO_2, there is a compelling body of evidence from experimental clinical studies of sensitive individuals where responses to short term exposures to SO_2 are well documented. Of special interest are individuals who are exercising asthmatics. We agree with the conclusion that the current daily and annual standards are not adequate, especially in relation to short term exposures to SO_2 (5-10 minutes). The recommendation for a 1-hour daily maximum standard of 50-150 ppb seems reasonable. The panel was divided about the upper limit. Some were comfortable with 150 ppb; others thought it should be reduced to 100 ppb.

We note some ambiguity as to whether the 1-hour daily maximum SO_2 standard would replace the 24-hour standard, or whether both would be in force. The merits of one versus two standards should be more clearly discussed. What staff recommends is not clear.

In conclusion, the rationale for a 1-hour daily maximum standard is convincing. We believe that more effective protection against short term effects is critical.

Other Comments:

This second iteration of the REA has been carefully prepared. The results and conclusions are presented in a systematic fashion, and the document as a whole is coherent and compelling.

Chapter 1 is an excellent introduction. The inclusion of the "policy relevant questions" in Chapter 1 is useful and provides direction for the entire document. I particularly like the "Key Observations" at the end of most chapters. The use of bullets is also appropriate and helpful. The REA logically proceeds from characterization of sources to exposure to dose to health outcomes.

Some of the models are complex and difficult to understand. Thus their role is less evident to the novice reader not familiar with the assumptions and structure of some of the models, such as APEX and AERMOD.

I am impressed that in this document the detailed analyses now cover forty counties. This is a reasonable sample of the United States. Moreover, they have been picked with attention to variability in climate, topography, demographic diversity, and the mix of pollutant sources.

Short term (e.g. five-ten minute) responses are particularly important. Supporting this focus is the critical role of exercise for SO_2 and for sulfates. With increasing levels of exercise, ventilation increases. Thus, larger amounts of SO_x are inhaled per minute. Probably more important is the shift in pathway from nose breathing to mouth breathing. SO_2 is a highly water soluble gas, and the majority of SO_2 is absorbed in nasal mucus during nose breathing. However, at higher inspiratory flows and when inhaling through the mouth, then uptake in the upper airways is greatly diminished, and the exposure of large and small airways is greatly increased. This can trigger bronchoconstriction in susceptible individuals.

An important issue is the extent to which the REA corresponds to the final ISA. They should be looked at together to ensure that there is correspondence. Does the REA build appropriately on the ISA? As someone looking at this second draft REA without having examined either the ISA or REA, I also note the difficulty of assessing the extent to which this current draft has been responsive to earlier CASAC comments on the 1st draft of the REA.

The “five-minute potential health effect benchmark values” has been adjusted to 100-400 ppb. This may not be adequate in protecting the most sensitive asthmatics. We are looking at the convergence of two susceptibility factors. One is exercise (increased ventilation and mouth breathing) combined with airway hyperresponsiveness. This range may not adequately protect sensitive asthmatics when they exercise.

Comments from Dr. Ellis Cowling

Before dealing with the details of my specific assignment during the April 16-17, 2009 CASAC Peer Review of the Second Draft Risk and Exposure Assessment (REA) for SO_2, I would like to offer a few general comments and suggestions for improvement of these periodic NAAQS Review processes and the changes that are being made in both the organization and focus of these reviews.

The Clean Air Act (CAA) of 1970 established two general goals for management of air quality in the United States -- protection of human health and protection of public welfare. Section 108 of the CAA directs the Administrator of EPA to identify and list "air pollutants" that "in his judgment may reasonably be anticipated to endanger public health and welfare " and to issue air quality criteria for those that are listed – hence the term "Criteria Pollutants."

As described on pages 1 and 2 of the Second Draft REA for SO_2, the CAA further directs the Administrator of EPA to "promulgate and periodically review, at five-year intervals, primary (public-health based) and secondary (public-welfare based) National Ambient Air Quality Standards for such pollutants. Based on periodic reviews of the air quality criteria and standards and promulgate any new standards as may be appropriate. The Act also requires that an independent scientific review committee advice the Administrator as part of the NAAQS review process -- a function now performed the Clean Air Scientific Advisory Committee (CASAC)."

A secondary standard, as defined in Section 109, must "specify a level of air quality the attainment and maintenance of which, in the judgment of the Administrator, based on such criteria, is required to protect the public welfare from any known or anticipated adverse effects associated with the presence of [the] pollutant in the ambient air …" The welfare effects of concern include, but are not limited to "effects on soils, water, crops, vegetation, man-made materials, animals, wildlife, weather, visibility and climate, damage to and deterioration of property, and hazards to transportation, as well as effects on economic values and on personal comfort and well-being."

So far, the several Administrators of EPA since 1970 have:

1) Identified six specific "Criteria Pollutants" – carbon monoxide, ozone and other photochemical oxidants, sulfur dioxide, oxides of nitrogen, particulate matter, and lead – which have thus been designated officially as requiring development and implementation of National Ambient Air Quality Standards;
2) Emphasized protection of public health as the principal (and overwhelmingly important) *de facto* focus of concern within the Agency, and public welfare as a (rarely openly acknowledged) but distinctly less important *de facto* focus of concern;
3) Established Secondary (public-welfare-based) NAAQS standards for all six criteria pollutants that almost always were identical in form (including level, indicator, statistical form, and averaging time) to the Primary (public-health based) NAAQS standards for each of these six criteria pollutants;
4) Developed a long-standing tradition of dealing with these six specific air pollutants mainly on a "one-at-a-time" basis rather than collectively – i.e., without strong attention

to the frequent interactions and simultaneous occurrence of some of these pollutants as mixtures within the air in various parts of our country;

5) Maintained a reluctant attitude about the concepts of ecologically based "Critical Loads and Critical Levels" developed in Europe as possible alternative or additional approaches to air-quality management in the US; and
6) Maintained a long-standing general focus on the related concepts of:
   a) "Attainment counties and non-attainment counties,"
   b) "Attainment demonstrations" based on mathematical modeling of a limited number of exceedance events under extreme weather conditions, and
   c) "Local anthropogenic sources" as opposed to "both local and regional biogenic and anthropogenic sources of emissions."

In recent years, in contrast to several of the six ideas listed above, EPA has shown increased willingness to think more holistically – and in more fully integrated ways – about both the policy-relevant science and the practical arts of air quality management aimed at protection of both public health and public welfare. These shifts in both emphasis and approach have included:

1) Participation with other federal agencies and international bodies in discussions about the "One Atmosphere," "Critical Loads–Critical Levels," and "Multiple-Pollutant–Multiple Effects" concepts;
2) Adoption of the "NOx SIP Call" in 1999 and both the "Clean Air Interstate Rule" (CAIR) and the "Clean Air Mercury Rule" (CAMR) in 2005 with their more balanced perspectives about both regional (interstate) and local sources of emissions and interactions among NOx, SOx, VOCs, "air toxics," and mercury in the formation, accumulation, and biological effects of "ozone and other photochemical oxidants," and fine, coarse, thoracic, and secondary aerosol particles;
3) Recognition of both fine and coarse PM as complex and geographically variable mixtures of sulfate-, nitrate-, and ammonium-dominated aerosols; natural biogenic and anthropogenic organic substances; heavy metals including cadmium, copper, zinc, lead, and mercury; and some other miscellaneous substances;
4) More frequent discussion about of the occurrence and both ecologically-important and public-health impacts of mixtures of air pollutants; and, most recently
5) Making the unprecedented decisions (at least in the case of the NAAQS reviews for oxides of nitrogen and sulfur) to:
   A) Separate the preparation and review of documentation, the required CASAC and public reviews, and the final decision-making processes for the Secondary (public-welfare-based) National Ambient Air Quality Standards from the (previously always dominating) Primary (public-health-based) NAAQS review processes, and
   B) Prepare and publish a single draft plan for integrated [simultaneous] review of two different criteria pollutants (NOx and SOx), and
6) Identifying in advance a set of key "Policy-Relevant Scientific Questions" that are to be used as the primary focus of attention in the design and completion of all four major components of the new NAAQS review processes:
   A) The Integrated Review Plan (IRP),
   B) The Integrated Science Assessment (ISA),
   C) The Risk/Exposure Assessment (REA), and an operative

D) Policy Assessment (PA) that historically has been developed in the form of an "EPA Staff Paper" and in the case of the last three Criteria Pollutant review processes (for lead, ozone, and PM) were developed in the form of an "Advanced Notice of Proposed Rule Making (ANPR)."

[As all of us in CASAC are well aware, the recent NAAQS review for lead provided the first opportunity for CASAC to make a direct comparison between a PA developed in the form of an "EPA Staff Paper" and one developed in the form of an ANPR. In this particular case, CASAC found the Staff Paper much superior to the ANPR as a basis for setting NAAQS standards.]

All six of these adjustments in focus of attention, documentation requirements, and sequential procedures are being undertaken with the intention to:"

"… improve the efficiency of the process while ensuring that the Agency's decisions are informed by the best available science and timely advice from CASA and the public" … and

"… help the agency meet the goal of reviewing each NAAQS on 5-year cycles as required by the Clean Air Act without compromising the scientific integrity of the process."

Need for Policy Relevancy as the Dominant Concern in NAAQS Review Processes

In a May 12, 2006 summary letter to Administrator Johnson, CASAC Chair, Dr. Rogene Henderson, provided the following statement of purpose for these periodic NAAQS review processes.

> "CASAC understands the goal of the NAAQS review process is to answer a critical scientific question: *"What evidence has been developed since the last review to indicate if the current primary and/or secondary NAAQS need to be revised or if an alternative level or form of these standards is needed to protect public health and/or public welfare?"*

During the past 3 years, CASAC has participated in reviews for all six criteria pollutants and has also joined with senior EPA administrators in a "top-to-bottom review" and the resulting recently-completed revision of the NAAQS review processes. These two experiences have led to a seemingly slight but important need for rephrasing and refocusing of this very important "critical scientific question:"

> ***"What scientific evidence and/or scientific insights have been developed since the last review that either support or call into question the current public-health based and/or the current public-welfare based NAAQS, or if alternative levels, indicators, statistical forms, or averaging times of these standards are needed to protect public health with an adequate margin of safety and to protect public welfare?"***

With regard to the important distinction in purpose of the primary (public health) and secondary (public welfare) NAAQS standards, it is noteworthy that in all five cases in which a secondary NAAQS standard has been established, the secondary standard has been set "Same as Primary."

Thus, a second very critical scientific question that needs to be answered for all six criteria air pollutants is:

> ***"What scientific evidence and/or scientific insights have been developed since the last review to indicate whether, and if so, what particular ecosystem components or other air-quality-related public welfare values, are more or less sensitive than the populations of humans for which primary standards are established and for this reason may require a different level, indicator, statistical form, or averaging time of a secondary standard in order to protect public welfare."***

I hope these two "critical scientific questions" will be borne in mind carefully as CASAC joins with the various relevant parts of the Environmental Protection Agency in completing the upcoming reviews of both the primary and secondary National Ambient Air Quality Standards for SO_2 and, for that matter, also the other five Criteria Pollutants.

We now have the considerable advantage that a much more complete focus can be achieved in the Integrated Science Assessment than has historically been achieved in the encyclopedic Criteria Documents that have been prepared during the years since 1970.

Thus, several of us in CASAC have recommended that every chapter of the Integrated Science Assessment, Risk/Exposure Assessment, and the Policy Assessment documents for all criteria pollutants contain a summary section composed almost entirely of a series of very carefully crafted statements of Conclusions and Scientific Findings that:

1) **Contain the distilled essence of the most important topics covered in each chapter, and**
2) **Are as directly relevant as possible to the two Critically Important Scientific Questions written in bold italic type above**.

In this connection, I call attention once again to the attached "*Guideline for Formulation of Statements of Scientific Findings to be Used for Policy Purposes.*" These guidelines were developed and published in 1991 by the Oversight Review Board for the National Acid Precipitation Assessment Program. They are the best guides that I know of for formulation of scientific findings to be used for policy purposes.

GUIDELINES FOR FORMULATION OF SCIENTIFIC FINDINGS TO BE USED FOR POLICY PURPOSES

The following guidelines in the form of checklist questions were developed by the NAPAP Oversight Review Board to assist scientists in formulating presentations of research results to be used in policy decision processes.

1) **IS THE STATEMENT SOUND?** Have the central issues been clearly identified? Does each statement contain the distilled essence of present scientific and technical understanding of the phenomenon or process to which it applies? Is the statement consistent with all relevant evidence – evidence developed either through NAPAP research or through analysis of research conducted outside of NAPAP? Is the statement contradicted by any important evidence developed through research inside or outside of NAPAP? Have apparent contradictions or interpretations of available evidence been considered in formulating the statement of principal findings?

2) **IS THE STATEMENT DIRECTIONAL AND, WHERE APPROPRIATE, QUANTITATIVE?** Does the statement correctly quantify both the direction and magnitude of trends and relationships in the phenomenon or process to which the statement is relevant? When possible, is a range of uncertainty given for each quantitative result? Have various sources of uncertainty been identified and quantified, for example, does the statement include or acknowledge errors in actual measurements, standard errors of estimate, possible biases in the availability of data, extrapolation of results beyond the mathematical, geographical, or temporal relevancy of available information, etc. In short, are there numbers in the statement? Are the numbers correct? Are the numbers relevant to the general meaning of the statement?

3) **IS THE DEGREE OF CERTAINTY OR UNCERTAINTY OF THE STATEMENT INDICATED CLEARLY?** Have appropriate statistical tests been applied to the data used in drawing the conclusion set forth in the statement? If the statement is based on a mathematical or novel conceptual model, has the model or concept been validated? Does the statement describe the model or concept on which it is based and the degree of validity of that model or concept?

4) **IS THE STATEMENT CORRECT WITHOUT QUALIFICATION?** Are there limitations of time, space, or other special circumstances in which the statement is true? If the statement is true only in some circumstances, are these limitations described adequately and briefly?

5) **IS THE STATEMENT CLEAR AND UNAMBIGUOUS?** Are the words and phrases used in the statement understandable by the decision makers of our society? Is the statement free of specialized jargon? Will too many people misunderstand its meaning?

6) **IS THE STATEMENT AS CONCISE AS IT CAN BE MADE WITHOUT RISK OF MISUNDERSTANDING?** Are there any excess words, phrases, or ideas in the statement which are not necessary to communicate the meaning of the statement? Are there so many caveats in the statement that the statement itself is trivial, confusing, or ambiguous?

7) **IS THE STATEMENT FREE OF SCIENTIFIC OR OTHER BIASES OR IMPLICATIONS OF SOCIETAL VALUE JUDGMENTS?** Is the statement free of influence by specific schools of scientific thought? Is the statement also free of words, phrases, or concepts that have political, economic, ideological, religious, moral, or other personal-, agency-, or organization-specific values, overtones, or implications? Does the choice of how the statement is expressed rather than its specific words suggest underlying biases or value judgments? Is the tone impartial and free of special pleading? If societal value judgments have been discussed, have these judgments been identified as such and described both clearly and objectively?

8) **HAVE SOCIETAL IMPLICATIONS BEEN DESCRIBED OBJECTIVELY?** Consideration of alternative courses of action and their consequences inherently involves judgments of their feasibility and the importance of effects. For this reason, it is important to ask if a reasonable range of alternative policies or courses of action have been evaluated? Have societal implications of alternative courses of action been stated in the following general form?:

"If this [particular option] were adopted then that [particular outcome] would be expected."

9) **HAVE THE PROFESSIONAL BIASES OF AUTHORS AND REVIEWERS BEEN DESCRIBED OPENLY?** Acknowledgment of potential sources of bias is important so that readers can judge for themselves the credibility of reports and assessments.

My Assignment in this CASAC Peer Review of the Second Draft Risk and Exposure Assessment (REA) for SO_2

My specific assignments for review of the Second Draft REA for SO_2 were to examine those aspects of Chapters 6 and 8 that relate to "Characterization of Exposure." This same assignment was also given to my CASAC colleague Ted Russell whose is even more experienced than I am with regard to "Characterization of Exposure" to gaseous and particulate forms of sulfur compounds in the ambient air – both through direct measurements of air concentrations and through modeling analyses of spatial and temporal variability in exposure to sulfur compounds. Thus, I am looking forward very much to Ted's responses to the same five Charge Questions outlined in Lydia Wegman's letter to March 20, 2009 to Angela Nugent.

As I began my examination of this Second Draft REA for SO_2, it was a pleasure to find that pages 4 and 5 in Chapter 1 do indeed contain a list of 10 very detailed "policy-relevant questions" that relate directly to the issue of the adequacy or inadequacy of the existing primary NAAQS for SO_2 to protect humans from the adverse health effects of ambient sulfur dioxide. These 10 questions relate very well within the framework of the general purposes of these NAAQS reviews as outlined earlier in these individual comments:

> ***"What scientific evidence and/or scientific insights have been developed since the last review that either support or call into question the current public-health based and/or the current public-welfare based NAAQS, or if alternative levels, indicators, statistical forms, or averaging times of these standards are needed to protect public health with an adequate margin of safety and to protect public welfare?"***

The next step in my review was to examine each of the 10 Chapters of this REA document hoping to find summary statements of "Conclusions and Scientific Findings" that could guide my thinking about many of the myriad of important topics covered in each of these 10 Chapters – and especially the five Charge Questions that Ted Russell and I had been asked to review. As indicated above, I was very please to find that bulleted summary statements of conclusions and scientific findings were provided:

1) In the form of 10 summary statements of "policy-relevant questions" in the "Introduction" of Chapter 1; these same 10 "policy-relevant questions were also repeated in the "General Approach" part of Chapter 10.
2) In the form of two separate lists and a detailed table (Table 4-1) on "Weight of Evidence for Causal Determinations" in the "Introduction" of Chapter 4,
3) In the form of five "Key Observations" listed at the end of Chapter 7, and
4) In the form of a detailed list of 13 "Key Uncertainties" and also five "Key Observations" listed at the end of Chapter 9.

In all the other Chapters and three Appendices, however, it was necessary to slog through the text, figures, and tables and thus find out for myself how to separate the proverbial wheat" from the "chaff" and then try to draw logical inferences regarding the important Conclusions and Scientific Findings that need to be drawn from the large body of scientific information covered in the remaining five Chapters of this REA document (Chapters 2, 3, 5, 6, and 8) – which, perhaps by chance, included the two chapters (6 and 8) that I was assigned! With these general

remarks in mind, let me turn to my specific assignments and the 5 Charge Questions that both Ted Russell and I were asked to address.

In the paragraphs below, please note my individual responses (written in normal type) following each of the five Charge Questions (**written in bold type**) for my particular parts of these two chapters as provided in Lydia Legman's March 20, 2009 transmittal letter to Angela Nugent.

1. **Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?**

   Yes, in my opinion (as a mostly public-welfare savvy but a not so experienced public-health savvy research scientist), the exposure analyses described in Chapters 6 and 8 appear to me to be technically sound and appropriately characterized. My major concerns with regard to clarity of communication have to do with my inability to figure out what is meant the frequently used term "public health benchmark values." Although this term is used in many places throughout this REA document, and seems to be very important, I have no idea what is meant by what I suppose may be either a "term of art" in the medical science literature, or a specialized term used in EPA NAAQS review documents.

2. **The second draft REA evaluates exposures in St. Louis and Gene County, MO. What are the views of the panel on the approach taken to model SO_2 emission sources?**

   The approach taken in efforts to model SO_2 emissions sources, dispersal, transport, and air-concentration exposures in and around the City of St. Louis, MO and the much less densely urbanized area of Greene County, MO appear to be very similar to those used in the Southern Oxidants Study's 1993 through 2003 ozone and PM exposures in the areas surrounding Atlanta, Georgia and Nashville Tennessee in which I served as an important leader. Thus, the modeling approach taken in this REA document appear to be generally appropriate for the kinds of analyses needed to understand spatial and temporal variability in exposure to gaseous SO_2 and particulate sulfate within the two Metropolitan Statistical Areas in Missouri that were selected for exposure determinations in this REA.

   To what extent does this approach help to characterize the public health implications of the current standard? Does the panel have technical concerns with this approach?

   I have only very limited experience in the field of public-health assessments, and thus have no special competence with which to offer an informed judgment about the "public health implications of the current PM standards."

3. **What are the views of the panel regarding the approaches taken to model SO_2 emissions sources?**

   See comments in response to Charge Question 2, above.

4. **What are views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterizations been addressed? To what extent has variability adequately been taken into account?**

   Both uncertainty and variability in with regard to exposure estimates seem to have been covered pretty well. With regard to the implications of variability and uncertainty for health risk characterizations, however, I must admit to having only very limited experience and thus have no special competence with which to offer an informed judgment.

5. **What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposures and risk estimates?**

   Judging from the kinds of analyses and interpretations that we had to make in making decisions about "where to go next" after we completed our two-year-long Southern Oxidants Study investigations of ozone and PM production and accumulation in the 17 counties surrounding the Atlanta metropolitan area and the 11 counties surrounding the Nashville, Tennessee metropolitan area, it seems to me that EPA staff have done a very adequate job of determining the representativeness of the St. Louis and Greene County Missouri areas for the purposes of establishing National Ambient Area Quality Standards for SO_2 – recognizing, of course, that there are not very many urban and nearby suburban areas where both long-term and very short-term SO_2 monitoring data of adequate quality are available.

One additional point not related to the issue of Characterization of Exposure

The "history" part of Chapter 1 makes clear that the 1996 suit brought by the American Lung Association and the Environmental Defense Fund after the 1996 review of the SO_2 primary NAAQS standard regarding the need for a short term (e.g. 5-minute) NAAQS standard, led to a decision by the District of Columbia Court of Appeals that EPA had "failed to adequately explain the rationale for its decision NOT to promulgate a 5-minute standard."

Chapter 7 is the part of this REA document where 5-minute exposures are given relatively thorough attention. But the explanatory parts of Chapter 10, where the difficulties of establishing and implementing a five-minute exposure NAAQS standard are described, make me wonder if EPA may not come across once again as not giving a really adequate explanation of its reasons – if, it decides, once again, NOT to promulgating a 5-minute kind of NAAQS standard for SO_2.

Comments from Dr. James Crapo

Policy Assessment (Chapter 10)

Overall the second draft REA for SOx is well written, thorough and comprehensive. The REA appropriately reflects the data contained in the ISA and provides a strong rational basis for the proposal to establish a new short-term standard (1 hr) for SO_x. A substantial amount of both human clinical data and epidemiologic data demonstrate that there are significant adverse health effects associated with short-term (5-10 minute) excursions in SO_2 levels that would not be addressed or controlled by the current 24 hr average standard. Most adverse effects due to SO_x exposures are associated with short-term excursions that would be better reflected by a 1 hr standard than by either 24 hr or annual standards. It also appears that there would be little advantage to retaining 24 hr or annual standards once an appropriate 1 hr standard is in place. There would be expected to be few or nearly zero conditions under which environments would meet an appropriate 1 hr standard and not fail to also meet appropriate 24 hr or annual standards. The health data would also suggest that correlations with adverse health effects are much stronger for short-term excursions than for long-term cumulative exposures.

The staff recommendations for a 1 hr daily maximum standard within a range of 50-150 ppb are appropriate. This recommendation is supported by controlled human exposure data. The discussion in Chapter 10 regarding the conditions that would favor an ultimate standard either at the high end of the recommended range or the low end of the recommended range is appropriate and identifies the assumptions and uncertainties that would argue for a choice within the defined range. I would concur that a final standard within the range of 50-150 ppb for a 1 hr standard is strongly supported by the controlled human exposure data and by epidemiologic studies. A standard within this recommended range would be expected to appropriately protect the public health based on currently available data.

Comments from Dr. Douglas Crawford-Brown

This review is formed entirely around the charge questions, or at least the ones I felt competent to answer. I will note at first, however, that this was an impressive analysis by the EPA staff, covering an array of health measures that will inform regulatory decisions. The authors have focused attention onto the most significant health metrics and have produced an assessment that is consistent with the primary conclusions of the ISA. While quite long, the document is fairly easy to follow due to a good scheme for organization, with the reader able to skip over sections where they have insufficient expertise to move on to later sections, all without loss of information that will prove crucial later. This is due in large measure to a clear separation between steps in the assessment. There is also a good discussion, and science-based recommendations provided, for the form, averaging time, indicator and level.

I note also that this document addresses the most significant concerns raised by the CASAC in the previous draft review. I won't speak for other CASAC members, who understand their own initial concerns better, but at least in the case of my own concerns, these have either been addressed directly or have gone away due to the reorganization of the material.

I now turn to the specific charge questions:

Air Quality:

1. I will leave this to others with more expertise in this area. I do note that I found it simple to follow the assessment here, and that it was consistent with the findings of the ISA.

2. My view here remains as it was in the first draft: that I believe the methodology is computationally sound but results in a simulation that will have little relationship to actual exposures that will occur. But as this is a scenario assessment, and not an assessment of actual historical exposures, I am comfortable with the methodology. At the least, I cannot propose a methodology that would be better (only different). So, I support the use of this methodology.

3. I will leave this to others with more expertise in this area.

4. I believe the authors have responded adequately to concerns raised in the first draft. There is still no real nested variability/uncertainty analysis to provide quantitative estimates of the PDFs for both distributions. But the report identifies the major sources of each; gives at least a qualitative and at times a semi-quantitative estimate of the impacts of different variables; and helps the reader understand which are significant and which are less so. The reader is provided a les detailed and systematic view of variability than of uncertainty, but it is probably as far as that component can be quantified. I am inclined, therefore, to say the EPA staff has done enough work on this topic to satisfy regulatory needs.

Health Effects Evidence

1. I found this section good on all counts. It properly reflected the findings of the ISA, and the summary was sufficiently short and concise to focus attention onto those effects and

subpopulations that would form the basis of the health risk assessment. I see no evident bias in the presentation, or in its use in subsequent calculations.

2. I feel this selection is adequate and well explained. There are many different values that could be assessed, but the ones chosen cover the "space" of such values adequately for later regulatory decisions. I would not propose a more detailed mesh across these values as it is unlikely that there will be discontinuities in the region between any two alternative scenarios assessed.

Characterization of Exposure

1. There are two kinds of assessment conducted here: one based on air quality compared against benchmarks, and one based on APEX styles of assessment. In regards to whether air quality has been adequately simulated, I have to leave that to others with more expertise in the interpretation of monitoring results. I found it rather easy to follow the argument in the document, and to understand the results that were presented, but I don't know enough about this issue to have recognized gaps that might have existed or alternative and better ways to interpret the data. On the larger assessment rooted in APEX, however, I found the discussion easy to follow and the computational steps to be current state-of-the-art. My concern remains, as in all past reviews, that this level of detail in the assessment may go beyond the capacity of the scientific community to produce accurate depictions of exposure and risk, but even with the caveat I note that the authors have applied the methodology correctly and summarized results clearly.

2. I will need to leave this to others with more expertise on city and region-specific ambient air concentrations. However, the rationale for the selection is at least cogently presented.

3. I will leave this to others with more expertise in this area.

4. I found this part of the assessment to be less than fully informative, but probably about as far as things can be pushed at the moment. This a very complex set of assessments, and so there will naturally be some mixture of quantitative and qualitative methods. The current uncertainty and variability analyses succeeds in pointing the reader to most significant sources of U/V and giving a sense of both the direction and magnitude of impacts on the final risk numbers. That is about as far as we can push this issue at present. I would have liked to see a little more quantification of the impact of specific sources of uncertainty on key results such as numbers of days with an exceedence, but I also am not convinced that such information would prove determinative or even especially useful in setting standards.

5. I will leave this to others with more expertise in this area.

Health Risks

1. I am fully comfortable with this range as it stands. It is likely to include the values to be considered in regulatory decisions, and I am unconvinced of effects at below 100 ppb (which doesn't mean they don't exist, only that I think the uncertainty in their existence is too large at these lower levels).

2. I found the health risk characterization to be well developed and clearly explained. It is a bit overwhelming to go through such a large body of results and try to find a consistent and compelling story to tell in a way that will guide later decisions. But at least all of the information is there and the authors have provided some summary remarks that help set the stage for subsequent decisions. The problem with having such an array of information to digest is that decision-makers are left somewhat free to focus on the results they want to use, rather than those the scientific community judge to be most sound as a basis for public health protection. But again, the authors have provided summary conclusions that will help guide this process.

3. I am completely comfortable with the methodology and the results generated, as it is a methodology we have seen applied in a number of these NAAQS assessments. I continue with my reservation that such a detailed assessment may be somewhat outside my comfort zone given the existing state of the science, but there is no step in the assessment at which I would say a debilitating error or approximation has been introduced. I simply note that such assessments require some pretty specific simulations of human behaviour within the ambient air concentration field, and I am sceptical of our ability to specify these behaviours fully. So long as we recognize that these are simulations of scenarios rather than actual human populations – and that is all we can do at the moment – then I am comfortable with the methodology.

4. My comments here are the same as earlier, although amplified by the fact that this part of the document integrates information from all of the sections and, hence, the problems in uncertainty characterization are even more pronounced. This document doesn't come close to a fully quantified nested U/V analysis, but I don't believe that would have been feasible anyway. As in other sections, I came away understanding where the authors believe the major sources of U and V are located, and with some idea of the magnitude and direction of uncertainty introduced by each variable or model. That is all I would expect at the present.

Policy Assessment

1. I was pleased to see this section in the report. It does exactly what one would hope from such a chapter: summarize the information at a level of detail and resolution sufficient for the policy side to pick up and run through to a decision. I was looking for a bit more specificity on the policy implications in the chapter, but would also understand if the EPA's argument is that this would be outside the remit of an REA. At the least, this chapter helps bound the range of information the decision-maker must reflect on.

I like the fact that the chapter integrated material from the ISA and REA. The reason I say this is that it gives the policy-maker two ways to consider a standard: one based purely on the health effects information from epidemiological and clinical studies, and one rooted in quantitative risk assessment. I have been involved recently in European Commission deliberations on these same air pollutants, and am struck by how much less computationally intensive the EC process is compared to that in the US. There is more reliance here on simply asking for the levels of SO2 and other compounds at which health effects have or have not been noted, and then going forward with regulation based on these data. So I was happy to see that Chapter 10 gives a decision-maker information directly from the ISA that might inform a decision, while also providing the more detailed and computationally intensive results of the REA.

2. I am comfortable with this discussion, Both the ISA information and these REA data suggest the current standard is inadequate, and this chapter makes that point directly without over-stating the science.

3. Again, I am comfortable with the characterization and the implications drawn. There is a vast amount of information in both the ISA and REA, and the authors have distilled this information and drawn what I find to be sound conclusions that will be clear to decision-makers.

4. I am comfortable with this range. The authors have presented their rationale in a way that can at least be fully understood. I would have preferred to see a bit more of a discussion of how the uncertainty in health effects below 50 ppb cause this to be the lower bound to be considered, but also realize it is a judgment call as to whether my claim about the uncertainty is correct. In any event, I believe the final standard is likely to fall somewhere within this range anyway, and the document presents a good case as to why this is a reasonable range to consider.

Comments from Dr. H. Christopher Frey

Note: These comments incorporate and revise my previous comments, plus include some new comments.

Chapter 7, Section 7.4 – Uncertainty

The ISA refers to WHO (2008) as the basis for the qualitative uncertainty analysis approach that is used by EPA. However, EPA should explain why it chose a qualitative approach rather than a more quantitative approach. As WHO (2008) explains (p. 31):

> *Determination of an appropriate level of sophistication required from a particular uncertainty analysis depends on the intended purpose and scope of a given assessment. Most often tiered assessments are explicitly incorporated within regulatory and environmental risk management decision strategies. The level of detail in the quantification of assessment uncertainties, however, should match the degree of refinement in the underlying exposure or risk analysis. Where appropriate to an assessment objective, exposure assessments should be iteratively refined over time to incorporate new data, information and methods to reduce uncertainty and improve the characterization of variability. Lowest-tier analyses are often performed in screening-level regulatory and preliminary research applications. Intermediate tier analyses are often considered during regulatory evaluations when screening-level analysis either indicates a level of potential concern or is not suited for the case at hand. The highest tier analyses are often performed in response to regulatory compliance needs or for informing risk management decisions on suitable alternatives or trade-offs.*

Hence, the Tier 1 (Qualitative) approach is not a default. It should be a justified choice that is consistent with the purpose and scope of the assessment.

WHO specifies a structured approach to qualitative assessment of uncertainty that includes

1) qualitatively evaluate the *level of uncertainty* of each specified source;
2) define the major sources of uncertainty;
3) qualitatively evaluate the *appraisal of the knowledge base* of each major source;
4) determine the controversial sources of uncertainty;
5) qualitatively evaluate the *subjectivity of choices* of each controversial source; and
6) reiterate this methodology until the output satisfies stakeholders

Hence, there are three dimensions to the qualitative approach, as depicted in Figure 6 of WHO (2008). EPA seems to have created a different approach in which the level of uncertainty and the appraisal of the knowledge base are combined, and it is less clear as to the role of subjectivity of choice in the framework. Given the significance of the ISA and the apparent differences in approach from that in the WHO Guidelines, further explanation is needed.

EPA has adopted an approach that seems to focus on "bias" and "uncertainty." However, these terms are not defined (at least not in Section 7.4). One could argue that uncertainty includes both bias and imprecision, and thus it is inconsistent to refer to bias and uncertainty as if they are different (if the former is a subset of the latter). Perhaps "uncertainty" is intended to refer to imprecision, or random error. This should be clarified.

There is some inconsistency in terminology. In Table 7-14, the terms "medium" and "moderate" are used. Presumably, "moderate" should be replaced by "medium" through Section 7.4 (including text) for consistency. Furthermore, the reader presumes that each entry in the Table 7-14 is supported with explanatory text in the various subsections of Section 7.4. However, many of the subsections do not end with a clear statement as to the bias direction and characterization of uncertainty. The use of consistent labels in the "source" column of Table 7-14 and subheaders or perhaps bold-face identifiers in the text would help the reader in connecting text with specific rows of the table.

Setting aside possible questions regarding the validity of the qualitative uncertainty analysis, the reader is left wondering to what use EPA could, should, or will put the results given in Table 7-14. Two recommendations: (1) add a discussion in which the uncertainty results are compared between "sources" to assess which ones are deemed to have the most significant effect on the air quality and risk characterization; (2) develop a plan of action to address those sources with the highest relative uncertainties. The plan of action could include efforts to quantify the uncertainty or steps to reduce uncertainty by collecting more or better information or implementing quality assurance procedures, and so on, as appropriate. For example, if interference associated with ambient measurement leads to a "medium" uncertainty, which would rank this possibly as the 2^{nd} to 6^{th} highest uncertainty among all of the sources listed in Table 7-14, then it might be significant enough that some action should be taken to further evaluate and perhaps try to reduce this source of uncertainty. In the case of interference, the text implies that this issue may need to be investigated in more detail in order to improve the knowledge base.

A general comment is that I found this section somewhat difficult to read. There seem to long sentences and paragraphs. I found that I had to reread many sentences two or three times to try to figure out the intended meaning.

Specific comments:

Table 7-14: Spell-out PMR. Use "medium" consistently in place of "moderate"

Section 7.4.3. which paragraph addresses "scale" as given in Table 7-14? Should be labeled more clearly.

Figure 7-22. The "98 monitors" is confusing. Figure 7-21 implies that there are approximately 25 to 60 monitors, depending on the year, that reported both 5 min and 1 hr average SO2 concentrations. Hence, it is unclear how there could be 98 such monitors in Figure 7-22.

p. 142, middle of page – a reference to distributions of "nitrogen dioxide concentrations." Is this relevant to SO2? Explain why.

p. 143. Last paragraph – example of a run-on sentence (1^{st} sentence, 5-1/2 lines of text).

p. 145, top of page. Would it not be possible to check the assumption that the data removed does not create bias. Since the data were removed on the basis of the Peak-to-Mean Ratio (PMR), there are 5 min and 1 hr average concentration data available. One could compare the frequency distribution of the 5-min concentration data that were removed to the frequency distribution of the 5-min concentration data that were retained to see if they differ in any important way. Similarly, a comparison could be done for the 1 –hour concentration. In general, it is better to conduct a quantitative analysis where possible, rather than rely on assumptions that could be tested but aren't.

p. 146. Goodness-of-fit tests become very sensitive to even small deviations from the hypothesized distribution for sample sizes that are large, such as n=1,000, or n=3,800. It would be reasonable to look at the deviation of the fitted distribution versus the data to make a practical judgment as to whether the results of the goodness-of-fit tests should be used without question. However, there is also nothing incorrect about using empirical distributions of data, especially with such large sample sizes, unless one were attempting to make predictions that would require extrapolating beyond the range of the observed data. In the latter case, a plausible parametric model with a strong theoretical basis that also is a good empirical fit to the data might be used. Perhaps it is not necessary in this situation.

p. 149. The "sensitivity" runs are unclear as to what was changed from run-to-run. Was this a Monte Carlo simulation from an assumed population distribution in order to assess the effect of random-sampling error on the number of benchmark exceedences? What exactly was changed from one to the next when running the "ten independent model runs"? If the goal is to assess reproducibility (repeatability) then comparing multiple runs of the same simulation sample size but different random seeds would be adequate. What is the purpose of introducing the "100 model simulation"? It seems implied that the latter is some kind of ground truth, and that comparisons of the fluctuation of the results from the "ten model runs" with the "100 model simulation" provide some kind of indication of lack of bias. This section is difficult to follow because the study design does not seem to flow from the purpose that was given, nor is the terminology easy to follow. Several scenarios should be defined in a table and the text should use simpler language to refer to each type of modeling scenario.

p. 151. I had trouble understanding the material at the bottom of this page and into the next page. A clearer presentation of what was done would benefit the reader.

p. 155. The term "ambient" is used as if it means "ambient air quality data" or "ambient concentration." The use of "ambient" in this way is very informal and should be avoided. Ambient is usually an adjective and not a noun. Similarly, top of p;. 156, what is "characterization of risk using the air quality"? Is the word "data" missing?

Section 7.5 provides key observations but there do not seem to be any that draw upon Table 7-14. What are the key findings and implications of the qualitative uncertainty analysis?

Chapter 8 – Exposure analysis

General comment: it seems unnecessary to use terms such as "Staff used…" This is almost like writing in the first person, which is not necessary or preferred in a technical document.

Section 8.2 Overview of Human Exposure Modeling Using APEX

p. 162. Whether the modeled individuals are a random sample of the population of the geographic area being modeled depends in part on how well the CHAD data represent activities in that particular area. There may be geographic differences in infrastructure that might lead to differences in activity patterns, such as for commuting by private car versus mass transit.

8.11 Uncertainty Analysis

See comments on Section 7.4 above regarding the WHO guidelines and the need to explain why a different approach was used.

In general, this section is very well written and covers a wide range of issues in a clear manner. There are some quantitative analyses to support the uncertainty characterizations, which is encouraging. There should be more of this wherever possible, and not just in this chapter.

Figure 8-20. Could these differences be due to inter-city variations in the age or type of housing stock. E.g., some of the extremes are RTP, which tends to have the lowest geometric mean and standard deviation, RedBluff, which has the highest geo. Std. dev, and New York City, which has the highest mean.

Figure 8-21. Cannot read if there are any "original data" in this graph – or does this refer to just one point (presumably). Could be more clear to the reader. Are these results for one city? If they are combined results from multiple cities, were these treated as just one distribution?

p. 238. The statement that "there may be uncertainty added to the exposure results" might be taken literally by some readers – is an actual "additive" relationship the intended meaning?

p. 238, 2nd paragraph. Not clear as to what "assumptions staff made" – are these given somewhere? What is meant by "effectively" generate a distribution of SO2 removal rates? (delete term?). Here again, the notion of "add to uncertainty" appears. Perhaps "increase uncertainty" not "add" to it.

Are 5-minute peaks uniformly dispersed in an area, or are they more like a roving puff? This might affect the timing of when such a puff arrives at a particular location, and may argue for difficulty in predicting the timing of a 5-min max peak in exposure associated with a 5-min max in concentration at the closest monitor, which may be many km's away.

For the exposure assessment, would it be possible to assign the highest 5-min concentration to the time of activity outdoors for each person in a given hour?

At the end of section 8.11, there should be a comparative discussion of the sources of uncertainty in Table 8-16 and a bottom line conclusion as to which are of greatest concern and what should be done about them. The results in Table 8-16 imply that AERMOD area source emissions profiles in time and space and APEX air exchange rates are of comparable importance in terms of uncertainty. On the other hand, the APEX multiple peaks analysis is the only source of uncertainty that is indicated to have a directional bias. Perhaps these three merit some additional discussion in terms of their implications.

Chapter 9, Section 9.3, Characterizing Uncertainty and Variability.

Table 9-10. Some of the cross references to other tables and sections need to be corrected.

Table 9-10 is nice in that it has a comment section. This should be adopted in Tables 7-14 and 8-16. However, the comments can be brief, and supported by lengthier text in the main body of the chapter.

The results imply that spatial representation is the largest source of uncertainty related to lung function response health risk assessment. Is this a correct inference by the reader? If not, why not? In terms of bias, there are several that are listed as overestimate or underestimate. Are the biases of more concern than the uncertainties?

Chapters 7-9 and uncertainty

What is the general finding from the uncertainty assessment in terms of priorities for data collection or research to reduce uncertainty between now and the next revision of the standard?

Chapter 10.

Is the use of first person and references to "staff" rooted in historical precedent? It would seem possible to write the document free of such references. For example, p. 283, 2nd paragraph, "We note that" could simply be deleted.

In general, this chapter is very helpful.

The implied decision process associated with considering an alternative 1-hr average standard and whether to change or revoke the 24-hr and annual average standards could be made more clear. The document seems to convey that a starting point for the decision is to determine the need for a 1-hour average standard, and set its form, level, and indicator. For indicator, SO_2 is clearly the preferred choice. The 1-hour averaging time is a compromise. The health effects data are on a 5-10 minute average basis.

The document implies that there may be a sequential process of deciding on whether there is a need for the 24 hour and annual average standards, given that compliance with the possible alternative 1 hour standard might imply 24-hour average and annual averages that are below the current standards.

The document implies that if a 1-hour standard is to be developed, as recommended, that the choice of level should be informed by keeping in mind that health effects are associated with 5-10 minute exposures. Hence, the analysis supporting inferring a 1-hour average level that offers protection in terms of peak 5-minute average concentrations might be explained a bit more and perhaps augmented.

For Table 10-1, for each of 42 monitoring sites, the basis of the ratios given for the "5-minute max:1-hour daily maximum" could be more clearly explained. Since these data are from a 3 year time frame, does this mean that the 99th percentile of 5-minute maximums, over the entire 3 year period (one number) was selected, and divided by a single number for the 1-hour daily maximum SO2 concentration observed over the 3 year period? This analysis would be useful if these were the specific forms selected for the revised standard. However, since the form of the standard has not yet been decided, it would be more useful to consider a more general situation, such as the distribution of the variability of the ratio of daily max 5 minute concentrations to daily max 1 hour concentrations on a day-by-day basis for each monitor. Furthermore, it would be useful to assess whether this ratio has a relationship to the magnitude of the daily max 1 hour concentrations. This information would enable a choice of an appropriate ratio to use given a particular choice of the level of a standard, and would allow some flexibility to evaluate the expected number of exceedances given the choice of form.

Setting aside questions of how the ratios were developed, their interpretation, and relevance, the following is an ***illustrative*** analysis of the ratios reported in Table 10-1. The purpose here is to demonstrate a way to visualize the data in order to assess its internal consistency and to have a basis for making inferences from similar types of data sets.

The ratios of the 5 minute maximum to the 1 hour daily maximum given in Table 10-1 appear to be approximately described by a 3 parameter lognormal distribution. For example, if one substracts the minimum value of the ratio from each of the 42 estimates of the ratio, the result is given in the Figure 1. These results were generated using AuvTool, which is a stand-alone

software tool for fitting distributions to data and conducting bootstrap simulation that was developed for EPA/ORD. This tool is available at: http://www.foodrisk.org/exclusives/AuvTool/

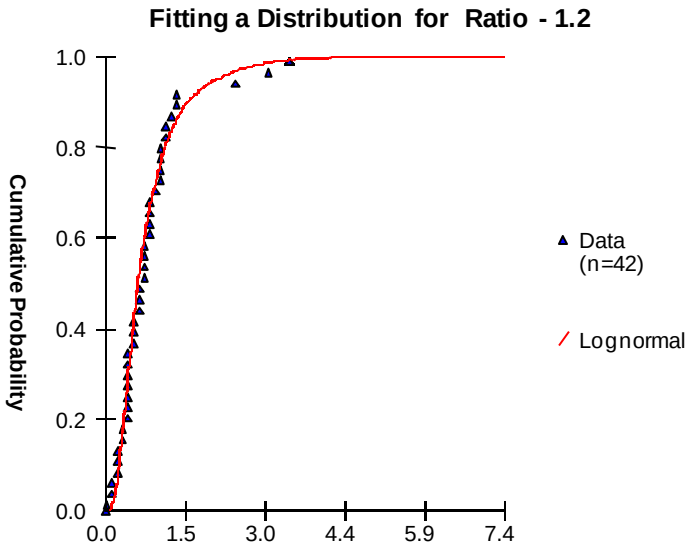

Figure 1. Empirical and fitted lognormal frequency distribution for variability in the ratio of daily maximum 5-minute versus 1-hour SO2 concentrations at 42 monitoring sites.

The implication of this figure is that the three data points that are described as being "the remaining 3 monitors" are not really inconsistent with the other data, in that they are approximately described by a continuous fitted distribution for all of the data. Hence, they are not outliers.

The lognormal distribution is perhaps not the best fit to these data, as indicated by the comparison of bootstrap confidence intervals of the fitted distribution with the data points. However, the fit appears to be better at the upper end of the distribution than at the lower end.

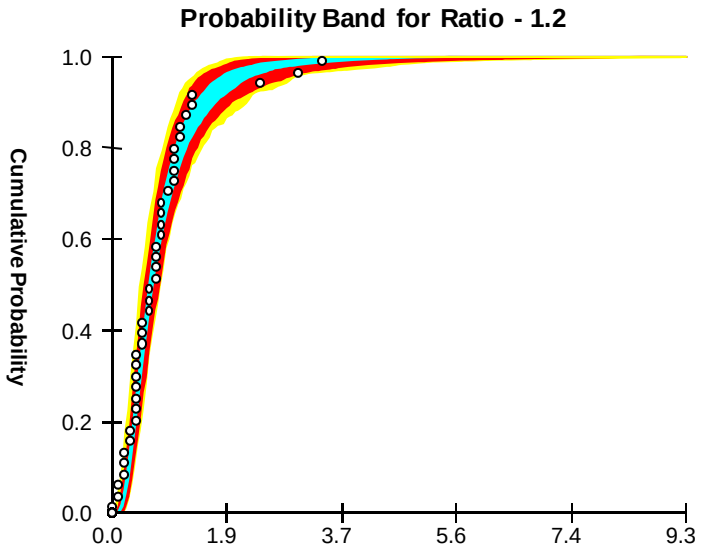

Figure 2. Results of bootstrap simulation of the fitted lognormal distribution. The confidence intervals shown are the 50 percent (light blue), 90 percent (red), and 95 percent (yellow).

In terms of policy implications, it may be more useful to consider the ratio of the 1 hour daily maximum to the 5 minute daily maximum. For example, if the goal is to achieve a 5-minute daily maximum concentration that is protective of public health, one can use this:

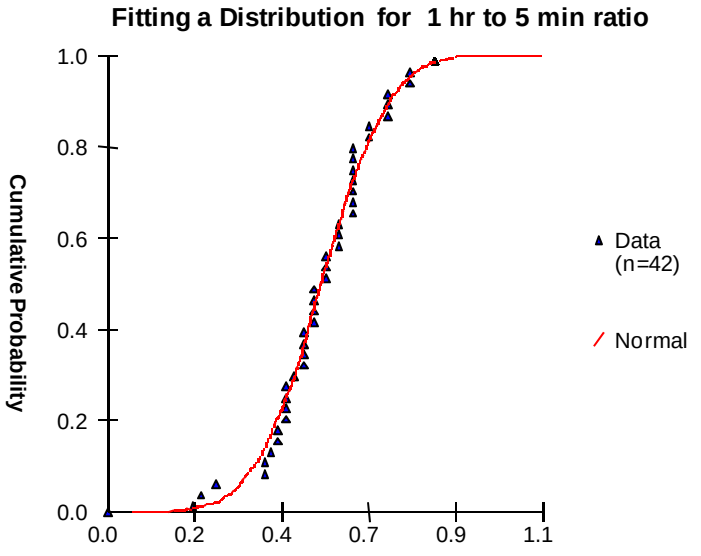

Figure 3. Empirical and fitted normal distribution for the ratio of daily max 1-hour concentration to the max 5-min concentration.

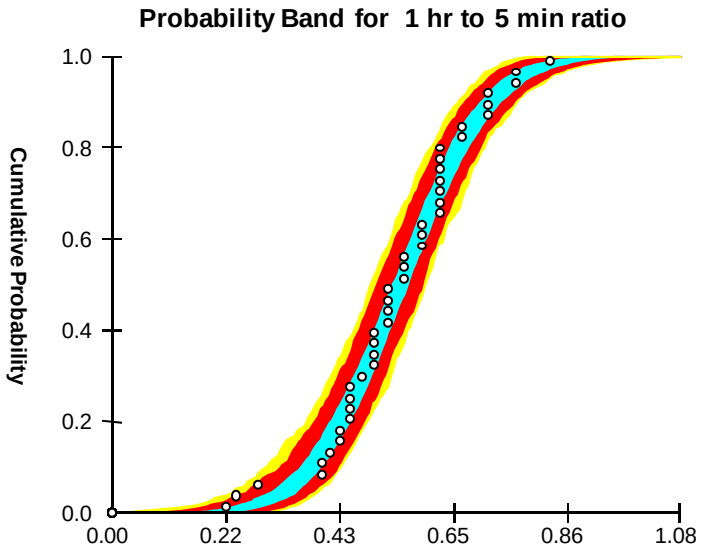

Figure 4. Comparison of Fitted Distribution with Bootstrap Confidence Intervals to Empirical Data

This analysis suggests that a normal distribution is an acceptable description of the distribution of the inverse ratios (daily max 1 hour to daily max 5 min average concentration) to those reported in Table 10-1. This inverse ratio is useful if one wants to start with a daily maximum 5-min average benchmark and infer what daily max 1 hour concentration would be equivalent. For example, if one wanted to choose a 200 ppb daily max 5-minute average, then one can infer that the corresponding daily max 1 hour concentration might vary from 40 to 180 ppb over a 99 percent frequency range for variability. If one wanted to select a protective level, in this case one would choose a low end of the distribution (e.g., the 5th percentile), which would be a inverse ratio of approximately 0.3 or a daily max 1 hour concentration of approximately 60 ppb. Such a selection would mean that the standard would be protective for 95 percent of the monitoring locations based on historical data. Of course, as is done in Chapter 7, the analysis can be stratified by factors that account for geographic or temporal variability in hourly concentrations, such as the relative variation and the average concentration. Hence, the example above is **merely illustrative of a methodology** and does not provide a specific level.

p. 311. The discussion of "concentration-based" form versus "allowing only a single exceedence" is confusing, although it was more clear when explained by Harvey Richmond at the CASAC SOx Review Panel Meeting. The explanation offered in the meeting should be included in the text. As pointed out on the next page, a 99 percentile form of daily maximums is equivalent to 4 exceedances per year. Hence, a key difference is in the number of exceedances.

Scientific Considerations in Selecting the Form, Averaging Time, and Level of the Regulatory Alternatives

The selection of the form, averaging time, and level of the standard(s) is informed by controlled experiments that demonstrate a significant frequency of adverse health effects in study participants for 5 to 10 minute exposures of as low as 200 ppb. However, because the study participants did not include the most highly sensitive asthmatics, a benchmark of concern might be lower than 200 ppb. The discussion of this point in the REA is of critical importance. Considerations based on margin of safety and setting a standard that is protective of public health could lead to strong consideration of a lower 5-minute averaging time benchmark than 200 ppb. In fact, in other health risk assessments, it is common to use uncertainty factors to make downward adjustments in benchmarks both for long-term (e.g., cancer slope factors) and acute health (e.g., thresholds in the form of reference concentration, RfC, or reference dose, RfD) effects. The use of a downward adjustment to account for uncertainty regarding sensitivity to the exposure would be reasonable and should be considered more strongly.

Chapter 7 does an excellent job of quantifying the relationship between air quality measured at a 5 minute averaging time versus a 1 hour averaging time. EPA is appropriately recommending the latter as the averaging time for consideration of alternative standards. Forms based on 3-year averages of the annual 99th or 98th percentile are analyzed.

Figure 7-18 implies that there is strong consistency between a 98th percentile and 99th percentile form, at least for a 1-hour daily level of 200 ppb. This comparison suggests that a 99th percentile form might not introduce variability that is substantially different from that of a 98th percentile form.

For the 40 counties whose air quality data were analyzed in detail, there are substantial differences in the rate at which the benchmark levels of 100 ppb or 200 ppb would be exceeded for a given choice of a 1-hour average, 99th percentile form standard when comparing levels of 50 ppb, 100 ppb, and 150 ppb. For example, suppose that the goal of a standard were to protect the public from 5 minute average concentrations of greater than 200 ppb such that this benchmark was not exceeded more than 4 times per year (or 1 % of the year). Further, consider as a regulatory option a 1-hour average, 150 ppb level, and 99% form. In this situation, based on results given in Table 7-11, it seems likely that 37 of the 40 counties analyzed could exceed the benchmark more than 1% of the days each year even if air quality was adjusted to just meet the specified standard. Hence, a "99-150 ppb" standard might arguably not be protective of public health if a benchmark of a 200 ppb 5-minute average is deemed to be the appropriate point of departure for setting a standard. Similarly, the "99-100" alternative would be associated with 14 of 40 counties having more than 1% frequency of exceedances of the 200 ppb health benchmark. The "99-50" alternative would have no counties with more than 1% frequency of exceeding the 200 ppb benchmark. The implications for a violation of a standard over a 3 year period might need additional analysis (i.e. to account for inter-annual variability in the number of exceedances). However, it is not likely that the insights would be substantially different.

If a 100 ppb, 5 minute average benchmark is assumed, then the "99-50" alternative would lead to 14 counties of 40 with more 1% frequency of exceeding the benchmark in a year.

The weight of evidence and the uncertainties associated with the state-of-science have implications for the decision making process. Weight of evidence involves a qualitative determination of causality and supports, in this case, strong conclusions that there are relationships between air quality, exposure, and adverse effects. Uncertainty implies that scientists are not entirely sure of the numerical values that precisely and accurate quantify these

relationships. However, in many cases these quantities can be bounded, and EPA is using the best available information to support its assessment. Based on quantitative analysis and reasonable and informed expert judgments, information regarding uncertainty can be used to inform explicit or implicit choices of the margin of safety with which to develop a standard that protects public health.

Comments from Dr. Terry Gordon

Characterization of Air Quality (Chapters 2, 5, 6, and 7)

1. *Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?*

The characterization of the air quality analyses was presented in a clear and balanced approach. The document is improved in style and clarity from the previous REA draft and is better in many respects, particularly clarity, than the final version of the NOx REA.

2. I*n order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?*

Although I don't have the expertise to consider the technical concerns, the adjustment approaches seem solid.

3. *In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?*

Of course, more is better and the broad comparison of U.S. cities should be considered appropriate.

4. *What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?*

The assessment of uncertainty and variability was very clear and seemed appropriate. One minor point of uncertainty did not appear to be addressed, that is: how long refractoriness to SO2-induced bronchoconstriction lasts after the initial exposure? The risk characterization seems to assume, however, that it is 24 hr and considers only a 1-hr max per day.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)

1. *The presentation of the SO2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO2 contained in the final ISA?*

Does the Panel find the presentation to be clear and appropriately balanced?

The draft REA appears to accurately reflect the final ISA for sulfur oxides. The repeated personalization of the ISA, by saying 'the ISA found…', might be avoided. More importantly, Chapter 4 is written unevenly and is less clear than other chapters. For example, certain sections (e.g., page 30) are merely paragraph-by-paragraph descriptions of study results with no clear synthesis of what they mean.

2. *The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?*

As mentioned above, I feel the appropriateness of the approach to alternate standards and the organization of the REA draft document is excellent. Obviously, the EPA staff are getting the hang of this new NAAQS process and have honed their skills. While I realize that time, money, and effort are limited, a semi-quantitative analysis of the epidemiology data may have more strongly supported the risk characterization which was based on the health effects observed in the controlled clinical trials.

<u>Characterization of Exposure (Chapters 6 and 8):</u>

1. *Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?*

Yes.

2. *The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?*

The approach is appropriate, although, of course, the inclusion of additional counties throughout the U.S. may have reduced uncertainties which might be attributed to extrapolating from 2 counties to the rest of the U.S.

3. *What are the views of the Panel regarding the approaches taken to model SO2 emission sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?*

I do not have the expertise to comment on this aspect of the REA.

4. *What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been*

identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

I do not have the expertise to comment on this aspect of the REA.

5. *What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?*

The staff's characterization was appropriate, but, as stated above, more counties would have reduced uncertainty surrounding the representativeness of the 2 counties.

Characterization of Health Risks (Chapters 7, 8, 9):

1. B*ased on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO2 exposures, we have adjusted our range of 5-minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO2 exposures evaluated in the ISA?*

The range of benchmark values is appropriate.

2. *Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?*

Yes, the results are clearly communicated.

3. *A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?*

The approach and interpretation are fine.

4. *What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?*

Actually, I got the impression that the discussions of uncertainty and variability were relatively on the mark but maybe repeated more often than necessary throughout the chapters.

Policy Assessment (Chapter 10):

1. *The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?*

The integration was excellent and Chapter 10 was clearly communicated – staff should be applauded for this Chapter.

2. *What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?*

The logic in the discussion for keeping or rescinding the current standards was excellent, although the staff's discussion of the adequacy of the current standards was somewhat unbalanced. The validity of the current standards should stand on their own merit and not be considered for retention or revocation based upon how well a proposed alternate standard may keep ambient concentrations controlled within the current standard(s).

3. T*o what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO2 standards?*

The policy chapter characterization of the alternative 1hr daily maximum standard was clear and appropriate.

4. *Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50- 150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?*

The chapter was excellent and some of the best work EPA staff has done during the new process for reviewing SOx and NOx NAAQS. The rationale is appropriate to justify this range of levels, although I would suggest limiting the range to 50 – 100 ppb.

Minor Comments:
Page 1, line 16 – Delete 'now'
Page 1, lines 20-22 – A strange sentence that appears to say the review plan was presented in the review plan.
Page 9, line 29 – extra space before 'ppb'.
Page 10, line 16 – Is an 'and' missing at the end of this line?
Page 12, line 13 – Add 'can' before 'be'
Page 12, line 5 – extra space before 'assessments'
Page 18, Table 3-1 – All of the susceptibility factors make sense except low birth rate. It implies that having a low birth rate makes one more susceptible to SO2. Low birth rate and adverse birth outcomes are a result but may not make sense as susceptibility factors such as age or gender.

Page 26, line 9 – I may have missed an earlier mention, but this first mention/definition of labeling 'moderate or greater bronchoconstriction' should be referenced or justified previously or here.
Page 27, lines 4-7 – It is strange and misleading to include the percentages in parentheses when these percentages are not for the entire 40 subjects but a subset of a subset. Only 1 of 40 had both PFT decrements and symptoms after 200 ppb, not 20%.
Page 29, line 27 – 'these' is unclear.
Pages 30-32 – these pages are just a listing of study results with no visible purpose or conclusion/synthesis. Even worse is the fact that they end the Chapter and no conclusion is provided.
Page 33, line 11 – add period.
Page 58, line 2 – Should be 'a' 2nd highest?
Page 65, lines 20-23 and footnote – 'to improve the temporal perspective' does not seem to warrant only reporting the number of times in a year that a daily 5 min concentration exceeds a benchmark rather than the total 5 min periods too. This approach ignores the possibility that an asthmatic could lose refractoriness and respond 2 or more times in a day.
Page 66, line 19 – 'in other instances is could as many' is unclear.
Page 68, line 16-17 – unclear sentence structure.
Page 69, Table 7-2 – The 'Combined Set Duplicates' is unclear and the open and shaded boxes are not defined.
Page 71, line 6 – Was a rationale given for using a 75% completeness criteria?
Page 82, lines 1-3 – This is a non-sentence.
Page 83, line 8 – Add 'minute' after 5?
Page 92, lines 7-9 - This is a non-sentence.
Page 102, lines 9-10 – This is not a strong rationale to use daily 5-min exceedences.
Page 114, line13 – 'at each to the' is unclear.
Page 135, Table 7-14 – Would ambient measurements, given EPA excellent QA program, really deserve a 'Medium' for level of uncertainty? I would say 'Low'.
Page 238, lines 22-31 – This is an excellent and important section that could have been included in an earlier chapter.
Page 248, line 5 – Should 'previous reviews' be 'ISA' instead?
Page 253, line 1 – 'who' or 'whose'?
Page 255, lines 14-16 and next page – Here is the discussion/rationale for focusing on the highest 5-minute period in a day. It could be expanded to discuss the uncertainty on the length of the refractory period and used in earlier chapters.
Page 256, line 6 – 'adjusting' or 'adjusted'?
Page 256, line 26 – Is the Table identified correctly? Seems it should be 9-3.
Page 257, line 5 - Is the Table identified correctly?
Pages 263 – 263 – Legend for Figures 9-4 and 9-5 are the same?
Page 265, line 3 – 'recent'? 7 years ago and will be 8 years before final ruling.
Page 276 – There is no definition for the X-axis labels regarding 99/100 (same for other tables).
Page 277, line 30 – Change 'are' to 'is'.
Page 313, Table legend – Is this 5-min data?

Comments from Dr. Rogene Henderson

Policy Assessment (Chapter 10)

Charge Question 3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hr daily SO2 standards?

My comments on Chapter overlap with Chapter 5.

The discussion of the continued use of SO2 as the indicator for ambient SOx was adequate to defend this choice.

The discussion of the appropriate averaging time was especially well done. The major evidence for short-term health effects of SO2 is from human clinical studies of exercising asthmatics for 5-10 min., while the supporting epidemiological studies were based on exposures for 1 to 24 hr.. The current standard is for a 24 hr average. As indicated in Table 10-1, a standard based on the 24-hour average would not be effective for addressing the effects of a 5-min peak in SO2 concentration. However, the same table indicates that a 1-hr daily maximum standard would be effective. The ds ata in Table 10-2 indicate that a 99th percentile 1-hour daily maximum standard set at a level of 50-100 ppb would limit 99th percentile 24-hr average SO2 concentrations observed in epidemiological studies where statistically significant results were observed in multi-pollutant models with PM.

The levels chosen for the alternative standards were based on evidence from human and epidemiology studies and the basis for the choices was clearly presented. The evidence that the current daily and annual standards are not protective of the health effects caused by short-term (5-10 min) exposures to elevated SO2 is clear and reasonable. The provisional recommendation is for a 1-hr daily maximum standard within the range of 50-150 ppb. This provides a margin of safety over the known human clinical evidence that exercising asthmatics show increased respiratory symptoms at 200 ppb and epidemiological studies show effects where 99ty percentile 1-hr daily maximum SO2 concentrations were as low as 200 ppb. Thus the 1-hr standard needs to be lower than 200 ppb.

The public health implications of the form of the standard were briefly discussed. There is adequate justification given to follow the recent approach used for ozone and PM, and to use a concentration-based form averaged over 3 years. There is a provisional suggestion to use the 99th percentile form to reduce the number of days allowed to exceed to standard level. I would like to hear more discussion by the Agency on the public health implication of choosing the 99th vs the 98^{th} percentile form.

General Comments:
The REA is appropriately based on the conclusion of the ISA that new information since the last review of this criteria pollutant provides sufficient evidence to infer a causal relationship between respiratory morbidity and short-term exposures to SO2. This is based on human clinical exposures for 5-10 minutes and is supported by epidemiological studies mostly using a 24-hr

average exposure. Thus a change in the current standards to reflect this new information is required. The REA provides a good review of the health effects of concern taken from the ISA and provides a reasonable approach to setting up a new short-term (1 hr) standard that will protect the public health better than the current 24-hr or annual standards.

The Agency has now expanded its exposure analysis cases to include 5 (up from 2) areas and that is a good step forward. As they point out, the Agency is still trying to work out the reasons for discrepancies between modeled predictions of SO2 exposures and monitored data. I agree that this is an important problem that must be addressed.

Chapter 5 is a key chapter and is especially clear in the explanation of the choice of form, averaging time, level and indicator.

Comments from Dr. Dale Hattis

My charge question#3:

"In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO_2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?"

Response: The two criteria the staff have chosen to use are both good options in the context of the Clean Air Act. First, selection by the lowest mean adjustment factor means selecting by relatively high pollution levels. Thus the analysis is biased to cases where SO2 is judged to be more of a problem relative to what would be observed elsewhere (in, say a representative sample of counties in the country). This choice sacrifices national representativeness for a releatively "worse" but still realistic case analysis. Sacrificing national representativeness prevents the staff (or others in the regulatory impact evaluation business) from accurately estimating national benefits from the alternative rules. The cost benefit analysts who may wish to review the results of alternative choices for the SO2 standard will not have the inputs they will wish to have, but the analysis does conform to the spirit of the act in evaluating regulations to allow protection of public health with an adequate margin of safety. If national estimates of impact are desired, the present analysis could be supplemented with a set of counties selected to be nationally representative.

Second, it is a defensible choice to select counties with at least two working monitors. This means that the analysis will be based on a more robust data set than would be the case if only a single monitor were used to characterize the whole county.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis

2. The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?

Response: The analysis in Section 5.5 makes a reasonable case for the range of standards to be considered. However it is ultimately pretty qualitative. It would be more satisfying to this reviewer if there were some attempt to do meta-analytic combination of the data to see how the effect size and confidence levels across epidemiological studies varied with 98^{th} and 99^{th} percentile levels. Ideally the results of such an analysis could be displayed in a single graph.

Characterization of Exposure (Chapters 6 and 8)

Characterization of Health Risks (Chapters 7, 8, 9)

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?
3. A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?
4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

Combined response to the three questions above: From my reading this analysis is basically sound but it can and should be improved in several ways. Most fundamentally the authors fail to provide the detailed results of their fancy Markov Chain Monte Carlo model fitting[*] (1) in ways that illuminate quantitative uncertainties and (2) in ways that can be quantitatively compared across the two models; across the two types of endpoints (increase in specific airway resistance and reduction in FEV1); and across the two levels of severity of each endpoint considered (doubling or tripling of specific airway resistance and 15% or 20% reductions in FEV1). The current quantitative presentation of results is limited to median estimates of a single endpoint (apparently doubling of specific airway resistance) derived from only one of the two model forms (the logistic) with only cursory qualitative comparison to the other model form (the probit).

I recognize that even in its current form the complexity of the presentation of the analytical results for the multiplicity of standards considered is already sufficient to try the patience of analytical reviewers, let alone executive decision-makers. Nevertheless I think that decision-makers must have major uncertainties called to their attention and at least approximately quantified where that is readily achievable. I think the discussion of the concentration-response modeling leading to the expression of results to five significant figures (see Table 9-2 on page 255) falls short in that respect.

[*] When I first saw that this advanced technique had been used to analyze the clinical dose response data, it reminded me of the under-used "cop equipment" applied to the case of littering in the classic 1967 Arlo Guthrie song, "Alice's Restaurant", resulting in the "27 eight-by-ten color glossy pictures with circles and arrows and a paragraph on the back of each one explaining what each one was." Nevertheless, as a way of integrating information from diverse studies and providing Bayesian posterior estimates of uncertainties in projected risks in the light of the correlated uncertainties in estimated parameters, Markov Chain Monte Carlo Modeling has excellent capabilities. Unfortunately those capabilities were not used in this case.

The current presentation of the choice between use of the logistic and probit model forms is couched only in terms of the goodness of fit of the two models. The closeness of the two models in the range of observed data is emphasized, and indeed I have never encountered a data set that is robust enough that it is capable of supporting a clear choice between these models on grounds of the statistical fit. I do think the decision-maker should be informed of three other facts about the choice:

- First, the two models arise fundamentally from different assumptions about the population distribution of thresholds among humans. The probit model (which I happen to prefer and have applied to a wide variety of data sets in the past—Hattis et al. 2002; 1999) is based on an assumption that the thresholds for effect for different people in the diverse human population are lognormally distributed, whereas the logistic model assumes a logistic distribution. The assumption of lognormality has a least a weak mechanistic justification: it follows from the central limit theorem that if different causes of human individual differences are many and if each tends to act multiplicatively, one expects the distribution of human thresholds to approach lognormality as the number of factors contributing to individual variability rises. By contrast, I am not aware of any mechanistic reasoning that would lead one to expect a logistic distribution of thresholds in the human population.

- As illustrated below, the lognormal distribution of thresholds derived from the application of the probit model can be readily characterized as having a geometric mean and a geometric standard deviation; and the geometric standard deviation is a measure of variability that can be compared across different chemicals and types of response. By contrast the parameters estimated using the logistic model do not have straightforward interpretations that lend themselves to comparisons across chemicals and effects.

- Second, the logistic model is known to have "fatter tails" meaning that projections of very low dose risks will generally be larger using the logistic than the probit model form. This is not a reason to prefer one model form over the other, but it is a fact that both analysts and decision-makers should know about. Moreover the difference between the models, while usually very small in the dose range of observable clinical experiments, becomes larger at low doses where, as it happens, most of the exposures and projected responses occur in this case.

To illustrate the differences I have done probit model fitting using the same data assembled by EPA in Appendix C, plus data from a source (Horstman et al. 1986) that was apparently excluded without a clear explanation or discussion in the document. I did the basic fitting using a simple Excel likelihood optimization routine that was published many years ago (Haas, 1994). The likelihood fitting allows me to either analyze different levels of effect separately or in combination using a common parameter for the geometric standard deviation of the distribution of human thresholds for responses. I will provide the analytical spreadsheets to interested investigators on request.

Table 1 summarizes the results of this fitting for the specific airway resistance endpoints in terms of the ED50's for different levels of effect, the geometric standard deviations for the distribution

of human thresholds, and 90% confidence limits for the latter (the range between the 5th and 90th percentiles of the statistical sampling error uncertainty distributions).

Table 1

Results of Probit Model Fitting Using Ordinary Likelihood Analysis—Showing Median ED50's for Different Effects and Medians + 90% Confidence Limits for Estimates of Lognormal Human Variability (Expressed as Geometric Standard Deviatons--GSDs)

Data Sets	Endpoint(s)	Median ED50 (ppm)	Median GSD	5th %tile GSD	95th %tile GSD
EPA Compilation	Double SRAW Only	0.767	2.84	2.26	4.16
EPA Compilation + Horstman (1986)	Double SRAW Only	0.835	2.50	2.14	3.13
EPA Compilation + Horstman (1986)	Double + Triple SRAW	.859 (doubling) 1.32 (tripling)	2.61	2.29	3.19
EPA Compilation	FEV1 Reduced 15% + 20%	0.600 (15% loss) 0.869 (20% loss)	2.39	2.03	3.06

It can be seen in this table that it is a close contest between the specific airway resistance (SRAW) and FEV1 reduction endpoints as to which will lead to greater projections of low dose risks. The 15% FEV1 reduction endpoint has a slightly higher ED50, but slightly smaller interindividual variability in the distribution of thresholds. Another preliminary conclusion is that addition of the Horstman data slightly raises the estimated ED50 and reduces the estimate of human variability, which will lead to reductions in estimated risks.

Table 2 compares the risk projections for St Louis that would be made using the full SRAW probit model (including the Horstman data) with those derived by EPA and presented in Table 9-2. It can be seen that there is a 20-fold difference between the two projections. This may well not be large enough to appreciably affect policy choices. However it is, I think, large enough to be communicated to the audience of decision-makers and the public, and contrasts sharply with the impression given in the current presentation in the document that there is no important difference arising from the choice between the probit and logistic models.

Table 2

Comparison of Risk Projections for a Doubling of Specific Airway Resisteance from Short Term Exposures of Asthmatics in St. Louis—Those Derived from EPA's the Logistic Model Fit to the Data Excluding the Horstman (1986) Observations vs Those Derived from a New Probit Model Fit to the EPA Compiled SRAW Data Plus the Horstman (1986) Observations

Midpoint of Exposure Bin (ppm)	Number of people exposed	Logistic Model Estimate of Fraction Responding	Logistic Model Estimate of Number Responding	Probit Model Estimate of Fraction Responding	Probit model est number responding

0.025	16519000	0.00406	67067	0.00011	1882
0.075	136621	0.02334	3189	0.00553	755
0.125	15760	0.05162	814	0.0223	351
0.175	3826	0.08563	328	0.0487	186
0.225	1051	0.123	129	0.0813	85
0.275	413	0.1622	67	0.1176	49
0.325	175	0.2021	35	0.1556	27
0.375	83	0.2419	20	0.1939	16
0.425	31	0.2806	9	0.2317	7
0.475	24	0.3183	8	0.2685	6
0.525	8	0.3543	3	0.3039	2
0.575	0	0.3885	0	0.3379	0
0.625	0	0.4209	0	0.3702	0
0.675	8	0.4515	4	0.4008	3
0.725	0	0.466	0	0.4299	0
0.775	0	0.4938	0	0.4573	0
		total	71673	total	3371

References

Haas CN 1994. Dose response analysis using spreadsheets. Risk Analysis 14:1097-1100.

Comments from Dr. Donna Kenski

General comments: This was a very impressive 2nd draft. I found it much easier to follow than the first draft, and I was pleased to see that EPA has been responsive to many of the CASAC SO2 panel's previous suggestions. The document as a whole does a great job providing a thorough basis for supporting the proposed range of alternative standards. Chapter 10 in particular was a great addition, with a welcome discussion of the adequacy of the current standard in light of the new evidence from the ISA and REA, and a thoughtful consideration of alternative standards. Bravo!

CHARGE QUESTIONS

Characterization of Air Quality (Chapters 2, 5, 6, and 7)

1. Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?

The air quality analyses portion of the REA was outstanding. I found it to be well written, well organized, and extremely thorough. It provides an excellent grounding for the subsequent analyses, and also for the policy discussion in Chapter 10.

2. In order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO_2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

The proportional roll-up has been pretty thoroughly vetted by now. I have no concerns with the approach, and the REA does a nice job demonstrating that the distributions of higher concentration years are generally linearly related to those of lower concentration years (Fig. 6-3, Section 6.5.1, and the Rizzo memo).

3. In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO_2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?

The selection methodology was explained clearly and, because it emphasized monitors where concentrations were highest, it was appropriate. A map would be a nice addition, since the geographic representativeness of the results is an issue that is raised repeatedly.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

The uncertainty analysis in Sec. 7.4 was quite thorough. I liked table 7-14 a lot.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)

1. The presentation of the SO_2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO_2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?

This was a concise, accurate, balanced summary of the much lengthier discussion in the ISA. It seemed to be just enough detail for the REA.

2. The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?

Chapter 5 was quite clear and logical in its rationale for the various alternative standards. While it makes a convincing case for a 1-hour standard, there clearly remains a need for more 5-minute data that could possibly support a future 5-minute standard. That need should be addressed by EPA in this document, at least to justify the need for such data. Probably that need/justification should be incorporated into the policy discussion in Chapter 10.

Characterization of Exposure (Chapters 6 and 8):

1. Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?

Yes, these sections were well written and easy to follow.

2. The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

My only concern with the approach is the limited geographic scope.

3. What are the views of the Panel regarding the approaches taken to model SO_2 emission sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?

The approach is appropriate. Model performance was better than I expected and adequate for this analysis.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

Like the discussion in Chapter 7, this uncertainty analysis (Section 8.11) was also clear and thorough, and I liked Table 8.16 and the accompanying discussion.

5. What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?

These weren't particularly compelling. In fact Sec. 8.10 was disappointingly qualitative. I will leave the technical assessment of the exposure assessment to others more qualified, but note that the geographic scope of the exposure and risk assessment was the major weakness of the REA. Understanding that time and resources are in short supply, perhaps it was the best that could be done, but it's hard to imagine that results from 2 counties in Missouri can adequately represent national exposures to SO2.

Characterization of Health Risks (Chapters 7, 8, 9):

1. Based on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO_2 exposures, we have adjusted our range of 5-minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO_2 exposures evaluated in the ISA?

The benchmark values are appropriate.

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?

It was clearly communicated. I can't comment on its technical merit.

3. A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?
No comments.

4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been

identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

Each of these uncertainty discussions has been very well done.

Policy Assessment (Chapter 10):

1. *The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?*

Yes, this chapter pulled together the health, risk, and air quality information in a straightforward, lucid manner. It provides a concise technical underpinning for the proposed alternative standards. Very well done.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?

The chapter nicely documents the inadequacies of the current 24-hr and annual standards and makes a convincing case for a 1-hr standard. Section 10.4.4 concluded appropriately that the annual standard should be revoked – it was good to see that plainly stated

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?

This is probably the weakest part of the chapter, but perhaps that is inevitable. As noted above, I'm uncomfortable with the geographic limitations of the exposure and risk assessment. Nevertheless, the public health benefits of the alternative standards are clear and compelling, even if they are not quantifiable nationally.

4. Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50- 150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?
I agree with the staff selection of alternative 1-hr standards in the range of 50-150 ppb, with the middle of that range probably sufficiently protective of public health with an adequate margin of safety, given the various uncertainties in the supporting analyses. I would have liked to see some additional documentation on the stability of the form (99^{th} vs. 98^{th} percentile) beyond the discussion of concentration based forms vs. expected exceedances, in order to look at the frequency with which monitors might flipflop in and out of attainment. The comment in section 10.5.3 about the influence of extreme meteorological events was a little odd, since they hadn't been mentioned anywhere else – did I miss an analysis of the impact of extreme events on SO2

concentrations? Do we need one? Other than those minor points, the draft chapter does an excellent job supporting the staff choices with good science.

Other minor comments, not in charge questions:
Appendix A was a useful compendium of supplemental information. I especially appreciated the responsiveness to CASAC's earlier requests for more information about the monitor proximity to sources and the analyses of duplicate values. Both of these efforts made the document stronger.
One important weakness with the whole approach is that the 5-min data are sparse and not geographically representative. Nothing shows that quite as well as a map – seems like an oversight not to have included one of the 5-minute data.
The key observations at the end of Chapters 7 and 9 provide a nice summary of these chapters. It wasn't clear why Chapter 8 didn't have a similar summary. The shorter chapters 1-6 probably don't need it.

Comments from Dr. Steven Kleeberger

My comments focus on the chapter (Policy Assessment, Chapter 10) and associated questions that were assigned to me.

Policy Assessment (Chapter 10)

1. The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?

I believe the integration was very clearly communicated and appropriately characterized. Staff did due diligence in consideration of the available evidence for consideration of current and potential alternative standards.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?

Staff discussion of considerations related to adequacy of the current standards was appropriate. Sufficient information was presented to understand the decision-making process for the current standard based on evidence that was available at the time. An important distinction was made between the relative paucity of clinical and epidemiological studies available to inform on the current standard relative to the number of studies now available to consider for alternative standards. Staff also appropriately, and I believe conservatively, characterized the public health implications of the current standards. For example, Staff pointed out that clinical studies of a susceptible population (asthmatics) do not include those who are likely most responsive to SO_2 (severe asthmatics) due to health/ethical concerns. Furthermore, although no mention was made in the REA, other populations may also be at increased risk of adverse health effects under the current standard, including infants and genetically predisposed individuals (although insufficient evidence is available currently to make recommendations based on these populations).

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?

As was done with the characterization of public health implications under the current standard, I believe Staff adequately characterized the potential implications on public health with the alternative 1 hour daily maximum SO_2 standards.

4. Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative daily maximum standard within a range of 50-150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?

Chapter 10 was extremely well-written and remarkably cogent in the presentation to support a potential alternative daily maximum standard. Staff appropriately identified the general approach used and the series of general questions raised to inform the range of options that may be applied to decision making. The presentation of the evidence- and air quality, exposure and risk-based considerations balanced with the key uncertainties during consideration of the adequacy of the 24-hour and annuals standard were presented evenly. Similarly, presentation of potential alternative standards was clearly presented. Appropriate representation from the ISA was described for the Indicator, Averaging Time, Form, and Level. My only minor criticism is that the "Conclusions regarding level" (section 10.5.4.3) was internally somewhat inconsistent. The beginning statement in the section "provisionally concludes that the evidence and exposure and risk information reasonably support a 1-hour daily maximum standard within a range of 50-150 ppb" and concludes "if the alternative standard selected is not expected to prevent ambient SO2 concentrations from exceeding the levels of the current standards, it would be appropriate to consider retaining the current NAAQS." I understand that due consideration needs to be made for recommendation/suggestion, but it seems the evidence presented throughout the "Potential Alternative Standards" (section 10.5) and the language used clearly fall in support of a 1-hour daily maximum standard and the conclusion should reflect this.

Comments from Dr. Patrick Kinney

Charge question 2: The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards?
Does the Panel have technical concerns with this approach?

Overall the second draft REA appropriately characterizes the public health implications of the current and alternative standards. The technical approach is sound given the time and logistical constraints faced by staff. Specific comments on chapter 8 are provided below.

The REA provides a helpful discussion and rationale for choosing the specific study areas for detailed exposure modeling, Greene County and the city of St. Louis, MO. The reasons for choosing these locations, availability of baseline air monitoring data and a large number of SO2 sources in the region, make sense. Another valuable aspect is the information provided comparing year 2002 meteorology to the 30-year climate normal period 1978-2007.

p. 170, line 13: what is meant by "all stations were considered at an airport"? If only airport surface characteristics were considered, this would likely underestimate surface roughness across the domain. Please clarify.
p. 172, line 21: which is rural and which urban?
p. 193, line 3: delete "of the exposure"
p. 199, lines 22-24: exact repeat of text a few sentences earlier.
p. 203, line 19: fix grammar.
Figure 8-13 etc. Re-label y axis without exponential notation – for the general reader.

Section 8.10 Representativeness of Exposure Results
This section is a helpful addition.

Section 8.11.2.9, starting page 238. "Occurrence of Multiple Exceedances Within an Hour.
This section strengthens the REA significantly. While one might ideally have liked to see an exposure assessment that takes account of all 5 minute averages, staff make a good case for the approach taken given the computational constraints. Further, their new analyses give a sense of the possible biases that may have resulted from analyzing only the single maximum 5-min avg in an hour, vs. analyzing all 5-min avgs. This is as much as one can expect to assess for this issue.

Comments from Dr. Timothy Larson

Characterization of Air Quality

1. Does the panel find the results of the air quality analyses technically sound, clearly communicated and appropriately characterized?

The staff is to be commended for including the 5-minute data in this analysis, given that there is strong evidence for effects from these short-term exposures above certain thresholds. The 5-minute data are limited in geographical scope, but the analysis of the relationships between the 5-minute, 1-hour and 24-hour levels is reasonable. The use of the 1-hour data as an integrative link between the shorter and longer term levels provides additional support for these analyses, given that the 1-hour data is more ubiquitous. Figure 6-1 is a useful addition that clarifies the overall approach.

2. In order to simulate just meeting potential 1-hour daily maximum standards, we have adjusted SO_2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. To what extent does this approach characterize the public health implications of the current standards? Does the panel have technical concerns with this approach?

The use of "as is" air quality data to establish the "just meeting" values has certain limitations that are discussed in the document. There are relatively few urban areas with multiple monitors and so it is difficult to assess intraurban spatial patterns based upon measurements. Adding to the problem is the potential for increased space-time interactions with 1-hour averages relative to the 24-hour averages used in the previous draft. In the final analysis, the use of a pure temporal adjustment based on one site applied equally to all sites in a given area is necessary, given the lack of spatial information needed in order to include a space/time interaction.

The multiple approaches used in this assessment make the particular assumptions from any one of them less critical than if only one approach had been used. The results summarized in Figures 7-5 through 7-9 provide support for the use of COV and GSD metrics as pdf categorization variables. The cross-validated results summarized in Table 7-4 support the use of the COV metric. The approach is clearly communicated.

3. In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO_2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?

There are relatively few urban areas with multiple monitors and so it is difficult to assess intraurban spatial patterns based upon measurements. Therefore the reliance on plume models to infer the smaller scale variations is the only reasonable approach that is available. Those areas with multiple monitors have been identified and given appropriate priority for inclusion in the

larger modeling exercise. Combining the multiple site criterion with the minimum mean adjustment factor also seems like one reasonable selection approach. An alternative philosophy might be to choose these sites based on the COV values of the 1-hour concentrations. This alternative approach might generate a slightly different set of results. It would be interesting to know how many of these counties are classified "c" with respect to their coefficient of variation (potential for relatively high peak to mean ratios), and alternately, how many were not included. This information is in the Appendix and could easily be extracted in a few sentences.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

Table 7-14 provides a good summary of the key sources of bias and uncertainty. The discussion of these error sources is very thorough.

The statement in Table 7-14 that the effect of spatial scale on the air quality adjustment is to overestimate the values is not supported by the text.

Characterization of Exposure

1. Does the panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?

Yes.

2. The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

The benchmark approach is useful in summarizing what would otherwise be a very complex and involved set of results. It is not relied upon in the detailed exposure assessment in St. Louis and Greene County, but provides a link to the monitoring data analyses.

EPA states that they are attempting to include several other locations in populated areas. If its possible to do so, this would be a useful addition.

3. What are the views of the Panel regarding the approaches taken to model SO2 emission sources? Does the Panel have comments on the comparison of the model predictions to the ambient data?

The choice of Aermod is reasonable. Given that the agreement between the predicted and measured SO2 levels in St. Louis and Greene County depends upon the approach used to adjust the diurnal variation in the area source emissions (page 226 of the draft document), some

discussion of the resulting diurnal profiles vs. profiles deduced from other information would be useful. In any case, it should be called a diurnal adjustment of the model, not a diurnal emissions profile because the lack of agreement between unadjusted model and measurement could be due to factors other than emissions.

Is the effect due to the emissions patterns alone, or are there meteorological influences? For example, does the dispersion model include an initial residual layer of SO2 aloft from the previous day that is brought down by growth of the daytime mixing layer? Are the non-point source plumes of sufficient height that they could be isolated from ground level except in daytime, convective boundary layers. Is there a possibility that the nighttime mixing of the non-point emissions is enhanced by the urban landscape (roughness effects) in a way that causes the current model to overestimate the nighttime downwind impacts, thereby requiring the diurnal adjustment that is used.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

Table 8-16 states that the uncertainty due to the Aermod algorithms is low and the direction of bias is unknown. However, the Aermod-based predictions did not include building downwash effects. This uncertainty and its associated bias is not discussed. The uncertainties in the algorithms applied to complex terrain (as in Greene County) are also not discussed. Finally, it is stated that the uncertainty in the SO2 emission rates for the major point sources is low. This conclusion should be included as a separate row in Table 8-16.

5. What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?

Regarding the air quality estimates, they are reasonable and the limitations are well described. One useful comparison to make is between the characteristics of the 40 county monitors (as provided in the pie chart summaries) and the monitoring sites in St. Louis and Greene counties. Are the monitors in these two locations classified the same as those in all 40 counties?

Comments from Dr. Kent Pinkerton

Overall Comments: The second draft of the Risk and Exposure Assessment to Support the Review of the SO_2 Primary National Ambient Air Quality Standards is well organized to provide a reasonable overview of human exposure to SO_2 as well as those individuals at greater risk (i.e., asthmatics) to help the reader arrive at logical conclusions for the need to consider 1-hour, 24-hour and annual SO_2 levels as well as the relevance of both temporal and spatial effects of 5-minute peak SO_2 levels. This second draft of the REA is well written to address issues regarding short-term peak exposure (based on 5 to 10 minutes or over a 24 hour period) to SO_2 and corresponding respiratory health effects to provide clear evidence and exposure/risk-based considerations for the primary SO_2 National Ambient Air Quality Standard. The staff at EPA should be commended for their efforts to produce a high quality and unbiased document that provides compelling evidence in their risk and exposure assessment to consider revising the current SO_2 NAAQS.

Characterization of Air Quality (Chapters 2, 5, 6, and 7)

1. *Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?*

Reply: The presentation of the air quality analyses is well presented and appropriately characterized. It is good to know the air quality data and database used is unlikely to contribute to uncertainty in the exposure analysis.

2. I*n order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?*

Reply: This approach appears to be highly reasonable. To simulate just meeting potential alternative 1-hour daily maximum standards seems to be the most effective approach for providing the greatest measure for health protection. I have no technical concerns with the approach used.

3. *In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?*

Reply: Expanding from 20 to 40 counties seems to be advantageous to demonstrate similar findings as well as some indication some limited discrepancies. The selection of counties and the SO_2 monitoring data for the indicated years seems to be quite reasonable and appropriate. Perhaps the greatest advantage for the approach of using an increased number of counties is to evaluate the relationship between short-term peak concentations and the level of the current

annual SO_2 NAAQS in these selected counties.

4. *What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?*

Reply: Although I am not an expert in assessing uncertainty and variability, the approach taken by the EPA staff seems highly logical. To evaluate uncertainty by the EPA staff was adapted from the World Health Organization (WH0) guidelines using low, medium and high levels of uncertainty seem to be highly responsible in judging how uncertainty would influence concentration estimates. Table 7-14 is helpful to better understand the multiple sources and types of uncertainty and bias direction that could be introduced.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)

1. *The presentation of the SO2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?*

Reply: The draft REA provides a brief reminder of those factors falling under susceptibility and vulnerability that could precipitate and/or enhance adverse health effects due to exposure. However, a clear definition for the terms susceptibility and vulnerability should be stated in the document. Examples of SO_2-related health effects based on epidemiological and human clinical studies are a nice summary of the more detailed information provided in the final ISA. The presentation of this data is brief, but good.

2. *The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?*

Reply: The selection of potential alternative standards is based primarily on susceptible asthmatic individuals and is completely justified.

Characterization of Exposure (Chapters 6 and 8):

1. *Does the Panel view the results of the exposure analyses to be technically sound, clearly*

communicated, and appropriately characterized?

Reply: The EPA staff has done an excellent job to thoroughly explain the methods used in exposure analysis in Chapter 8. The human exposure modeling used and the characterization of ambient hourly air quality data are both appropriately characterized and reasonably well communicated.

2. *The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?*

Reply: Exposure evaluation in St Louis and Greene Co, MO is extremely well characterized in Chapter 8. This evaluation is based on both temporal and spatial variation in SO_2 levels, while also simulating human contact to these various SO_2 levels. The selection of receptor locations to represent the location of the residential population, coupled with the locations of the available ambient SO_2 monitors seems highly reasonable. Using these and other approaches described should greatly help to better characterize the public health implications of the current SO_2 standards.

3. *What are the views of the Panel regarding the approaches taken to model SO2 emission sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?*

Reply: The approaches to model SO_2 emission sources are well described and logically presented.

4. *What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?*

Reply: An analysis of the uncertainty of an applied model includes model algorithms and model inputs. Each is discussed in Chapter 8 with some limited summary remarks, however, an overall summary for the section under uncertainty analysis (8.11) would be useful.

5. *What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?*

Reply: The benefit of St. Louis and Greene County, MO to perform exposure and risk estimates lies in the power of the monitoring data available as well as the characterization of the population living in these two areas. It is also reassuring that additional preliminary analysis from other regions appears to confirm the findings from the St. Louis and Greene County, MO analyses.

<u>Characterization of Health Risks (Chapters 7, 8, 9):</u>

1. B*ased on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO2 exposures, we have adjusted our range of 5- minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO2 exposures evaluated in the ISA?*

Reply: The authors have been highly conscientious to use the conclusions in the ISA to appropriately adjust the range of 5 minute potential health effect benchmark values. This range of potential health effect benchmark valves from 100 to 400 ppb have been carefully characterized in this REA by using clearly appropriate parameters and clearly detailed applications of models. This range is particularly important based on recently published studies to demonstrate decreases in lung function among asthmatics exposed to SO_2 to 200 to 300 ppb for 5-10 minutes.

2. *Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?*

Reply: The results of the risk characterization are in general clearly communicated and appropriately characterized. The lung function quantitative risk assessment is also clearly communicated and appropriately characterized.

3. *A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO.*
What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?

Reply: The approach taken to perform a quantitative risk assessment for these two regions (St Louis and Greene Co., MO) with respect to lung function in exercising asthmatics appears to be highly appropriate to provide a reasonable interpretation of the results for this analysis.

4. *What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?*

Reply: The discussion of uncertainty in the document seems to be quite reasonable. To identify key sources of the assessment contributing to uncertainty as well as a qualitative characterization for the types and components of uncertainty (Table 7-14) seems good, but at times difficult to completely follow. Regardless, reasonable conclusions seem to have been made following appropriate consideration of each point contributing to uncertainty and variability. Table 9-10 in Chapter 9 is excellent in presenting key uncertainties in lung function response health risk assessment in terms of direction of bias, level of uncertainty and accompanying clarifying comments.

<u>Policy Assessment (Chapter 10):</u>

1. *The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?*

Reply: This chapter does an outstanding job to consider the scientific evidence found in the ISA to allow for logical and highly relevant risk and exposure assessment parameters to be fairly characterized and communicated. The integration of ISA findings within this draft of the REA document greatly facilitates reaching relevant conclusions and recommendations to be made.

2. *What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?*

Reply: This chapter has done an excellent job to adequately characterize the public health implications of the current standards.

3. T*o what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO2 standards?*

Reply: Extremely well.

4. *Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50-150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?*

Reply: I think the chapter provides very compelling data to support a potential alternative 1-hour daily maximum standard within a range of 50-150 ppb for SO_2. This chapter, in combination with the data found in earlier chapters, provides more than sufficient rationale to justify this range of SO_2 level for recommendation to establish a new SO_2 standard to more effectively protect public health. The EPA staff provides reasonable justification of how a 1-hour standard should adequately protect (or reflect) possible 5-minute peak SO_2 levels.

Minor Comment:
1) List of Acronyms/Abbreviations: PMR (peak to mean ratio) is not found in this list.

Comments from Dr. Armistead Russell

Overall, I am pleased with the condition of the second draft of the SO2 REA. It has improved since the last version, and with some modification can serve as a role for supporting EPA's, and CASAC's, evaluation of the need for modification of the SOx NAAQS. An executive summary chapter of about 30 pages would be helpful.

I like the analysis of PMR binned by COV, GSD and concentration. This is very valuable for adjusting 1-hr average concentrations. The description of PMR_{ij} in Eq. 7-1 should explicitly state that it was sampled from the appropriate concentration/variability (COV/GSD bin) distribution. I would concur that the alternative model (excluding the max and min) should be used. The future use of the results should note the biases in the prediction of above benchmark levels found in Table 7-5, though it is notes that the agreement is relatively good, so little additional quantitative analysis is required. The limitation of having at least 30 samples in a bin might bias your results. Has this been explored?

The AERMOD/APEX application is a good choice for conducting the exposure analysis. The presentation of AERMOD results could be improved. For example, the CDFs in Figs 8-6 to 8-12 can be difficult to differentiate at the extremes. The evaluation of the AERMOD results should be more thorough in regards to presenting likely biases, particularly at the high end since that is where the exceedences of certain benchmark levels in the following APEX modeling of exposure will be most sensitive. The statement "some difficulty in reproducing maximum concentrations" is too vague, and should be replaced with specifics. In particular, a bias at predicting the highest levels should be noted and further analyzed.

Details:

Fig. 7-7: Should be GSD in what should be panel (d)

Characterization of Air Quality (Chapters 2, 5, 6, and 7)
1. Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?

For the most part, it is clear enough and sufficiently sound. I found Section 6.5.2, and the use of the figures, still a bit less clear than it could (and should) be, given the role this step plays in the resulting analyses. Overall, I thought the extent of the characterization of appropriate length.

2. In order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

There is no perfect way to simulate just meeting the standards as any approach to ramp up the concentration distribution entails certain assumptions. The current approach is reasonable. The approach probably does not really capture the public health implications of the current standards exactly as current SO2 controls are being driven by factors beyond just the meeting the SO2 NAAQS standard. Thus, a proportional role up will not be a reversal of controls applied to meet the current standards. However, a better approach is not apparent, and there are many benefits to a simple, transparent approach.

3. In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?

Doubling the number of counties being used in the analyses is, of course, a good step. The years chosen are relevant, the period long enough and the locations reasonable. The rationale is sufficient.

.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

The sources of uncertainty have been identified, and I appreciate the mention of likely bias. Again, I would prefer a more quantitative analysis (we always want more), with particular emphasis on identifying the major contributors. Though EPA has done an admirable job of qualitatively assessing individual uncertainties in this REA, they have not propagated the uncertainties all the way through.

Characterization of Exposure (Chapters 6 and 8):

1. Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?

The analysis discussed in Chapter 6 is sound and generally well communicated (see comments above), though I would not call it an exposure analysis given that it is aimed at assessing the potential ambient SO_2 levels.

Chapter 8's discussion of APEX and its application is thorough, though while it contains an evaluation of the dispersion model (AERMOD) results, it does not evaluate APEX results due to the lack of appropriate data. The lack of data to evaluate APEX results is not limited to SO_2, and suggests a general need to be addressed by EPA. AERMOD is a good choice for providing concentration fields.

The current version of the REA needs to be more specific about how well AERMOD simulates the more extreme (in this case at the high end) concentrations. Biases in simulating the 95th %ile and above concentrations can play a key role in the ensuing exposure analysis. Further, to the extent practical, the ability of the modeling system to simulate 5-minute peaks should be conducted. Please use probability scales when showing cdfs. It would be good to conduct a sensitivity analysis of APEX model results to adjustments in AERMOD concentration fields. In other REAs, the distribution of source contributions to exposure has been given, and similar information here would be useful. Further, the distribution of how exceedences link to AERMOD concentration levels would be insightful. For example, if most of the exceedences of a specific benchmark level come from very high simulated AERMOD concentrations, this has implications as to the potential sensitivity of the results to peak simulated levels.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

The assessment is extensive and many of the uncertainties identified. Staff have also done a good job in suggesting potential biases in the results due to the uncertainties discussed. The lack of evaluation of the APEX results suggests a potentially large uncertainty as to how those results represent actual exposures, particularly since the exposures of concern are at the high extremes. A key concern here is the need to assess the potential bias in the results given the biases in the AERMOD simulations of the maximum concentrations. A sensitivity analysis of how the APEX results would be impacted by increases/decreases in the simulated concentration fields would be insightful.

Comments from Dr. Richard Schlesinger

I am limiting my major comments to health risks and associated conclusions (Chapters 7, 8,9, 10). The exposure analysis seems appropriate, although it is important to clearly state the justification for the two cities used as models in terms of relevance to other urban areas.

The document is well organized and follows to a logical conclusion regarding the justification that the annual standard does not provide adequate protection and that a new short term standard is needed. The evidence for health outcomes related to exposure is well characterized and clearly reflects the conclusions in the ISA.

The range of potential benchmark values (100-400 ppb) for 5 minute exposures seems adequate based upon the information summarized from the ISA. However, it may be useful to reduce the lower limit of the range to 50 ppb since adverse health outcomes will likely occur at levels below 100 ppb in sensitive groups.

One of the uncertainties in the analysis involves interaction between SOx and other pollutants. The risk assessment was based largely upon controlled exposure studies of humans who were exposed only to SO2. Any assumption that adverse health outcomes noted would be the same regardless of any co-exposure to other air pollutants cannot be made with any degree of certainty, but that seems to be the case in this document.

Finally, some clarity needs to be shed on the statements made on page 325 in the last paragraph. Throughout the document, it is clearly noted that the annual standard is not protective. However, it is stated here that, "….if the alternative standard selected is not expected to prevent ambient SO2 concentrations from exceeding the levels of the current standards, it would be appropriate to consider retaining the current NAAQS." This seems to contradict the strong comments elsewhere that the current NAAQS is in fact not protective and a new shorter term standard is needed. I think this is a matter of rewording.

.

Comments from Dr. Christian Seigneur

My comments pertain to the "*Characterization of the Air Quality and Exposure*". Overall, I find the air quality analysis to be technically sound. My main concern is the emphasis on industrial point sources and the small contribution of ship-related emissions in the area used for the exposure analysis (i.e., St Louis).

Charge question 1: Are the results of the air quality analyses technically sound?

Industrial point sources have historically been a major source of SO_2 and accordingly have been subjected to emission control regulations. Recently, SO_2 emissions from ships have become of concern and, in some areas, may be the major cause of significant SO_2 exposure. This issue is being addressed through the set up of Sulfur Emission Control Areas (SECAs), within which the sulfur content of the fuel will be constrained.

The 2nd draft REA correctly singles out ship-related emissions in the exposure analysis (e.g., port emissions in Table 8-5 on p. 178 and supporting text). However, the port emissions in St Louis are a small fraction of total SO_2 emissions in the area (about 3%). Such emissions may constitute a larger fraction of total SO_2 emissions in other areas (e.g., large sea ports such as Long Beach or Oakland in California). It would be useful if a discussion of this source of variability were included in the REA, perhaps in the uncertainty/variability section.

I found the model performance evaluation to be satisfactory, i.e., within the range of uncertainty expected from current atmospheric dispersion models. Among all the monitors where the model simulation results are compared to the available measurements, model performance appears to be poor only at monitor ID 290770040. The model reproduces the temporal evolution and magnitude of the measured SO_2 concentrations fairly well at the other eleven monitors. This satisfactory performance is not unexpected as point source emissions dominate the SO_2 emission inventory and the dispersion model used here, AERMOD, was designed for simulating atmospheric dispersion from point sources.

Charge question 4: Is the assessment of the uncertainty and variability adequate? To what extent has variability adequately been taken into account?

My main criticism of the uncertainty analysis (Section 8.11) is that it pertains mostly to an uncertainty analysis of the St Louis case study and fails to address variability among various urban areas. There is some discussion of the interurban variability of air exchange rates for example, but there is no discussion of the variability of emission sources among urban areas in the United States. Some discussion (at the minimum, a qualitative discussion) of the variability of SO_2 exposure among various areas (see comments on ship-related emissions above) is warranted.

Comments from Dr. "Lianne" Elizabeth Sheppard

Overall this document represents substantial improvements since the last version. I commend EPA staff for their hard work, excellent progress, and thoughtful analysis.

Characterization of Air Quality (Chapters 2, 5, 6, and 7)
Q1.

- My major concern is that there is an assumption that the universe of monitoring data represents a reasonable sample for analysis because the available monitors are representative of the exposure to a defined underlying population. However, the monitoring data are far from a probability-based sample of the spatial locations in the United States and there is no complete characterization of what locations they represent. (This characterization would not only be over geographic space (e.g. as a function of state or latitude and longitude), but also important "design" features such as urban/rural, proximity to sources, typical terrain, etc.) Furthermore, there is an assumption that site-years are exchangeable. There should be a description of the monitoring network design in Chapter 2 and further analysis of the network features in Chapter 7. (There is evaluation of the 5-minute vs. all monitors that suggests that while the two have some similarities, there are some key differences that may be important, particularly with respect to source orientation and neighborhood scale. Yet the text suggests more similarity in these networks than I would conclude (e.g. p 140 line 25)). Network features should be characterized not only from variables stored in AQS, but also based on relevant geographically defined characteristics as calculated from a GIS. Relevant characteristics should be defined based on understanding of SO_2 sources and the key features of their temporal and spatial variation, particularly w.r.t. features that will lead to high concentrations. Some work has already been done to characterize the monitoring network (and this represents a strong and needed enhancement of the report), so the further additions will provide better insight into how well the existing dataset represents population exposure. See also comments under 7.2.2. below.
- My second concern is that only one 5-minute exceedance per day was counted. While from some policy perspectives this choice is reasonable, this approach is not mentioned in the early part of the document. (Much to my chagrin, I didn't figure it out until the middle of Chapter 7 and after I had spent considerable time critiquing the approach and coming up with a different analysis approach to address the multiple exceedances!) Make sure the exceedance definition is stated and justified early in the document. For instance, Figure 6-1 outputs refer to "number of times per year" without defining the unit as days; the probability (output 2) does refer to days, but during my initial reading this was confusing because it wasn't consistent with my assumption that each 5-minute period would be counted separately. Also consider using the wording "number of days with at least one 5-minute concentration above potential health effect benchmark levels" instead of "numbers of daily maximum 5-minute concentration exceedances" or similar language.
- Note that many analyses are done within separate strata (bins); in some cases a smooth function might actually perform better. For instance, Table 7-8 makes it very clear the perils of calculating probabilities from very small sample sizes. The analysis solution was to exclude bins with less than 30 observations; a better approach would be to fit a regression model (such as a logistic model) to borrow strength across neighboring bins. I have no

problem with a nonparametric approach per se; sometimes it is better to not impose distributional assumptions on the structure of relationships.

- There should be at least a brief discussion of the reactivity of SO_2 and its implications for indoor exposures in Chapter 2.

Q2. The adjustment seems appropriate given logistical constraints. Capturing additional sources of variation in the adjustment would be desirable, but much more complex.

Q3. The expanded set of locations and the shorter time period both appear reasonable. However, the entire exercise relies on an underlying assumption that the existing monitoring network represents SO_2 concentrations in the U.S. that are relevant to the population.

Q4. Generally there is good progress here. I appreciated the use of WHO guidelines. A few categories/sources are omitted: representativeness of the monitoring network (both the full network and the two subsets with 5-minute data), and that site-years are exchangeable. I think there is more uncertainty in the spatial representation than is implied by the discussion in 7.4.4. I think the opening sentence in that section is flawed in its assertion that because monitors are used to determine whether areas are in compliance with the NAAQS this implies that they are representative. In section 7.4.6 the spatial representation of data used in the statistical model may be more important than the temporal representation. It is also important to learn from monitor features for determining adequacy of the analysis – for instance analysis on p. 148 demonstrates that local terrain is an important contributor to monitor exceedances. This has important implications for human exposure.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)

Q1. Yes, in general. The text in chapter 4 still appears to confuse statistical significance with scientific importance. Improve the definitions of vulnerability and susceptibility. Incorporate dose into the concepts since certain behaviors actually increase dose rather than exposure (e.g. increased exercise).

Q2. This was clear and convincing. There was an argument in chapter 10 that the epidemiological effect estimates are stronger for locations with higher long-term averages. Ideally the chapter 5 figures should be reworked to bring this point out, perhaps by ordering results as a function of the 99^{th} percentile value in each location.

Characterization of Exposure (Chapters 6 and 8):

Q1. The beginning of chapter 6 needs to address the number of exposure events counted per day and why. The exposure characterization effort can be divided into two parts: ambient concentration prediction and exposure prediction for individuals.

- Ambient concentration prediction: It is important that there was an effort to characterize population exposures to SO_2, particularly when these sources are not well captured by the existing monitoring network. The difficulty reproducing maximum concentrations and the evidence that some monitors fall outside the prediction envelopes suggest additional improvements to the modeling would be beneficial. However, the work presented represents a considerable effort to improve the AERMOD model, a great deal of progress, and I

generally agree with the decision to use the unadjusted AERMOD predictions in the exposure assessment.

- Exposure prediction for the simulated population: This is based on application of the APEX model modified to capture 5-minute periods. APEX is well-tested and has been used successfully in previous exposure analyses for other pollutants, leading to my confidence in the results. (However much of the testing was not specific to SO2, perhaps an important source of uncertainty.) The uncertainty analysis indicates the estimated number of persons exposed may be underestimated by as much as 35% due to the single peak per hour approach to analysis. This is a large percentage and should be addressed in the interpretation of results (and ideally the approach revised to address this limitation in future work).

Q2. It is valuable to be able to contrast exposure exceedances from a more rural with a large urban area. Additional characterization of features that lead to such different exposure results in the two counties and putting these in the context of the U.S. population would be helpful. We should not lose sight that the areas screened as likely to have elevated SO_2 data also had to have sufficient monitoring data. This is a feature that was necessary logistically but should not lull us into believing that the selected areas are necessarily representative of high SO_2 areas in the U.S.

Q3. See responses to Q1 above. I was surprised by the high percentage of non-road emissions represented by the port.

Q4. I appreciate the new summary based on WHO guidelines. Add in "selected receptor locations as a basis of predictions" as a source of uncertainty. This source of uncertainty should be less of an issue for SO_2 (vs. e.g. NO_2) because the census block centroids are not systematically farther from sources than a percentage of the exposed population, but it should be documented and discussed nonetheless. The multiple exceedances uncertainty analysis highlights the importance of local sources since one monitor contributed heavily to the percentage of hours with multiple exceedances.

Q5. The characterization focused on time spent outdoors and distribution of asthma prevalence. These were reasonably characterized although the higher prevalence of asthma in the northeast suggests future analyses should focus on that region. Also areas with lower air conditioning prevalence may find a higher proportion of exceedances indoors. The discussion of representativeness of these two areas should also consider other spatial locations in the U.S., regardless of presence of SO_2 monitoring data. For instance, locations near major ports may have very high exposures.

Characterization of Health Risks (Chapters 7, 8, 9):
Q1. This is appropriate.

Q2. Minor point: Clarify the differences between 1) and 2) on p. 252; to me they appear identical. Can the columns in Table 9-1 be matched up with 1) – 3)?

Q3. The approach is appropriate given constraints and underlying assumptions. The recognition and discussion of the different impact on results on annual number of occurrences vs. percent of individuals affected is valuable. The sources of differences in the two counties are noted.

Q4. Does this uncertainty characterization fully reflect the impact of the multiple exposures per day or hour? (i.e. is it sufficiently brought forward from the previous chapter?) A great deal of effort has been made to evaluate assumptions with data and/or simulations. Additional assessment of assumptions and sensitivity of the results to these should be encouraged. The additional comments in Table 9-10 is helpful but may need to be expanded in the text.

Policy Assessment (Chapter 10): Chapter 10 is well-done and thoughtful. Please remove references to statistical significance in this chapter. The assessment of the evidence by now should have proceeded well beyond statistical significance. In the context of discussion of specific effects, as an alternative to mention of statistical significance, it is appropriate to discuss the inference from confidence intervals and whether (or not) these intervals rule out no increased risk.

Specific comments:

- Figure 1-1 is a good addition. I particularly appreciated the distinctions between risk and evidence-based considerations.
- P 16 13-16: This statement is correct only if the epi studies being referred to are time series studies. I suggest rewording to refer to that design specifically.
- P 16 23-24: The increased error is relative error not additive error, correct? For epi studies it is the magnitude of the additive not relative error at low concentrations that will dominate.
- P 17 8-10: Which study design?
- P 17 23: Accounting for exposure errors typically leads to *more* uncertainty in the effect estimates.
- P 20 1: Typically genetics are not associated with exposures and thus are not confounders.
- P 30 8: Does "negative" mean not statistically significant or an estimate with a negative sign?
- Chapter 6: Include key definitions in an easily referenced section in this chapter (or in a glossary elsewhere in the document). Examples of terms that need to be defined are "exposed asthmatics at elevated ventilation rates" to clarify the denominator as well as the numerator when percents are reported (e.g. p. 296).
- P 47 9-11: Also experiments give better evidence for causal effects than observational studies.
- P 48 16-18: That the full set of monitors represents a "broad characterization" is true only if the monitors are appropriately sited.
- P 48 21-22: and the siting of monitors
- P 53 28: Add "with monitoring data" after "States"
- P 55 7: Move "(2007)" to after "year" on line 5?
- Figure 6-4: Consider rescaling so readable
- P 63 7-9: I could think of as many reasons that the opposite statement would be true. Try to reword. The next sentence makes sense. The final sentence (12-13) is unclear.
- P 63 24-25: Here is an example where it is important that the reader knows that only one exceedance per day is being counted.

- P 65 14-15: Is population density enough stratification given that some monitors are sited to be population-oriented and some source-oriented? Consider additional stratification or at least an analysis to assess the effect of orientation on this approach to categorization.
- P 65 18-20: Perhaps I missed it, but the comparison of estimated vs. actual number of exceedances per year for monitors with both would be helpful.
- P 66 footnotes: Here is where we learn what is counted. It is easy to miss as a footnote and also belongs in chapter 6.
- Table 7-1: Add a column with the number of hours represented by these measurements for additional comparison.
- P 70 20-22: Say what is done with the screened measurements.
- Section 7.2.2:
    - Not only do we want to know how comparable the monitor subsets are, but we also want to know important features of the network and how these relate to population exposure. Analyses stratified by population density start to get at the latter question, but it is unclear how much this variable is just serving as a proxy for local sources (since relatively more of the rural monitors may be sited near local sources). Acknowledge that available information will limit how well we can use the existing network data to infer its representativeness to human exposure.
    - Additional GIS features that could be explored for determining if they distinguish sites are: proximity to ports, road density or other road-related characteristics (e.g., due to sulfur in diesel fuel, perhaps stratify by truck traffic volume), total amount of emissions in a buffer of fixed size (summed across multiple types of sources as appropriate), number of sources within a buffer. Consider varying buffer size. Analysis of COV or GSD summaries for sites by geographic characteristics may yield additional insights. (Consider determining whether there are correlations between siting features (continuous ones) and COV or differences in distributions of COV by categories (categorical siting features.) Insights from more careful assessment of sites close together with very different concentration distributions may suggest additional variables to explore.
    - I think weighting by site is more informative than by site-years since the goal is to understand the effect of location attributes on conclusions.
    - P 78 10: Unclear how the different types of sources for each monitor were captured.
    - P 80 17-18: Instead of just stating data were stratified by population density, show some analyses to demonstrate the utility of the stratification (beyond the obvious utility that monitors close to lots of people are probably more representative). (Later it is evident distributions vary by population density.)
- P 83 26: I believe the "not a direct linear relationship" wording should be replaced with "constant" (or am I missing the point?).
- Figure 7-5 is highly collapsed. I wonder what can be learned by looking at monitor-specific scatter plots of 1 hour average vs. any of the following: 5-minute maximum, 5-minute $11/12^{th}$ percentile, 5-minute mean (GM or AM), 5-minute SD (GSD or SD), 5-minute COV. These will be noisy but smooth curves will help show typical relationships.
- Figure 7-8 Make axes identical. Add N's to boxplots. The unit is number of monitors, correct? Also at this point I have lost track of whether the "broader" network includes the subset with 5-minute data or not. "Broader network" is another term to be defined in a glossary section.

- P 88 1: The PMR summary is at the site-hour level, correct?
- P 88 22: For clarity, insert "as summarized across monitors" at the end of the sentence.
- Figure 7-7: The figures are difficult to discern. Try reconfiguring to get at key features.
- P 90 21: It would be helpful to have available an analysis that shows how many and which monitors switch categories. Do we learn anything useful from that information?
- P 124 27: In Figure 7-19 I think the curve pairs have generally the same steepness, but some are offset along the x-axis. (exception is low pop density >400)
- P 139 Figure 7-21: Is there a trend over time in spatial representation of these monitors?
- P 139 Figure 7-22: Add N's
- P 140 9: I think "high" not moderate. The next line refers to the SO_2 monitoring network design, but the document does not describe the regulatory network design or siting criteria.
- P 140 15-18: Here are criteria that may cause bias and could be assessed with further analyses.
- P 144 Figure 7-23: Force all axes to be identical.
- P 146 20-21: Agree
- P 149 Figure 7-27: In both the 300 ppb and 400 pbb plots there is a cluster of points below the 1:1 line. Is there anything special about these points? Insert the number of monitors in the title.
- P 153, 154: Is there an easy way to make it clear how the two figures are different? Perhaps boldface or italicize the distinguishing "by each" words in the titles?
- P 155 17-18: This comment gives a hint that the spatial representation of the ambient monitoring network isn't well aligned with population exposure. Can this type of observation be exploited to further understand the representativeness of the monitoring network?
- P 156 14: Sentence fragment
- P 159 17: Clarify that the number of 5-min daily max exposure events in an entire year is counted as days.
- P 161 Figure 8-13: Figure 8-13: Clarify the definition of "exposure concentration estimates", ideally within the figure.
- P 163 16: Clarify these concentrations are the maxima
- P 163 21: Define "exposure event" here or provide a reference to the section where more details will be given.
- P 171 Table 8-3: Why call out "no snow"? It is not mentioned in the text.
- P 203 16-17: Explain (or refer to another section where it is addressed) the county-specific adjustments.
- P 171 Table 8-3: Why the no snow designation?
- P 261 7-9: Incorporate insights from Hattis' comments here.
- Tables 9-4 – 9-9: Add words to the titles to indicate the characterization is over a one-year period.
- P 277 22-24: Fix the wording
- P 287 27: insert "or that average population exposure to SO2 is less well captured by the existing monitoring network (i.e. measurement error)."
- P 288 21-22: Why is this sentence needed?
- P 289 20: Insert "at monitored locations" after "levels"

- P 290 bullets: should clarification be added that the asthmatics are by definition those with mild to moderate disease?
- Table 10-5: Define “*” and “**”.

Comments from Dr. Frank Speizer

Characterization of Air Quality (Chapters 2, 5, 6, and 7)

1. Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?
Chapter 2, Human Exposure: As indicated on page 15, SO2 except in areas of high volcanic activity the PRB for SO2 is generally less than 1% of SO2 concentration and therefore the decision to ignore PRB seems appropriate. With regard to potential for indoor exposure on Page 17, line 18-19, I would argue that more is known about kerosene heater use than is indicated in this sentence. There are very few states or districts where the use of kerosene stove are allowed indoors (because of fire risk) as a source of heat and there this sentence could be stronger
Page 130-133, Tables 7-11-7-13 summarized modeled 5 minute max-days/year with various 1 hour max standards. What comes across to me is that there is "comparability" between As is, 98%200 and 98%250. There is a modest improvement with 99%200 and a substantial improvement going to 99%150. This seems to be the workable range, with which to begin to look at the health data.

2. In order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO_2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?
Chapter 7, page 66, footnote. The justification of why 5 minute exceedances are counted only once per day is not clear. This is a significant change from the REA draft 1. There may also be within day variations of max 5 minutes that could be important. An asthmatic child sleeping in an air conditioned room at 3 am does not have the same exposure as the same child playing outside at 3 in the afternoon.
.
3. In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO_2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?
Page 80. Not sure this goes here but I have some concern about the definitions of low, mid and high-population density. In the east there would be considerable differences between communities of 50,000 plus and 500,000 plus. Some might look upon what is being called high as "green suburbs" in contrast to urban heat islands of the much larger communities. Should there have been a 4^{th} category that separated the 50, 000 into an even larger grouping?
4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

Chapter 7, page 84-85. For the not technical reader need to provide a more intuitive definition of "concentration variability" (even though formal words are given) if this variable is to be use to extrapolate 5min to 1 hr.
Page 91-93: Logic for calculating PMR seems good and simplified formula on page 93 for estimating 5 min max seems justified.
Table 7-14 and subsequent description is a thoughtful qualitative summary of factors affecting certainty. I like the way the qualitative categories are described and then used as justifications for the summary category for each component. However, what comes through is that the characteristics of uncertainty seem appropriate, but the directionality of the potential biases with regard to concentrations/exceedances is essential 'random' (or unknown). This doesn't seem very useful, except to point out that more research and more measurements are needed.
Page 156, Health Benchmarks. This is the one category where I found a discrepancy between the text and the table. The text, seemingly rightly, judges the uncertainty between a 5 and 10 minute controlled human exposure as similar and thus the effects seen as overall uncertainty as low (table says moderate). In fact older studies at considerably higher levels of exposure to SO2 showed tendency for airways resistance to start to improve during the second 5 minutes of continued exposure. I would disagree with the fact that if the health effect may be underestimated (as discussed in the paragraph that follows in the text), that it would change the uncertainty to moderate. It really speaks to the potential population at risk.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)

1. The presentation of the SO_2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO_2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?

Chapter 3: Page 18, Table 3-1. It is not clear how "Adverse birth outcomes" is a susceptibility factor. It may be an outcome from exposure but unless the authors mean that prematurity put the infant at greater risk it makes no sense. Notably, "low birth **rate**" probably should be low birth **weight.** On the vulnerability factors side geographic location as indicated seems a bit broad.
Page 20, para beginning line 12: This might have to be re-written. It is not clear that the same mechanism is operative for those less than 18 and those over 65. There is the potential as indicated below that those under 18 simple spend more time outdoors and thus are more vulnerable, rather than more susceptible (as seems to be the case for the elderly).

Chapter 4: Having just returned from the PM CASAC meeting the selection of only evidence sufficient to infer a causal relationship for SO2 is no consistent with the staff decision (concurred upon by CASAC) for that pollutant. This leads to a dilemma in that either there will be inconsistency on how the various pollutants are handled, or we may surprised that in the next round of PM instead of seeing Risk Assessments for categories of "suggestive of causation" that only results for "sufficient to infer" will turn up. That would be disappointing. The only other outcome that reached a level of risk for SO2 was the suggestive risk for Respiratory mortality. For consistency it might be worth doing a calculation or two for this risk, with appropriate caveats added. On the other hand, since alternatives are being considered in the range down to

100 ppb for short term respiratory morbidity it may not make any difference and it can just be commented upon.

2. The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?

Chapter 5: Page 34, lines 1 & 2. I think this is an important sentence (along with rest of the paragraph) that directly answers the first question in this section with regard to **Indicator** and **averaging time** and justifies the approach. However, I think it would be worth repeating here some greater detail of the correlations between 5 min and 1 hour measures. With regard to **form** presenting 98 and 99%iles over 3 years as alternatives seems appropriate. An important point not discussed here, but perhaps to come up later is in discussing the max and min **levels** to be considered in the risk assessment no mention of margin of safety is indicated. It appears in the selection of each level a residual of 5-10% of subjects (generally mild to moderate asthmatics or elderly) remain at risk. I would think it worth mentioning the concept that margin of safety would need to be taken into account for these subjects if the Administrator is to be compliant with the Clean Air Act that says …margin of safety for the sensitive individuals (the number of asthmatics in these categories is not trivial).

Characterization of Exposure (Chapters 6 and 8):

1. Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?
Page 51, Figure 6.2: This figure demonstrates some of the difficulties in the selection of the 1 hour as a surrogate or estimator of 5 minutes excess exposure. If we suppose we are trying to control 99% of the time getting to 200ppb for the 5 minute exposure, then if we use 65 ppb for the one hour (the lowest level that reached 200ppb for the 5 minute periods) than this could occur on 86.4 x 3(years)= 259 times before the monitor would suggest out of compliance. Surely this would lead to a significant number of asthmatics hitting emergency room floors.

2. The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?
Descriptions of St. Louis and Greene County seem like a reasonable comparison of rural to urban area, with relative similar "climate" variables. How representative of US is another issue, but probably not of concern here.
Pages 207-215, does point to the contrast between the two sites. In fact the contrasts are striking and these therefore become an excellent example to use for contrasting the "potential extremes" to consider.

3. What are the views of the Panel regarding the approaches taken to model SO_2

emission sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?
Others better qualified to comment

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?
See below

5. What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?
Page 199 and see above. Here is an example of why there are problems with generalizing from these sites. Assuming a 95.5% air conditioning prevalence rate for these two communities is may be too high since this lumps central and room a.c. (probably even for these communities, an one considers some of the more urban older parts of St. Louis). Certainly room a.c. (as well as central a.c.) must depend on usage and that can't be 95+%. Table 8.10, page 201 suggests about a 5-10 fold difference in SO2 dependent on usage.
Page 218-219, although the time spent outdoors seems reasonably uniform from the CHAD study, and the prevalence rates of asthma in children were similar in 3 of 4 regions, these cannot be the sole criteria for suggesting the sites in MO are representative. The contrasts just within the two sites chosen, in terms of percent exposure to given scenarios indoors, in cars, and outdoors points to some of the potential differences that might be expected. That said, it is not clear that staff could have done more than they did, and certainly what was done seems quite informative.

Characterization of Health Risks (Chapters 7, 8, 9):

1. Based on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO_2 exposures, we have adjusted our range of 5-minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO_2 exposures evaluated in the ISA?
This is a reasonable choice of parameter to test. One issue not discussed (although implicit in the data) is that there is a subgroup within each of the primary studies who appear to be susceptible at any given dose of exposure. Thus it is not a straight forward phenomena that if the dose were to increase from 100-400 ppb that this would simply illicit a greater number of responders, although it does. If one studies non-responders at the lower does they may continue to be non-responders at the higher doses. Too few studies have studied the same individuals at differing exposures to sort this out. We simply do not know what makes an individual sensitive to SO2., albeit true that dose is one part of the cause.

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?

Chapter 9 summarizes clearly and effectively the estimated change in airways resistance to be expected under a variety of scenarios and is clearly presented. What is not said is that a doubling (100% increase) in airways resistance in exercising asthmatic children does not necessarily result in a perceived health effect (it depends upon the baseline level of sRaw). For this reason it might have been useful to present similar data for a 15% decline in FEV1 that are in the Appendix as this is more intuitive measure of lung function effect.

3. A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?
See comment above.

4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?
Discussion of the AERMOD Algorithms uncertainties best by others with more expertise than I. With regard to estimates of population exposure, is it true as indicated on pages 226-228 that the data on commuting is taken from 2000 Census and includes only adult home to work only? If this is so than it is not clear how estimates are being made for school children. To say that most exposures are occurring outdoors, and not to account for approximately 1.5 hours/day (two ways) for the very large segment of the at –risk population that is spending time in poorly ventilated school buses 5 days a week, seems a source of uncertainty that needs to be discussed. Although sources of variability are well discussed there are significant limitations as to how well they are treated in the estimates. For example, page 239 indicates the potential frequency of multiple exceedances if 5 minute SO2 in 1 hour by different benchmark levels within an hour. One gets a consistent picture of what might be happening, along with some insight into the uncertainty that might result from this phenomena. However, in regard to variability of the final estimates (as is demonstrated in Figure 8-23 on page 242) there are no error bars on the histogram. This is not to fault staff, as I do not think it possible to quantitatively deal with the variability, except to discuss the sources. Further with regard to the estimates used for asthma, page 246-7 points out the potential variability in the diagnosis of asthma within the St. Louis site and suggests the potential for uncertainty in the estimates made.
Table 9-10, Page 279, I think the issue of 5 vs 10 minutes of exposure is overstated as to it being a potential overestimate of bias. As discussed in the comment section much of the response in those in whom it has been measured show a response within 5 minutes. It was the protocol of the studies that resulted in the measure being recorded at 10 minutes. In fact, if anything the 10 minute measures may be an underestimate since in some of the studies recovery from the initial response was already underway after 5 minutes, in spite of continued exposure.

Policy Assessment (Chapter 10):

1. The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be

technically sound, clearly communicated, and appropriately characterized?
Integration of data from both the modeling and actual data from controlled studies is well done and lays the groundwork quite effectively for the summary of findings on page 290.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?
Logical and complete presentation of data that leads to the conclusion that the current annual standard does not provide sufficient protection for the short term effects and alternatives must be considered. Not clear, at least to page 300 if the 24 hr and or 1 hr alternative would replace or be added to the annual, however concur with the continued use of SO2 as the **indicator**. Page 310 staff suggests that the **averaging time** that controlled would best predict both acceptable levels of 5 minute and 24 hour averages and keep annual in line would be 1 hour averaging times. I would concur. Next with regard to **form** agree with the 1hr daily max standard with a 99^{th} percentile form. No indication here is given to whether this is to be averaged over 1 or 3 years, but I would favor 1 year, since the number of measures are so much greater at the 1 hour level, the numbers of observations that would be in excess over a 3 year period, would greatly increase potential risk, particularly to asthmatic children, were a run of excess to occur in one of 3 years. With regard to **level** the discussion as presented justifies a range of 50-150 ppb.

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?
Staff expressed concerns, with which I concur, that only if a 1 hour standard is within the range suggested is implemented then it would be inappropriate to allow this new standard to replace the existing 24 hr and annual standard. Given the current standard does not protect against the short term effects than the current standard would have to be lowered and since there is greater uncertainty in how the longer term standards affect the 5 minute averages, the standards would have to be lowered even more than might be anticipated to maintain any margin of safety.

4. Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50- 150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?
Well justified. After presenting the evidence staff has added a discussion of the uncertainty which if anything leads to the conclusion that the top end of the range may be too high, particularly as the evidence that is used suggests the findings are in less than the potentially most susceptible populations (children with moderate to severe asthma) who were not studied in the clinical human exposure studies.

Comments from Dr. George Thurston

The document is generally in excellent shape, and the EPA Staff and their collaborators should be commended for an admirable job. However, I do have remaining issues, primarily regarding the lack of any quantitative risk analyses based upon the epidemiological literature.

My comments address all Charge Questions, as appropriate, but focus primarily on my assigned Charge Question regarding Characterization of Health Risks (RE; Chapters 7, 8, 9).

Characterization of Air Quality (Chapters 2, 5, 6, and 7)
1. Does the Panel find the results of the air quality analyses to be technically sound, clearly communicated, and appropriately characterized?

Yes.

2. In order to simulate just meeting potential alternative 1-hour daily maximum standards, we have adjusted SO_2 air quality levels using the same approach that was used in the first draft to simulate just meeting the current standards. What are the Panel's views on this approach? To what extent does this approach characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

No concerns. This is a very useful approach.

3. In this second draft document, the locations selected for detailed analyses were expanded from twenty to forty counties, using ambient SO_2 monitoring data for years 2001-2006. What are the views of the Panel regarding the appropriateness of these locations and time period of analysis? To what extent is the rationale for selection of these locations and time periods clear and sufficient to justify their use in detailed air quality and exposure analyses?

These seem a valid choice, however, it is unfortunate that later in the document, as noted on page 216: "Due to time and resource constraints the exposure assessment evaluating the current and alternative standards was only applied to the two locations in Missouri." This limits the ultimate usefulness of the work done to characterize all 40 counties in this chapter. Alternatively, if EPA's BENMAP model were to be applied to these data in conjunction with the epidemiological literature for SO_2, then all 40 Counties could be considered quite quickly for use in this document.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

This seems well done. I especially like that the EPA staff has noted the likely bias direction, if any, in Table 7-14.

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis (Chapters 3, 4, 5)

1. The presentation of the SO_2 health effects evidence is based on the information contained in the final ISA for Sulfur Oxides. Does the draft REA accurately reflect the overall

characterization of the health evidence for SO_2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?

Yes, it is brief, but balanced.

2. The specific potential alternative standards that have been selected for analysis are based on both controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential alternative standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alternative standards for use in conducting the air quality, exposure, and risk assessments?

These appear to be appropriate choices, based upon the clinical study evidence. However, given that Table 7-10 indicates that there is only a reduction in the number of modeled exceedances at 50 ppb for the highest counties (e.g., Hudson, Tulsa, and Wayne), consideration should also be given to also evaluating a 50 or 75 ppb benchmark, as well.

Characterization of Exposure (Chapters 6 and 8):

1. Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?

Yes, it is appropriate and state-of-the art.

2. The second draft REA evaluates exposures in St Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

This is likely the best that can be accomplished when basing assessments on clinical studies. However, because such clinical studies do not consider populations representative of the full distribution of the public, and because their data are not collected in the "real world", numerous exposure modeling assumptions must be made to extrapolate from these controlled exposure conditions to what is actually happening in the real world to real people in this approach. These assumptions regarding dispersion modeling (which is accurate only within a factor of 2), population time-location-activity patterns, meteorology, outdoor-indoor permeation rates, air conditioning, indoor decay rates, etc. are piled one upon the other, leading to potentially large errors in exposure assessment in this process. In contrast, the use of epidemiological studies based upon central site modeling can avoid these problems because they have already adjusted for all of these factors inherently through their original design, by using central site data and real populations, which controlled exposure studies have not. Controlled exposure studies are most appropriate for testing biological plausibility, but epidemiological studies offer many advantages over them when conducting a quantitative exposure-health effects evaluation.

3. What are the views of the Panel regarding the approaches taken to model SO_2 emission sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?

No comments.

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for

the risk characterization been addressed? To what extent has variability adequately been taken into account?

This section has done an excellent job of laying out the many layers of uncertainty involved in using clinical studies in a quantitative risk assessments. However, I would like to see a summary table added in Section 8.11 that is similar to Table 7-14.

5. What are the views of the Panel regarding the staff's characterization (in Section 8.10) of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?

I think it hits the key points, and does as well as possible, given that they only have analyses for two counties from which they are trying to draw generalized conclusions. A 40 county analysis would be preferable (as would be possible via an parallel epidemiology-based risk assessment).

Characterization of Health Risks (Chapters 7, 8, 9):

1. Based on conclusions in the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO_2 exposures, we have adjusted our range of 5- minute potential health effect benchmark values to 100 – 400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO_2 exposures evaluated in the ISA?

This range appropriately reflects the health effects evidence provided by controlled exposure studies reported in the ISA. However, as noted above, the exposure analyses in the earlier chapters suggests that consideration should also be given to a benchmark as low as 50 ppb.

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?

The quantitative risk assessment conducted in this REA are technically sound, but could, on occasion, be more clearly characterized and communicated. In general, there is a need to succinctly explain how to interpret the key results in each figure, and to provide illustrative examples as to how to read the figures, when possible.

Discussion of Figure 7-16 is a case where EPA staff has done a good job culling out the underlying message of the results presented by stating: "There are a decreasing number of exceedances with increasing benchmark concentrations, though there is a greater proportion of monitors with exceedances when considering concentrations adjusted to just meeting the current standard than when using the *as is* air quality (e.g., see Figure 7-13)."

However, I find the discussion of Figure 7-19 to be less clear, and suggest the insertion of a statement saying (if I am understanding this figure correctly) that:

"Figure 7.19 shows the relationship between the probability of a 5 minute exceedance as a function of a given ambient 1-hr daily maximum concentration. For example, in the high population density locales, there is roughly a 40% chance of exceeding 100 ppb 5-min concentration benchmark when the prevailing 1-hr maximum

daily concentration is limited to 50 ppb, but a 0% chance of exceeding 200 ppb 5-min concentration at this same 1-hr maximum daily concentration limit."

On page 128, the statement that: "Most counties have fewer mean estimated 5-minute benchmark exceedances of 100 ppb using air quality adjusted to just meeting the 99th percentile daily 1-hour maximum concentration of 100 ppb, than estimated using the *as is* air quality." is clear and concise, but needs to also address the results for the highest counties by adding text something like:
"*, but this is not the case in the counties with the greatest number of benchmark exceedances (i.e., Hudson, Tulsa and Wayne), which must go to a 50 ppb limit to achieve any reduction in the number of SO_2 exceedances vs. the "as is" case*".

Also, EPA should consider adding a column for 75 ppb to Tables 7-10 through 7-13, for comparison with 50 and 100 as a policy option in chapter 10..

3. A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of this analysis?

The EPA staff has done an excellent job at conducting the selected quantitative risk assessment for the two chosen indicators in these specific two counties. However, I still feel that this is too narrow a scope for CASAC to fully evaluate and inter-compare the various alternative short-term benchmarks for SO_2. As noted in Appendix C of this report: The SOx ISA concludes that the health evidence "is *sufficient to infer a causal relationship between respiratory morbidity and short-term exposure* to SO_2" (ISA, p. 3-33). It goes on to state that:

> *".A larger body of evidence supporting this determination of causality comes from numerous epidemiological studies reporting associations with respiratory symptoms, ED visits, and hospital admissions with short-term SO_2 exposures, generally of 24-h avg. Important new multicity studies and several other studies have found an association between 24-h avg ambient SO_2 concentrations and respiratory symptoms in children, particularly those with asthma… Collectively, the findings from both human clinical and epidemiological studies provide a strong basis for concluding a causal relationship between respiratory morbidity and short-term exposure to SO_2.*

While this REA does address the first category (clinical studies) very well, it does not make quantitative risk evaluations using the second, much broader, category (epidemiological studies). I feel consideration of both would bring differing perspectives and insights into the potential health implications of the various possible short-term benchmarks presented, but only one type is quantified in this document. The application of epidemiological-based risk assessment using the EPA's BENMAP model to the 40 selected counties considered in this report could have also been accomplished, and would have improved the usefulness of this document. Indeed, such an epidemiology-based risk assessment should be always be conducted for this and all REAs during the Criteria Pollutant standard setting process.

While the use of BENMAP has its own limitations and uncertainties (e.g., how best to consider other co-pollutants), the application of the clinical studies are not without their own

limitations and uncertainties, as elaborated upon on pages 3-12 through 3-14 of Appendix C of this REA. Most notably, on page 3-15 (as well as in Table 9-10 of the REA) it is noted that a "main uncertainty" includes:

> *Interaction between SOx and other pollutants. Because the controlled human exposure studies used in the risk assessment involved only SO_2 exposures, it was assumed that estimates of SO_2-induced health responses would not be affected by the presence of other pollutants (e.g., $PM_{2.5}$, O_3, NO_2).*

However, it is known from the literature that the co-presence of particles enhances the penetration of sulfur oxides into the lung, and, therefore, that the impacts estimated based on the clinical studies using pure SO_2 is an underestimate of the effects of these same concentrations in the real world,.

Furthermore, Appendix C (and Table 9-10) also points out that another main uncertainty in this analysis is that:

> *As indicated in the ISA (p. 3-9), the subjects studied represent the responses "among groups of relatively healthy asthmatics and cannot necessarily be extrapolated to the most sensitive asthmatics in the population who are likely more susceptible to the respiratory effects of exposure to SO_2."*

Thus, this analysis only includes two counties, is limited only to effects among asthmatics, and, even then, it doesn't include the most sensitive members of the asthma population.

Overall, while this risk analysis based on clinical studies is one appropriate assessment approach that has been well executed by the EPA staff, it has its own limitations that, overall, tend to understate the health benefits of lowering SOx pollution in the general public throughout the nation. In contrast, the application of the epidemiological study results to all 40 counties using the EPA's BENMAP model would provide an alternative perspective of this issue for a much larger and more representative population, and would seem an essential analysis to <u>also</u> be completed as a part of this and all future REA documents.

Finally, I'd like to make comment on the multi-pollutant issue. As noted above, the consideration of the use of epidemiological studies for risk assessment raises the issue regarding how best to estimate the effects of individual pollutants, but the issue is broader than this one application. How to best consider possible co-effects, effect modification, or confounding by other co-pollutants is a concern in every one of the documents, and cuts across the purview of any one pollutant committee. For this reason, I'd like to suggest that SAB consider appointing a separate CASAC committee to examine just this one specific issue (rather than a specific pollutant) with the aim of giving general guidance to all future individual pollutant evaluations as to how to deal with this multi-pollutant issue in a consistent manner across all pollutants.

****

In addition, with regard to this particular SOx REA, I also have some additional specific comments/suggestions regarding the quantitative risk assessment in this document, as follows:

Pg. 157. EPA should consider inserting a summary sentence on line 20 that says:

> *"Thus, the current standards are seen to be ineffective in protecting the public against the adverse health effects of short-term (e.g., 5 minute average) peaks in SO_2 concentration."*

Pg. 158, line 9: the EPA should consider adding a summary sentence, something like:

Therefore, if a new 1-hour daily maximum standard is to protect the public against short-term peaks better than the existing annual standard does, it will have to be at a level below the 200 ppb benchmark level.

Page 158, line 9. After the above, EPA should consider adding another bullet discussing the informative results in Table 7-10, and a statement saying something like:

"Thus, the 40 county analysis of exceedances in Table 7-10 indicate that, if the public is to be consistently protected against 5-minute peaks in in SO_2 better than the "as is" case, then the 1-hour 99th percentile maximum limit will need to be set lower than 100 ppb.

Page 247, line 3. This sentence does not make sense. I think it should read "Therefore, in St. Louis City"

4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

This seems reasonable. Table 9-10 summary of uncertainties is really helpful.

Policy Assessment (Chapter 10):
1. The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?

Yes, it is.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of the current standards?

It does this adequately. However, the risk-based considerations are really only exposure-based, and the health effects implied should *also* be discussed. Again, an epidemiological study based risk assessment would be helpful in considering the risks associated with the current standards.

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO_2 standards?

This is well done from a clinical study perspective, but would benefit from a discussion of the public health impacts implied by the epidemiology studies for the various options. I disagree with the gist of the discussion at the top of page 305, which implies that clinical studies are superior for this purpose than epidemiological studies, and then (at line 5) lists the limitations of epidemiology-based risk assessments, while never mentioning the limitations of applying clinical study-based results to real-world situations. A more balanced discussion is needed that presents the strengths and limitations of each. At a minimum, the word "greater" should be removed from line 5, as there are many uncertainties associated with applying clinical studies to

health risk assessment, as well.

4. Staff believes that the evidence presented in the final ISA and the exposure and risk information presented in this second draft REA supports a potential alternative 1-hour daily maximum standard within a range of 50- 150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?

I feel the chapter makes a compelling case for this conclusion. However, at line 27 on page 319, I feel more quantification is needed as to exactly how many fewer exceedance days are associated with each option.

Also, on page 320, line 12, it would be helpful to refer to Table 7-10, especially if a 75 ppb case were added to that table. In addition, Figure 7.19 would be useful to refer to here, as it appears to me to indicate that, for the high population density cities, setting a 100 ppb maximum 1-hr daily maximum limit would still allow about a 40% chance of a day with a 5-minute peak greater than 100 ppb.

In addition, on page 312, line 7, from a grammatical perspective, should read: “allows fewer days per year” (not “allows less days per year”).

Comments from Dr. James Ultman

This document, that now includes the health risk assessment in chapter 9 and policy-enabling suggestions in chapter 10, is more clearly written and easier to follow than the first draft. Moreover, the earlier chapters have been improved in several important respects: the range of health benchmarks investigated has been sufficiently lowered to provide a margin of safety, particularly for children and sensitive asthmatics that were not studied in clinical tests; a cross-validation of the available 5-minute monitoring data has been used to estimate the error associated with the proportional roll-up/roll-down method of adjusting air quality; the equivalence of scaled-down benchmarks to scaled-up air quality has been validated using simulations from existing 5-minute monitoring data; and the comparison between monitors reporting 5-minute maximum SO_2 concentrations and the broader nation-wide monitoring set was expanded to include land-use, setting, objective and scale.

Most importantly, chapter 10 provides a good summary of the key points covered in the earlier chapters and thoughtfully-written recommendations for improving the current standard.

Characterization of Health Risks – Charge Question 2: Does the panel view the results of the risk characterization in Chapters 7 and 8 and the lung function quantitative risk assessment in chapter 9 to be technically sound, clearly communicated, and appropriately characterized.

The overall answer to this question is YES. However, I believe that there is still a need for the following improvements:

1) The risk assessment uses an "equivalent ventilation value" of 22 L/min-m^2 as a threshold for activity levels of moderate or greater intensities in people of different ages. An explanation of how this factor was arrived at should be added. According to the discussion at the CASAC review meeting, there is past research that can be used to support this approach.

2) Both the ISA and the REA indicate the pulmonary response observed clinically in asthmatics is more sensitive to SO_2 concentration than to exercise intensity. This assertion is the basis for using a ventilation threshold to compute exceedances in the exposure analysis (Chapter 8) and to determine the number of responders in the health risk analysis (Chapter 9). Yet the assertion appears to be based on only one study (Gong, 1995) carried out at 0.5and 1.0 ppm SO_2 at three exercise levels. And I cannot find the data from that study in either the ISA or the REA. If possible, Henry Gong's data and any other relevant clinical data from other labs should be analyzed to justify the use of a ventilation threshold (see above point).

3) It would be useful to overlap "confidence limits" on the exposure-response functions. If staff feels it would be confusing to implement this in figures 9-2 to 9-5, then the reader should be pointed to the appropriate figures in the appendix.

4) The equivalence of the benchmark scale-down technique to the air quality scale-up is a key assumption that has been validated in this document. However, I think that the presentation of these results on pages 61 and 62 could be much improved with relatively little effort (see my specific comments below)

Specific Technical and Editorial Comments

Page 25, line 20-23. This statement is certainly true during quiet breathing and light exercise. As exercise becomes more intense, however, subjects do switch from nasal to oronasal to oral breathing. At the minute ventilation rates used in the exposure assessment (Ch. 40-50 Lpm: pg. 195), adults on the average breath about 50% of their minute volume through their mouth (Niinimaa V, Cole P, Mintz S, Shephard RJ. Oronasal distribution of respiratory airflow. Respir Physiol. 43:69-75, 1981). Moreover, obligate oral breathing may occur in some individuals because of anatomical abnormalities or nasal congestion. It would be informative to discuss the issue of "portal of entry" effects in more detail.

Pages 38-42 (Figs. 5.1-5.5). It would be helpful to place a marker on each entry on the figures to indicate statistical significance in single and in multipollutant models where available).

Page 56, line 13-14. Proper wording should be "With a rounding convention applied to the third significant figure…"

Page 57, line 15. For clarity, change add wording "whichever is the controlling standard (*i.e.*, results in the smallest upward adjustment).

Page 60, lines 8-9. This phasing that is repeated several times in the REA is not correct. A proportional change in the benchmark (an output of the model) will be equivalent to a proportional change in the air quality (an input to the model) when the model is linear, not "because the adjustment procedure is proportional."

Page 60, line 26. Change wording to clarify to "were input to the statistical PMR model…"

Page 61-62, Fig. 6-5 and 6-6. It is important that these results be presented as clearly as possible in order to justify the equivalence of the scale-down of the benchmark values to the scale-up of current air quality. Please read the following for some possible ideas on how to clarify and possible strengthen your discussion:

As I understand it, the distributions in both graphs will be the same no matter what benchmarks are examined; the left-hand distribution is the result of imposing the PMR model using "as is" air quality and the right-hand distribution is the result of applying the PMR model to the scaled-up air quality.

You chose to focus on comparing cumulative percentile at the 400 ppb *actual benchmark* on the right-hand distribution to the 78.4 ppb *hypothetical benchmark* (obtained by applying the adjustment factor of 5.10 found for Cuyahoga County—400/5.10=78.4) on the left-hand distribution. You indicate that if both methods predict the same cumulative percentile of 5-minute daily maximum SO_2 concentrations, then the number of exceedances of the actual benchmark resulting from both methods will be the same.

What about treating the 100, 200 and 300 ppb actual benchmarks in the same manner? The cumulative percentages corresponding to each of these values on the right-hand distribution should be equal to the cumulative percentages corresponding to 100/5.1, 200/5.1 and 300/5.1 on the left-hand distribution. I think that making all four comparisons (perhaps in a short table) would strengthen your case.

Page 61-62, Fig. 6-5 and 6-6. If you choose to retain these figures please make the following changes. Fig. 6-5: eliminate the horizontal interrupted line; label the 78.4 ppb point on the abscissa; and fix the symbol for the "Adjust Benchmark Down" in the legend. Fig. 5-6: Correct from "Figure 6-4" to Figure 6-5" in the legend.

Page 89, Fig. 7-7. It would be useful to give the reader some idea of how many data points were used to construct each distribution. Based on my reading of appendix A.3, it appears as if there are only about 5 points.

Page 92, line 8. Delete "staff characterized."

Page 95, line 8. Here and at many other places in the document one is written as a number (*i.e.* 1 exceedance). It would be preferable to spell it out (*i.e.* one exceedance).

Page 96, Table 7-5. May be helpful to mention in a footnote to the table that prediction error refers to the actual exceedance less the median exceedance.

Page 97, line 11-12. Reorder sentence as "(4 and 2 years out of 8 total site-years did not meet the completeness criteria for each of …)"

Pg. 107, line13. Shaded area is not visible on table 7-8.

Pg. 112, line 6. "greater number of monitors"

Pg. 168, line 5. "…, 2002 temperatures were similar to the 30-year normal…"

Pg. 173, line 23-24. Please give a short explanation of the how and/or why the different values of chemical decay were arrived at for rural and urban environments.

Pg. 181, line 21. The phrase "…concentration distribution measured at each ambient monitor…" is a bit confusing. Since a single monitor cannot measure a spatial distribution, I assume that your phrase refers to a distribution obtained by binning measurements taken at different times.

Pg. 197, line 8. The PMR model was developed based on outdoor monitoring measurements. How were the 5-minute indoor averages determined?

Pg. 207, line 9. "…were different from those observed in Greene…"

Pg. 253, Table 9-1. What are the units applicable to the first three columns? ppm?

Pg. 262-265, Figs. 9-2 to 9-5. Is it possible to put "standard deviations" on the data points to reflect variations between the different studies.

Pg. 272, table title. "Concentrations…"

Pg. 275, Fig. 9-7. Coding in histogram subdivisions is illegible.

Pg. 295, lines 10-11 Change two phrases to be more precise: "…predicted yearly mean number…" and "…from 1-102 days when…"

Pg. 295, lines 13-14. Change two phrases "…experiencing a yearly mean of at least 20 days…" and "…from 22-171 days, with about…"

Pg. 295, line 16. Change 400 to 200.

Comments from Dr. Ronald Wyzga

Characterization of Health Effects Evidence and Selection of Potential Alternative Standards for Analysis

1. The presentation of the SO2 health effects evidence is based largely on the information contained in the final ISA for Sulfur Dioxides. Does the draft REA accurately reflect the overall characterization of the health evidence for SO2 contained in the final ISA? Does the Panel find the presentation to be clear and appropriately balanced?

My biggest concern is the discussion on pp. 24-25 about adversity and the selection of health endpoints to be used in the subsequent risk analysis. It is very brief and provides little justification for the subsequent analyses. The ATS (2000) guidelines for adversity are presented, but not adhered to in the subsequent analyses. They are only guidelines, but the rationale for deviation from them should be more fully discussed. The given rationale makes little sense given the description of subject responses to the lower SO2 exposures. There were a small number of asthmatics who asked for medication after some SO2 exposures; however, a small number also asked for medication after exercise-only with no SO2 exposure. Work output was if anything less after exercise only exposure than for low SO2 (200ppb) plus exercise exposure. These results appear to contradict the current rationale for endpoint selection given on p. 25, ll. 2-5. There also needs to be more discussion about the choice of Sraw over FEV-1.0.

This information would be useful in developing dose-response functions from the human clinical evidence to the epidemiological evidence. For example, how would the dose-response function differ if lung function results plus symptom response were considered.

It is correctly noted that medication use is an effect modifier (p. 27), and there is evidence that several asthmatics take medication regularly as a prophylactic. To assume that no asthmatics use asthma medication regularly will result in an overestimate of effect, a bias that is not cited in the document. I do not know the percentage of asthmatics who regularly take medication; a quick search on the web found one paper based on a sample emergency room visits, hardly a good sample population. Reliable data may exist elsewhere. The argument that mild asthmatics are less likely to use medication needs support and clear definitions of mild, moderate, and severe. Could asthmatics be "mild" because of regular medication use? Clarity here would aid the arguments in the REA.

I believe there is greater independence between the human clinical study results and the epidemiological results than is acknowledged in the document. For example, on p. 28 the Mortimer et al., 2002, study reported an association between morning SO2 exposures and symptoms 1-2 days later. The human clinical studies report symptoms, when they occur almost immediately, and they subside after exposure, especially at low levels. (See Linn et al., 1987, for example.) In many cases the epidemiological results are potentially confounded by other pollutants, making the linkage between the human clinical studies and the epidemiological studies awkward. The REA cites some examples of confounding (e.g., Mortimer et al., 20002); in other cases it ignores the published impacts of potential confounding (e.g., Peel et al, 2005 where ozone appeared to be the pollutants of greatest

concern); in other cases the studies did not consider confounding or consider it in a sufficiently systematic manner to allow a conclusion.

2. The specific alternative standards that have been selected for analysis are based upon controlled human exposure and epidemiological studies. To what extent is the rationale for selection of these potential standards clear and sufficient to justify their use in the air quality, exposure and risk analyses? What are the views of the Panel regarding the appropriateness of these potential alterative standards for use in conducting the air quality, exposure, and risk assessments?

I would like to see less grouping of levels than is presented in section 5.5; in particular, given comparable studies (same investigator, very similar protocols), there appears to differences between 0.2 and 0.3ppm; hence I would like to see the REA discuss these separately.

I am not convinced that an examination of the epidemiological results is informative in considering an appropriate level. First of many of the epidemiological results are not statistically significant, indicating uncertainty in their results. Secondly the estimated dose-response curves were linear with little no consideration of thresholds. There was little consistency in the consideration of confounders across studies, and if one plots percent of excess risk for a given endpoint against 1hr SO2 maxima across studies, there is no indication of a consistent pattern. I find the arguments given in this section vis-à-vis the use of epidemiological data to be insufficient. Focus on the human clinical studies is appropriate.

Characterization of Exposure

1. Does the Panel view the results of the exposure analyses to be technically sound, clearly communicated, and appropriately characterized?

In general, this analysis is thorough and well-communicated. My biggest concern is the assumption that exercise patterns are similar for asthmatics and non-asthmatics. There is scant evidence presented to support this assumption; the Gent et al. paper cited is a European study where activity patterns may be very different from the US; moreover is considers "(un)diagnosed asthmatics", whatever that may be. One who has not been diagnosed by a physician may have very activity patterns than an asthmatic that has been diagnosed by a physician. Secondly, there are other some sources of data that could help inform this analysis. Shamoo et al. (1984) (Journal of Exposure Assessment and Environmental Epidemiology, 4(2):133-148) studied the exercise patterns of 49 asthmatics aged 18-50; many of these were subjects in the human clinical studies discussed and used in the REA. This group spent 0.2% of waking hours engaged in outdoor fast activity and 2% of waking hours engaged in outdoor moderate activity. I suspect that this a considerably lower activity rate then is embedded in APEX and CHAD. This needs to be discussed. A second study is described in an old EPRI Report (EPRI TR-101396, November 1992), in which a random sample of 136 asthmatics was studies in Cincinnati in August of 1987. This study accompanied one of healthy individuals which I believe Ted Johnson used in the development of the CHAD database. That study found asthmatics (adults and children) to be outdoors exercising at a strenuous (jogging or more extreme) levels 3% of the their waking

hours and at a mode4rate level (brisk walking or more) 11% of their waking hours. There may be additional data sources as well. Clearly these need to be considered in justification of the above assumption.

2. The second draft REA evaluates exposures in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken? To what extent does this approach help to characterize the public health implications of the current standards? Does the Panel have technical concerns with this approach?

It is important to have undertaken exposure and risk assessments for areas with real 5-minute averaging data. It adds credibility to the document as a whole. The specific areas considered are reasonable as large and small urban areas. I am unaware of any characteristics of these areas that would make them outliers.

3. What are the views of the Panel regarding the approaches taken to model SO2 emissions sources? Does the Panel have comments on the comparison of the model predictions to ambient monitoring data?

4. What are the views of the Panel regarding the adequacy of the assessment of uncertainty and variability? To what extent have sources of uncertainty been addressed? To what extent has variability adequately been taken into account?

See the response to the first question here. The uncertainty embedded in the assumption about similar exercise patterns for asthmatics and non-asthmatics need be addressed.

5. What are the views of the Panel regarding the staff's characterization of the representativeness of the St. Louis and Greene County, MO exposure and risk estimates?

See above.

<u>Characterization of Health Risks</u>

1. Based on the conclusions of the ISA regarding decrements in lung function in exercising asthmatics following 5-10 minute SO2, we have adjusted our range of 5-minute potential health effect benchmark values to 100-400 ppb. To what extent does this range of benchmark values appropriately reflect the health effects evidence related to 5-10 minute SO2 exposures evaluated in the ISA?

I believe that this range is appropriate and well-informed by the human clinical study literature.

2. Does the Panel view the results of the risk characterization in Chapters 7 and 8 and the lung function qualitative assessment in Chapter 9 to be technically sound, clearly communicated, and appropriately characterized?

I have several issues concerns about this characterization and would like to see it supplemented by additional analyses that would inform decisions about the NAAQS. First of all the REA alters the ATS definition of "adverse impacts" to the existence of decrements in lung function and/or respiratory symptoms. This consideration is useful, but at the very least it should be accompanied by additional analyses which strictly follow the ATS definition of decrements in lung function and symptoms. I know there are CASAC Panel colleagues who have endorsed the current approach; I accept that, but it is important to augment this approach with one based upon the ATA guideline. It would be important to see whether the conclusions drawn would be robust given alternative dose-response functions.

In terms of specifics, I wonder if the discussion on p. 108 and Figure 7-11 about the influence of population densities on the probability of exceedances really reflects proximity to point sources, many of which are in low population density areas.

3. A quantitative risk assessment has been conducted with respect to two indicators of lung function response in exercising asthmatics in St. Louis and Greene County, MO. What are the views of the Panel on the approach taken and on the interpretation of the results of the analysis?

This exercise was valuable and is helpful in interpreting the results. My concern is that I these analyses be extended to consider alternative dose-response as indicated above. These should consider different health endpoints (lung function decrements plus symptoms, and possibly symptoms only). I would also like to see more transparency in the assumptions made about activity levels; the impact of alternative estimates should be investigated.

4. What are the views of the Panel regarding the adequacy of the discussion of uncertainty and variability? To what extent have sources of uncertainty been identified and the implications for the risk characterization been addressed? To what extent has variability adequately been taken into account?

This is my greatest concern and the area of the REA that needs considerably more articulation. There are several sources of uncertainty, potential variability that are not addressed: the ignoring of the potential influence of habitual medication use on response, the consideration of alternative responses (e.g., lung function decrements plus symptom changes), the consideration of alternative estimates of asthmatic exercise levels.

<u>Policy Assessment</u>

1. The policy chapter has integrated health evidence from the final ISA and risk and exposure information in this second draft REA as it relates to the adequacy of the current and potential alternative standards. Does the Panel view this integration to be technically sound, clearly communicated, and appropriately characterized?

See my above comments; I would like to see additional analyses before making any conclusions.

2. What are the views of the Panel regarding the staff's discussion of considerations related to the adequacy of the current standards? To what extent does the draft policy chapter adequately characterize the public health implications of current standards?

See above; I find the existing analyses to be incomplete to assess the current 24-hour standard. I see little support in the REA for the current annual standard.

3. To what extent does the draft policy chapter adequately characterize the public health implications of the potential alternative 1-hour daily maximum SO2 standards?

Without additional analyses, I believe the current chapter does not adequately characterize these implications. There is also the issue of whether these responses should be placed in perspective, by, for example, contrasting the implications with those of exercise alone or of daily activities independent of air pollution..

4. Staff believes that the evidence presented in the final ISA the exposure and risk information presented in this second draft REA supports a potential 1-hour daily maximum standard within a range of 50-150 ppb. To what extent does the draft policy chapter provide sufficient rationale to justify this range of levels?

Clearly SO2 is the only indicator that can be considered. I think one of the best arguments for a 1-hour standard as opposed to a 5 minute standard is that of expediency and the awkwardness of promulgating a 5-minute standard. I think we need a much better understanding of variability of SO2 levels across 5 minute intervals in an hour before, for example, considering what percentile of 5-minute exposures is most relevant.

I reserve any judgment on the level of the standard until a more comprehensive risk assessment is completed.

Minor comments:
p. 10, l. 9: "Allegheny"
p. 248, l.4 Introduction
p. 256, ll. 23, 26 Table 9-3.