

Compilation of Individual Panel Member Comments on EPA’s second draft *Risk and Exposure Assessment (REA) to Support the Review of the SO₂ Primary National Ambient Air Quality Standard*

This enclosure contains pre-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	2
Comments from Dr. John Balmes	7
Comments from Dr. Ellis Cowling	10
Comments from Dr. Douglas Crawford-Brown	18
Comments from Dr. Terry Gordon.....	22
Comments from Dr. Rogene Henderson.....	28
Comments from Dr. Dale Hattis.....	30
Comments from Dr. Patrick Kinney	36
Comments from Dr. Timothy Larson	37
Comments from Dr. Richard Schlesinger	40
Comments from Dr. Christian Seigneur.....	41
Comments from Dr. "Lianne" Elizabeth Sheppard.....	43
Comments from Dr. Frank Speizer.....	46
Comments from Dr. George Thurston.....	53
Comments from Dr. Ronald Wyzga.....	60

Comments from Prof. Ed. Avol

SOx 2nd Draft REA Comments
Ed Avol, initial comments

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 one 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 appropriately 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 5th percentile and 95th percentile, or a 10th and 90th) 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.

General Comments on REA 2nd Draft

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).

Specific Comments on REA 2nd Draft Sections

Chapter 1:

1. In the 2nd 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

Comments of John Balmes Re: Second Draft of the Risk and Exposure Assessment to Support the Review of the SO₂ Primary National Ambient Air Quality Standards

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 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 accurately reflects the characterization of the evidence regarding health effects of SO₂ 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 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?

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 SO₂ 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 SO₂ 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 SO₂ standards.

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?

The draft of Chapter 10 adequately characterizes the public health implications of the potential 1-hour daily maximum SO₂ 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 SO₂ standards.

Comments from Dr. Ellis Cowling

Second Draft Risk and Exposure Assessment (REA)
To Support the
Review of the Primary National Ambient Air Quality Standard for SO₂

Very General Comments on the New NAAQS Review Process and Suggestions for Improvement in Development of Integrated Science Assessments and Risk and Exposure Assessment Documents

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 SO₂, 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 SO₂, 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 advise 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 “NO_x 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 NO_x, SO_x, 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 (NO_x and SO_x), 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 SO₂ 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 SO₂

My specific assignments for review of the Second Draft REA for SO₂ 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 SO₂, 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 SO₂ 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 SO₂ emission sources?

The approach taken in efforts to model SO₂ 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 SO₂ 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 SO₂ 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 SO₂ – recognizing, of course, that there are not very many urban and nearby suburban areas where both long-term and very short-term SO₂ 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 SO₂ 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 SO₂.

Comments from Dr. Douglas Crawford-Brown

Review of the Risk and Exposure Assessment document for SO₂

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 less 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. 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 NO_x REA.

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?

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.

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.

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?

It appears that EPA staff have adequately addressed uncertainty and variability.

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. Based 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 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 standard was somewhat unbalanced. More emphasis was placed on discussing the potential inappropriateness of the annual standard than the 24 hr standard. I wasn't sure if this was because more data was available to evaluate the annual standard's (in)appropriateness in protecting health or the 24 hr standard discussion was just added to the chapter at the last minute.

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?*

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 SO_x and NO_x 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, line 13 – ‘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

Comments on the 2nd draft REA for SO_x

Rogene Henderson

Assigned charge question:

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 SO_x 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 data 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 99th 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 98th 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

Pre-meeting Responses to Charge Questions—Dale Hattis

Characterization of Air Quality

My charge question#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?”

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 relatively “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 98th and 99th 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

* 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

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.

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 at 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.

parameters, Markov Chain Monte Carlo Modeling has excellent capabilities. Unfortunately those capabilities were not used in this case.

- 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 Deviations--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 Resistance 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 estimate 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. Patrick Kinney

Comments of Pat Kinney on Exposure Charge Question 2.

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

Comments by Tim Larson on the Second Draft of the SO₂ REA

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 SO₂ 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 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?

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.

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.

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 Aermoc algorithms is low and the direction of bias is unknown. However, the Aermoc-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.

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 SO_x 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

Comments on the 2nd draft REA for SO₂

Christian Seigneur
Cerea, Université Paris-Est

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 SO₂ and accordingly have been subjected to emission control regulations. Recently, SO₂ emissions from ships have become of concern and, in some areas, may be the major cause of significant SO₂ 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 SO₂ emissions in the area (about 3%). Such emissions may constitute a larger fraction of total SO₂ 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 SO₂ concentrations fairly well at the other eleven monitors. This satisfactory performance is not unexpected as point source emissions dominate the SO₂ 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 SO₂ exposure among various areas (see comments on ship-related emissions above) is warranted.

Comments from Dr. "Lianne" Elizabeth Sheppard

Lianne Sheppard preliminary comments

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

In my preliminary comments I focus primarily on air quality and exposure characterization. In my final comments I will append additional detailed comments by page to address details that can be improved.

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 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), 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 SO₂ sources and the key features of its temporal and spatial variation.
- 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 many 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.
- There should be at least a brief discussion of the reactivity of SO₂ 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 SO₂ concentrations in the U.S.

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.

Q2. This was clear and convincing.

Characterization of Exposure (Chapters 6 and 8):

Q1. The beginning of this chapter 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 SO₂, 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, 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. 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 a more rural with a large urban area. The areas screened as likely to have elevated SO₂ 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 SO₂ 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 SO₂ where 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 SO₂ monitoring data. For instance, locations near major ports may have very high exposures.

Characterization of Health Risks (Chapters 7, 8, 9): To be added.

Policy Assessment (Chapter 10): To be added.

Specific comments: To be added.

Comments from Dr. Frank Speizer

Pre-meeting Comments on REA Draft 2 for SO₂ and Charge Questions; dated March 2009

Submitted by Frank E. 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 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?

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 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?

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 4th 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 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?

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 (concurrent 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 be 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 $86.4 \times 3(\text{years}) = 259$ 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 SO₂ 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 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?

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 99th 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 SO₂ 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

RE: Risk and Exposure Assessment to Support the Review of the SO₂ Primary National Ambient Air Quality Standard: Second Draft

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 pre-meeting comments address all Charge Questions, as appropriate, but focus primarily on my assigned Charge Question regarding Characterization of Health Risks (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 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?

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 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?

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 SO₂, 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 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?

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 a 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 5minute 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 5minute 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 713, 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 SO_x 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 been easily accomplished (and can still be accomplished in the remaining time available), and would greatly improve 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 SO_x and other pollutants. Because the controlled human exposure studies used in the risk assessment involved only SO₂ exposures, it was assumed that estimates of SO₂-induced health responses would not be affected by the presence of other pollutants (e.g., PM_{2.5}, O₃, NO₂).*

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 SO₂ 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 SO₂."

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 SO_x 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 also be completed as a part of this and all future REA documents.

In addition, 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 SO₂ 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 SO₂ 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 SO₂ 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.

Finally, 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. Ronald Wyzga

Comments on the
Second Draft Risk and Exposure Assessment to Support the Review
of the SO₂ Primary National Ambient Air Quality Standard

Ronald E. Wyzga, Sc. D.

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?

More information would be useful in developing dose-response functions and in relating the human clinical evidence to the epidemiological evidence. For example, on page 26, it is noted that individual data from the Linn et al. (1987) study were obtained from Smith, 1994 and used to construct a dose-response curve. It would be informative to indicate the response of subjects to exercise alone (0 SO₂ exposure). For example, it is shown that 3/40 subjects had a decrease in FEV₁ of 20% or more for round 1 of exposures. This would illuminate how “adjustments for effects of exercise in clean air” were undertaken as well as provide some perspective about the severity of impacts.

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 alternative 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, but if one plots percent of excess risk 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.

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?
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.

Characterization of Health Risks

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 formal definition. It would be important to see whether the conclusions drawn would be robust given alternative dose-response functions. A paper by Shamoo et al. (1994) (Journal of Exposure Assessment and Environmental Epidemiology, 4(2):133-148) reports on the time activity patterns and diurnal variation of respiratory status in a group of asthmatics. The paper suggests that it is important to understand these in order to understand the health risks from pollution exposure. 49 asthmatics subjects 18-50 years old were studied in the Los Angeles area. Interestingly diurnal variation in FEV1 demonstrated average changes of 7%, 12% or 18% depending upon whether the asthmatics were mild, moderate, or severe.

This type of result argues that joint consideration of lung function changes and symptoms should be considered.

The Shamoo et al. paper also provides some data that should be considered estimating the risks of response. For those sample (n = 49), ventilation rates were estimated for men and women at different levels of activity. Outdoor fast activity was reported for 0.2% of waking hours and medium activity was reported for 2% of waking hours. These estimates should be compared with those used in the risk assessment and implications of differences should be discussed.

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; alternative estimates, such as those, obtained by Shamoo et al. 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.

Policy Assessment

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 24-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.