UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

DATE

EPA-CASAC-13-XXX

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

Subject: CASAC Review of the EPA's *Health Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)* and *Welfare Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)*

Dear Administrator Jackson:

The Clean Air Scientific Advisory Committee (CASAC) Ozone Review Panel met on September 12-13, 2012, to peer review the EPA's *Health Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)* and *Welfare Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)*, hereafter referred to respectively as the Health REA (HREA) and Welfare REA (WREA). The CASAC's consensus responses to the agency's charge questions on both documents, the individual review comments from the CASAC Ozone Review Panel, and the CASAC Ozone Review Panel's prioritization of additional analyses under consideration by the EPA for inclusion in the Second Draft REA documents are enclosed. The CASAC's key points on both documents are highlighted below.

The draft HREA and WREA documents are incomplete works in progress. Advice is provided on which analyses to perform and which directions to take as these documents are revised. However, the CASAC will need adequate opportunity to review the final products.

Health REA

The HREA does not present an adequate conceptual model or a framework. The conceptual steps in an exposure and risk assessment need to be presented, identifying the important elements and types of methods/tools that are used. A clear description and definition of ozone background should be included

in this conceptual model, which would describe how background ozone is considered in the risk estimates.

Risk estimates can be developed with reference either to the lowest measured level (LML) of epidemiologic studies or to zero ozone concentration. The advantage of developing risk estimates relative to the LML is that the concentration-response (C-R) functions are not extrapolated beyond the range of data from which they were developed. However, there is little difference between risk results based on the LML versus those based on zero ozone concentrations. Furthermore, the actual LML is not known for many of the epidemiologic studies. The advantage of not including both is that lengthy exposition of methods and surrogate measures of LML, and repetition in providing two sets of results, could be avoided. However, for completeness, EPA should compare results based on LMLs and zero ozone concentrations for a few cases, to quantify to what extent results are sensitive to this choice. An approach based on zero ozone concentration can be used for all cases.

Regarding the methods for simulating just meeting the ozone standard, the CASAC prefers a model-based approach over the quadratic rollback approach due to its significant limitations. The CASAC supports the use of the Higher-order Direct Decoupled Method (HDDM). The HDDM has the advantage of providing a protocol for the minimum emission reductions needed to achieve just meeting the ozone standard, and it does not require estimates of "background" ozone.

The overall methods used to conduct the exposure assessment are sound and seem to make the best use of available data. However, the description of the exposure assessment methods would be improved substantially with greater clarity and specificity such as by providing specific information on Air Pollution Exposure (APEX) model performance and validation, and discussion of the relevance and representativeness of the Consolidated Human Activity Database (CHAD) and other APEX data inputs.

The description of exposure analysis results can be enhanced substantially with clearer and more organized presentation and additional interpretation, and discussion. The latter should include implications for later analyses. The planned uncertainty and sensitivity analyses are appropriately focused on most of the key uncertainties of the exposure modeling process. Inclusion of additional contributors to uncertainty, such as ozone averting behavior and unaccounted for variability, should also be considered. A description of the methods used to conduct the uncertainty analysis should be included.

EPA has appropriately considered use of the 2010 McDonnell et al. model in the risk analysis. However, the threshold model described in the 2012 McDonnell et al. study would provide more accurate risk estimates for ozone-induced forced expiratory volume in one second (FEV_1) effects. EPA should obtain the necessary clincal data from the study authors to adequately conduct the risk analysis and should discuss the strengths and limitations of the model.

Although FEV_1 currently appears to be the best respiratory response indicator to be used in the risk assessment, careful consideration should be given to other respiratory responses. A strong and clear rationale should be given for why FEV_1 is considered as the principal metric of response rather than other repiratory health responses reported in the controlled human exposure studies.

The epidemiological studies and the corresponding C-R functions that were selected are appropriate. The use of multi-city studies, where available, for the health endpoints selected is justified. The discussion of uncertainty and variability is rigorous; however, exposure measurement error could also be a source of variability in effect estimates and should be included in the discussion on sources of variability in effect estimates. The presentation of results can be improved by reducing repetition, and prioritizing and consolidating results.

The synthesis chapter is currently a summary of the other chapters. Summaries should be given at the end of each chapter. The synthesis chapter should focus on the most critical findings from the prior chapters, followed by interpretation, comparison, integration, and implications. The most critical findings and insights from the synthesis should be provided in figures, tables, or both.

Welfare REA

The current scope of the WREA can be covered more succinctly but should be expanded to include much more attention to crop yield losses. More attention needs to be given to: (1) effects of ozone on competition among ozone sensitive species and less sensitive species in mixed-species stands; (2) scaling from tree seedlings to mature trees; (3) assessing impacts for individual sensitive species; and (4) assessing impacts for regions with different degrees of ozone exposure, in additional to overall national assessments.

It is not clear why exceptional events are excluded from the air quality data in Chapter 4, as they would have a welfare impact, even if they cannot be controlled. The overall approach to developing a national scale surface of W126 is suitable. The quadratic rollback method should not be used for the WREA. Given that a more comprehensive approach (Community Multi-scale Air Quality with HDDM) is being used in the Health REA, its use should be examined for application for the WREA. The HDDM sensitivities could be used to reduce ozone at specific monitors, which are then applied in the Modeled Attainment Test Software (MATS).

The WREA adequately presents the ecological effects and the broad range of ecosystem services that are potentially impacted by ozone. Consideration should be given to combining the chapters on ecological effects and ecosystem services, so that the underlying ecological impacts can be more directly linked to the services they provide. This could allow for deeper understanding of these linkages, a more brief and focused presentation, and a clearer indication of which impacts and associated services may be amenable to quantitative or qualitative risk assessment.

Crop yield loss needs greater attention. As summarized in the ISA, there is very strong scientific evidence that ambient ozone exposure in many locations and years is causing yield loss of sensitive annual and perennial crops. It is clear that production of food, feed and fiber is an important provisioning service whose impairment presents clear evidence of an adverse welfare impact. Effects on economically important, ozone-sensitive crops such as soybean, cotton and others, should be analyzed at the same level of detail as are individual tree species in the current draft. This should include projection of crop yield loss estimates for individual crops in specific production areas under current and

alternative projected ozone scenarios. Reduced yields on a regional basis could constitute an adverse welfare effect for affected growers, while reduced yields in the United States would decrease global competitiveness of U.S. agriculture relative to farmers in production areas subject to lower ozone concentrations.

The chapter on synthesis is more of a summary of results than a true synthesis of the WREA document. These summaries should be placed at the end of each chapter. This change would result in a consistent organizational structure common to both the WREA and HREA documents. The content of this synthesis chapter should also make a logical progression to the *Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards* (PA), so it will be useful to keep the goals of the PA in mind when developing the synthesis. It would help to integrate each of the Risk sections (Biomass Loss, Foliar Injury, and Ecosystem Services) to provide a comprehensive and synthetic analysis of the complete welfare impact and associated risks. The order of presentation here is also important: a progression from foliar injury risks, to biomass, and then ecosystem risks moves logically up an organizational scale from the best characterized risks with the largest datasets to the least characterized risks with highly complex interactions, fewer datasets, and qualitative risk factors. Expanding the discussion of qualitative risk factors would strengthen this chapter.

The CASAC appreciates the opportunity to provide advice on the HREA and WREA documents and looks forward to reviewing the revised draft documents.

Sincerely,

Dr. H. Christopher Frey, Chair
Clean Air Scientific Advisory Committee

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

Enclosures

NOTICE

This report has been written as part of the activities of the EPA's Clean Air Scientific Advisory Committee (CASAC), a federal advisory committee independently chartered to provide extramural scientific information and advice to the Administrator and other officials of the EPA. The CASAC provides balanced, expert assessment of scientific matters related to issues and problems facing the agency. This report has not been reviewed for approval by the agency and, hence, the contents of this report do not necessarily represent the views and policies of the EPA, nor of other agencies within the Executive Branch of the federal government. In addition, any mention of trade names or commercial products does not constitute a recommendation for use. The CASAC reports are posted on the EPA website at: http://www.epa.gov/casac.

U.S. Environmental Protection Agency
Clean Air Scientific Advisory Committee

CHAIR

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

MEMBERS

Mr. George A. Allen, Senior Scientist, Northeast States for Coordinated Air Use Management (NESCAUM), Boston, MA

Dr. Ana Diez-Roux, Professor of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI

Dr. Jack Harkema, Professor, Department of Pathobiology, College of Veterinary Medicine, Michigan State University, East Lansing, MI

Dr. Helen Suh, Associate Professor, Bouve School of Health Sciences, Northeastern University, Boston, MA

Dr. Kathleen Weathers, Senior Scientist, Cary Institute of Ecosystem Studies, Millbrook, NY

Dr. Ronald Wyzga, Technical Executive, Electric Power Research Institute, Palo Alto, CA

SCIENCE ADVISORY BOARD STAFF

Dr. Holly Stallworth, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board (1400R), 1200 Pennsylvania Avenue, NW, Washington, DC 20460

U.S. Environmental Protection Agency
Clean Air Scientific Advisory Committee
Ozone Review Panel

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

IMMEDIATE PAST CHAIR
Dr. Jonathan M. Samet, Professor and Flora L. Thornton Chair, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA

CASAC MEMBERS
Mr. George A. Allen, Senior Scientist, Northeast States for Coordinated Air Use Management (NESCAUM), Boston, MA

Dr. Ana Diez-Roux, Professor of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI

Dr. Jack Harkema, Professor, Department of Pathobiology, College of Veterinary Medicine, Michigan State University, East Lansing, MI

Dr. Helen Suh, Associate Professor, Bouve School of Health Sciences, Northeastern University, Boston, MA

Dr. Kathleen Weathers, Senior Scientist, Cary Institute of Ecosystem Studies, Millbrook, NY

Dr. Ronald Wyzga, Technical Executive, Electric Power Research Institute, Palo Alto, CA

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

Dr. Michelle Bell, Professor, School of Forestry and Environmental Studies, Yale University, New Haven, CT

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

Dr. David Chock, Independent Consultant, Bloomfield Hills, MI

Dr. William Michael Foster, Professor, Pulmonary and Critical Care Medicine, Duke University Medical Center, Durham, NC

Dr. David A. Grantz, Department of Botany and Plant Sciences, University of California at Riverside, Kearney Agricultural Center, Parlier, CA

Dr. Daniel Jacob, Professor, Atmospheric Sciences, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

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

Dr. Frederick J. Miller, Independent Consultant, Cary, NC

Dr. Howard Neufeld, Professor, Department of Biology, Appalachian State University, Boone, NC

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

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

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

Dr. Peter Woodbury, Senior Research Associate, Department of Crop and Soil Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, U.S.A.

* Immediate Past CASAC Member

SCIENCE ADVISORY BOARD STAFF

Dr. Holly Stallworth, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board (1400R), 1200 Pennsylvania Avenue, NW, Washington, DC 20460

Mr. Aaron Yeow, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board (1400R), 1200 Pennsylvania Avenue, NW, Washington, DC, 20460

Consensus Responses to Charge Questions on EPA's Health Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)

HREA Chapter 1: Introduction

1. Does the Panel find the introductory and background material, including that pertaining to previous reviews of the ozone standards and the current review, to be clearly communicated and appropriately characterized?

The Introduction is generally clear in its presentation of the background and introductory material, though there are concerns about how some of the topics are presented, particularly concerning the events preceding the 2008 promulgation, and the following reconsideration. For example, the text should be particularly clear on the range being considered for the standard. In particular, when the document states "within the proposed range of 0.070 to 0.075 ppm," it should state the source of the proposed range. Further, the text on current levels of the standard should include the fact that the secondary standard is the same as the primary standard and that the Administrator did not follow the advice of CASAC to have a different index for the secondary standard. The section should also be more comprehensive in its discussion of the CASAC deliberations and communications. It is important that this section be accurate in its representation of the history.

HREA Chapter 2: Conceptual Model / Framework

2. To what extent does the Panel find that the discussions accurately reflect and clearly communicate the currently available health effects evidence, and the relevance of that evidence for quantitative exposure and risk assessment, as characterized in the 3rd Draft ISA?

Chapter 2, entitled "Conceptual Framework" (note: in the index, this chapter is called "Conceptual Model") is not really a conceptual model, nor a framework. Although it provides information on the first few topics (e.g., chemistry, exposure, endpoints) that are important to developing a conceptual model for an ozone risk assessment, it does not provide a working conceptual model. Further, there is no framework presented. This chapter could benefit from figures that describe the conceptual model and lead the reader through the framework. The chapter needs to present the conceptual steps in an exposure and risk assessment, identifying the important elements and types of methods/tools that are used.

This chapter does a good job of presenting information on the sources of ozone, the various microenvironments to consider for exposure, at-risk populations, and which health endpoints are most suitable for inclusion in a risk assessment.

HREA Chapter 3: Scope

3. Does the Panel find the scope of the health risk and exposure analysis is clearly communicated?

This chapter does a good job of laying out the scope of the health risk and exposure analyses that have been conducted in the HREA. Further, it discusses most of the options, caveats, and considerations that are proposed for the 2nd draft. However, while it omits mention of a long term, national mortality assessment, such an assessment is planned. A table of the analyses conducted for the current HREA as compared to those of the 2008 HREA should be added.

4. Based on information in the 3rd draft ISA indicating lack of evidence for a threshold in ozone concentration-response functions, we have included risk estimates down to zero ozone concentrations. Based on further discussion in the ISA regarding the decreased confidence in the shape and magnitude of population health response at very low ozone concentrations, we have also included risk estimates based on applying concentration-response (C-R) functions only down to the lowest measured level (LML) in the underlying epidemiology studies.

a) To what extent does the Panel support the use of two different risk estimates, one applying the C-R function down to zero, and one applying the C-R function down to the LML, to characterize the range of risk estimates to balance comprehensiveness of the estimates with confidence in the estimates?

It is valuable to include risk assessment results down to ozone concentrations of zero. There was little difference between risk results down to the LML and risk results down to zero ozone concentrations. By not including risk estimates down to the LML, lengthy exposition of methods and surrogate measures of LML could be avoided as well as the repetition entailed in presenting two sets of parallel results. However, by presenting risk results down to the LML, risk is not being estimated beyond the range of data used in the health effects models and there is more confidence in C-R functions when applied to the actual data range from which the C-R functions were generated. However, for completeness, EPA should compare LML and zero concentration based results for a few cases, to quantify to what extent results are sensitive to this choice. An approach based on zero concentration can be used for all cases.

b) What are the views of the Panel on alterative cutoffs based on other points within the distribution of ozone concentrations used in the underlying epidemiology studies?

The CASAC has no clear recommendations on alternative cutoffs, such as background ozone concentrations. Including risk estimates down to ozone background concentrations is useful because this is the range that can be potentially modified by implementation of the ozone National Ambient Air Quality Standards (NAAQS). However, this is an implementation issue and not a scientific/technical issue. Additionally, background ozone is already incorporated in estimates of change in risk from existing levels to levels that meet the current standard, and is planned to be incorporated in estimates in the next draft involving change in risk for several alternative levels of the standard.

HREA Chapter 4: Air Quality Considerations

5. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The two overlapping 3-year periods of 2006-2008 and 2008-2010 are appropriate. Although it is desirable to use the most recent 3-year data period, 2009 and 2010 were relatively clean ozone summers in the eastern U.S., making 2008-2010 a period that may not represent more recent ozone trends. It is difficult to designate which years are more or less representative than others because both emissions and climate continue to evolve. However, 2006-2008 is relevant in that it is the most recent 3-year period where GEOS-Chem model run data for background ozone data are available - an important component of the REA process.

6. Regarding the methods for simulating just meeting the ozone standard:

a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard (including application of US background as a lower-bound on rollback) is a reasonable approach?

With the lower-bound constraint, the quadratic rollback is a simple observational approach that has been used, but has significant limitations inherent in its simplicity and lack of consideration of nonlinear ozone chemistry. Many of its shortcomings are described in the EPA's Higher-order Direct Decoupled Method (HDDM) rollback technical memo (Simon et al., 2012). Given that there is now a more science-based rollback approach, it is time to retire the quadratic rollback scheme.

b) To what extent does the Panel support using an air quality model based approach for simulating just meeting the standard in future drafts as a replacement for the current quadratic approach?

A model-based approach is preferred because that is in practice how state implementation plans (SIPs) are developed to meet an air quality objective. However, there are different combinations of emission reduction strategies possible to achieve just meeting the ozone standard and these may have different impacts on overall ozone distributions. Thus, a protocol is needed for choosing one combination in particular.

c) What are the views of the Panel on the strengths and limitations of the proposed approach using the Higher-order Direct Decoupled Method?

HDDM would have the advantage of providing a protocol for the minimum emission reductions needed to achieve just meeting the ozone standard, and it does not require estimates of "background" ozone. The analysis done by Simon et al. (2012) is well done, though the figures (e.g., Figures 8 and 9) and tables (e.g., Table 1) should be more explicit as to what emissions are being changed. Sensitivities need to be expressed in terms of emissions from particular sectors so that the effects of these emission reductions on ozone can be propagated to the whole domain. The HDDM approach is self-consistent

only if the sensitivities are applied to the modeled concentration fields. However, past practice has been to use the ratio of observed-to-modeled concentrations as a relative response factor to scale the modeled concentrations. By judicially selecting a set of intermediate-emission scenarios as a "brute-force" simulation base case, the HDDM may be replaced by the simpler first-order direct decoupled method (DDM) while retaining the boundary conditions of the "brute force" simulation without incurring significant errors in determining the "just meeting the standard" concentration distribution. There was some discussion of performing a "classic" quadratic rollback for one urban area as a basis for comparison between the old and new rollback approaches.

This chapter also briefly discusses the use of a hierarchical Bayesian model to fuse the 2006-2008 monitored ozone data with the Community Multi-scale Air Quality (CMAQ) model-predicted ozone concentrations to create a spatially-resolved and time-dependent ozone concentration field nationwide. The spatial distribution of the fused ozone field is highly sensitive to the variances assumed in the prior distributions for the measurement errors of the monitored and model-predicted ozone concentrations. The basis of these prior distributions and their implications should be discussed.

HREA Chapter 5: Characterization of Population Exposure

7. To what extent does the Panel find the methods used to conduct the exposure analysis technically sound? Does the Panel have any recommendations on the methods used?

The overall methods used to conduct the exposure assessment are sound and seem to make the best use of available data. However, the description of the exposure assessment methods would be improved substantially with greater clarity and specificity.

In this regard, five specific areas need increased clarity and specificity. First, Section 5.1 should include a figure showing the various models, inputs, outputs, cities and years for the population exposure assessment. Such a figure would provide a clear, concise overview of the modeling approach and provide context for the rest of the chapter. Second, the methods used to perform the uncertainty and longitudinal exposure analyses should be described, with their objectives and their incorporation into other components of the REA also discussed. Third, given the Air Pollutants Exposure Model's (APEX's) central role in the lung function risk assessment, specific information on model performance and validation should be provided as possible, for example for estimated exertion levels. Fourth, a description of the relevance and representativeness of the Consolidated Human Activity Database (CHAD), and other APEX data inputs, should be included specifically with regard to their: (1) age;, (2) relevance to children, older adults, and outdoor workers; and (3) relevance to the 16 cities to be modeled. For CHAD, concerns may be addressed in part by comparing the distributions of percent time spent outdoors for children and other age groups for different time periods (for example 1980s, 1990s, and 2000+s). If available, inclusion of information from more recent time-activity or exertion databases would be helpful as well. An example of such a database is Brochu et al. (2006), which could be either added to the APEX database or used to validate the model outputs. Finally, more precise word choice and labeling of figures and tables would substantially improve the clarity of the chapter.

8. To what extent does the Panel find the assessment, interpretation, and presentation of the results of the exposure analysis as presented technically sound, appropriately balanced, and clearly communicated?

The presentation and interpretation of results from the exposure analysis is a good start. The presentation and interpretation of results can be enhanced substantially with a clearer and more organized presentation of the results and with the addition of text that interprets and discusses how these results relate and impact later analyses.

Specifically, the presentation of results would be improved with: (1) the addition of a written summary of key patterns observed in Figures 5-1 to 5-15 and Tables 4-5 to 4-19; (2) more specific referencing and labeling of the above figures and tables; and (3) inclusion of results that show the distribution of exposure estimates for each city (in addition to current presentation of the percent of people with exposures above benchmark values). Further, although generally well-written and presented, Section 5.6.1 would be substantially improved with the addition of a sentence clarifying that the findings are based on simulated exposures and an additional paragraph explaining whether and how these findings will be incorporated into the planned uncertainty or health analyses.

9. Regarding the characterization of uncertainties and variability:

a) To what extent does the Panel find that the uncertainties associated with the exposure analysis are clearly and appropriately characterized?

b) To what extent does the Panel find that the uncertainty assessment is technically sound? Are there other important uncertainties which are not covered?

10. What are the views of the Panel on the sensitivity analyses that EPA plans to conduct as part of the second draft REA to evaluate the influence of uncertainties in the exposure analysis?

The planned uncertainty and sensitivity analyses are appropriately focused on most of the key uncertainties of the exposure modeling process. Additional contributors to uncertainty, such as ozone averting behavior and unaccounted for variability, should also be addressed.

A description of the methods to be used to conduct the uncertainty analysis was not provided. There should be a clear description of the methods used to assess uncertainty, even if the analysis is qualitative in nature. Possible approaches to consider include the use of city-specific lower and upper bound estimates of key model input parameters to reflect uncertainty.

HREA Chapter 6: Characterization of Health Risk Based on Controlled Human Exposure Studies

1. To what extent does the Panel find the methods used to conduct the risk analysis to be technically sound? What are the views of the Panel members on the methods used?

EPA has appropriately considered use of the McDonnell et al. (2010) model in addition to the probabilistic method that is currently being used. Use of the McDonnell model for individual response functions is more likely to capture what is really going on in the population compared to other previous models. However, it would be better to fit controlled human exposure/response data to the new model described in McDonnell et al. (2012). This model contains a threshold below which no response occurs. The McDonnell et al. (2012) study states that this new threshold model is likely to provide more accurate estimates of risk in future risk assessments of ozone-induced FEV_1 effects.

The EPA should obtain the necessary clinical data from the study authors to adequately conduct the risk analysis using this threshold model. Papers by McDonnell et al. (2012) and Schelegle et al. (2012) should be appropriately documented in the ISA and referenced in Chapter 6 of the Health REA. The strengths and limitations of the threshold model (McDonnell et al., 2012) with respect to the other models discussed in the chapter should be carefully documented.

When considering the application of the model to predict risk of a health effect(s) (e.g., a decrement in a spirometric outcome index), differences in reported susceptibility between ethnic groups (see Third Draft ISA, p. 6-22) as well as influence(s) of obesity upon susceptibility to develop alteration in spirometric indices (see 3rd draft of ISA, p. 6-24) should be taken into consideration. In addition, when considering the data from controlled exposure studies to be used for the risk analyses (e.g., Schelegle et al., 2009), the EPA should carefully consider the dropout percentage of participants prior to completion of the study, and the reasons underlying unusually high dropout rates. Finally, the EPA should clearly describe how spirometric data will be collated from studies with different *square wave* and *variable* concentrations, with different induced minute ventilations due to varying exercise regimens (some studies were done under extremely elevated minute ventilations), or with different exposure durations (e.g., 8 hours versus 6.6 hours).

A stronger and clearer rationale needs to be provided for why FEV_1 should be used as the principal metric of response rather than other respiratory health responses reported in the controlled human exposure studies.

2. To what extent does the Panel find the assessment, interpretation, and presentation of the results of the risk analysis as presented in Chapter 6 to be technically sound, appropriately balanced, and clearly communicated?

Overall this first draft is appropriately organized, documented, and clearly written. A flow chart or diagram outlining the methodological approach and what key data will be input into the analysis would be helpful. There should also be a clear explanation of how background levels of ozone will be handled,

and if and how FEV_1 changes in filtered air will be used in the computation of the percent change in FEV_1.

3. To what extent does the Panel find the focus of the assessment on lung function decrements in the quantitative risk assessment to be appropriate and informative?

Although FEV_1 currently appears to be the best respiratory response data set to be used in the risk assessment, careful consideration should be given to other respiratory responses. As mentioned above, there needs to be a strong and clear rationale for why FEV_1 should be used as the principal metric of response rather than other repiratory health responses reported in the controlled human exposure studies.

FEV_1 may not be the most important response to ozone (e.g. inflammation may be equally or more important), but it is the response that is easiest to measure. It may be useful to include a statement at the end of section 6.1.1 that the multiple phenotypes elicited by ozone exposure likely have different mechanistic underpinnings and do not necessarily correlate or co-segregate. Therefore, risk assessments may be different depending on which phenotype is used in the assessment.

4. What are the views of the Panel on the use of the two different modeling approaches for specifying the exposure-response function linking the change in FEV_1 to ozone exposure?

Please see the responses above.

5. What are the views of the Panel on the treatment of the relationship between age and $dFEV_1$ in the McDonnell-Stewart-Smith model?

The treatment of age and FEV_1 response appears appropriate for both the McDonnell-Stewart-Smith model as well as the recommended McDonnell et al. (2012) threshold model.

Risk estimates will vary among the models since they use different approaches. It is important to understand why the differences exist. There needs to be an explanation of the model-specific differences compared to plausible alternative models and why they occur.

6. To what extent does the Panel find that the qualitative discussion of uncertainty and variability has covered important sources of uncertainty and variability and has appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates?

The HREA identifies crucial sources of uncertainty and variability, but the discussion should extend beyond the qualitative to more quantitative when appropriate.

7. What are the views of the Panel on additional sensitivity analyses or other approaches to addressing uncertainty and variability?

It would be useful to provide an explanation of any differences that are found.

HREA Chapter 7: Characterization of Health Risk Based on Epidemiological Studies

11. Regarding the epidemiologic studies used in the analysis:

a) What are the Panel's views on the set of epidemiological studies selected for use in specifying C-R functions and on the set of C-R functions specified for use in the risk assessment?

The epidemiological studies and the corresponding concentration-response (C-R) functions that were selected are appropriate. The use of multi-city studies, where available, for the health endpoints is justified. The use of Bayesian shrunken estimates is also justified.

b) To what extent does the Panel find the detailed descriptions of rationales for the selection of the epidemiological studies and the selection of the set of C-R functions specified using those studies to be appropriate and complete?

The descriptions of rationales for selection of studies and C-R functions are clear.

12. To what extent does the Panel find that the qualitative discussion of uncertainty and variability have covered important sources and appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates?

The discussion of uncertainty and variability is rigorous and thorough. However, a great source of uncertainty in ozone epidemiology studies, especially those that involve sensitive and susceptible populations (who are susceptible to ozone mortality), is exposure measurement error. Although this source of uncertainty is qualitatively different from the sources itemized in this draft of the risk assessment, some attempt should nevertheless be made to address the impact of measurement error on the C-R functions. Exposure measurement error could also be a source of variability in effect estimates and should be included in the discussion on sources of variability in effect estimates.

13. Regarding the results of the risk analysis:

a) What are the views of the Panel on the presentation and discussion of risk estimates, including the key observations presented in section 7.6.2?

- There is a great deal of repetition in this draft. This results partly from parallel presentation and description of findings from analyses of risk assuming existing concentrations and risk on just meeting the current standard, something that will be exacerbated as the number of parallel analyses increases in the next draft, and partly from the summaries that go into a level of detail not much different from the original presentation of results.

- More descriptive, simple and standard terminology should be used, especially in the tables and figures. For example, using “total incidence” or “total prevalence” or “mortality” is confusing when what is really meant are the total number of incident cases, total number of prevalent cases or

total number of deaths. The EPA should consider using percent attributable risk (or percent of attributable deaths) as opposed to "total risk," and percent reduction in risk or percent reduction in absolute number of cases as opposed to "risk delta."

- Better consolidation in displaying the results is recommended. Specifically, percent reductions and absolute reductions in numbers could be combined in the same table or figure. Consider showing total number of deaths attributable to ozone, percent of total deaths attributable to ozone, and percent and absolute reduction in number of deaths expected if a given standard were met. These data could be shown side by side in tables or figures for the different urban areas, and for all areas combined. The same applies to the morbidity findings.

- An attempt should be made to prioritize the presentation of results and the sensitivity analyses. This will become a more acute issue in the next draft that will include more results from more risk assessments as well as more sensitivity analyses.

b) What are the views of the Panel on the presentation of the distribution of ozone-related mortality across daily ozone levels for each city as "heat maps"?

The CASAC is generally supportive of the "heat map" presentation concept; however the utility and clarity can be increased by including a more detailed explanation of how to interpret the heat maps, as well as more description of the heat map findings. An expanded color scale to present the heat map findings should be considered.

14. To what extent does the Panel agree with the characterization of overall confidence, including the degree to which the conclusions reached regarding overall confidence are supported by available information?

The CASAC agrees with the characterization of overall confidence.

15. What are the views of the Panel on EPA's discussion of potential refinements to the REA for the second draft, including the plans for quantitative sensitivity analyses, additional refinements to the core risk estimates, and plans for assessment of long-term mortality and morbidity (i.e., plans to model risk for mortality and the decision not to model risk for morbidity endpoints, given data limitations)?

- Regarding the Second Draft HREA, the CASAC agrees with the proposed refinements to the core estimates and the plans for the quantitative risk analyses.

- <u>Long-term exposure and respiratory morbidity</u>. There is not a clear recommendation on whether to include long-term exposure to ozone and respiratory morbidity endpoints in the next draft of the HREA. Out of all the potential respiratory morbidity endpoints, only respiratory symptoms and new-onset asthma are identified in the ISA as endpoints that contribute to the likely casual determination of the relationship between long-term exposure and respiratory effects. Based on CASAC comments on the ISA regarding the evidence on new-onset asthma, a risk assessment

for new-onset asthma is not recommended. Further, a risk assessment for respiratory symptoms is unlikely to be informative. However, because there is a likely to be causal relationship between long-term ozone exposure and respiratory effects, there is an argument to be made in favor of carrying out this risk and exposure assessment for the next draft. Because effects on respiratory mortality are included as one of the "respiratory effects" being considered in this designation, limiting the risk and exposure assessment of long-term exposure and respiratory effects to respiratory mortality may be the most reasonable approach.

HREA Chapter 8: National Scale Risk Assessment and Representativeness Analysis

16. What are the views of the Panel on the overall approach used for the national scale risk analysis, including the ozone concentration methods and metrics, the use of city-specific and national average concentration-response relationships derived by Bell et al. (2004) and Zanobetti and Schwartz (2008)?

Although there are many detailed comments on Chapter 8, as enumerated in comments from individual panel members, the general approach seems reasonable. The chapter describes the national scale risk assessment but does not provide much detail from which the reader can fully understand how it was done. Furthermore, many of the terms, concepts, and quantities used are not clearly defined or used consistently.

This chapter would be benefit from better organization and clearer writing, as per detailed individual comments of the panel members. The Introduction seems to be an abstract, which is confusing because it is framed as if it is background material. The Introduction does not set up the whole chapter as well as it should. For example, it does not discuss the use of the Jerrett et al. (2009) effect estimates. The Introduction also does not discuss the results of the assessment of the county distribution, though it does discuss the results of the risk assessment using the short-term risk estimates. This chapter should also start out with the findings from the ISA that state that there is likely to be a causal relationship between short-term exposures to ozone and all-cause mortality. It should also present findings from the ISA that evidence is suggestive of a causal relationship between long-term ozone exposures and total mortality. The chapter should re-emphasize why the assessments are being performed and also highlight the strength of the findings from the ISA.

The overall rationale for using or not using lowest measured levels (LMLs) should be made clearer. The actual LML is study-specific but often is not reported. The EPA should develop risk estimates based on applying the concentration-response functions "down to zero." A limited number of sensitivity analyses can be done "down to the LML" and interpreted. This issue applies not just to Chapter 8 and should be implemented consistently throughout the document.

The REA appears to assume that "policy relevant background" concentration of ozone is 29 ppb. What is the basis for this concentration? The REA should be consistent with the ISA in terms of how background is defined, and should provide explanations for which of several possible definitions are used (e.g., U.S. background, North American background, etc.).

The REA should more clearly justify the use of national average concentration-response functions applied to locations that were not part of multi-city studies as opposed to, for example, using some type of regional assignment or inverse distance weighting approach. If it is not feasible to apply these latter types of approaches broadly, then either a quantitative sensitivity analysis (highly preferred) or a qualitative discussion of the implications and potential errors of using national averages should be further discussed.

With regard to whether a threshold-based model might be better, the lack of observed thresholds in the epidemiologic studies could be reflective of measurement error or a more realistic assessment of the effect of ozone exposure under actual ambient exposure conditions. At the current time, insufficient data is available on which to differentiate between these two possibilities.

The EPA should quantify the correlation between exposure metrics, such as the correlation between the May to September average 8-hour daily maximum concentration versus the June to August average 8-hour daily mean concentration. The correlation of each of these with the annual 4th highest maximum daily 8-hour average and with the three-year average of the annual 4th highest values would also be informative.

17. What are the views of the Panel on the approach identified for quantifying long-term mortality using the Jerrett et al. (2009) two-pollutant model national respiratory mortality effect estimate?

The brief description of the plan for the Second Draft REA on page 8-9 seems reasonable. The REA can and should be clearer as to the relationship between the short-term and long-term mortality concentration-response models that will be used – i.e. are they mutually exclusive or does one subsume the other? The document is not clear as to whether the focus will be the ozone season only or whether it will be based on year-round data. If the former, a clearer rationale should be given.

18. Regarding the representativeness analysis:

a) What are the views of the Panel on the methods and presentation of results for the representativeness analyses?

b) Does the Panel have suggestions for additional risk characteristics that would be useful to include in the analysis?

The representativeness analysis seems reasonable in terms of the scope of variables identified, the analysis of each, and the conclusions reached. The text should be better organized to more clearly state inferences and findings definitively, rather than offer a tentative finding and then contradict it later (e.g., should the number of urban areas be expanded to include higher mortality rates or older populations? – answer: no, as explained later in the same chapter).

HREA Chapter 9: Synthesis

19. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall health exposure and risk analysis?

Integration Versus Summary

This document, a work in progress, is a nice start, but could be improved. Currently, the chapter is not so much a synthesis, but a summary. Since an executive summary will be provided separately, this chapter should focus more on integrating key points from the other chapters. Rather than presentation of findings, the synthesis should focus on how key findings relate to each other (interpretation, implications, and integration) and their importance.

Rationale for Key Concepts

Many of the concepts need further explanation. In particular, the policy relevance of risk estimates for ozone concentrations at zero needs to be better described. Although each chapter is not written as a stand-alone document, this concept will be very confusing to readers who are not very familiar with the earlier chapters. The rationale could be briefly provided in the synthesis.

Clarity of Language

There are several places where the language is unclear and confusing throughout this chapter. In fact, there are a few places where the text may not imply the intended meaning or an accurate meaning. As a few examples, two studies having 3 of the same 4 cities with the highest impact does not seem to be a "significant" difference. Care should be taken with the use of the word "significant," as it has statistical implications. The language that health endpoints "remain" (see page 9-8), is odd and potentially confusing. The text of "20 to approximately 930 deaths" is potentially misleading with respect to uncertainty. This text requires context that results are for the various cities, which, of course, will have different numbers of deaths because of population size. Results described as having "considerable variation" (see page 9-8) actually appear quite similar. The statement that the urban study areas provide a good representation of the overall distribution of risk (see page 9-9) is unclear as risks may differ in rural environments. These are provided as examples; the EPA should review the entire chapter for this issue. In addition, the EPA should consider referring back to earlier tables and chapter sections to help clarify statements.

Presentation of Findings

The chapter includes a useful summary of key findings from risk assessments, including the exposure and controlled human experiment-based assessment, the short-term epidemiological assessment for individual cities, and the national scale epidemiologic-based assessment. However, in its current format, the key points could be a bit difficult to discern because of the use of text. In the next version, The EPA should consider providing key findings in a figure or table. This should not include every finding, but

just the most important points. The current version seems to list findings, but may not focus on those that are central to the EPA's findings.

Additional Analyses

The chapter briefly mentions sensitivity analyses (co-pollutants, additional health endpoints, improved approach to adjust for ozone levels), but does not give enough information for the CASAC to assess them. This will likely receive more attention in the next version.

References Cited

Brochu, P., Ducré-Robitaille, J.-F., and Brodeur, J. (2006). Physiological daily inhalation rates for Free-Living individuals aged 2.6 months to 96 years based on doubly labeled water measurements: Comparison with Time-Activity-ventilation and metabolic energy conversion estimates. *Human and Ecological Risk Assessment: An International Journal*, 12(4):736-761.

McDonnell W.F., Stewart P.W., Smith M.V. (2010). Prediction of ozone-induced lung function responses in humans. *Inhalation Toxicology*, 22(2):160-168.

McDonnell, W. F., Stewart, P. W., Smith, M. V., Kim, C. S., and Schelegle, E. S. (2012). Prediction of lung function response for populations exposed to a wide range of ozone conditions. *Inhalation Toxicology*, 24(10):619-633.

Schelegle, E. S., Morales, C. A., Walby, W. F., Marion, S., and Allen, R. P. (2009). 6.6-Hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans. *American Journal of Respiratory and Critical Care Medicine*, 180(3):265-272.

Schelegle, E. S., Adams, W. C., Walby, W. F., and Marion, M. S. (2012). Modelling of individual subject ozone exposure response kinetics. *Inhalation Toxicology*, 24(7):401-415.

Consensus Responses to Charge Questions on EPA's Welfare Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)

WREA Chapter 1: Introduction

1. Does the Panel find the introductory and background material, including that pertaining to previous reviews of the ozone standards and the current review, to be clearly communicated and appropriately characterized?

The CASAC finds this chapter to be useful and recommends retaining it. However, there are advantages and disadvantages of having overlap among the related ozone NAAQS documents. There is a need for documents to be understandable as standalone documents. However, material summarizing prior documents should be brief and focused, with references to specific sections of prior documents.

WREA Chapter 2: Conceptual Model

2. To what extent does the Panel find that the materials accurately reflect and clearly communicate the currently available welfare effects evidence, and the relevance of that evidence for quantitative exposure and risk assessment, as characterized in the 3rd Draft ISA?

This chapter is useful and should be retained. However, it can be improved to briefly clarify the risk assessment approach, including the tools and approaches to be used and the methods for interpreting the results. Although there is a lot that is not known about ozone effects on vegetation, there is also a lot that is known, and it is critical to distinguish what is well-known from what is less well-known. The next draft should make such distinctions clearer.

WREA Chapter 3: Scope

3. To what extent does the Panel find the scope of the welfare risk and exposure assessment to be clearly communicated?

This chapter is useful and should be retained. However, the scope can be covered more succinctly, as is done in the draft *Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards* (PA). Additionally, the scope should be expanded to include much more attention to crop yield losses (for more detail, see responses to Questions 13 and 16). Also, more attention needs to be paid to: (1) effects of ozone on competition among ozone sensitive species and less sensitive species in mixed-species stands; (2) scaling from tree seedlings to mature trees; (3) assessing impacts for individual sensitive species; and (4) assessing impacts for regions with different degrees of ozone exposure, in additional to overall national assessments.

WREA Chapter 4: Air Quality Considerations

4. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The two overlapping 3-year periods of 2006-2008 and 2008-2010 are appropriate. While it is desirable to use the most recent 3-year data period, 2009 and 2010 were relatively clean ozone summers in the eastern US, making 2008-2010 a period that may not represent more recent ozone trends. The 2006-2008 time period is relevant in that it is the most recent 3-year period where GEOS-Chem model run data for background ozone data are available - an important component of the REA process.

It is not clear why exceptional events are excluded. They would have a welfare impact, even if they can't be controlled. The 2008 California wildfires (which accounted for all the exclusions according to the text) seem like a particularly odd exclusion since Singh et al. (2010) showed that these fires did not produce significant ozone, unless mixed with urban plumes.

5. What are the views of the Panel on the approach used to develop a national scale surface of W126?

The overall approach to develop a national scale surface of W126 seems fine. However, Figure 4-3 shows surprisingly high W126 values over the intermountain west. This must largely reflect Community Multi-scale Air Quality (CMAQ) information, because there are not many monitors there. However, the quality of the CMAQ simulation in that region is uncertain. For example, the W126 hotspot in central Idaho is due to wildfires, but that just reflects the simulation of large ozone production from wildfires in CMAQ. There is evidence from the literature that models overestimate ozone production from wildfires (Singh et al., 2010; Alvarado et al., 2010). There is a need to better establish the reliability of the fused data on the national scale.

6. Regarding the methods for simulating just meeting the ozone standard:

a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard is a reasonable approach?

A problem with the quadratic rollback method is that it does not allow for lower ozone levels to actually increase in response to controls. Only having the monitors that experience exceedences go down is problematic, with a specific example captured in the text (if two sites are close, and one is exceeding, only that monitor is reduced: page 4-7). The quadratic rollback method produces other inconsistent results, such as reducing ozone concentrations in central Idaho even though they are of wildfire origin. For these and other reasons, the panel considers that the quadratic rollback method should not be used for the Welfare REA.

b) Does the Panel have suggestions for alternative approaches for simulating just meeting the current secondary standard or alternative standards based on the W126 metric?

Given that a more comprehensive approach (CMAQ with HDDM) is being used in the Health REA, its use should be examined for application here. The HDDM sensitivities could be used to reduce ozone at the specific monitors, which are then applied in MATS.

WREA Chapter 5: Ecological Effects

7. Regarding the assessment of relative biomass loss (RBL) for individual species:

a) What are the views of the Panel on the use of the linear model forced through the origin to assess the proportional relationship between the relative biomass loss (RBL) values for each species comparing the RBL at recent ambient conditions to the RBL under the scenario modeling ozone just meeting the current standard?

Either a better justification of the linear approach is needed in the next draft, or the linear approach should be dropped in favor of the less demanding and less assumption-ridden approach based on the mean of the ratios.

b) To what extent does the Panel find that this an appropriate analysis to compare the proportional changes in RBL? Does the Panel have suggestions for alternative approaches that provides a comparable result and maintains the cell-by-cell approach to help control for environmental variability?

Some members find that comparison of proportional changes is appropriate, while others do not find the comparison appropriate. See the individual comments for specific suggestions. For the next version of the document, a regional analysis should be included.

c) To what extent does the panel agree with the approach used to combine the 11 tree species into one analysis?

Combining species gives an “average” view, which some members consider to be appropriate and necessary. However, there is also significant concern that the averaging obscures important information, especially in regard to protecting the most sensitive species.

8. Regarding the assessment of RBL for combined species:

a) To what extent does the Panel support the use of the Importance Values from the U.S. Forest Service to weight the RBL values in extrapolating from individual trees to larger ecosystem level effects?
Weighting the relative biomass loss (RBL) is a useful way to spatially extrapolate the data and to begin to get semi-quantitative indices of regional effects. It is an imperfect solution, but in the absence of more extensive data, it is a reasonable approach. However, please see the response to 8b below.

b) What are the views of the Panel on the use of the summed-RBL as a metric to use for assessing effects at the larger ecosystem scale?

There is the need to account for sensitive and insensitive species, especially for forest species that occur in mixed-species stands. Total stand response to ozone cannot be estimated using a weighted average RBL due to competition among sensitive and insensitive species. "The larger ecosystem scale" is not well defined—this should be rephrased to something more specific, such as effects across hundreds of kilometers.

c) Does the panel have any recommendations for methods to include a wider range of tree species (beyond the 8 species included)?

In the absence of data, the CASAC does not have any recommendations for methods to include a wider range of tree species. However, the Forest Health Monitoring and Forest Inventory and Analysis (FHM/FIA) data should be fully explored.

9. What are the views of the Panel on the use of federally designated Class I and Critical Habitat areas as endpoints for this analysis? Does the Panel have any suggestions for additional parks beyond Rocky Mountain National Park and Sequoia/Kings Canyon National Park that should be analyzed? Does the panel have recommendations for additional or alternative geographic analysis areas that could be used?

There is a mandate to protect the biological resources within these systems. However, because they are unique, heterogeneous and/or sensitive landscapes relative to the rest of the US, how will they be used as endpoints? More emphasis and analysis should be devoted to cropland and other vegetation. Because coastal sites along the eastern seaboard are subject to high ozone, Class I or Critical Habitat areas in these regions might also be considered.

10. To what extent does the Panel find that the vegetation mapping data from USGS and NPS used to generate a scaled-RBL surface in Great Smokey Mountain National Park is appropriate?

For these case studies, it is appropriate and useful to have finer spatial scale data for extrapolation and synthesis. However, see individual comments about overall issues with the scaled-RBL approach (e.g. Dr. Woodbury's comments).

11. Regarding the screening level foliar injury risk assessment:

a) To what extent does the Panel find the updated assessment of foliar injury risk in national parks originally performed in Kohut (2007) to be an appropriate screening level risk assessment?

The updated analysis is useful and appropriate, however, only W126 should be used in the analysis.

b) What are the views of the Panel regarding the potential methods for estimating ozone exposure at additional parks?

If this question is referring to estimating exposure at additional parks without any baseline monitoring of ozone, then spatial extrapolation/empirical-GIS modeling, and/or "data fusion" approaches are reasonable. Some parks have conducted passive sampler monitoring campaigns and some of these data could possibly be used, in combination with monitoring data, to estimate exposures.

c) What are the Panel's views regarding the appropriateness of requiring that two criteria must be satisfied (i.e., based on both W126 and N100) in order to receive higher risk ratings?

It would not be useful or appropriate to require satisfaction of two criteria to receive higher risk ratings.

d) Is the Panel aware of any assessments of foliar injury in national parks conducted between 2006 and 2010 that could potentially be used to validate the updated risk ratings?

The EPA should check with the park natural resource managers and FHM/FIA monitoring results.

12. Regarding the assessment of cover of ozone sensitive species:

a) To what extent does the Panel find the preliminary analysis of sensitive species cover to be an appropriate and useful approach to highlight areas of potentially higher risk due to the presence of sensitive species?

This topic is important and the analysis has some utility, but there are problems with the current analysis as discussed elsewhere.

b) To what extent does the Panel find the vegetation mapping data appropriate to assess the cover of ozone sensitive species in GSMNP?

It is a useful tool primarily because the spatial resolution is high.

c) What are the views of the Panel on the decision to not distinguish between vegetation strata (i.e. herb, shrub, tree)? To what extent does the Panel agree with this methodology relative to analyzing the strata individually?

Although both sensitivity and ozone exposure have been shown to vary with strata, it is likely to be unfeasible to examine and extrapolate by strata.

d) What are the views of the panel on using benchmarks, similar to those used in the Kohut analysis of foliar injury risk, to allow estimates of change between exposure scenarios?

These benchmarks may be useful to estimate change, although the analysis is semi-quantitative, at best.

WREA Chapter 6: Ecosystem Services

13. To what extent does the Panel find that EPA has adequately characterized the range of ecosystem services that are potentially adversely affected by ozone?

The use of the Millennium Ecosystem Assessment of 2009 is appropriate. The chapter adequately presents the broad range of ecosystem services that are potentially impacted by ozone. Consideration should be given to combining Chapters 5 and 6, so that the underlying ecological impacts can be more directly linked to the services they provide. This could allow for deeper understanding of these linkages, a briefer and more focused presentation, and a clearer indication of which impacts and associated services may be amenable to quantitative or qualitative risk assessment.

There is a serious omission from the chapter, in that much more attention is required for crop yield loss. As summarized in the ISA, there is very strong scientific evidence that ambient ozone exposure in many locations and years is causing yield loss of sensitive annual and perennial crops. It is clear that production of food, feed and fiber is an important provisioning service whose impairment presents clear evidence of an adverse welfare impact. Effects on economically important, ozone-sensitive crops such as soybean, cotton and others, should be analyzed at the same level of detail as are individual tree species in the current draft. This should include projection of crop yield loss estimates for individual crops in specific production areas under current and alternative projected ozone scenarios. Reduced yields on a regional basis could constitute an adverse welfare effect for affected growers, while reduced yields in the U.S. would decrease global competitiveness of U.S. agriculture relative to farmers in production areas subject to lower ozone concentrations. Anecdotal evidence that growers are currently observing increasing yields cannot be considered counter-evidence regarding negative impacts of ozone on yield. Modest yield losses are often not discernible without comparison with on-site controls which are not available to producers. It is likely that improved average yields reflect improved or additional management inputs which are costly to growers and the environment, and improved germplasm with greater tolerance of ozone which is costly to society to develop even if ozone tolerance is not an explicit objective.

14. To what extent does the Panel agree with EPA's ecosystem services framework, connecting ozone exposure, through ecological effects to ecosystem services?

The CASAC is strongly supportive of the use of an ecosystem services framework. A more direct linkage between ecological impacts and specific ecosystem services is desirable for the Second Draft WREA.

15. Does the panel agree with EPA's use of combined ozone exposure data with other data sources (e.g. fire data, bark beetle maps, trail maps) to link areas of concern or interest with areas of higher vegetative risk due to ozone? Does the Panel have any recommendations for additional datasets and ecosystem services that could add to or improve these analyses?

The CASAC has no suggestions of additional datasets to overlay with ozone concentration data to identify areas of specific concern. The CASAC is supportive of the use of the spatial information to link elevated ozone with both intended use (trails) and secondary threat (bark beetle) information in an effort to identify such areas for further focus.

Overall, the CASAC is very supportive of inclusion of secondary effects that may be partially attributed to ozone, or whose likelihood or severity is increased by ozone. The examples considered in the WREA appeared to be the strongest available, specifically bark beetle infestation and wild fire. Nevertheless, there is concern that the data supporting such interactions are often weak or indirect. The EPA is encouraged to attempt quantitative or semi-quantitative risk analyses of these indirect impacts, with explicit fractional attribution to ozone, to the extent possible. This is likely to be difficult and uncertain, and EPA has largely made this effort in the current draft, but any initial attempts would be useful for the second draft and to provide direction for future review cycles.

Possible feed-forward effects of ozone on vegetation, including on volatile organic compound (VOC) emissions leading to greater ozone concentrations, may be important impacts that should be carried through the analyses.

16. Regarding the analysis of forest yield impacts:

a) To what extent does the Panel agree that the Forest and Agricultural Sector Optimization Model (FASOM) model is appropriate to assess timber and crop yield changes and the effects of those changes on additional ecosystem services?

FASOM output can be useful at a national scale, but the implicit averaging of effects in different regions with different ozone environments could mask important welfare effects occurring at those scales. Moreover, the market compensation mechanisms and grower choices that are included in FASOM simulations may not be appropriate as part of a prescriptive estimate of risk. For example, an avoidance strategy might be to plant more resistant but possibly less profitable crops to avoid ozone-induced injury. Indeed, if the economic cost of an avoidance strategy could be quantified, it would represent an unambiguous adverse impact.

A potentially underlying weakness of the national FASOM approach is that it ignores effects of ozone on competition among unmanaged species. For herbaceous species and woody species, there are data on differential sensitivity and results from observation, theory, and physiologically-based modeling showing differential effects on sensitive species grown in competition with other species exacerbates effects of ozone. However, the effects depend on many competitive factors rather than just the ozone-sensitivity of each species grown in isolation.

The abundant science underlying adverse ozone impacts on agricultural yields should be considered more prominently. The depth of these data sets, and the recent confirmation of earlier data using newer exposure protocols, would suggest these impacts are well established and should be given equivalent weight relative to forest species, in the second draft WREA. This could be considered in a new section parallel to the current section 6.2.2, including figures similar to 6-4 and 6-5 and Table 6-1. Yield loss in regions with high ozone exposure should represent clear evidence of an adverse welfare effect on growers in these areas, a conclusion that is somewhat obscured by sector-wide economic analyses using FASOM-GHG. An unbiased evaluation of currently available science, including decades of previous experimental results and significant recent confirming evidence, requires that the welfare risk assessment reflect that major crop and tree species currently experience growth and yield loss due to ambient ozone in many regions and years.

b) What are the views of the Panel on the extrapolation of concentration-response functions across similar species?

It is not feasible to obtain concentration-response (C-R) functions for all species, yet there is no reliable mechanism to infer C-R relationships in a novel species even from knowledge of a closely-related species. Attributes such as C_3 and C_4 metabolic groupings, functional groups, and growth environments could be considered. Consideration of early versus late succesional status is a promising classification. Due to a lack of specific recommendations to be made, the EPA is urged to make such extrapolation with care, and to fully document the rationale for each such action. The CASAC is concerned about assigning a C-R relationship of the most sensitive species to other, unknown, species. A more nuanced approach would be to assign a moderate C-R function, and to attempt to propagate uncertainties through the analyses.

The CASAC is supportive of the effort by the EPA to aggregate species to achieve a clearly communicated and easily understood risk. However, such aggregation would serve to weaken the influence in the relationships of the most sensitive species, and thereby fail to protect them. Failure to protect the most sensitive crops could have adverse economic impact and failure to protect the most sensitive native species could lead to endangered status, extirpation, or even extinction.

17. Regarding the analysis of urban forest impacts:

a) To what extent does the Panel feel that the i-Tree model is appropriate for assessing changes to urban forest ecosystem services based on ozone exposure?

The use of i-Tree for urban forest species is appropriate, particularly because the model has been broadly parameterized and supported for use in diverse environments. However, the degree to which the i-Tree model can represent responses to ozone must be addressed for each endpoint.

b) In order to increase the number of tree species covered by the iTree model, does the Panel have recommendations for additional species that could be included, based on estimates from similar species?

The CASAC is not aware of additional species, although research underway at this time may provide additional data for the next review cycle.

18. Regarding the use of PnET-CN:

a) What are the views of the Panel on the potential use of the PnET-CN model in the 2nd draft to assess impacts on larger scale ecosystem services (e.g. hydrologic changes, c sequestration)?

PnET-CN should be fully parameterized for the environments in which it will be employed. Its use will not allow full description of ozone damage throughout the canopy, nor of competition between species differing in ozone sensitivity, as it is a big leaf model. However, its use may allow impacts of ozone that are well known at a species level to be scaled throughout the ecosystem.

b) Does the Panel have recommendations of other models that are accessible to EPA that could be used instead of PnET-CN?

The EPA previously used the TREGRO and ZELIG models, which together are capable of extrapolating from available physiological data on ozone effects on net photosynthesis, to impacts on whole-trees and mixed species forest stands. The DayCent GHG model might be considered for ecosystem services such as greenhouse gas production and carbon balance. The CASAC is also supportive of a reconsideration of the DLEM model.

19. Regarding ecosystem services related to foliar injury:

a) To what extent does the Panel agree that potential visible foliar injury is appropriate to use as a metric to assess potential loss of cultural services associated with recreation in national parks?

The CASAC is strongly supportive of using visible injury because of its long history of use as a metric, and because of the Kohut (2007) study and follow-on studies by the EPA. However, the actual loss of aesthetic value is less clear. Tourists and nature enthusiasts, when told that visible damage is due to

ozone, are unanimously disturbed. However, unless they are told, they are often unaware of the damage, or assume it to be part of the natural landscape. It is unclear whether this is a detrimental effect if they do not know it is due to ozone. Clearly future studies could address this impact, with willingness to pay type studies. However, given the current lack of quantitative relationships with adversity, inclusion of the endpoint is warranted, without over-emphasizing it.

The bark beetle-killed conifers in California are considered to be clearly detrimental to viewer pleasure, whether or not they know the reason for the tree death. Therefore, to the extent that bark beetle infestation can be linked to ozone, this is a clear loss of cultural value. The linkage to ozone appears to be strong, but quantitatively difficult to assess. The EPA is encouraged to attempt semi-quantitative assessment of the contribution of ozone to the overall mortality.

b) Does the Panel feel that there are ozone benchmarks that could be used to assess changes in foliar injury potential between exposure scenarios similar to those used by Kohut (2007)?

The Kohut approach to visible injury would be much more important to the risk assessment if the EPA could find a way to express it as a C-R function. However, the data are not easily amenable to this treatment. Only one ozone metric, presumably the W126, should be used to define the risk of visible injury, rather than the combination with N100 used in the current draft. Additionally, the risk assessment as applied to a specific locale, park, etc., should incorporate contemporaneous drought, through the Palmer Drought Index or other metric, as this may have large impacts on sensitivity to ozone.

WREA Chapter 7: Synthesis

20. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall welfare exposure and risk analysis?

This particular chapter is more of a summary of results than a true synthesis of the WREA document. The EPA staff have noted that they will work to produce a true synthesis in the next version. Although the summaries bring together the main accomplishments and results of each previous chapter, they should be placed at the end of each chapter rather than in the synthesis chapter. This change would result in a consistent organizational structure common to both the WREA and HREA documents. The content of this synthesis chapter should also make a logical progression to the PA document, so it will be useful to keep the goals of the PA in mind when developing the synthesis.

The scope of the risk analysis should be expanded to include much more attention to crop yield losses given the extent of scientific understanding on this topic (for more detail, see responses to Questions 13 and 16). Also, as mentioned in previous comments, more attention needs to be paid to (1) effects of ozone on competition among ozone sensitive species and less sensitive species in mixed-species stands, (2) scaling from tree seedlings to mature trees, (3) assessing impacts for individual sensitive species, and (4) assessing impacts for regions with different degrees of ozone exposure, in addition to

overall national assessments. Once these analyses are completed, they should be included in this synthesis chapter.

It would help to integrate each of the Risk sections (Biomass Loss, Foliar Injury, and Ecosystem Services) to provide a comprehensive and synthetic analysis of the complete welfare impact and associated risks. The order of presentation here is also important: a progression from foliar injury risks, to biomass, and then ecosystem risks moves logically up an organizational scale from the best characterized risks with the largest datasets to the least characterized risks with highly complex interactions, fewer datasets, and qualitative risk factors. The current WREA contains a clear delineation of those risk factors that can and cannot be quantified in this version of the WREA, and both sets of risks are important in the final risk assessments. Expanding the discussion of qualitative risk factors would strengthen this section.

References Cited

Alvarado, M. J., Logan, J. A., Mao, J., Apel, E., Riemer, D., Blake, D., Cohen, R. C., Min, K. E., Perring, A. E., Browne, E. C., Wooldridge, P. J., Diskin, G. S., Sachse, G. W., Fuelberg, H., Sessions, W. R., Harrigan, D. L., Huey, G., Liao, J., Case-Hanks, A., Jimenez, J. L., Cubison, M. J., Vay, S. A., Weinheimer, A. J., Knapp, D. J., Montzka, D. D., Flocke, F. M., Pollack, I. B., Wennberg, P. O., Kurten, A., Crounse, J., Clair, Wisthaler, A., Mikoviny, T., Yantosca, R. M., Carouge, C. C., and Le Sager, P. (2010). Nitrogen oxides and PAN in plumes from boreal fires during ARCTAS-b and their impact on ozone: an integrated analysis of aircraft and satellite observations. *Atmospheric Chemistry and Physics*, 10(20):9739-9760.

Kohut, R. (2007). Assessing the risk of foliar injury from ozone on vegetation in parks in the US National Park Service's Vital Signs Network. *Environ Pollut*, 149:348-357.

Singh, H. B., Anderson, B. E., Brune, W. H., Cai, C., Cohen, R. C., Crawford, J. H., Cubison, M. J., Czech, E. P., Emmons, L., Fuelberg, H. E., Huey, G., Jacob, D. J., Jimenez, J. L., Kaduwela, A., Kondo, Y., Mao, J., Olson, J. R., Sachse, G. W., Vay, S. A., Weinheimer, A., Wennberg, P. O., and Wisthaler, A. (2010). Pollution influences on atmospheric composition and chemistry at high northern latitudes: Boreal and california forest fire emissions. *Atmospheric Environment*, 44(36):4553-4564.

Appendix A

Compendium of Individual Comments by CASAC Ozone Review Panel Members on EPA's Health Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012) and Welfare Risk and Exposure Assessment for Ozone (First External Review Draft - Updated August 2012)

Mr. George A. Allen

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 4 - Air Quality Considerations

5. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The two overlapping 3-year periods of 2006-2008 and 2008-2010 are appropriate. While it is desirable to use the most recent 3-year data period, 2009 and 2010 were relatively clean ozone summers in the eastern US, making 08-10 a period that may not represent current ozone trends. Ozone during the summer of 2012 has reinforced that concept. 2006-2008 is relevant in that it is the most recent 3-year period where GEOS-Chem model run data for background O3 data are available - an important component of the REA process.

6. Regarding the methods for simulating just meeting the ozone standard:
a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard (including application of US background as a lower-bound on rollback) is a reasonable approach?

With the lower-bound, the quad-rollback is a reasonable approach, but does have limitations inherent in its simplicity and lack of any chemistry.

b) To what extent does the Panel support using an air quality model based approach for simulating just meeting the standard in future drafts as a replacement for the current quadratic approach?

There is potential for improvement in rollback estimations using air quality model-based approaches as noted in the Simon et al. memo; see c) below. I encourage EPA to continue to explore such alternatives.

c) What are the views of the Panel on the strengths and limitations of the proposed approach using the Higher-order Direct Decoupled Method?

To the extent that the emission inventories used for input are reasonably useful at the potentially not very large (urban) spatial scales and sub-daily time-scales under consideration, air quality model based rollback approaches such as HDDM have the potential to capture important features that are lost with a simpler rollback method. It appears that HDDM does not need any external estimate of ozone background.

The Simon memo shows interesting examples from Atlanta where rollback concentrations at core urban sites are similar using both the quadratic and HDDM methods, but are substantially lower at non-core urban sites using HDDM. For Detroit, HDDM and quad rollbacks are similar for VOC reduction scenarios for both core and non-core sites, but HDDM rollbacks are lower for NOx reduction scenarios.

Comments on the Welfare Risk and Exposure Assessment

Comments on Chapter 4 - Air Quality Considerations

4. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The two overlapping 3-year periods of 2006-2008 and 2008-2010 are appropriate. While it is desirable to use the most recent 3-year data period, 2009 and 2010 were relatively clean ozone summers in the eastern US, making 08-10 a period that may not represent current ozone trends. Ozone during the summer of 2012 has reinforced that concept. 2006-2008 is relevant in that it is the most recent 3-year period where GEOS-Chem model run data for background O3 data are available - an important component of the REA process.

5. What are the views of the Panel on the approach used to develop a national scale surface of W126?

Section 1.2.3 of the Wells memo describes the data fusion approach using CMAQ and monitoring data to create fused spatial 12x12 km grid surfaces for each of 2006, 2007, and 2008 as well as the 3-year average. This is likely to work reasonably well in the eastern US where the ozone monitoring network is more dense. In the western US, there are large areas with no monitors as well as more complex topography; both of these present challenges for any attempt to model a national ozone surface.

6. Regarding the methods for simulating just meeting the ozone standard:
a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard is a reasonable approach?

With the application of US background as a lower-bound, the quad-rollback is a reasonable approach, but does have limitations inherent in its simplicity and lack of any chemistry.

b) Does the Panel have suggestions for alternative approaches for simulating just meeting the current secondary standard or alternative standards based on the W126 metric?

Not at this time. The Higher-order Direct Decoupled Method or other possible air quality model based approaches have not yet been evaluated for this metric.

Mr. Ed Avol

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 4 Charge Questions

5. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The decision to use 2006 to 2010 data does raise some issues. As Figure 4.2 shows, there is a marked decrease in most of the urban areas being assessed in 2008 and 2009. The root causes for these declines – be it national economics, meterological changes, successful emission reduction strategies, or something else-- is arguable. However, this dramatic drop raises questions about how representative this time period is for long-term estimates of risk and exposure. The regulatory five-year review cycle prevents waiting for another year or two of data to see if this was an anomaly, so I understand why it is being used, but I worry a bit that the "dip" may be an artificial one, and using it to assess risk may lead to under-estimates of public health effects.

General Comments on Chapter 4

The document lays out a thoughtful and broad approach to addressing the risk issues associated with short and long-term ozone exposures. The use of a large number of city-specific data, drawn from across the country, regions, and meteorologies is encouraging, although there are still some nagging concerns about under-representation of higher-exposure areas.

Given the high rates of asthma and poor air quality in California's Central Valley, for example, it's a shame that the information either wasn't available or that Fresno CA was not selected for study inclusion. This is a community with a history of high asthma rates, low socio-economic status, a substantial population of Hispanics, and very poor air quality (both with respect to the rest of the country and in the context of non-compliant regions within CA). Fresno will continue (for years) to be a community impacted by a challenging array of environmental issues. If it were possible to review the criteria for urban area inclusion, the inclusion of Fresno (even in place of Sacramento) would be worth considering, in my judgment. Alternatively, potential confounding by agricultural exposures (pesticides), wind-blown dust (Valley Fever), and health care access in this region would have made this a challenging (but nevertheless valuable) area to consider.

Specific Comments on Chapter 4

1. Introduction

1. Pg1-5 to Pg 1-6: This justification, as to what the previous Administrator "place primary consideration on" or "recognized" or "placed less weight on" or … seems inappropriate in the context of the document. The previous accumulated data was presented to the previous Administrator and a decision was rendered. How can we say (or do we need to) what made the Administrator decide the way he did? This reads like an attempt to justify the Administrator's actions. This revisionist history/justification of a previous controversial decision that arguably should not be a part of this REA presentation of facts and analyses. Therefore, in my opinion, this section should either be severely edited or removed.
2. Pg1-6, line 9: font size error
3. Pg 1-7, lines 1 to 3: awkward sentence (run-on); change to "…might be considered, including…"

2. Conceptual Framework

4. Pg2-1, line 26: "Titration is usually short-lived…" is this what is really meant, or do you mean to say that it is usually a spatially-limited phenomena, due to nearby NOx sources (as the sentence goes on to say)? It's not really "short-lived", since the reduced local ozone concentrations will persist in that location, so long as the proximal NOx source continues to provide NOx for titration.
5. Pg2-5, line 19: delete "that"
6. Pg2-6, line 2: change "environments" to" locations"
7. Pg2-6, lines21 to 23: change to read "…lifestage (children less than 18 years of age, adults older than 65 years of age), diets…'
8. Pg2-6, Line 24: change "*pound* of body weight" to "*unit* of body weight" (…and shouldn't this be kilograms, anyway?)
9. Pg2-6, line 26: change from "…and their lungs continue to develop until they are fully grown…" to "…and are in a critical time period of rapid lung growth and organ development…"
10. Pg 2-7, lines 4 to 5\: change to read "…including children, older adults, people with asthma, and people with low socioeconomic status…"
11. Pg2-7, line 5 – how will risk assessments be focused on low SES segments of the population? There is lots of discussion about children and asthmatics, some comments about older adults, but relatively little in the way of specifics regarding low SES.
12. Pg 2-7, line 16: add a comma so that line reads "…suited to risk assessment, because…" and remove the comma on the next line, so that it reads "…air pollution exposure and include responses…"
13. Pg2-7, line 25 – change to read "…controlled human exposure studies are generally focused on small numbers of individuals in good *or moderate* health…"; the two central issues here are (1) that only a small number of people can participate, and (2) that ethical concerns dictate that those severely compromised or in poor health not participate in such studies.
14. Pg2-7, line 27: this is incorrect; the issue is NOT that chamber studies are biased away from highly susceptible individuals (many susceptible sub-groups are, in fact, sought out for study

participation); the issue is that those with the most compromised health status cannot ethically be asked to participate in these studies.

15. Pg 2-7,line 27-29: This is an overly simplistic, naïve, and incorrect perspective on what health outcomes can be evaluated in controlled chamber studies; a wide range of inflammatory, blood, neurological, cardiovascular, respiratory, etc endpoints can and have been used to quantify observable changes, well beyond collection of symptoms or forced exhalation (although those seemingly simple observations continue to be important, as well).
16. Pg2-8, line 3: delete one of the "human" references from the phrase "…*human* controlled human exposure studies…"
17. Pg2-8, line 10: change to read "…more serious *or chronic* health endpoints…"
18. Pg2-8, lines 12 to 15: change these lines that presently read "…includes both more sensitive and less sensitive individuals, and thus may be able to identify more serious health effects in at-risk subpopulations which cannot be evaluated in controlled human exposure studies which generally exclude individuals likely to experience significant adverse health effects from O3 exposure…" to "…and is therefore more likely to include a broad range of susceptibilities and sensitivities, compared to controlled human exposures, which involve a smaller number of individuals over a more limited health status range."
19. Pg2-8, line 20: add comma after "…controlled human exposure…"
20. Pg2-9, Figure 2-1: change legends to read "Short-term O3 exposures" and "Long-term O3 exposures".
21. Pg2-9, line 16: add comma after "…hospital admissions"
22. Pg2-9 line17 to Pg 2-10 line 2: This is awkwardly phrased – OF COURSE the evidence is not consistently supportive of a relationship between short-term O3 metrics and asthma medication use in children who don't have asthma…!

<u>3. Scope</u>

23. Pg3-2 line 20 – replace "…done as part of…" with "…included in…"
24. Pg3-3 line12 to 13: add parentheses, so sentence reads "…counts of person-occurrences (which accumulate occurrences…over an O3 season)."
25. Pg3-3 line22: run-on sentence needs punctuation; change to read "…moderate or greater exertion. Health effects observed…"
26. Pg3-4, line 1 change to read "…based on both controlled human exposure studies and epidemiological studies." (..the current sentence seems incorrect, since it refers to "both" but lists <u>three</u> items…)
27. Pg3-4, line 22: add hyphen between "location" and "specific"
28. Pg3-4, lines 27 to 30 : sentence seems redundant.
29. Pg3-5, line 18: change to : "…in the general and susceptible populations, respectively,.."
30. Pg3-5, line26: add comma, so that sentence reads "... be considered, including…"
31. Pg3-7, lines 17 to 21: run-on sentence needs editing, breaking down into shorter sentences with focused statements.
32. Pg3-8, line 23: insert comma to read "…quality, as evaluated…" and remove comma at end of line (following "…(U.S. EPA, 2008a)"

33. Pg3-8, line 24: add commas to read "…This information, along with additional analyses,…"
34. Pg3-8 line27: change sentence to read "…alternative standards, to be presented in the second draft REA…"
35. Pg3-9, lines 1 to 3: change sentence to read "This information is then used to place the relative comparative attributes of the selected study areas into a broader national comparative context."
36. Pg 3-10, line 6: delete "…, including….", and replace with ":
37. Pg3-10, lines 10 to 14: the sentence as presented does not make sense and needs to be broken apart and re-worded.
38. Pg3-10, line 15: "…controlled human exposure studies have only examined markers of short-term reversible lung responses…' is an incorrect statement, since there have been numerous controlled exposure studies evaluating cardiovascular as well as neurological outcomes, to mention two broad classifications.
39. Pg3-13, line23: Either a word is missing from this sentence, or the order is incorrect; re-word so line reads "Risk measures for lung function risk assessment estimated…"
40. Pg3-14, line 1: age parenthetical comment needs correction.
41. Pg3-16, line 12: change to read "…the 12 *case study* urban areas."
42. Pg3-16, line29: rewrite so line reads "…for which we *presently* only provide…"
43. Pg3-17, line 29: text says blue text in Fig3-3 summarizes data outputs, but Figure 3-3 caption says blue identifies user=selected inputs…???
44. Pg3-18, Figure 3-3: for consistency and readability, consider adding "Black identifies…"
45. Pg3-19, line7: change "of just" to "restricted to"
46. Pg3-19, line 8: add "effects" after the phrase "short-term"
47. Pg3-19, line 9: replace "higher confidence" with "greater confidence"
48. Pg3-22, line 26: change "risk" to "risks"
49. Pg3-22, line 27: reword so line reads "…as well as core risk estimates *for* ozone…"
50. Pg3-22,line28: remove "zero" and insert "0 ppm"
51. Pg3-22, line 28: change end of sentence to read "…(LML) *observed* in the epidemiology.studies."
52. Pg3-22, line 31: change line to read "…a *no-effects* threshold have indicated a generally linear C-R function, with no indication of a *no-effects* threshold in analyses *examining* the 8-hour…"
53. Pg3-23, line6: add comma to read "…ozone is reduced, because…"
54. Pg3-23, line 8; change "higher confidence" to "greater confidence"
55. Pg3-23, line 15: insert Figures reference at end of sentence (since comments refer to vertical lines on specific plots without reference), so it reads "…(see Figures__)."
56. Pg3-25, line 7: is there a reference missing or an extra space?

<u>4. Air Quality Considerations</u>

57. Pg4-1, line 3: change "Chapters 5-7" to "Chapters 5 to 7" (since pages are identified by the chapter-dash-number designation)
58. Pg4-1, line 9 and 10: punctuation corrections to sentence, so that it reads "The four urban areas evaluated for this first draft were: Atlanta GA; Denver CO; Los Angeles CA; Philadelphia PA."

59. Pg4-1, lines27 to 29: re-word to read "…from zero to four, from areas with a population…of the NAAQS, to areas with a population…"
60. Pg4-2, line1: change "must be designed to record" to "must be located to capture"
61. Pg4-3, Figure 4-2 illustrates an observation, which may be relevant – the potential impact of the economic downturn and climate change on air quality trends. Since the 2006 to 2010 time period is identified for consideration in this document, should there be "some consideration" of these "exceptional events" (i.e., economic effects, atypical summer weather)?

5. Characterization of Human Exposure to Ozone

62. Pg5-1, lines 23 to 26: It is appropriate that there be a strong emphasis on children, given their at-risk designation, but the lack of focus (or perhaps information) on children age 0 to 5 years is a gap of concern. Given the importance exposures during this time window have been demonstrated to have, in terms of subsequent health outcomes, the absence of any attempt to model exposures for children 0 to 5 years of age is a missing part that needs to be filled.
63. Pg5-11 to Pg5-12, Table 5-1: The CHAD data set has been invaluable in providing useful activity data, but the patterns of activity of children are likely changing. The emergence of electronic media (television, electronic games, electronic entertainment via cell phones, time spent on Facebook and other social media sites, etc) and its availability to broad groups over a wide age and socioeconomic range, as well as social shifts in time spent outdoors due to safety concerns, availability of parks and outdoor sport venues, or employment and child-care considerations, have potentially shifted children to more sedentary and indoor activities in the past decade. CHAD data relies primarily on data collected in the 1980s and 1990s, based on Table 5-1, so there is the concern that available exposure assignments may not well-represent the current population. This is another data gap that should be evaluated and potentially re-visited and addressed.
64. Pg5-20, lines 2 and 3: Is the decision to "not (take) these younger (than five years old) children into account in our analysis…" due to lack of such data in the CHAD data base, or the general lack of such information anywhere? In other words, is this an issue that might be addressed by some effort to update CHAD, or is there a need for original data collection?

7. Characterization of Health Risk Based on Epidemiological Studies

65. Pg7-34 Section 7.4: the discussion on variability and uncertainty is worthwhile and extended (a page and a half), but it is repetitive with previous sections. In fact, it appears to be almost a verbatim repetition, so some editing to place this discussion in one section, then refer to it from others, would seem more appropriate.
66. Pg7-76, title of section 7.7, "refinements" is mis-spelled.

8. National Scale Assessment of Short-term Mortality Related to O3 Exposure

67. Pg8-4, Figure 1.2: it's surprising that none of the Los Angeles, Houston, or Atlanta metropolitan regions show up on this figure as having elevated May-September average 8hr daily maximum levels of concern...is this a problem with the model, the mapping, my interpretation,…?
68. Pg8-4, Figure 1.3: …still surprising that the Houston area does not show anything of note here…?
69. Pg8-7, lines 22 to 27: This sentence about what has been done in this first draft, and what will be done in the 2nd draft REA, is worded awkwardly and somewhat confusing to the reader. Are you proposing that in the second draft document, you will only estimate risks for May to September, and not consider the ozone mortality impact outside of that time window, because "…the higher effect estimates derived from year-round O3 data ***may*** yield an equivalent O3 mortality impact as the lower effect estimate derived from the warm season O3 data only?" (lines 23 to 25)? Is this the test threshold being applied? What if the estimates are not equivalent?
70. Pg8-8, lines1 to 3: Is applying a "national average to grid cells outside of the urban areas included in the studies the best we can do? Would some regional assignment or inverse-distance weighting approach be any better? I raise this point to address the potential concern that larger populations in certain states (populations near locations of higher ozone but not in the cities for which data exist) may be "under-assigned" through the use of a national average, and that the magnitude of that under-assignment (due to population density) might be larger than the over-assignment (through the use of the national average) in other parts of the country.
71. Pg8-8, lines18-19: is the "concentration threshold assumption intact, in light of the recent publication of McDonnell et al (Inhalation Toxicology Aug 2012) which argued that the threshold model was superior in fitting the available data? One published study should not necessarily "undo" all previous work, but it does open the door to discussions about alternative considerations. The McDonnell et al study is based on controlled human exposure research, and chamber work provide critically important insights into human health effects. The lack of observed thresholds in the epidemiologic studies could be reflective of measurement error or a more realistic assessment of the effect of ozone exposure under actual ambient exposure conditions. At the current time, insufficient data is available on which to differentiate between these two possibilities.

<u>9. Synthesis</u>

72. Pg9-1, line 17 – change "…*pound* of body weight…" to "…*kilogram* of body weight…"
73. Pg9-1, line 18 – awkwardly phrased; suggest rewording to read "…and are in a period of rapid growth and development, placing their tissues in a state of increased sensitivity."
74. Pg9-1, line 27: incorrectly phrased, should read "…but *which* cannot be evaluated…"
75. Pg9-2, line3: change to read "…the percent of children with multiple ozone exposures."
76. It doesn't seem like the following several pages are well-used, if the intent is to *"summarize"* the <u>*key*</u> findings; the bullets seem to go on endlessly… This needs to actually be summarized. This could probably be more effectively communicated in a summary table, with carefully-constructed columns and rows to delineate health status, city, and number or percentage estimates.
77. Pg9-2, line9 and all subsequent bullets where this phrasing is used: change to read, "…the average percentages of school-age children (across 2006 to 2010) estimated to …"
78. Pg9-2, line 36: replace "term" with "terms"

79. Pg9-3, line22: "In the *worst* O3 year" is poorly phrased and should be changed for several reasons; 1) what does "worst" mean? Highest observed levels? 2) the phrase "worst O3 year" needs to be anchored to something – in the 2006 to 2010 time period, for example.
80. Pg9-3, line 27: "In the *best* O3 year," has the exact opposite issue as item#71 above.
81. Pg-9-6, line 29 to pg9-7, line23 could be summarized more clearly, as suggested immediately above.
82. Pg9-9, lines23 to 26: The logic and reasoning here is not clear – please re-consider the logical links you think have been made here. Why do the observations that urban areas do not capture the highest mortality rates, the oldest populations, or those with the lowest AC prevalence necessarily lead to conclusions that urban areas are not missing high risk populations with high ozone concentrations, or those with greater ozone susceptibility?
83. Pg9-10, line 2: Some careful thought needs to be taken here to differentiate short-term from long-term implications and effects. The observation that recent ozone concentrations have generally been declining makes it likely that short-term risks and exposures may be declining as well, but that is not necessarily a certainty. The chapter discussion does note that there has been some variability. In addition, changes in temporal, spatial, and physical patterns of activity will change the exposure patterns of certain susceptible sub-groups. These changes could undercut the reductions and create scenarios of higher exposures and risk, in the face of generally declining ambient concentrations. Longer working hours under more temperate conditions, or more exercise by asthmatics outdoors because restrictions are lifted, could create scenarios leading to longer exposures at lower ambient concentrations, and potentially higher cumulative risks.
84. Pg9-10, line20: add a hyphen between "epidemiology" and "based"

Dr. Michelle Bell

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 5 - Characterization of population exposure

7. To what extend does the Panel fid the methods used to conduct the exposure analysis technically sound? Does the Panel have any recommendations on the methods used?

Overall, the methods seem appropriate and well described. It would be nice to have a figure showing a flow chart of the various models, inputs, and outputs for the exposure assessment. In general, the methods could be described in more detail, with particular attention to various assumptions that are made. As an example, the text describes the differences between the transfer factors model and the mass balance model, but does not sufficiently explain why the mass balance model was used for indoor air, or note this explicitly (see Table 5-4). Provide detail on what "just meet the current standards" means, as this concept is critically important to the methods. I suspect that many readers will understand this, but it would be helpful to be explicit. As another example, there are a few vague terms that could be better described, such as "ozone season" or "sufficient" air quality data. For instance, for Table 5-3, explain why the period modeled differs for each urban area and what these time periods are intended to represent. Do these reflect ozone monitoring seasons, a relationship to temperature, or anticipated ozone seasons? Provide references for statements on methods and assumptions, such as "The lack of a better treatment of indoor air chemistry is not considered to be a significant limitation of APEX for modeling O3." In section 5.4.1, the text nicely describes the underlying premises for each of the three ozone benchmarks (e.g., 0.070 ppm based on asthmatics' sensitivity); however, references (either to articles or to other sections) would be useful to help provide the basis. A key part of the methods that were unclear to me was the incorporation of averting behavior and exposures. The text describes that certain elements will be "considered" in future work, but it is not clear how this will be conducted (see text beginning on page 5-13). Similarly, the general methods described for the elements of uncertainty are not clear (see page 5-24). An analysis using city-specific values does not necessarily capture the variability. This is another case where the methods may be perfectly reasonable, but are difficult to understand. In general, these suggestions relate to better description of the method. Although the methods do appear sound, they will be easier to evaluate with more detail.

8. To what extend does the Panel find the assessment, interpretation, and presentation of the results of the exposure analysis as presented technically sound, appropriately balanced, and clearly communicated?

The underlying concepts of the presentation of results work well, and I found the authors' interpretation of results to be sound. The presentation of results could be a bit better. There are some minor improvements that would greatly aid readability. For a table on multiple pages, repeat the row of column headings. There are too many abbreviations on the figures as they are unnecessary (e.g., spell out cities in Figures 5-1 to 5-15, do not abbreviate "75 6-8" for 75 ppb 2006-8, year rather than myear in Table 5-

5, more clear column headings such as in Table 5-5). There is plenty of room to avoid some of these abbreviations. In cases where abbreviations are necessary, please give their definition (e.g., ME in Figure 5-19, PREC in Figure 5-21). Consider combining Figures 5-1 to 5-15 into fewer figures (this may not work well, but is worth trying). These figures could be bar charts or some other format rather than 3-dimensional bar charts. Adjust the presentation of percentages to have another significant digit as currently there are bars of obvious different sizes, but the same percent results (e.g., 0% is at least 3 different heights in Figure 5-6). There is a typo in Table 5-10 (Atlanta is tlanta). There are a few interesting results in the footnotes that are very informative, and the authors may wish to move this information to the main text (footnote 9 on page 5-41, footnote 14 on page 5-43). For Figures 5-22 and 5-25, I think estimates were only generated for specific exposure levels in 0.01ppm increments. To help aid interpretation, avoid a linear fit through the estimates, or add point estimates to emphasize that the full spectrum of ozone exposure values was not modeled (if this is correct).

9. Regarding the characterization of uncertainties and variability:
a) To what extend does the Panel find that the uncertainties associated with the exposure analysis are clearly and appropriately characterized?

The document does highlight several key uncertainties and usefully separates uncertainties and assumptions for various parts of the exposure assessment process (e.g., APEX, CHAD). EPA should be commended for the extensive attention that has been paid to uncertainties. It is clear that the authors have given considerable thought to this issue. It would be useful to have even more discussion of some of the uncertainties. In particular, differences in housing structures could be discussed in more detail, such as how housing structures are likely to have regional patterns; therefore the influence on exposure assessments may also be regional.

b) To what extend does the Panel find that the uncertainty assessment is technically sound? Are there other important uncertainties which are not covered?

This text gives the impression that the authors have given considerable thought to uncertainties of this process. There are some uncertainties that will be addressed quantitatively, although many will not. The text could explicitly note uncertainties that the authors acknowledge, but for which current scientific methods do not exist for sensitivity analysis. To the degree possible, the anticipated direction of uncertainties on results should be discussed, for uncertainties that are not considered quantitatively (e.g., omission of outdoor workers).

10) What are the views of the Panel on the sensitivity analyses that EPA plans to conduct as part of the second draft REA to evaluate the influence of uncertainties in the exposure analysis?

The methods proposed for sensitivity analysis are not very specific (e.g., use city-specific diaries) and seem more another way of performing the calculation rather than an actual way of incorporating uncertainty. In addition to the base method and proposed sensitivity analysis, another approach would be to generate lower and upper bounds, such as by using the lowest and highest values from any city.

Comments on Chapter 9 – Synthesis

19. To what extend does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall health exposure and risk analysis?

The synthesis is useful and insightful, but could be improved. There are a few places where the key message is unclear. In particular, the policy relevance of risk estimates for ozone concentrations at 0, a value that is unfeasible for many regions, needs more justification, even in a synthesis section. In general, please review the linage to make sure the word choices convey the right meaning. For example, two studies having 3 of the same 4 cities with the highest impact does not seem a "significant" difference. Alternatively, if this is what is meant, please clarify. The language that health endpoints "remain" (see page 9-8, is odd and potentially confusing. Please reword. Rather than bullet points, some of the key points could be provided in a figure or table. The statement that the urban study areas provide a good representation of the overall distribution of risk (see page 9-9) is unclear as risks may differ in rural environment. The other discussion points on the urban focus are clear. It is not clear why some aspects, such as alternative lag structures and copollutants, could not be incorporated in the analysis (see page 9-10). I was not sure if the authors mean that this will be done in later versions or if for some reason such analysis is not possible. As a minor point, I suggest changing the language of "the need to specify values for U.S. background concentrations is not necessary, as it is incorporated in the modeling directly" (see page 9-11). While technically true, the background concentrations are in fact specified, just by model results not the user. Better wording may be that the background concentrations are modeled and therefore do not need to be selected by the analysis. The text notes that the second draft RES will incorporate an improved approach to adjust for ozone levels, but the nature of the improvements are not clear. Perhaps this will become very clear in the next version.

Dr. Joseph D. Brain

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 1 – Introduction

1. Does the Panel find the introductory and background material, including that pertaining to previous reviews of the O3 standards and the current review, to be clearly communicated and appropriately characterized?

The committee has some concerns about the list of acronyms and abbreviations early in the report. On the one hand, it's useful to have these abbreviations defined. On the other hand, in some cases the abbreviations or acronyms needs further explication. For example, yes, EvNA stands for "enhanced voronoi neighbor averaging," but what does that mean? Similarly, the critically important "policy relevant background" is given as the definition of PRB. This tells what words correspond to which letters, but it does not describe how the PRB is measured and how it might be used in the regulatory process. As ozone levels diminish, the impact of the PRB looms larger. In these instances and many others, it would be useful if a link was provided so that a further definition or the nature of the methodology used was easily available to the reader. An example of the latter is MATS which means "model attainment test software." Those four letters are defined, but the reader is still ignorant of what this software is and what it accomplishes.

Yes, the panel finds that this first chapter, "Introduction," provides a useful review of the process leading to the previous ozone standards. With a few exceptions noted below, the history and relevant background is "clearly communicated and appropriately characterized." It does raise the issue of whether there should be a generic introduction/preface/history as we found extremely useful in the ISA. There could be some standard boilerplate, although it would clearly be less extensive and critical than for the ISA. Thus, we recommend that there be a generic piece of the REA. Perhaps this is an example of "no good deed goes unpunished." The ISA example is so effective, it raises the issue of whether it should be replicated in both the REA and the policy assessment document.

One question is whether the current Chapter 9, "Synthesis and Integration of Results," should come earlier in the document. In the ISA, this final chapter corresponds to the executive summary and the integrative summary, which come early in the document. Should this have the same construction? The history, as given on pages 1-3 through 1-6, is interesting and useful. However, as it reaches its conclusion on pages 1-5 and 1-6, conspicuous by its absence is the role of CASAC in terms of reconsideration. We did have meetings and did offer advice about the reconsideration decision in response to multiple suits filed against the EPA. A sentence or so with appropriate links should be added.

Also conspicuous by its absence is the PRB. I think an understanding of the policy relevant background is important. It should also appear in telling the history. Where did this concept originate, and when was

it introduced into EPA deliberations. This is another example of an important concept which should be adequately recognized.

Comments on Chapter 2 - Conceptual Model

2. To what extent does the Panel find that the discussions accurately reflect and clearly communicate the currently available health effects evidence, and the relevance of that evidence for quantitative exposure and risk assessment, as characterized in the 3rd Draft ISA?

Again, I am concerned that the PRB is not adequately defined or discussed. Particularly in sections 2.1 and 2.2, dealing with ozone chemistry and sources, it seems essential to discuss the PRB. In fact, aren't there three sources of ozone and precursors? There is domestic versus international, and even within domestic, there are natural events such as lightening and forest fires, and then sources that are definitely related to human activity. I suppose one could discuss to what extent forest fires are influenced by human activity.
Section 2.4, "At-Risk Populations," is another piece that should be generic and common to all EPA documents. In conclusion, the panel does find that the discussions in chapter 2 "accurately reflect and clearly communicate the current available health effects evidence." Moreover, those conclusions flow rationally from the information contained in the third draft ISA.

Comments on Chapter 3 - Scope

3. Does the Panel find the scope of the health risk and exposure analysis is clearly communicated?

The panel finds that the scope of the health risk and exposure analysis is appropriate and is presented with clarity and economy. EPA staff has done an excellent job in laying out the process, the sources of data, and how various models are used to make predictions.
I draw attention to the absence of discussion of adaptation. One of the hallmarks of oxidant injury, especially ozone, is the phenomenon of adaptation. There are levels of ozone, or hyperoxia, which produce serious injury or even death in naïve animals. However, in animals chronically exposed to lower levels of ozone or oxygen, there is morphologic and biochemical adaptation. Subsequent exposures to ozone produce a far lower response. This is important in understanding ozone toxicology. It also relates importantly to different patterns of ozone exposure. Citizens, who rarely see significant ozone levels and then suddenly have a two to three day episode of high ozone, may be much more affected than those who enjoy steady state ozone exposures all the time.

4. Based on information in the 3rd draft ISA indicating lack of evidence for a threshold in O3 concentration-response functions, we have included risk estimates down to zero O3 concentrations. Based on further discussion in the ISA regarding the decreased confidence in the shape and magnitude of population health response at very low O3 concentrations, we have also included risk estimates based on applying concentration-response (C-R) functions only down to the lowest measured level (LML) in the underlying epidemiology studies.

a) To what extent does the Panel support the use of two different risk estimates, one applying the C-R function down to zero, and one applying the C-R function down to the LML, to characterize the range of risk estimates to balance comprehensiveness of the estimates with confidence in the estimates?

The panel believes that the presentation of two different risk estimates will be confusing. We hope that CASAC and EPA staff can identify the one which is optimal. Personally, I would favor the C-R function down to the LML. The C-R function which goes down to zero makes little sense. First of all, such levels are never obtained, except in unusual chamber situations where ambient ozone is deliberately scrubbed. Secondly, this zone has little value since it cannot be influenced by the regulatory process.
For an analogous situation, consider ionizing radiation. Can we or should we look at the effects of radiation at levels below background radiation exposures? Do we know what cancer incidence might be in the absence of ionizing radiation? What's the point?

b) What are the views of the Panel on alterative cutoffs based on other points within the distribution of O3 concentrations used in the underlying epidemiology studies?

As mentioned before, I believe we should present the most rational and useful cut off. At this point in the regulatory process, we should have a vision of what levels/cut offs are scientifically sound and contribute to standard setting in a practical way.

Dr. David Chock

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 4 - Air Quality Considerations

Charge Question 5: To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The EPA chose five most recent consecutive years where air quality data are available at this time --- 2006-2010 --- for risk and exposure analysis. These five years most definitely encompass a high degree of variability in meteorology and emissions, and are sufficiently recent to serve as a valid starting point to preview the near-term future before the next round of CASAC ozone review. This choice, in my mind, is reasonable. It also involves the least number of assumptions that may be subject to questions or challenges.

There may be situations where resources may impose a constraint on the number of years available for more in-depth analysis. In these cases, I would be in favor of selecting the most recent possible years because they serve as the best initial conditions to look down the road into the near-term future. Cases with a strong resource demand may include the use of the hierarchical Bayesian model to fuse the monitor data and the model predictions to establish a time dependent nationwide ozone concentration field.

Charge Question 6: Regarding the methods for simulating just meeting the ozone standard:
a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard (including application of US background as a lower-bound on rollback) is a reasonable approach?
b) To what extent does the Panel support using an air quality model based approach for simulating just meeting the standard in future drafts as a replacement for the current quadratic approach?
c) What are the views of the Panel on the strengths and limitations of the proposed approach using the Higher-order Direct Decoupled Method?

a) The quadratic rollback scheme was devised in the absence of more scientifically rigorous modeling efforts in the early days of risk and exposure assessments. The steps assumed in moving the ozone concentration distribution downward toward attainment for a given area cannot be readily verified and are lacking scientific underpinning. Many of its shortcomings are already described in the HDDM-rollback attachment. Among other things, I am particularly concerned about the tendency of the rollback scheme to excessively suppress the ozone concentration distributions for areas that require more drastic reduction in ozone concentrations to just meet the present standard. This unphysical artifact is a result of using identical rollback coefficients for all concentrations less than the highest concentration for which the coefficients are first determined in a given area. The concentration regions that are most noticeably suppressed are the high concentration regions whose rollback does not trigger the use of the background

floor values. The artifact is quite apparent when we look at Tables 4-1 and 4-2. The two study areas --- Los Angeles and Sacramento --- that had the highest design values during 2006-2008 have the highest percentages of cases where the roll-backed concentrations have to be replaced by some background "floor" values. (Note that the background referred to is the U.S. background, not the NA Background.) Figure 2-4 on p. 2-14 of the Health REA first draft attachment prepared by Wells et al. (2012) shows a comparison between the 2006-2008 observed ozone concentration distributions and those after quadratic rollback for Detroit and Los Angeles. Note the excessive suppression of the high concentration region in the case of Los Angeles compared to the case of Detroit. Obviously, the quadratic rollback scheme will create biases in its assessment of ozone exposure and the associated health risk. Given that there is now a more science-based rollback approach, it is time to retire the quadratic rollback scheme.

b) and c) Using an air quality model to simulate the scenario of "just meeting the standard" would be a valuable and worthy effort. In fact, the benefit can go far beyond just demonstrating attainment for certain areas, especially when the model domain actually covers at least the regions where regional transport of ozone precursors is relevant. Regional-scale modeling can also help develop cost-effective emission control strategies that cut across state and local boundaries, resulting in overall cost savings in the long run. The effort here, however, is more about the applications of the higher-order direct decoupled method (HDDM) for establishing the ozone concentration distribution of an area when it just meets the present ozone air quality standard. The approach is scientifically sound and the computation burden is not excessive, especially for applications to a given urban area. The attachment prepared by Simon et al. (2012) describes the rationale and illustrates the application with examples. Use of multisteps in the case of NOx reduction to achieve the ozone standard is sensible. For the second draft REA, EPA is developing the 2007 modeling platform using the 2007 meteorology and a combination of 2007-2008 emissions. The CMAQ version 5.0 may also be used. These are all very encouraging developments. It would be helpful if CMAQ's background ozone concentration profiles had been compared well against those of GEOS-Chem and CAMx for the same modeling conditions. There are two concerns here in the applications of the calculated sensitivities. One is their applications to the monitored rather than modeled ozone concentrations. The single-step and multistep justification test cases were done based on modeled concentrations. The resulting justification may not necessarily be transferable to the monitored ozone concentrations. However, the past practice has been to use the ratio of observed-to-modeled concentrations as a relative response factor to scale the modeled concentrations to observed concentrations. The second concern is the choice of emission reductions. Should it be NOx only, or VOC only, or a combination of both? If the resulting distributions are comparable anyway, it may not matter much. In that case, it may be economical to choose an emission reduction scheme that avoids the multistep procedure. In fact, by judicially selecting a set of intermediate-emission scenarios as a "brute-force" simulation base case, the HDDM may be replaced by the simpler first-order DDM while retaining the boundary conditions of the "brute force" simulation without incurring significant errors in determining the "just meeting the standard" concentration distribution. On the whole, I support this rollback methodology as a replacement for the quadratic rollback.

The Chapter also briefly discusses the use of a hierarchical Bayesian model to fuse the 2006-2008 monitored ozone data with the model (CMAQ)-predicted ozone concentrations to create a spatially-resolved and time-dependent ozone concentration field nationwide. This is a highly computation-

intensive but otherwise commendable effort. The relatively high sensitivity of the spatial distribution of the fused ozone field to the variances assumed in the prior distributions for the measurement errors of the monitored and model-predicted ozone concentrations calls for caution in the choice of these variances.

Dr. Ana Diez-Roux

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 5

7. To what extent does the Panel find the methods used to conduct the exposure analysis technically sound? Does the Panel have any recommendations on the methods used?

Overall the methods used to conduct the exposure analysis appear sound and seem to make the best use possible of available data.

A major predictor of exposure levels is time activity patterns. The APEX model utilizes CHAD, the most complete source of human activity data currently available. Despite the use of this data and new methods to reflect time-location-activity in simulated individuals, estimates will only be as good as the time-activity data utilized. Time-activity patterns are likely to vary substantially by region and a range of social and economic factors. These patterns are also likely to have changed substantially over time. In this regard it would be helpful if Table 5-1 included some information on how the individuals in the various studied were sampled and the extent to which the studies included a range of occupations, regions, and socioeconomic backgrounds.

For example, it may be possible to provide even some basic characterization of the extent to which these samples are likely to represent the activity patterns of the areas being modeled.

The document could better discuss the implications of using these samples to characterize time activity patterns. For example, variations in time-activity patterns may lead to very large inter-individual differences in exposures with very high exposures in some population groups (many of which may be "vulnerable" groups) which are not captured by the approach used. In addition, these high exposure groups may also be the ones with less ability to implement exposure averting behaviors in response to information, as noted in the review provided. If possible, some estimation of "extreme " exposures (or distribution of exposures) for population subgroups derived from the model (as opposed to global population metrics) could also be useful in describing the population impact.

8. To what extent does the Panel find the assessment, interpretation, and presentation of the results of the exposure analysis as presented technically sound, appropriately balanced, and clearly communicated?

In the next draft, EPA may consider providing a more complete description of the key exposure levels predicted by the model for various population subgroups (as tables/figures as well as by summarizing key patterns in the text). Sometimes the rationale for the selected results presented is unclear. Key patterns need to be succinctly summarized. For example, it would be helpful to summarily describe the key patterns observed in Figures 5-1 to 5-15 and tables 4-5-4-19. The tables in particular are

sometimes difficult to interpret, some column headings are unclear (as are the titles). A figure presentation might be better. (A relatively minor point is that three-dimensional figures are often misleading and probably should be avoided).

As noted above, and to the extent permitted by the simulation model, it would be helpful to illustrate some of the variability in exposures in addition to overall population averages. For example, it might be of interest to show the distribution of the population across exposure levels (e.g. distribution of the population across categories of #of 8 hour exposures across a certain level). [This may be limited by the limted ability of the APEX model to capture repeated exposures, if so this seems like a key limitation]. Some descriptives of continuous exposure metrics (rather than just #of 8 hour exposures above a level) would also be useful and could perhaps be linked to the health risks assessment which models continuous exposures.

The section characterizing factors influencing high exposures was generally well presented. However an important caveat is that these are analyses based on simulated exposures and therefore the factors that explain variability are the ones that were by design input into the modeling. This makes it difficult to draw very firm conclusions about sources of variability across the cities or even across individuals (especially in light of the fact that the time activity data input into the model may not captures all variability between and within cities).

Figure 5-18 needs to be labeled and explained more clearly. I found it cryptic.

Some of the descriptives are also discussed in the conclusion section of the chapter but they seem to belong earlier in the chapter. The rationale for showing figures 5-22-25 (especially at the very end of the chapter is unclear). In general this last section should discuss the key patterns observed in the population exposures predicted by the simulation model.

9. Regarding the characterization of uncertainties and variability:
a. To what extent does the Panel find that the uncertainties associated with the exposure analysis are clearly and appropriately characterized?
b. To what extent does the Panel find that the uncertainty assessment is technically sound?Are there other important uncertainties which are not covered?

Although I see the distinction between variability and uncertainty that the document attempts to make, it is also true that unaccounted for variability leads to uncertainty. In fact several of the sources of uncertainty would be minimized if additional data on variability (such as variability in time-activity patterns or microenvironment levels) were available. It may be helpful to recognize this.

10. What are the views of the Panel on the sensitivity analyses that EPA plans to conduct as part of the second draft REA to evaluate the influence of uncertainties in the exposure analysis?

The planned sensitivity analyses appear sound and cover the main points. If at all possible, it would also be useful to evaluate sensitivity of results to utilizing time-activity data that matches the social and economic characteristics of the areas being modeled.

Comments on Chapter 7 - Characterization of Health Risk Based on Epidemiological Studies

11. Regarding the epidemiologic studies used in the analysis:
a. What are the Panel's views on the set of epidemiological studies selected for use in specifying C-R functions and on the set of C-R functions specified for use in the risk assessment?
b. To what extent does the Panel find the detailed descriptions of rationales for the selection of the epidemiological studies and the selection of the set of C-R functions specified using those studies to be appropriate and complete?

The set of studies selected seems reasonable. The criteria for selecting the studies are clearly described and also seem reasonable. Consider adding information on confounder adjustment to all the studies shown in Table 7-4, as well as some brief summary of the strength associations reported in the study (e.g. RR estimates for most important exposures studied or equivalent).

12. To what extent does the Panel find that the qualitative discussion of uncertainty and variability have covered important sources and appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates?

The discussion of variability and uncertainty is clear and coherent. Major sources are appropriately considered. Overall I found this section accurate and balanced.
Table 7-6 provides an excellent summary of the issues.

13. Regarding the results of the risk analysis:
a. What are the views of the Panel on the presentation and discussion of risk estimates, including the key observations presented in section 7.6.2?
b. What are the views of the Panel on the presentation of the distribution of O_3-related mortality across daily O_3 levels for each city as "heat maps"?

Overall the chapter presents very useful information. However I found some of the results presented difficult to understand, perhaps in part due to some inconsistencies in the way in which terminology is used. The terminology used throughout the chapter and in table headings is confusing. Consider using a consistent terminology throughout the chapter to describe the various metrics being calculated.

Avoid saying "total incidence" or "total prevalence" or "mortality" when you refer to total number of

incident cases or total number of prevalent cases or total deaths (pg 7-7 and many of the tables). Consider using percent attributable risk (or percent of deaths attributable to…) as opposed to what is sometimes referred to as "total risk" and percent reduction in risk or percent reduction in absolute number of cases as opposed to "risk delta" (see page 7-3).

Tables 7-7 to 7-10 appear to show the number of annual deaths attributable to ozone across categories of daily max ozone level. The same applies to Tables 7-11 to 7-15. Labeling these tables as "total deaths attributable to ozone" would help.

Table 7-16 to 7-18 show percent of total deaths attributable to ozone (percent attributable risk in epidemiologic terms). Table 7-20 shows percent reduction in mortality associated with gong from existing conditions to meeting the current standard. Simplifying and clarifying table titles would help readers follow.

It is not clear what lags were used in estimating these attributable deaths (same day?). The lag structure used (and the simultaneous consideration of multiple lags) could substantially affect estimates.

The terminological issues described above also apply to tables 7-22 to 7-24. These tables were hard to understand.

The description of the main findings could be streamlined . For example section 7.5.1 reports on deaths and % of deaths (as well as morbidity) attributable to ozone generally, and section 7.5.2 reports on deaths and % of deaths attributable to ozone even if the current standard were met. This basic fact sometimes gets lost in the details.

As an alternative consider showing total number of deaths attributable to ozone, % of total deaths attributable to ozone, and percent and absolute reduction in deaths expected if a given standard were met. This could be shown side by side in table or figure form for the different urban areas as well as for ALL areas combined. (the same approach could be used for morbidity).

14. To what extent does the Panel agree with the characterization of overall confidence, including the degree to which the conclusions reached regarding overall confidence are supported by available information?

What is in the document seems very reasonable for a first draft and will no doubt be elaborated on as part of the planned sensitivity analyses.

15. What are the views of the Panel on EPA's discussion of potential refinements to the REA for the

second draft, including the plans for quantitative sensitivity analyses, additional refinements to the core risk estimates, and plans for assessment of long-term mortality and morbidity (i.e., plans to model risk for mortality and the decision not to model risk for morbidity endpoints, given data limitations)?

Plans for sensitivity analyses and refinements of core risk estimates are appropriate and address key issues.

Given the very high population impact of possible long term effects of ozone on respiratory morbidity, and especially given the fact that these outcomes often affect vulnerable groups, I would urge some risk assessment of these outcomes even if limited compared to the mortality analyses.

Dr. William Michael Foster

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 6 - Characterization of Health Risk Based on Controlled Human Exposure Studies

1. To what extent does the Panel find the methods used to conduct the risk analysis to be technically sound? What are the views of the Panel members on the methods used?

As stated on pg. 6-5, li 14-17, " the health effect included in this portion of the risk assessment is lung function decrement (e.g., ≥ 15% reduction in FEV_1) is an estimate of the expected number of people who will experience that lung function decrement."

I do not have a clear understanding of the rationale for the selection of the range of decrements in FEV_1 that were chosen/evolved (i.e., ≥ 10, 15, and 20%) for comparison. For change in FEV_1 function for this portion of the risk assessment, it would be helpful to understand if there is a rationale for the selection of the hypothetical decrements in FEV_1 (e.g., what are these decrements considered to represent as a type of metric for health risk?).

2. To what extent does the Panel find the focus of the assessment on lung function decrements in the quantitative risk assessment to be appropriate and informative?

Please see response above to #1.

3. To what extent does the Panel find the focus of the assessment on lung function decrements in the quantitative risk assessments to be appropriate and informative.

Please see response above to #1.

4. What are the views of the Panel on the treatment of the Panel on the use of the two different modeling approaches for specifying the exposure response function linking the change in FEV_1 to ozone exposure.

I do not have sufficient expertise to evaluate the essential differences between the 2007 ozone NAAQS review based upon a Bayesian Markov Chain Monte Carlo approach versus the McDonnell-Stewealrlt-SmithFEV_1 model.

However, I do have issues with the intent to add new controlled human exposure study data and potentially add these data to each of the above identified modeling approaches for the 2nd Draft REA. The issues I have with respect to adding the identified controlled human exposure studies (Table 6-2, pg. 6-8) are the following:

a) "square wave" and "variable" concentration were utilized, and thus given the demonstrated differences in the respective time course of spriometric responses to sqw vs. var concentration profiles, a description of how the spriometric data will be collated would be helpful.

b) Relevance of 6.6 h exposure with participants utilizing unrealistic elevated minute ventilations during exercising periods of the exposure (50 min/h) as suggested by Schelegle and co-authors (one of the new data studies) that overall ventilations are ≥ mean ventilations that might be encountered during a day of heavy severe manual labor and represents the higher end of ventilations that might be encountered in the normal population for this prolonged period (6.6 h). This description by Schelegle and co-authors identifies the response data as being perhaps more relevant to occupation workers exposed to ambient background ozone, and less relevant to general population.

c) For the Schelegle et al report, for each of the ozone profiles investigated (i.e., 60, 70. 80, and 87 ppb) on average less than half of the subjects completed the full 6.6 h of the exposure period. Some explanation of this high dropout should be provided and how the time line of spirometric response data collected during the exposure periods will take this into consideration.

d) For the Hazucha et al report, the exposure periods for sqw vs. var were of 8 h durations, whereas the reports by Schelegle et al, and Kim et al, were of 6.6 h durations. It would be helpful to understand how these differences will be considered when response data are being analyzed. In addition and due to study design, the overall minute ventilations during the exercise periods of the exposures varied per hour between the Hazucha study (30 min/h) vs. the other two studies (50min/h).

5. What are the views of the Panel on the treatment of the relationship between age and $dFEV_1$ in the McDonnell-Stewart-Smith model.

I am not so certain of what this question is referring to, by "treatment of the relationship" between age. The age range is considerably narrow, and thus how an understanding of risk for 18-35 yr old, would translate in a meaningful manner to subjects ≥ 60 yr of age, does not seem evident.

6. To what extent does the Panel find that the qualitative discussion of uncertainty and variability has covered important sources of uncertainty and variability and has appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates.

This question is outside of my expertise and has less to do with the topic of Controlled Human Exposure Studies. However under the sub-section of Characterization of Uncertainty (starts on pg. 6-46) the statement in the section of *Exposure History* (pg. 6-50) and contributing to uncertainty: "..approach used in calculating risk *assumes* that the ozone-induced response on any given day is independent of prior day ozone exposure" may require further qualification. A statement on the specific ozone-induced

response (e.g,. are these spirometric changes, or inducements of systemic or pulmonary inflammation, etc.) needs to be specified. This assumption for various host responses to ozone is largely incorrect.

Perhaps under the topic of key sources of variability with respect to ambient ozone exposure, the seasonal potential for bio-mass burning to contribute to the background levels of ozone, could be given some consideration.

7. What are the views of the Panel on additional sensitivity analyses or other approaches to addressing uncertainty and variability.

This question is outside of my expertise and has less to do with the topic of Controlled Human Exposure Studies.

Dr. H. Christopher Frey

Comments on the Health Risk and Exposure Assessment

My charge question assignments pertain to Chapters 8 and 9. However, I also offer comments on some other parts of the document since the scope for the analysis of Chapter 8 is given in Chapter 3, Chapter 8 references extensive material from Chapter 7, and Chapter 9 is a synthesis of the entire document.

Comments on Chapter 3

This chapter is clear that the scope of the risk and exposure assessment includes:

- estimation of the number of people with exposure above 'benchmark' O3 levels;
- estimates of the number of exposed people with impaired lung function resulting from O3 exposure,
- estimates of the potential magnitude of premature mortality 'and/or' selected morbidity health effects associated with recent ambient O3 levels, ambient O3 levels at the current standard, and ambient O3 levels associated with possible alternative standards
- better characterization of sources of variability to be considered when evaluating possible alternative standards
- insight regarding distributions of risks and risk reduction
- 'understand' (characterize? Estimate?) the national mortality burden associated with recent ambient O3 levels, and how the selected urban area estimates compare with national distributions of mortality risk.

Given the content of Chapter 8, it seems clear that it was not a goal to evaluate the alternatives of just meeting the current standard or just meeting possible alternative standards as part of the national mortality risk estimate. The third bullet above (from text on page 3-5, lines 21-24) and the last bullet (from page 3-6, lines 1-3) are partially contradictory. The text in Chapter 3 should be clarified to point out that the first 4 to 5 bullets are to be evaluated based on analysis of risk for 12 selected urban areas. The last bullet is evaluated in a separate national scale analysis, which does not attempt to evaluate just meeting the current or alternative standards.

Later parts of the report mention that 16 cities will be included for the risk assessment of impaired lung function, but apparently only 12 cities will be used for the epidemiological-based risk assessments. Some discussion of the criteria that led to choices of different sets of cities is needed at an appropriate location in the document.

The issue of policy relevant background (PRB) concentration is mentioned on page 3-8. In Chapter 8, the "lowest measured level" (LML) is used in some risk estimates. The 3rd draft of the Integrated Science Assessment provides information regarding North American background levels of

approximately 25 to 40 ppb. Clarity is needed regarding whether or how background levels were taken into account in the epidemiologic-based risk assessments.

The text on page 3-8 states that background O3 is assessed in the ISA and will be 'considered in the Policy Assessment.' This text implies but does not really explain that background is not intended to be considered in the REA (and why) and does not really state as to why background will be considered in the PA – e.g., even a short statement to the effect that background will be considered in the PA as a constraint on how low a standard could be set might be useful to the reader. However, is it clear from the Clean Air Act that background can or should be considered this way?

Why is the focus on mortality? Some reason is given. However, a statement could be made regarding the infeasibility of conducting a national assessment of morbidity at this time, and what factors limit this feasibility.

Epidemiological studies have been conducted for endpoints such as school absences, emergency room visits, hospital admissions, respiratory symptoms, and premature mortality (Page 3-14, lines 19-21). The risk analysis for the 12 urban areas includes the following selected endpoints (p3-16, lines 4-14): mortality, hospitalization for chronic obstructive pulmonary disease and pneumonia, additional hospitalizations in a subset of urban areas, and emergency room visits and respiratory symptoms in one or two cities for which data were available. The assessment in Chapter 8 of whether the 12 selected cities are representative of national distributions of key exposure factors focuses on mortality, but should also better take into account representativeness for morbidity end points, at least in the discussion and interpretation of information and data provided.

EPA plans to use the WHO framework for addressing variability and uncertainty more fully in the 2nd draft REA. Where quantification is possible, a Tier 2 or Tier 3 approach is strongly preferred.

Page 3-717, lines 23-30: "the risk assessment is implemented using BenMAP…" - does this include morbidity estimates based on exposure estimates developed using APEX? (more clarity is needed – the term 'the risk assessment' seems to be all encompassing).

Section 3.2.5 on the national scale mortality risk assessment should have more discussion on the following points:

- Why is the focus only on a short-term mortality, and why are morbidity endpoints not considered? (please explain)
- The current version only deals with current air quality levels during 2006-2008, and thus does not consider levels just meeting the current standard or possible alternative standards – will this be added? Or why is it not included?
- Currently, there is no analysis of variability or uncertainty along the WHO guidelines – presumably, this will be added?

Page 3-21, lines 15-26: with regard to evaluating alternative C-R functions, it is straightforward to do sensitivity analysis in which each is used separately and then compared. This would be a Tier 2 type of analysis.

A Type 3 uncertainty can be incorporated in the exposure estimates produced using APEX.

Use of confidence intervals for effects estimates is mentioned in a footnote. However, this information should be discussed in the main body, since it is important.

Comments on Chapter 5

I suggest that there should be more overview of CHAD, including a summary of the distribution of the diaries by age, gender, and other selection criteria (i.e. how many diaries of each type).

Even though APEX inputs describe variability, it should be possible to quantify uncertainty in the parameters of the variability distributions.

The method for estimating in vehicle exposure should be discussed in more detail.

For clarity, which microenvironments are based on mass balance and which are based on transfer factors?

Page 5-10: lines 3-7: ‘many assumptions are strengthened by the manner in which the data are used’ – does not make sense as written. Not sure what this is trying to say.

Page 5-13: the analysis of ‘averting behavior’ also raises the question of what is the purpose of the NAAQS. Is the purpose to protect at risk subgroups so that they can make use of the outdoors without significant adverse impact, or is the purpose to take advantage of averting behavior to therefore allow the standard to be less stringent than it otherwise would have been. That is, should averting behavior be taken into account, or is it an inherent right of each individual to use the outdoors without having to engage in such behavior? While certainly it is important to account for averting behavior when interpreting health effects data or developing risk models, a policy question is whether averting behavior should be assumed or mandated (probably not) as an alternative to lowering the ambient concentration on high concentration days.

Page 5-23: although true exposure may be unknown, to say that it is ‘largely unknown’ is ambiguous and not helpful. (This phrase appears also in the appendix). While it is certainly the case that the true value of exposure for an individual or of the distribution of inter-individual variability in exposure is not known exactly, it can be estimated. The estimation error depends on the level of aggregation of the estimate over time and space and to what extent sources of variability have been appropriately taken into account. The list of the most influential elements of uncertainty seems appropriate.

AER sensitivity analysis can also involve a straightforward comparison of how results vary with point estimates of AER.

Figure 5-1 and similar (through Figure 5-15) should be reconsidered. It is strange to have a vertical axis that is not defined or shown, and to have vertical bars that have numerical values associated with them but no vertical axis. Alternatives to these figures would be to use bar charts with multiple categories for each city that include the base and alternative case. The nomenclature "75 8-10" is not defined in the charts and must e. These charts should be self-explanatory.

The planned analysis of repeated exposures (pages 5-51 and 5-52) will be interesting.

Comments on Chapter 6

A table, perhaps in the scoping chapter, that lists all of the causality determinations from the ISA, and also lists which ones are addressed in the REA and how, would be very useful.

Figure 6-1 – what are S, K in the figure? Figures should always be self-explanatory.

Table 6-7: some numbers are given in percent, and others are not. The units for any numbers in a table should be clearly defined in the table header, and not in footnotes. The labeling in the columns under each city is not very clear and may need footnotes. E.g., "Current 0.075 ppm, 2006-2010" refers to an analysis conducted for the current REA based on the current 0.075 ppm standard, based on 2006 to 2010 air quality data. If figures and tables are not very clear, then there is potential for misinterpretation and confusion. Thus, it is critically important that figures and tables be clearly and completely labeled and that they are self-documenting.

Figure 6-21: here again, the notation used is not defined in the figure itself – e.g.,, what is "75 6-8", etc.?

Likewise, even FEV1 should be spelled out and defined in a note.

Comments on Chapter 7

Pages 7-2 and 7-3: it would be helpful to the reader if the discussion of LMLs also included discussion of background concentrations, and the relationship between the two. Would it not be the case that the LML would be expected to be at or above background levels? Yet, the LMLs used appear to be lower than the North American background concentrations reported in the ISA.

Page 7-11: please provide equation(s) that show how the LML enters into the analysis, so that it is more clear as to why LML-based risk estimates are lower. The idea seems to be that the same beta value is used, even though the increment used is smaller if the LML is included. Is beta independent of the LML?

Page 7-12 mentions the U.S. background ozone levels. This is another opportunity to compare with LML (or there could be a section added that compares LML and background).

Does the proposed alternative rollback procedure (Decoupled Direct Method) require positing an emissions reduction scenario?

Table 7-4, spell out HA for Hospital Admission… similarly, spell out ED and ER.

Section 7.4: it helps to have tables that summarize and discuss the key sources of variability and uncertainty (not just in this chapter but throughout the document).

Page 7-36: it is important to do quantitative sensitivity or uncertainty analysis for what are hypothesized to be the most significant sources of uncertainty. If the purpose of the analysis is to supplement the point estimate analysis as a way to help identify and rank sources of uncertainty, then the epistemic status threshold need not be so high. The enumeration of sources of uncertainty and justification for the Tier approach used should be more thorough and clear.

Page 7-47 "heat map" is jargon and not very useful. Either explain the analogy or use another term.

Comments on Chapter 8 - National Scale Risk Assessment and Representativeness Analysis

Charge Question 16. What are the views of the Panel on the overall approach used for the national scale risk analysis, including the O3 concentration methods and metrics, the use of city-specific and national average concentration-response relationships derived by Bell et al. (2004) and Zanobetti and Schwartz (2008)?

Although there are many detailed comments on Chapter 8, as enumerated below, the general approach seems reasonable. The chapter should much more carefully and clearly define concepts and quantities, to more clearly communicate the key input assumptions, results, and findings (details below).

Charge Question 17. What are the views of the Panel on the approach identified for quantifying long-term mortality using the Jerrett et al. (2009) two-pollutant model national respiratory mortality effect estimate?

The brief description of the plan for the 2nd REA on page 8-9 seems reasonable. The REA can and should be more clear as to the relationship between the short- and long-term mortality C-R models that will be used – i.e. are they mutually exclusive or does one subsume the other?

Charge Question 18. Regarding the representativeness analysis: a) What are the views of the Panel on the methods and presentation of results for the representativeness analyses?; b) Does the Panel have suggestions for additional risk characteristics that would be useful to include in the analysis?

The representativeness analysis seems reasonable in terms of the scope of variables identified, the analysis of each, and the conclusions reached.

Detailed Comments on Chapter 8

Chapter 8 describes the national scale risk assessment but does not provide much detail from which the reader can fully understand how it was done. Furthermore, many of the terms and concepts used are not clearly defined or used consistently. It would help to develop some shorthand notation for complex concepts that are repeatedly applied, such as the 4th highest daily maximum 8-hour average concentration in one year, and the annual 4th highest daily maximum 8-hour average concentration averaged over 3-years. With a clear definition and a short-hand notation, it will be easier to clearly and consistently refer to this or other concepts throughout the chapter and the document.

The introduction to this chapter appears to be summarizing results that are given later in the chapter. This is confusing to the reader. If the goal of the introduction is to be an abstract or executive summary of the chapter, this should be clearly stated.

Page 8-1, lines 30-31: the results given here have an unclear interpretation at this stage in the document. The reader has to carefully go back and forth in the text to try to figure this out, but it appears as if the numbers given here are the cumulative totals over a three year period, rather than annual averages. For clarity, state that these are cumulative effect estimates for May through September for the sum of three years from 2006 to 2008. Alternatively, present the numbers as annual averages rather than three year totals (which would be more policy relevant). Also, the reader is curious as to what were the mortality estimates in each of the three years (2006, 2007, and 2008) to gain insight into inter-annual variability in estimated mortality. Similar comments apply to the summary based on Zanobetti and Schwartz (2008).

Page 8-3, text: please quantify the correlation between these exposure metrics, such as the correlation between the May to September average 8-hour daily maximum concentration versus the June to August average 8-hour daily mean concentration. The correlation of each of these with the annual 4th highest maximum daily 8-hour average and with the three year average of the annual 4th highest values would also be informative. On lines 10-12, the text is confusing because it does not mention that the NAAQS is based on a 3-year average of annual 4th highest values. Are the 'seasonal means' also on a 3-year average? This is an example of the need for extreme clarity in this chapter regarding the quantities being used.

Figure 1.1 (8.1?) is confusing because air quality modeling is shown twice – why?

Figure 1.2 and similar figures: are the estimated concentrations given here based on 12 km by 12 km grid cells? This should be defined in each figure that shows this type of data. 'fused with average 2006-

2008 observations' – does this mean that the fusion was done on three year averages, rather than, say, on individual dates? Not very clear.

Figure 1.4: the top panel should be shown as a bar chart and not a continuous curve. This is really a histogram, not a probability density function. The vertical axis has units of fraction and not probability density. The fraction shown applies only to bins or ranges of values, which are discrete, not continuous. The wording of the caption is unclear – what constitutes a 'sample' in these summaries? Is it one monitor site for one day? One monitor site for an average per year? One monitor averaged over three years? One grid square? Etc. The basis for these and any other graphs and tables must be crystal clear.

Page 8-8, lines 1-3: could conduct sensitivity analysis based on effects estimates for individual cities with highest or lowest beta values.

Page 8-8, lines 10-11: delete "It should be noted, however,"

Page 8-8, lines 24-26: Delete "In order." Please compare LML to U.S. background and explain the comparison.

Page 8-9, lines 24-27: is this in reference to cardiovascular disease? Please clarify.

Page 8-10, lines 1-7: does this mean that the long-term health effects estimate is assumed to include short-term effects? This also needs to be more clear in the Integrated Science Assessment.

Page 8-10, lines 28-34: what percent of the estimated mortality is in other urban areas? Rural areas?

Page 8-11, line 1: delete "It is important to note that"

Page 8-11, line 23: Bell et al. (2004) is based on non-accidental mortality, not all cause total mortality.

Page 8-11, paragraph on lines 20-26. As context, what is the national total mortality and national non-accidental mortality in 2007? i.e. by how much do these base rates differ, on average?

Page 8-12, Table 1.2: 2^{nd} and 3^{rd} column headers should make clear that the numbers given are three year cumulative estimates, not annual averages. Footnote 2 – could be more clear by adding 'applied to all 12 km by 12 km grid cells nationally' For clarity, what is the number of grid cells for city-specific versus national estimates?

Explain why LML is only 7.5 ppb if PRB is 29 ppb and if U.S. background is typically 25-40 ppb.

Also, explain and provide insight regarding the magnitude of the percent reduction from no concentration cut-off.

Page 8-15, Table 1.3. The basis is not clear and also is not the same for the two sources of risk estimates. Are both of the sets of results based on Bell et al. and Zanobetti and Schwartz on the basis of total mortality? The Bell et al. (2004) study was based on non-accidental mortality, not total mortality. Clarify the basis. Also, why are minimum and maximum values given, rather than the bounds of a 95 percent range (i.e. 2.5^{th} and 97.5^{th} percentiles of U.S. counties). Either in the table caption or footnote indicate the number of counties.

Page 8-18, lines 10-15: This text should be a new paragraph. There needs to be a transition from the previous paragraph explaining the purpose of making comparisons to the 4^{th} highest 8-hr daily maximum concentration. Furthermore, it needs to be crystal clear as to what exactly is being compared here. Is this a 4^{th} highest value in just one year? Is it the three year average of the 4^{th} highest values? If not the latter, then the subsequent inferences about the ppb ranges are not based on the form of the current standard.

Page 8-18, lines 15 –end of page: again, what is the basis here? Is this a 3-yr average of 4^{th} highest values? Or just the 4^{th} highest value in one year?

Page 8-20, Figure 1.11: similar to previous comment, what exactly is the basis of the numbers shown on the y-axis in this figure and in Figure 1.12? Is this just the 4^{th} highest value in one year, or is it the three year average of 4^{th} highest values from 2006 to 2008? Likewise, is the x-axis based on May to September averages for one year only, for each of three years, or is it a three year average?

Page 8-21, middle of paragraph – clarify that these numerical estimates are based on cumulative mortality over three years.

Page 8-22, line 5 – again, clarify exactly what is the form of this quantity.

Page 8-22, line 8 – this seems to be a one-year estimate, in contrast to numbers reported elsewhere in this chapter that appear to be three year estimates. Be clear as to the basis of these numbers.

Page 8-22, middle of page – this is the first time in this chapter that mention is made of the 'concentration cut-off' of 29 ppb. What is the basis for 29 ppb, and how does it enter into the analysis and how does it affect the results reported in Table 1.2. What other results are based on this number?

Page 8-22, end of long paragraph. After reading this paragraph, differences between two estimates are explained. Presumably, EPA staff deem their own estimates to be more valid or policy-relevant than those reported by Fann et al., 2012. Some synthesis of information that leads to this conclusion would be helpful to the reader: i.e. which analysis is better (or more relevant) and why?

Page 8-22, section 8.2: the way that the first few paragraphs are written, it appears as if this section is regurgitating material in Chapter 7. The tone of the text should be modified to make clear the purpose, the scope of this section, which aspects of the scope are drawn upon material in Chapter 7, and which aspects are in this section.

Page 8-23, middle of last paragraph: some discussion of why better characterization at the high end of risk is desirable would be… well, desirable.

Page 8-24 "impacted" is not the right word to use here… "affected" or 'influenced' seem more appropriate.

Page 8-25: end of first paragraph: needs clarification and discussion – i.e. what are the parameters, which are proxies and how, etc. – or refer to more details given later so reader knows this is coming.

Page 8-25, end of 2nd paragraph, cite the figures that include the city-specific mortality risks.

Page 8-30, Table 1.5: the basis of some of the numbers is not clear or could be more clear. For example, population of what (per county?)? Income – per capita?

Page 8-31: O3 4th highest maximum 8-hr average unit is given as ppb but the numbers in the table appear to be ppm. Indicate for what time period are these 4th highest values, and additional information on the form – is this the 4th highest value in just one year, or is it the three year average of annual 4th highest values?

The text should explain why the respiratory mortality O3 C-R and cardiovascular mortality O3 C-R estimates for the urban study areas and U.S. data set are higher than the all cause mortality C-R estimates – presumably because these are based on a smaller baseline population.

Page 8-32, and similar figures: the first part of these captions should be changed – "Comparison of County-Level Populations of Urban Case Study Area Counties to the Frequency Distribution of Population in 3,143 U.S. Counties" The title over the chart can be deleted. Similar comment for Figures 1.14 through 1.27.

Figure 1.15: The numbers here appear to be based on just one year – this should be clear in the figure caption.

Page 8-44: 6 to 2 lines from bottom: this text gives the reader the impression that the urban study areas should be expanded to include higher mortality rates and older populations. However, on next page, it becomes more clear that this is not really necessary or useful – the text should be better organized to raise a point and dispatch it and not leave this kind of lingering doubt. I.e. write a paragraph on the issue of high baseline mortality that introduces the observation but provides the counter factual information that leads to the conclusion that the current set of urban areas is sufficient. Then do the same for population age.

Page 8-49: the first paragraph appears to be a summary and therefore can be deleted. The second paragraph seems to provide a good discussion.

Comments on Chapter 9 - Synthesis

Charge Question 19. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall health exposure and risk analysis?

The synthesis is a useful part of the document in that it provides a summary of the key findings from the several types of risk assessments, including the exposure and controlled human experiment-based assessment, the short-term epidemiological assessment for individual cities, and the national scale epidemiologic-based assessment. This is clearly a work in progress. For the next version, please develop summary reporting tables that contain the key numerical results. The text is a bit tedious to read because it is essentially writing a table in text format. It would help to clearly organize the text so that the reader who understands the table can skip the tedious text that describes the table (i.e. have subsections on "Description of Results") to get to new sections that should be labeled as "Discussion of Results" – these latter sections should provide interpretation, implications, and integration.

Detailed Comments on Chapter 9:

Page 9-1, lines 2-6. The intro paragraph fails to mention the national scale assessment of Chapter 8 or the assessment of the representativeness of the selected urban areas, also given in Chapter 8.

Page 9-1, Section 9.1. The elderly are also considered to be an important at-risk group. Explain why this group was not the focus of the risk assessment.

Page 9-1, lines 33-34: It is not really the case that APEX lacks 'proficiency' to model repeated (longitudinal?) exposures, but rather the root cause for this situation is the lack of activity diary data that would enable such repeated exposures to be quantified. (The same language appears on page 9-11, lines 22-23). Thus, this is really more of a limitation of CHAD and of diaries in general than it is a limitation of APEX per se. The lack of such longitudinal information and its effect on results can be explored in sensitivity analysis in which the same diary is used every day for a given simulated child and compared to results in which diaries are sampled at random every day. The bounds of these two analyses would illustrate to what extent the results might depend on repeated exposures and to what extent the base case estimates given here might underestimate the number of acute adverse effect outcomes.

The summaries on Pages 9-2 through 9-5 are reasonably well-written, considering that this is very 'dry' and repetitive material. As noted above, please also include a summary table. Clearly identify the portions of the text that are merely describing the results and clearly identify parts that go beyond mere summaries to more interesting discussions of interpretation, implications, and integration.

For the epidemiologic bases studies, please address the following topics:

- LML versus North American background
- Why the LML-based effect estimates are lower than those for which the LML is not considered (this should be further detailed in Chapters 7 and 8). Is beta independent of LML?

- For consistency among all endpoints, please report mean values and the probability ranges of the effect estimates (not just the ranges), and include all of these quantitative results in a summary table.
- To support the later 'observations' (in section 9.4) that 2007 and 2009 represent worst and best case years with in the 5 year period, also include quantitative summary information regarding air quality in each year from 2006 to 2010 (e.g., averages and ranges for each city for the exposure-based estimates, and other useful metrics for the epi-based estimates). This would also support observations about year to year variability in ozone concentrations.
- The 'observations' should include discussion of what was NOT modeled – i.e. what health effect outcomes identified in the ISA with causal or likely to be causal determinations were not modeled.
- The 'observations' should include limitations of what was modeled – e.g., that the exposure-response relationships from clinical experiments might not address the most at risk subpopulations or the most severe outcomes.
- What can be said qualitatively about the possible biases in the quantitative risk assessment or about risks that were not estimated quantitatively?

Dr. David A. Grantz

Comments on the Welfare Risk and Exposure Assessment

Comments on Chapter 5 - Ecological Effects

7. Regarding the assessment of relative biomass loss (RBL) for individual species:
a) What are the views of the Panel on the use of the linear model forced through the origin to assess the proportional relationship between the relative biomass loss (RBL) values for each species comparing the RBL at recent ambient conditions to the RBL under the scenario modeling O3 just meeting the current standard?

I think the use of a linear (is it regression or correlation?) analysis is a poor choice because it implies a relationship that has no real basis. The text suggests several alternative approaches, including a mean ratio which would be superior. In the end, the only number retained from the linearization approach is the slope, which is taken as the mean ratio.

b) To what extent does the Panel find that this an appropriate analysis to compare the proportional changes in RBL? Does the Panel have suggestions for alternative approaches that provides a comparable result and maintains the cell-by-cell approach to help control for environmental variability?

As above, the mean of within-cell ratios is more straightforward.

c) To what extent does the panel agree with the approach used to combine the 11 tree species into one analysis?

For purposes of risk assessment it is essential to aggregate across species and landscapes as done in this document. The combination in Figure 5-7 appears straightforward and meaningful. The calculation of the statistics in this figure was not clear in the text, but I am assuming this was the distribution of ratios from the linearized plots discussed above.

8. Regarding the assessment of RBL for combined species:
a) To what extent does the Panel support the use of the Importance Values from the U.S. Forest Service to weight the RBL values in extrapolating from individual trees to larger ecosystem level effects?

I think this is very important, and well realized in the document.

b) What are the views of the Panel on the use of the summed-RBL as a metric to use for assessing

effects at the larger ecosystem scale?

I see this as a first cut at the problem. There are certainly more precise measures that could be devised, but current data limitations restrict their application. At the current time I think this approach is a good place holder for aggregating these types of impacts, and a potential spur to relevant research.

c) Does the panel have any recommendations for methods to include a wider range of tree species (beyond the 8 species included)?

The limitation to including more species is availability of C-R functions. When existing C-R functions are extrapolated to novel species, it is important to provide a range of possible sensitivities, rather than assume the sensitivity of the most sensitive species. This will provide a range of possible impacts, and is more conservative and more credible when uncertainty is propagated. An exception is if for taxonomic or biochemical reasons one can suggest similarity to the C-R relationship with a known species. The only alternative to the preceding approach is to find existing or create new C-R relationships with new species.

9. What are the views of the Panel on the use of federally designated Class I and Critical Habitat areas as endpoints for this analysis? Does the Panel have any suggestions for additional parks beyond Rocky Mountain National Park and Sequoia/Kings Canyon National Park that should be analyzed? Does the panel have recommendations for additional or alternative geographic analysis areas that could be used?

Class I areas by their definition deserve as much scrutiny as possible. They are very appropriate areas for analysis (not sure they are “endpoints”). Critical habitats require much more justification and analysis to justify their inclusion—both as to target species and nature of criticality. This may not be possible in this document.

The choice of parks seems excellent, but isn’t Great Smoky Mtn. NP in the analysis now? Acadia would be the most likely next one to add.

10. To what extent does the Panel find that the vegetation mapping data from USGS and NPS used to generate a scaled-RBL surface in Great Smokey Mountain National Park is appropriate?

Conceptually very appropriate.

11. Regarding the screening level foliar injury risk assessment:
a) To what extent does the Panel find the updated assessment of foliar injury risk in national parks originally performed in Kohut (2007) to be an appropriate screening level risk assessment?

I think it is limited, but as good as anything now available. It is an efficient way to screen a large number of sites. I think the use of multiple ozone metrics to screen is not useful. For this document I

recommend that EPA make a choice (presumably W126) and stick with it throughout the analysis.

b) What are the views of the Panel regarding the potential methods for estimating O_3 exposure at additional parks?

Use of the highest concentration monitor in a park is not very desirable. The original kriging method makes some sense, and the use of the data-model fusion approach suggested here may be better if validated. In future, the impacts of elevation and terrain will be required, and national parks provide a good set of locations to test new approaches.

c) What are the Panel's views regarding the appropriateness of requiring that two criteria must be satisfied (i.e., based on both W126 and N100) in order to receive higher risk ratings?

Not appropriate. See above.

d) Is the Panel aware of any assessments of foliar injury in national parks conducted between 2006 and 2010 that could potentially be used to validate the updated risk ratings?

No.

12. Regarding the assessment of cover of O_3 sensitive species:
a) To what extent does the Panel find the preliminary analysis of sensitive species cover to be an appropriate and useful approach to highlight areas of potentially higher risk due to the presence of sensitive species?
Unclear. Is this Figure 6-8?

b) To what extent does the Panel find the vegetation mapping data appropriate to assess the cover of O_3 sensitive species in GSMNP?

c) What are the views of the Panel on the decision to not distinguish between vegetation strata (i.e. herb, shrub, tree)? To what extent does the Panel agree with this methodology relative to analyzing the strata individually?

I think there is a high priority in this document to aggregate results for ease of communication to end users. Therefore I grudgingly approve of the use of single stratum analyses, even though they are clearly and demonstrably wrong. Ozone is depleted vertically through canopies, understory species are in dim light but improved water status leading to uncertain and site-specific differences in stomatal conductance to ozone, and the species are different with contrasting leaf morphologies. There is little reason to believe they will respond as do overstory trees.
d) What are the views of the panel on using benchmarks, similar to those used in the Kohut analysis of foliar injury risk, to allow estimates of change between exposure scenarios?

Unclear. Benchmarks or air quality thresholds for injury are a short-hand means of moving types of

vegetation between risk classes. If continuous functions can be developed this would be superior, but data may not be sufficient. The specific thresholds (benchmarks) suggested by Bob Kohut may not be invariant on a national scale, whereas a continuous function might be evaluated locally.

Comments on Chapter 6 - Ecosystem Services

13. To what extent does the Panel find that EPA has adequately characterized the range of ecosystem services that are potentially adversely affected by O3?

I think this is a pretty good job and consistent with previous analyses.

14. To what extent does the Panel agree with EPA's ecosystem services framework, connecting O3 exposure, through ecological effects to ecosystem services?

The framework is conceptually fine. I found the presentation of Chapter 5 and Chapter 6 to be confusing and somewhat tedious to read. Is it possible to combine, so that the thread between exposure-effects-services is accomplished all at once for each combination of factors?

15. Does the panel agree with EPA's use of combined O3 exposure data with other data sources (e.g. fire data, bark beetle maps, trail maps) to link areas of concern or interest with areas of higher vegetative risk due to O3? Does the Panel have any recommendations for additional datasets and ecosystem services that could add to or improve these analyses?

I think incorporating anything linked to ozone is appropriate. I disagree with the conclusion that bark beetle damage (monetized) cannot be disaggregated into an ozone-induced component. It may not be quantitative at this point, but it would be a useful placeholder for further analysis. It is not accurate to state that because total damage is included in the bark beetle estimate, that this precludes teasing out the ozone-induced component.

16. Regarding the analysis of forest yield impacts:

a) To what extent does the Panel agree that the Forest and Agricultural Sector Optimization Model (FASOM) model is appropriate to assess timber and crop yield changes and the effects of those changes on additional ecosystem services?

b) What are the views of the Panel on the extrapolation of concentration-response functions across similar species?

See comments above. I think as much generalization as possible is useful, but with uncertainties propagated through the analysis. Use of similarities among species to assign highest likelihood sensitivities could be useful.

[The appendix detailing the FASOM analysis will be submitted for review in August]

17. Regarding the analysis of urban forest impacts:
a) To what extent does the Panel feel that the i-Tree model is appropriate for assessing changes to urban forest ecosystem services based on O3 exposure?
b) In order to increase the number of tree species covered by the iTree model, does the Panel have recommendations for additional species that could be included, based on estimates from similar species?

[The appendix detailing the i-Tree analysis will be submitted for review in August]

18. Regarding the use of PnET-CN:
a) What are the views of the Panel on the potential use of the PnET-CN model in the 2nd draft to assess impacts on larger scale ecosystem services (e.g. hydrologic changes, c sequestration)?
b) Does the Panel have recommendations of other models that are accessible to EPA that could be used instead of PnET-CN?

19. Regarding ecosystem services related to foliar injury:
a) To what extent does the Panel agree that potential visible foliar injury is appropriate to use as a metric to assess potential loss of cultural services associated with recreation in national parks?

The types of canopy death attributed to ozone-enhanced bark beetle infestation are clear aesthetic degradations noticeable to visitors. I think visible ozone injury is also negative but perhaps not so much so. It is however, a reasonable endpoint for quantifying loss of cultural services. It has been used for a long time as a surrogate for many ozone impacts, even though the quantitative correlations are known to be only moderate.

b) Does the Panel feel that there are O3 benchmarks that could be used to assess changes in foliar injury potential between exposure scenarios similar to those used by Kohut (2007)?
See above.

Comments on Chapter 7 - Synthesis

20. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall welfare exposure and risk analysis?
The synthesis was mercifully brief and to the point. I thought it provided a clear wrap-up for the reader. I would have liked some numbers to appear, in effect highlighting the few quantitative results that were obtained in the document.

1
2

Dr. Jack Harkema

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 6

Question 1

The methods used to conduct the risk analysis are technically sound. However, in light of the recent papers by McDonnell et al. and Schelele et al., Inhal Toxicol 2012; 24:619-633 and 24:401-415, respectively, controlled human exposure/response data should also be fit to a third model that contains a threshold below which no response occurs using a statistical program for fitting nonlinear mixed models (i.e., threshold model). McDonnell et al. reported that this new threshold model is likely to provide more accurate estimates of risk in future risk assessments of ozone-induced FEV1 effects.

Question 2

The outline of this chapter appears appropriate. The approach used is clearly and concised presented and adequately defended by the authors. The data used in the previous review and the current review are clearly presented in the tables. As suggested above, the threshold model should be compared with the other two models in the second draft.

Question 3

The assessment on lung function (FEV1) decrements in the quantitative risk assessment appears to this panelist to still be the most appropropriate and informative single parameter for estimating overall health risk to ozone exposure.

Question 4

See responses to questions 1 and 2. The model that best fits the data, particularly at low levels of exposure, should be identified by the authors' in the second draft.

Question 5

The authors' defense of the application of E-R function for all lifestages on 6-49-50 appears justified based on the scientific literature to date.

Question 6

The qualitative discussion of uncertainty and variability is very important. The authors' have identified the major areas of uncertainty. The strengths and limitations of the each of the human exposure-response

models could be summarized both in the text and in a table for better clarification for those who are not experts in this area.

Question 7

Others more familiar with this field may have some suggestions on additional sensitive analyses or other approaches to addressing uncertainty and variability.

Dr. Daniel Jacob

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 4 - Air Quality Considerations

5. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

The 5-year period seems very reasonable.

1. 4-5, lines 15-17: I don't understand the logic for excluding exceptional events. From a health perspective it doesn't matter if an event is "exceptional" or not.

2. 4-6, lines 10-11: What is the rationale for spatial averaging? Some monitors in the urban area are likely more representative of population exposure than others.

3. 4-8, lines 1-13: the use of GEOS-Chem background-to-base ratios to derive background from observations is a little strange, as it assumes that this ratio is the most robust result from the GEOS-Chem simulation. I see no basis for this assumption, as errors on the background and errors on the US pollution enhancement are likely not correlated. It may lead to odd results for any sites where the model makes large errors on the US pollution enhancement. I would have suggested just taking the background values from GEOS-Chem since the ISA finds them to be overall unbiased. However, it may not make much difference in practice.

4. 4-13: I understand that section 4.3.2 shows national distributions of ozone as may be used to diagnose the effect of minimum emission reductions to meet the NAAQS. But then I'm puzzled about the Figures shown, which are seasonal averages of no direct relevance to the NAAQS. Wouldn't it be better to show the design values (4^{th} highest annual MDA8)? It would seem much more relevant.

5. 4-14: I'm concerned about the fused data map shown in Figure 4-6 and how much it differs from Figure 4-7 (which seems more reasonable). Figure 4-6 doesn't show any evident enhancements in urban areas (why?) and seems to overestimate background in high-elevation regions. It would be useful to document here the relative weight of observations and CMAQ in this fusion. It would also be useful to show the difference between the fused data and the actual observations, either as a map or (maybe better) as statistics for selected sites both remote and urban. Some explanation for why Figure 4-7 is so different from 4-6 would also be helpful.

6. Regarding the methods for simulating just meeting the ozone standard:

a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard (including application of US background as a lower-bound on rollback) is a reasonable approach?

It has the advantage of simplicity and transparency, and of being based on observations, but it involves simplifying assumptions that may be hard to defend and that could produce some odd results.

b) To what extent does the Panel support using an air quality model based approach for simulating just meeting the standard in future drafts as a replacement for the current quadratic approach?

I think that would be a better approach because that is in practice how SIPs are developed to meet an AQ objective. However, there are different combinations of emission reduction strategies possible to achieve just meeting the ozone standard and these may have different impacts on overall ozone distributions. So one would need a protocol for choosing one in particular.

c) What are the views of the Panel on the strengths and limitations of the proposed approach using the Higher-order Direct Decoupled Method?

HDDM would have the advantage of providing a protocol for the minimum emission reductions needed to achieve just meeting the ozone standard. The sensitivities would have to be expressed in terms of emissions from particular sectors so that the effects of these emission reductions on ozone can be propagated to the whole domain. The HDDM sensitivities should be applied to the observed concentrations (not the model fields). I think that this is the best approach available to EPA.

Comments on the Welfare Risk and Exposure Assessment

Comments on Chapter 4 - Air Quality Considerations

4. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

I think that the use of 2006-2010 is appropriate.

1. 4-4, line 1: I don't understand the logic for excluding exceptional events. They would have a welfare impact, even if they can't be controlled. The 2008 California wildfires (which accounted for all the exclusions according to the text) seem like an odd exclusion since Singh et al. (Atmos. Environ. 2010) showed that these fires didn't make significant ozone unless mixed with urban plumes.
2. 4-4, line 1: Figure 4-3 shows a W126 hotspot from the wildfires in central Idaho. Why weren't they removed as exceptional events, like the California wildfires? That seems inconsistent but I think I

know why – that's because there were no observations to exclude! The W126 hotspot in central Idaho is in my opinion a model artifact and I elaborate on that below.

5. What are the views of the Panel on the approach used to develop a national scale surface of W126?

1. I would like some more information on the data fusion approach, because some aspects of the results seem weird (see my comments to charge question 5 on the REA Health Draft 1). I think that it would be important to show comparisons of W126 in the fused data and in the actual observations for specific sites, and I would recommend using the CASTNet sites for that since they are so relevant to the secondary standard.
2. In particular, I'm surprised that Figure 4-3 shows such high W126 values over the intermountain west. This must largely reflect CMAQ information since there aren't many monitors there. However, I have some doubts as to the quality of the CMAQ simulation in that region. For example, the W126 hotspot in central Idaho is due to wildfires, but that just reflects the simulation of large ozone production from wildfires in CMAQ. There is evidence from the literature that models overestimate ozone production from wildfires (Singh et al., Atmos. Environ. 2010; Alvarado et al., JGR 2010). My own work analyzing the CASTNet observations shows no ozone enhancements from wildfires. For example, I see no ozone enhancements at Glacier National Park from the fires in central Idaho. I'm ready to acknowledge the controversy over ozone production from wildfires, but that would translate into large uncertainty in the national scale surface of W126.
3. Following on the above point, I think that there is enough uncertainty in CMAQ simulations of background ozone that using CMAQ to extrapolate ozone data to the national scale may be inappropriate for a W126-based standard. This can be assessed by comparing W126 in the fused data to observations at the CASTNet sites, as per above comment.

6. Regarding the methods for simulating just meeting the ozone standard:

a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard is a reasonable approach?

1. If the quadratic rollback approach is used then it should be implemented in the same way as for the Health REA. Right now it doesn't seem to be, notably in the correction for background. Background correction is probably more important for the welfare assessment than for the health assessment.
2. The quadratic rollback approach seems to do weird things. For example, it decreases the W126 in central Idaho even though the high W126 there is from wildfires that are not evidently controllable.

b) Does the Panel have suggestions for alternative approaches for simulating just meeting the current secondary standard or alternative standards based on the W126 metric?

1. The HDDM was proposed as a better alternative to the rollback method for the Health REA, and I endorsed this in my response to the charge questions for that REA. I think it would be a better alternative for the Welfare REA as well, and would address for example the above-mentioned weird result of the rollback method in central Idaho. See my responses to the charge questions on chapter 4 of the Health REA for further comments on application of the HDDM.

Dr. Steven Kleeberger

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 6 – Characterization of Health Risk Based on Controlled Human Exposure Studies.

1. To what extent does the Panel find the methods used to conduct the risk analysis to be technically sound? What are the views of the Panel members on the methods used?

Two models were used in the analyses. The first method was also used in the previous review to provide population level estimates of percent and number of people at risk. The second analysis is an apparently improved model to estimate FEV1 responses in individuals with short-term exposures. The rationale to use both methods appears appropriate, and differences and similarities between the analyses provided by both approaches are discussed.

2. To what extent does the Panel find the assessment, interpretation, and presentation of the results for the risk analysis as presented in Chapter 6 to be technically sound, appropriately balanced, and clear communicated?

Results of the risk analyses are largely sound, balanced, and clearly communicated. The authors indicate where deficiencies occur, and in some instance indicate that additional analyses will be done for the next REA. Some other comments are as follow:

a. Table 6-1. The cause for considering as outliers the data with double asterisks in the last line of the table was not clear. They do not seem to be different from 30/12 or 30/13. Perhaps some clarification would help.

b. Table 6-4. The data presented in the table were, at first read, somewhat confusing. I now understand what the numbers represent, but percentages of percentages were not intuitive. Perhaps a different representation could be used?

c. Page 6-47, lines 15 and 17. I believe the reference to Figure 6-7 should be Figure 6-8.

d. Some of the figures do not have adequately labeled X and Y axes (e.g. 6-5, 6-6, 6-7), and they should be added.

e. Page 6-45, line 28. Do the authors mean “inter-individual” instead of “intre-individual”?

f. Page 6-49, line 1. Authors suggest that lung function response appears to level after 6 hours of exposure, and that it is unlikely that longer exposures (8 hr) would change exposure-response

relationship. However, if one looks at the figures in the Schelegle paper, one could argue in fact the responses are still increasing and they come down only when the individuals discontinue exposure.

g. Page 6-49, line 29. I believe the authors refer to Figure 6-7, not 6-4. Also, instead of "which might indicate" I would rephrase the sentence to hypothesize that responsiveness in the age range of children would continue to increase.

h. Page 6-51. The Graham, 2012 citation is not listed in the references (section 6.5).

3. To what extent does the Panel find the focus of the assessment on lung function decrements in the quantitative risk assessment to be appropriate and informative?

I believe the focus on lung function decrements is the appropriate endpoint for the assessments because this endpoint is the most reported measure across many studies. It may not be the most important response to ozone (e.g. inflammation may be equally or more important), but it is the response that is easiest to measure. It may be useful to include a statement at the end of section 6.1.1 that the multiple phenotypes elicited by ozone exposure likely have different mechanistic underpinnings and do not necessarily correlate or cosegregate. Therefore, risk assessments may be different depending on which phenotype is used in the assessment.

4. What are the views of the Panel on the use of the two different modeling approaches for specifying the exposure-response function linking the change in FEV1 to ozone exposure?

I think the use of two very different modeling approaches is important, and the authors adequately justified both for their assessments.

5. What are the views of the Panel on the treatment of the relationship between age and dFEV1 in the McDonnell-Stewart-Smith model?

As indicated by the authors the risk estimates are considerably higher using the M-S-S model compared to the other model, and perhaps to be expected since the models use different approaches. However, it is important to understand why the differences exist, and the authors indicate that they will present analyses of the difference in the next draft REA.

6. To what extent does the Panel find that the qualitative discussion of uncertainty and variability has covered important sources of uncertainty and variability and has appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates?

I think the authors offered a very reasonable discussion of the sources of uncertainty and variability. I noted some comments about this discussion in my response to question 2.

7. What are the views of the Panel on additional sensitivity analyses or other approaches to addressing uncertainty and variability?

While this is not my area of expertise, I am not certain that additional sensitivity analyses would be particularly useful.

Dr. Frederick J. Miller

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 1 – Introduction

1. Does the Panel find the introductory and background material, including that pertaining to previous reviews of the O_3 standards and the current review, to be clearly communicated and appropriately characterized?

While the introductory chapter usually communicates clearly the information relevant to the previous and current O_3 reviews, there are some sentences that are misleading or inaccurate. For example, on page 1-5 line 17, the document sates that the proposed range considered for the primary standard was 0.070 to 0.075 ppm. This was clearly not the range proposed by CASAC. The current text implies that this was the case and does not clarify the range until line 2 of page 1-6. In addition, the statement on page 1-6 is not correct as to "…The Administrator explained in part that CASAC appeared to place greater weight on the results of the risk assessment as a basis for its recommended range, while he more heavily weighed the implications of the uncertainties associated with the exposure and risk assessments." The range proposed by CASAC was the result of the scientific evidence concerning the nature and severity of effects reported for O_3 at levels below the current standard, which at that time was 0.08 ppm, and was not heavily influenced by the exposure and risk estimate analyses.

There is no mention in Chapter 1 of any consideration of the secondary standard other than to state that it is the same as the primary standard. To appropriately characterize the previous discussions on the secondary standard, the authors should have noted that a different form of the O_3 standard was proposed by CASAC, but the Administrator chose to ignore CASAC's recommendation and set the secondary standard equal to the primary standard.

Comments on Chapter 2 - Conceptual Model

2. To what extent does the Panel find that the discussions accurately reflect and clearly communicate the currently available health effects evidence, and the relevance of that evidence for quantitative exposure and risk assessment, as characterized in the 3rd Draft ISA?

This chapter does an excellent job of presenting information on the sources of O3, the various microenvironments to consider for exposure, identifying the at-risk populations, and discussing which health endpoints are most suitable for inclusion in a risk assessment. There are a couple of minor points that should be clarified

On page 2-7 line14, the authors talk about appropriate concentration-response functions in the case of epidemiological studies and exposure-response functions for controlled human exposure studies. To this reviewer, they are both exposure response studies, and, if anything, the epidemiology studies could be

called exposure-response and the controlled human studies concentration-response. The authors need to make clear how they are establishing a distinction in terminology for epidemiology studies versus controlled human studies. To this reviewer, the terms defined in Zartarian et al. (J. Exposure Analysis & Environ. Epi. 15:1-5, 2005) do not imply a clear distinction for when exposure-response is appropriate to use instead of concentration-response. Since the intermingling of the two terms is used throughout the Health REA, the Agency needs to better define the meaning and applicability of the two terms. Also, on beginning on line19 on page 2-7, the authors talk about at-risk populations and contend that controlled human exposure studies are in fact clearly biased away from highly susceptible individuals. While in general this is the case, controlled human studies have been done on at-risk populations such as asthmatics and individuals with COPD. The text here should be reworded to better convey that at-risk subjects are sometimes studied but this is not the norm.

Figure 2-1 is confusing because of the placement of the terms “short term” and “long term” exposures. This reviewer suggests adding brackets in front of the listing of the health endpoints or some other way to convey that the top part of the figure relates to short-term exposures and the bottom portion relates to long-term exposures.

Comments on Chapter 3 - Scope

3. Does the Panel find the scope of the health risk and exposure analysis is clearly communicated?

This chapter does an excellent job of laying out the scope of the health risk and exposure analysis together with all of the options, caveats, and considerations that are presented in the 1st draft compared to what will be done in the 2nd draft. This reviewer has only a few points that need changing or clarifying.

On line 10 of page 3-8, the authors state that CASAC recommended that EPA move away from using PRB in calculating risk. The CASAC stated in 2007 “Finally, with respect to policy-relevant background (PRB), the Ozone Panel wishes to point out that the Final Ozone Staff Paper does not provide a sufficient base of evidence from the peer-reviewed literature to suggest that the current approach to determining a PRB is the best method to make this estimation. One reason is that part of the PRB is not controllable by EPA. It would require international cooperation beyond the bounds of North America. A better scientific understanding of the PRB and its relationship to intercontinental transport of air pollutants could serve as the basis for a more concerted effort to control its growth and preserve the gains in air quality achieved by control efforts within the U.S. In any case, there is no apparent need to define PRB in the context of establishing a health-based (primary) ozone NAAQS. The effects of inhaled ozone on decreases in respiratory function have been seen in healthy children exposed to ozone within ambient air mixtures in summer camps (1–6). Furthermore, the concentration- response functions above 40 ppb are either linear, or indistinguishable from linear. Thus, PRB is irrelevant to the discussion of where along the concentration-response function a NAAQS with an 8-hour averaging time that provides enhanced public health protection should be.” In the view of this reviewer, this statement was made at that time because there did not appear to be an ability to identify a threshold for pulmonary function effects due to O_3 exposure. Such may not be the case now.

On line 8 of page 3-13, the ≥ 24 should be ≥ 15 to be consistent with what is stated on line 19. In addition, some places the authors use ≥ and in others >, so there is also a need for consistency in this usage. On line 10 of page 3-23, the authors state that the use of both estimating risk down to zero and to the LML provide a reasonable bound on estimated total risk. It would be worth rewording this sentence to note that estimating risks down to zero O_3 exposure mathematically provides a "greatest upper bound" on total risk

4. Based on information in the 3rd draft ISA indicating lack of evidence for a threshold in O_3 concentration-response functions, we have included risk estimates down to zero O_3 concentrations. Based on further discussion in the ISA regarding the decreased confidence in the shape and magnitude of population health response at very low O_3 concentrations, we have also included risk estimates based on applying concentration-response (C-R) functions only down to the lowest measured level (LML) in the underlying epidemiology studies.

a) To what extent does the Panel support the use of two different risk estimates, one applying the C-R function down to zero, and one applying the C-R function down to the LML, to characterize the range of risk estimates to balance comprehensiveness of the estimates with confidence in the estimates?

Given the background levels of O_3 that cannot be controlled by U.S. regulatory actions, this reviewer endorses applying the C-R function down to the LML and does not support obtaining risk estimates down to zero. However, surrogate estimates of the LMLs should only be used if EPA cannot obtain the actual LMLs associated with the studies underlying the C-R functions. Moreover, the recent paper by McDonnell et al. (2012) clearly establishes the statistical significance of a threshold model for O_3 FEV_1 responses compared to a non-threshold model. While developed using data from practically all of the chamber clinical studies that have been conducted, the model would also be directly applicable to functional changes seen in future epidemiology studies that measure such changes or in putting some of the children's camp study results into perspective.

Just because the epidemiology studies are not able to define a threshold for O_3 effects for the mortality, hospital admissions, and other effects does not mean that a "biologically effective threshold" does not exist. This issue becomes a statistical one that epidemiology studies have a difficult time trying to establish. However, most biomedical scientists would argue that there is a threshold.

b) What are the views of the Panel on alterative cutoffs based on other points within the distribution of O_3 concentrations used in the underlying epidemiology studies?

There is a high probability that the LMLs will often be lower than the NA policy relevant background (PRB) levels estimated from the CEOS-Chem/CAMs model. Since a national map of the NA levels are available as model outputs, this reviewer would like to see risk estimates for exposures above the NA PRB levels incorporated into the city-specific risk characterization analyses for the various health endpoints the Agency is intending to conduct.

Comments on Chapter 6 - Characterization of Health Risks Based on Clinical Studies

General Comments

Chapter 6 is well written and reasonably easy for the reader to follow. Use of the McDonnell et al. (2010) model is appropriate in addition to using the current probabilistic method. The McDonnell model for individual response functions is more likely to capture what is really going on in the population. Importantly, the Agency is aware of the recently available online paper by McDonnell et al. (2012) that clearly established there is a threshold for FEV_1 changes in adult humans exposed to O_3. At the earliest time point with low exposure levels, the McDonnell et al. (2012) threshold model tended to fit the data better than the non-threshold model. Moreover, using the AIC values in Table 2 of their paper, the non-threshold model is 0.00002754 times more probable (i.e., only 1 chance in 36,311) as the threshold model to minimize the information loss.

Staff indicated that they plan to use the more recent McDonnell et al. (2012) threshold model. However, for some of the analyses they want to conduct, one investigator is hampering their efforts because they are reluctant to allow the Agency to use their data. Such obstructionism serves no useful purpose.

Specific Comments

Page, line	Comment
6-5, 8	The reference is to Figure 6-1 not Figure 3-1.
Fig. 6-1 to 6-3	While the text identifies what the letters in these figures mean, the figure legends should also define them because the figures should be able to "stand alone".
6-10, 9	The "x" in the numerator and denominator of the equation needs to be a substantially larger font.
6-10, 19	To this reviewer, the 50/50 probabilities scenario is not needed. Given that one is toward the lower end of the population response curve, an equal splitting of the two response probability models is extremely unlikely to occur. McDonnell et al. make this observation even more defensible in light of their 2012 paper where they have extended their model to incorporate a threshold.
6-11, 13	The authors state that they selected the last 1000 parameter values from the linear model simulations to combine with the last 9000 sets of values for the logistic model. Why not select the 1000 sets of values randomly from all of the linear model runs? Is there any reason to believe that the process of sampling from the parameter distribution may yield a trend in the outcomes if the last 1000 iteration values are selected?
6-13, 7	On this line, Equation 6-1 should be "Equation 6-3".
6-15, 9	Additional equations or definitions of variables may be needed here because y_{ijk}, U_i, and ε_{ijk} do not appear in Equation 6-6 but are defined in the listing after the equation.
Fig. 6-8 to Fig. 6-19	To this reviewer, it would be easier for the reader to follow the expected lung function responses in the various cities if the information in the current figures were rearranged

	in a series of 3 figures for each city followed by a set of 3 figures for the next city and so on. For a given city, the first figure would show the results for School-aged Children, Asthmatic Children, and All people for a FEV_1 Decrement ≥ 10 % followed by the same set of panels for ≥ 15 % and then for ≥ 20 %. Then move on to the data for the next city.
Table 6-9	This reviewer does not understand how the ratios for the MSS to E-R models were obtained. They certainly are not direct ratios of the % responses listed in the rows of the table (the first entry is a ratio of 3.3 but 15%/5 % = 3. So the authors need to describe at least in the text exactly how the ratios are derived.
6-46, 18	The authors state 3 topic areas for which they intend to perform additional analyses for the 2nd draft REA and compare results to those obtained earlier. The first area involves updating the E-R functions with new clinical study data and compare the prior function and risk results. This seems unnecessary – the expanded database is more reflective of the true situation and is bound to give different results. So why waste resources on comparing the findings to earlier results. The 2nd area listed involves determining the relative importance of low-dose extrapolation by looking at the number and percentage of people experiencing adverse responses to low O_3 exposures compared to the total response for all exposures. If the Agency uses the recent McDonnell et al. threshold model, such an exercise is unnecessary. And the 3rd area of investigation concerning age-related response dependencies will be easy to do using the McDonnell et al. (2012) model.
6-50, 9	This paragraph on subjects in clinical studies having possibly expressed either an enhanced or an attenuated response based on prior exposure is misleading. The true responses for a single day of exposure to O_3 in the real world are always at least as great as what was measured on any given day in the controlled chamber study. We know that repeated daily exposures result in a lessening of the magnitude of FEV1 response. Thus, only "truly naive" exposure subjects can respond maximally to an O_3 exposure.
6-51, 25	This reviewer is highly skeptical of using the CHAD database for children's activity patterns. Unless very recent activity data have been added to this database, the activity patterns in CHAD are not very representative of those for the children of today. The paper by Brochu et al. (Human & Ecological Risk Assessment 12:736-761, 2006) may be useful for comparing the percentiles of daily minute ventilation arising from the probabilistic sampling of CHAD to recent data on physiologic daily inhalation rates in a study that included children and that used double labeling water measurements to compare time-activity-ventilation and metabolic energy conversion estimates. This article does not currently appear in Chapter 4 of the ISA and would need to be added to that chapter if the HREA authors determine that the Brochu et al. (2006) study would be useful for their analyses.
6-51, 27	The Graham (2012) citation is not in the Chapter 6 reference list.

Comments on Chapter 7 - Characterization of Health Risks Based on Epidemiological Studies

General Comments

This reviewer has 3 major concerns about the analyses in Chapter 7 and the proposed path forward for the 2nd draft. The first concern relates to the consistent use of the LMLs from epidemiology studies or surrogate values for the LMLs. Almost all of the LML values in Table 7-5 are below the PRB levels discussed in the ISA even though it is recognized that these values are not reflective of the overall air quality distribution. To this reviewer, little is gained by presenting percentage reductions in risk associated with LMLs. Moreover, the prioritization discussions at the September 11-13, 2012 CASAC O_3 Panel meeting reflected this view by giving a low priority ranking for the Agency's assignment of time and resources for activities related to the 2nd draft of the HREA.

My 2nd concern relates to the rollback procedure for simulating just meeting the current standard and its implications for risks above PRB levels. The Higher-Order Decoupled Direct Method (DDM) capabilities in the Community Multi-Scale Air Quality model are proposed for use in the 2nd draft REA. While using this approach eliminates the need to specify values for U. S. background values because they are incorporated directly in the modeling, it was not clear to this reviewer prior to the September 11-13 meeting whether the Agency could present risk results for values above these levels in addition to or instead of the no cut-off or the LML cut-off. The ability to do so was made clear at the meeting. Moreover, the advantages of using the DDM method in the risk analyses for alternative levels of the standard was endorsed by CASAC as being superior to the use of the quadratic rollback method.

The 3rd major concern relates to the need for the Agency to prioritize the various analyses and additional topics that are being proposed for inclusion the 2nd draft of the HREA. In the opinion of this reviewer, the Agency has neither the resources nor the time to explore adequately all of the areas mentioned in the chapter. During the September 11-13, 2012 meeting, staff prepared a list of activities being considered for the 2nd draft HREA encompassing all of the chapters. This reviewer is in agreement with the priority ranking feedback that the Panel provided the Agency.

Whenever there is a series of tables showing the results for different urban areas, it would make sense to move all of them to an appendix and just present in the main body the results for one area. However, that area's results should still appear in the appendix as well. This would allow the reader to more easily follow the development of the issues that the Agency is addressing in this chapter.

Specific Comments

Page, Line	Comment
7-37, 35	The use of regionally specific rather than national level effect estimates is clearly a step in the right direction. Thus, the recommendation by Smith et al. (2009) is a good one and should be pursued in the 2nd draft.
Table 7-6	The Agency cannot possible examine all of the cases listed in this table. EPA needs to identify a few of the uncertainty questions that are the most important ones to

	address? Some of the uncertainty questions would appear to be better handled by simulation analyses where the true distribution can be specified and sampling errors are imparted to examine the impact on model estimates.
7-46, 31	The authors state that the health effects of interested are listed at the beginning of this section, but Section 5 begins at the top of the page and does not list the health effects.
Table 7-7, 7-8	The heat maps are useful, but they are also confusing. The recent conditions data in Table 7-7 sometimes have lower mortality values that those in the "meeting the current standard" in Table 7-8. This gives the reader the impression that deaths are now occurring more frequently at lower O_3 levels. A more detailed explanation of how to interpret the heat maps is needed. It would also be worth noting that the differences in mortality numbers in the "Delta" column come from a subtraction of the values in Table 7-8 from Table 7-7 and are the result of rounding of the numbers in the separate tables.
7-47, 12	How is the Agency going to show risk via percentage reduction in mortality above the Policy Relevant Background levels?
7-67, 27	Why does the writing here come across as the authors being surprised by the finding that the sum of cardiovascular and respiratory does not equal total mortality for most of the urban areas? There is more to total mortality than just these two aspects. What am I missing here?
7-68, 18	The authors state that the O_3-attributable hospital admissions for asthma vary depending upon whether PM2.5 is included in the model. To this reviewer, no results should be presented that have not taken into account $PM_{2.5}$ at a minimum. Even though most studies show that O_3 is still significant even when PM is included in the model, I seem to recall that the magnitudes of the O_3 risk estimates are reduced.
7-72, 32	Obtaining the actual LMLs associated with the underlying C-R functions must be given a high priority for the 2nd draft HREA if staff get to this lower ranked activity. The accuracy of any analyses will be greatly reduced using the actual LMLs, and thus the variability in the estimates greatly increased, if surrogate values have to be used.
7-74, 26	**The findings described in this bullet provide a clear reason why the current standard is not adequately protective of public health.**
7-75, 1-8	Here, the authors convey that the incidence data is reflective of a certain number of events, with examples being listed of 20,000 to 29,000 vents for chest tightness or shortness of breath and 55,000 events for asthma exacerbation. The authors might consider conveying these finding also by assuming a number of days such as 120 as reflective of an O_3 season and then reporting the number of extra events per day. This might better help convey the shear magnitude of the extra burden on the health care system. However, if the recommendation of the O_3 Panel to model the entire year is followed, then the length of the O_3 season becomes moot.
Section 7.7.1	**A number of sensitivity analyses that may be looked at in the 2nd draft HREA are discussed in this section. Unless the Agency knows that they have the resources to address all of the items being considered, a prioritization of the types of analyses should be done. This reviewer would prioritize short-term**

	exposure related mortality followed by short-term exposure –related morbidity and lastly by interpolation of missing air quality data.
7-77, 40	Generating CIs for the delta risk estimates would be extremely useful and needs to be done.
7-81, 5	This reviewer agrees that there currently are too many limitations for attempting to generate risk estimates for long-term exposure-related respiratory morbidity.

Comments on Chapter 8 - National-Scale Risk Assessment and Representativeness Analysis

General Comments

While the material is generally well written, the chapter would benefit from redrawing some of the figures, cross-referencing the text to specific figures, and correctly numbering the figures and tables as 8.x with x taking on values 1, 2, 3, etc. Also, there is duplication of the major findings that could be eliminated. For example, on page 8-1, two paragraphs are devoted to presenting the results of the pre-mature O_3-related deaths, and then these same results are described again in Section 8.1.3.

No treatment of PRBs is again a concern to this reviewer because LMLs and down to zero exposure do not tell the whole story for the regulatory issues associated with this NAAQS pollutant. The percentage reduction in risk associated with different choices for alternative levels of the standard is useful information and background levels influence such calculations.

Specific Comments

Page	**Comment**
Table 1.2	This table could easily be eliminated. Most of the aspects of the table are already stated in the text, and the ones that are not currently in the text could easily be added.
Figure 1.9	This figure should be redrawn using for the no cutoff findings the solid blue line for Bell et al. (2004) and the solid red line for Zanobetti and Schwartz (2008). Then the dashed colored lines could be used to present the 7.5 ppb cutoff results. Currently, the figure is somewhat confusing.
8-23 last ¶	In this paragraph, the authors state that they find that the urban study did not capture areas with the highest baseline, those with the oldest populations, and those with the lowest prevalence of air conditioning, but there is no support given or a link to the subsequent section of the chapter where the definitive analysis is presented. These kinds of statements made without adequate support or referencing leave the reader puzzled.
8-25	No adequate explanation is given for why the authors deleted the two highest cities found by Bell et al. (2004). An explanation needs to be provided.
Table 1.4	The source of data for the population/square mile is given in the table. Is the value for this variable calculated using “as the crow flies” data or from topographical

	maps?
8-38	Appendix 4-A is referenced here but where is it? Here also is a good example of where statements are made without linking them to the figures that support the statements. If Figure 1.21 was inserted in parentheses in the sentence stating the urban areas do not capture the higher end of the risks for 65-year olds, the reader would not be left wondering how the authors arrived at this conclusion.
8-39	The reader has to take "on faith" all of the conclusions stated on this page. Are the supporting data in an appendix?
8-45	This reviewer would not call 0.08 ppm a low ozone level. Also, the authors may be overstating their case in the 2nd paragraph on this page.
8-45	Given the number of problems with using the ATUS data, the effort to be made in the 2nd draft on the use of these data should be given a low priority for resource allocation.

Dr. Howard Neufeld

Comments on the Welfare Risk and Exposure Assessment

Comments on Chapter 4 - Air Quality Considerations

4. To what extent does the Panel consider the years of air quality data to be appropriate for use in the exposure and risk assessment?

4-4, line 15 – not sure that "photochemically" is the appropriate word to use here.

Line 25 – you need to define "*N*"

4-5, lines 9-10: I think there is a mistake in the equation here, unless you defining N100 differently from how I have in the past. Should N100 be for hours $\geq$ or just >? With respect to the SUM06, isn't this definition an AOT06? Traditional SUM06 accumulate all hours $\geq$ 60 ppb, but don't subtract 60.

Figures throughout chapter, and all subsequent chapters: all of the figures showing kriged ozone surfaces, and also those showing biomass losses (any figure with a scale) are lacking units. All such figures should have units explicitly shown either on the figure itself or in the legend.

5. What are the views of the Panel on the approach used to develop a national scale surface of W126?

I agree with the methodologies used in this chapter.

6. Regarding the methods for simulating just meeting the ozone standard:
a) To what extent does the Panel find that the quadratic rollback approach used in the first draft REA for simulating just meeting the current standard is a reasonable approach?

Based on the explanations in the chapter, I think this is a most reasonable approach to use.

b) Does the Panel have suggestions for alternative approaches for simulating just meeting the current secondary standard or alternative standards based on the W126 metric?

No.

Comments on Chapter 5 - Ecological Effects

7. Regarding the assessment of relative biomass loss (RBL) for individual species:
a) What are the views of the Panel on the use of the linear model forced through the origin to assess the proportional relationship between the relative biomass loss (RBL) values for each species comparing the RBL at recent ambient conditions to the RBL under the scenario modeling O3 just meeting the current standard?

There are times when forcing a regression through the origin makes sense, especially when you have a situation where a value of zero for the independent variable mandates a zero value for the dependent variable such as crop yield vs acreage – you can't have any yield if there is no land. However, such regressions do not yield interpretable r^2 values, although one author (Eisenhauer 2003) suggests several alternative substitutions to get around this. The more important issue to address is whether or not RBL is linear all the way to the origin, or if it increases non-linearly near the origin, and then becomes linear the further away it is. If so, then perhaps a non-linear approach is called for. Citation: **Eisenhauer, J.G. 2003. Regression through the Origin. Teaching Statistics 25: 76-80**.

It is my understanding that the RBL is based on biomass losses relative to an ozone exposure of zero. Since there are no areas where background ozone is this low for extended periods of time, is this the relevant baseline on which to base the RBL? What about selecting a robust estimate of background from which to deduce RBL?

b) To what extent does the Panel find that this an appropriate analysis to compare the proportional changes in RBL? Does the Panel have suggestions for alternative approaches that provides a comparable result and maintains the cell-by-cell approach to help control for environmental variability?

I found the approach used for determining RBL to be logical and pragmatic, given the databases they had to work with. However, there is no citation of this paper**: Matyssek, R., H. Sandermann, G. Wieser, F. Booker, S. Cieslik, R. Musselman and D. Ernst. 2008. The challenge of making ozone risk assessment for forest trees more mechanistic. Environmental Pollution 156:567-582**. The authors make an important distinction when doing modeling which is that one must distinguish between uptake (flux density) and sensitivity per unit uptake. Thus, two trees could take up equivalent amounts of ozone, but differ in sensitivity, due to a variety of internal mechanisms, so it is the combination of these two parameters that ultimately determine the risk to trees from ozone.

Furthermore, the modeling does not seem to take into account the range of genetic variation within a species. Perhaps some probabilistic approach could be taken to account for this through the use of some weighting factor, so that when the RBL is scaled up to the landscape, it is modified somewhat by this accounting.

Also, was there any discussion of the quality of the data used to construct the C-R curves? For example,

I thought sugar maple was highly insensitive to ozone. The studies by Topa seemed to indicate that biomass losses occurred primarily at 3X ambient exposures. In her 2001 article in Plant, Cell and Environment, she states that there was no statistical difference in dry weight of seedlings in the 1.0X and 1.7X ambient treatments. This does not suggest a very sensitive species to ozone, yet it seems highly sensitive in the REA analysis. Why is this? How confident are we that the data used really do give us realistic RBLs? I was not sure of the citations or databases used to develop the 11 tree RBL curves – can that be provided?

c) To what extent does the panel agree with the approach used to combine the 11 tree species into one analysis?

I think combining the analyses of all 11 trees into one analysis obscures individual tree responses, and puts too much weight on extremely insensitive species. If the goal is to protect the most sensitive, this would tend to bias against that. Nonetheless, it is informative to see the range of responses among these species and the box diagram does a good job expressing that variation to the reader.

8. Regarding the assessment of RBL for combined species:
a) To what extent does the Panel support the use of the Importance Values from the U.S. Forest Service to weight the RBL values in extrapolating from individual trees to larger ecosystem level effects?

I think this is critical to the analysis. It's important to first get the individual tree responses, and then to expand up the organizational scale to account for the abundances of each of these species in the forest (but see above discussion 7.b about genetic variation). This helps adjust the impact analysis to account for the situation where you might have a very sensitive tree species to ozone, but which is a minor component of the forest. From a RBL standpoint, this is the way to proceed. From a purely biodiversity point of view, this would de-emphasize such species, so one would have to make some provision for rare species that might be in danger of being extirpated, either directly from exposure to ozone, or indirectly through competitive interactions under ozone stress.

b) What are the views of the Panel on the use of the summed-RBL as a metric to use for assessing effects at the larger ecosystem scale?

As noted above, I think this is an appropriate metric to use. However, it does not seem to take into account competitive interactions (e.g., ZELIG model), so the question does arise as to what might be the magnitude of difference in the metrics when competition is not accounted for.

c) Does the panel have any recommendations for methods to include a wider range of tree species (beyond the 8 species included)?

I currently have unpublished data on several more tree species from the NPS-EPA funded OTC study done in Great Smoky Mountains National Park. The only way I know to include those response functions would be to have them peer-reviewed in some way, even if not in a formal journal publication. I am currently trying to complete these analyses with a statistician from the EPA. Other than this, I don't

know of any other data that might be used. What about earlier data using OTCs? Are there any that might be included?

9. What are the views of the Panel on the use of federally designated Class I and Critical Habitat areas as endpoints for this analysis? Does the Panel have any suggestions for additional parks beyond Rocky Mountain National Park and Sequoia/Kings Canyon National Park that should be analyzed? Does the panel have recommendations for additional or alternative geographic analysis areas that could be used?

Given that Class I areas are designated not to have significant declines in air quality, using them as endpoints seems quite reasonable, even if they constitute a small fraction of the landscape. It's a little greater stretch to justify using all critical habitats since some of these are not terrestrial habitats, and the linkages between the aquatic and terrestrial habitats are either unknown or difficult to quantify.

It might make sense to include Great Smoky, Shenandoah or Arcadia National Parks so that Parks in the east are included. And the western portion of Pennsylvania and eastern portions of Ohio seem hotspots that could use additional scrutiny. Western Pennsylvania contains the largest and most productive black cherry sites, and this species is one of the most sensitive species to ozone that we know of. It would seem reasonable to concentrate on forests where this species makes up a substantial proportion of the individuals, and where it is also under the greatest stress from ozone.

10. To what extent does the Panel find that the vegetation mapping data from USGS and NPS used to generate a scaled-RBL surface in Great Smokey Mountain National Park is appropriate?

The vegetation mapping is sufficient, but the ozone data seem overly coarse. In addition, these analyses don't seem to take into account the fact that ozone exposures increase substantially with elevation, and that the form of the exposure also changes (flat profiles at high elevations, exposing plants to ozone early in the morning when their stomata have a greater chance of being open).

11. Regarding the screening level foliar injury risk assessment:
a) To what extent does the Panel find the updated assessment of foliar injury risk in national parks originally performed in Kohut (2007) to be an appropriate screening level risk assessment?

I think the Kohut assessments are an efficient way to screen national parks for potential impacts on plants due to ozone. I think the updated analysis used by EPA, which did not differ that much from Kohut's earlier analysis, but which did account for the recent declining trends in ozone, is highly appropriate. It is very interesting to note that no Park showed an increase in sensitivity to ozone with this analysis, while several moved to a lower sensitivity due to declining ozone exposures in recent years.

b) What are the views of the Panel regarding the potential methods for estimating O_3 exposure at additional parks?

I am not clear what these other potential methods are. If Parks don't have their own monitoring data,

then some sort of kriging methodology using data from nearby monitors would be suitable I think. Using POMs data seems a reasonable thing to do also.

c) What are the Panel's views regarding the appropriateness of requiring that two criteria must be satisfied (i.e., based on both W126 and N100) in order to receive higher risk ratings?

I think this somewhat conflicts with previous analyses that emphasize the cumulative nature of plant responses to ozone. The REA and PA both go to great lengths to justify using cumulative exposure indices, such as the W126, whereas here, a Park is moved down an entire sensitivity category based solely on whether or not the N100 index is exceeded (since both indices need to be exceeded, as I understand this). It is entirely possible to have a Park at high risk due to chronic, moderate ozone (say between 60 and 99 ppb), which could cause significant biological impacts, yet never exceed the N100. I would suggest that if the W126 is exceeded in all three years, even when the N100 is not, that the Park be classified as at high risk for adverse effects. I'm not as sure what to do if the converse is true, but I'm guessing that when the N100 is exceeded, so is the W126.

Given the inverse relationship that sometimes exists between soil moisture and ozone exposure, it is entirely possible that a plant could be at higher risk with slightly lower ozone (or years without peaks > 100 ppb) than in years with high ozone and where there are peaks > 100 ppb. Perhaps a three dimensional space needs to be developed, with soil moisture, W126 and N100 as parameters and specific volumes within that space designated as High, Moderate and Low risk.

d) Is the Panel aware of any assessments of foliar injury in national parks conducted between 2006 and 2010 that could potentially be used to validate the updated risk ratings?

I would ask that the EPA incorporate the findings of Gretchen Smith's recent paper which summarizes the past 16 years of biomonitoring in the Forest Health Monitoring Program. While these analyses were not always done in national parks, the conclusions might be useful for interpreting foliar injury assessments, especially her insights regarding the relationship between wet and dry years and incidence of visible foliar injury. Here is that citation: **Smith, G. 2012. Ambient ozone injury to forest plants in Northeast and North Central USA: 16 years of biomonitoring. Environmental Monitoring and Assessment 184:4049-4065.**

12. Regarding the assessment of cover of O_3 sensitive species:
a) To what extent does the Panel find the preliminary analysis of sensitive species cover to be an appropriate and useful approach to highlight areas of potentially higher risk due to the presence of sensitive species?

These analyses are not yet done, so I cannot comment on them at this point in time.

b) To what extent does the Panel find the vegetation mapping data appropriate to assess the cover of O3 sensitive species in GSMNP?

I'd like to hear more from the people who developed this map to hear exactly what it means in terms of assessing sensitivity to ozone. One thing that concerns me is that herbaceous species, which are often in the forest understory, are exposed to significantly reduced amounts of ozone compared to foliage out in the open or at the top of the canopy. Thus, one should consider modifying exposure of these plants, perhaps in several ways, such as including one factor to account for lower light (and hence lower stomatal conductances perhaps) and lower ozone (see Neufeld et al. 1992 which shows that ozone at 1 m in dense, cove forests in the Smokies can be as much as 50% less than that above the canopy). Perhaps herbs and shrubs can have a dummy variable that indicates whether they are beneath a forest canopy or not, and which would then act as a surrogate for reduced exposure.

c) What are the views of the Panel on the decision to not distinguish between vegetation strata (i.e. herb, shrub, tree)? To what extent does the Panel agree with this methodology relative to analyzing the strata individually?

See comment in section b above.

d) What are the views of the panel on using benchmarks, similar to those used in the Kohut analysis of foliar injury risk, to allow estimates of change between exposure scenarios?

I am assuming that "benchmarks" means "thresholds". Given the coarseness of the analysis, I think benchmarks provide a good estimate of the risk to an area from ozone. Anything more detailed might be over-analyzed with the result that there is no more confidence in those conclusions than those from a benchmark analysis.

Comments on Chapter 6 - Ecosystem Services

13. To what extent does the Panel find that EPA has adequately characterized the range of ecosystem services that are potentially adversely affected by O3?

It is my feeling that the EPA has appropriately characterized the range of ecosystem services. Instead of re-inventing the wheel they are using the Millennium Ecosystem Assessment of 2009, which more than adequately summarizes the services that ecosystems provide.

14. To what extent does the Panel agree with EPA's ecosystem services framework, connecting O3 exposure, through ecological effects to ecosystem services?

Yes, I agree with the framework and have no substantive recommendations for change.

15. Does the panel agree with EPA's use of combined O3 exposure data with other data sources (e.g. fire

data, bark beetle maps, trail maps) to link areas of concern or interest with areas of higher vegetative risk due to O3? Does the Panel have any recommendations for additional datasets and ecosystem services that could add to or improve these analyses?

Yes, these are important linkages. Some are tenuous at best, but in the interests of covering all possible linkages, should be included here. The only addition I would consider is some new data that show beetle infested pine forests produce more VOCs than healthy forests, and may actually contribute to additional ozone formation as a result – sort of a vicious feedforward loop. A similar situation has been postulated for areas infested with Kudzu.

16. Regarding the analysis of forest yield impacts:

a) To what extent does the Panel agree that the Forest and Agricultural Sector Optimization Model (FASOM) model is appropriate to assess timber and crop yield changes and the effects of those changes on additional ecosystem services?

I have to admit not having much expertise with FASOM, so I don't feel qualified to comment on the appropriateness of its use. The write up in this section seemed to make sense to me though.

b) What are the views of the Panel on the extrapolation of concentration-response functions across similar species?

I think it is highly speculative to consider that similar species will necessarily react in the same manner to ozone. Consider black cherry and pin cherry, two closely related species: black cherry is considered one of the most sensitive to ozone, while pin cherry is perhaps more tolerant. However, given that researchers cannot develop ozone response functions for every single species, extrapolation is probably the best we can do at this point in time. Grouping species by certain functionalities, such as early successional vs late, or by those species that are determinate vs indeterminate, may help: e.g., early successional species tend to respond more than late successional species on average, and perhaps those with indeterminate growth can compensate through an ozone season for decreased leaf function whereas determinate ones cannot.

17. Regarding the analysis of urban forest impacts:
a) To what extent does the Panel feel that the i-Tree model is appropriate for assessing changes to urban forest ecosystem services based on O3 exposure?

Correct me if I am wrong, but wasn't this model developed using mostly southern tree species. Has it been re-parameterized for additional tree species? If so, I would not have a problem. If not, perhaps that should be done.

b) In order to increase the number of tree species covered by the iTree model, does the Panel have recommendations for additional species that could be included, based on estimates from similar species?

I do not have any suggestions for additional species at this time.

18. Regarding the use of PnET-CN:
a) What are the views of the Panel on the potential use of the PnET-CN model in the 2nd draft to assess impacts on larger scale ecosystem services (e.g. hydrologic changes, c sequestration)?

My only question is whether PnET-CN is parameterized well enough to use in forests that differ greatly in structure, species composition, and climate from the ones it was first developed for: i.e., Harvard Forest and Hubbard Brook. Has it been validated in any other forests aside from these two?

My other concern is whether this model accounts for within canopy variation in ozone deposition and structure? That is, does it have any sort of weighting factors for leaves at different depths in the canopy, or is it a "big-leaf" type of model?

b) Does the Panel have recommendations of other models that are accessible to EPA that could be used instead of PnET-CN?

The only other one that seems to be getting a lot of publicity lately is DLEM, and the REA noted that it was not practical to use this one. Is there a particular reason or set of reasons why DLEM cannot be used?

19. Regarding ecosystem services related to foliar injury:
a) To what extent does the Panel agree that potential visible foliar injury is appropriate to use as a metric to assess potential loss of cultural services associated with recreation in national parks?

I don't think there is as clear a relationship between foliar injury and loss of cultural services as implied in the REA. My experience is that most people would not recognize foliar ozone injury unless it was pointed out to them. Further, if there is foliar stipple on large trees, most of those leaves are out of sight of the general public, and hence, not a concern. Are there any studies that demonstrate that people "see" foliar injury, and more importantly, that they "recognize" it to be caused by ozone. And further, do they consider it, once pointed out, a noticeable detriment to their experience in the Park? Without those studies, I am not confident that foliar injury is a really good metric for assessing cultural impacts.

b) Does the Panel feel that there are O3 benchmarks that could be used to assess changes in foliar injury potential between exposure scenarios similar to those used by Kohut (2007)?

I don't believe I have enough expertise to answer this particular question at this time.

Comments on Chapter 7 - Synthesis

20. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall welfare exposure and risk analysis?

Generally, the synthesis is well done. On page 2 though, there is a bullet explaining the biomass losses in Great Smoky Mountains National Park (GRSM), which implies that the magnitude of response is quite large for a given reduction in ambient ozone, yet in the earlier chapter where the Kohut assessments are discussed, there is explicit mention of how GRSM ranks in the low risk category. How do we reconcile these two disparate ratings?

Dr. Armistead (Ted) Russell

Comments on the Health Risk and Exposure Assessment

Executive and Integrative Summary: I hope both are coming.

Comments on Chapter 1

The Minor issues:

1-6, l30: Remove "any"

Comments on Chapter 2

Overall, the chapter adequately conveys the first parts of the conceptual framework for conducting a risk and exposure assessment for ozone, but not the latter parts, which is what most of the REA actually is about. The chapter needs to lay out the conceptual steps in an exposure and risk assessment, identifying the important elements and types of methods/tools that are used.

The Minor issues:

2-1, l6: Add carbon monoxide: "…and carbon monoxide (CO) …

2-1, l21,22 (and elsewhere): Use chemical subscripts correctly

2-1, l25 Use of "valleys" here can be misleading (some will think topographically). Change to … local decreases where ozone…"

L27: …, and the NO_2 formed can lead to O_3 formation…" (also remove a period at the end)

L29: Likewise, don't use "valleys" in this context.

2-2, l16 replace relatively insensitive with less sensitive., and "both" with "either"

2-2 l26 add "VOCs, as well as CO, are…"

2-3, l28: Do you mean intrusions?

2-4, l31 … predicted Air Quality Index…

2-5 "communicate information on the levels of O3 and other pollutants. The predicted AQI … a set of potential actions…"

2-10 Remove a period.

Comments on Chapter 4

1. Question 6abc. First, it would be good to show the rollback method mathematically, though if it is not going to be used in the second draft REA, this is not as big of a deal. As pointed out, the use of the quadratic rollback method, as formulated, does not include the potential for lower ozone levels to increase as higher levels come down, though the ISA suggests this can be the case. Thus, I definitely applaud going to an air quality-based approach, or some other approach, that can capture this. My thoughts on using HDDM are likely known. Like using any air quality model-based approach, the results will then depend upon the choice of simulated controls, and the REA should be more explicit as to how this will be chosen though it would appear that the difference is only a few ppb between VOC and NOx in Detroit). One might recommend examining a few directions. This should be discussed more here. I liked the analysis done by Simon et al., (2012), though the figures (e.g., 8,9) and tables (e.g., 1) should be more explicit as to what emissions are being changed.

4-8 l17: This paragraph should refer the reader to Table 4-2..

4-8, footnote 8: Again, how often did this occur?

4-23, l1-4. This sentence seems to say you have evaluated methods by using HDDM. Do you mean this, or that you have evaluated using an HDDM-based approach? If it is the former, then some more information would be valuable as to the result of that evaluation.

Table 4-3: Curious, where was the average MDA 19.7 ppb It would be good to identify each city associated with the four extreme values given.

Comments on Chapter 8

The chapter is informative, and lays out the approach used for conducting the national scale risk assessment for premature mortality (note: the chapter title might be changed to include "for premature mortality"). The approach is to be comprehensive, using both short term and long term effect estimates, and it also analyzes the distribution of where the more detailed urban case studies fall within the national distribution, showing that those cities (actually counties) capture the range of ozone levels and population demographics well.

While the chapter is generally understandable, there are areas where clarity can be improved. The Introduction does not set up the whole chapter as well as it should. For example, it does not discuss the

use of the Jerrett et al., effect estimates. The Introduction also does not discuss the results of the assessment of the county distribution (though it does discuss the results of the risk assessment using the short-term risk estimates.

This chapter should also start out with the results from the ISA stating that there is likely to be a causal relationship between short term exposures to ozone and all-cause mortality and that evidence is suggestive of a causal relationship between long-term ozone exposures and total mortality, i.e., re-emphasize why this is being done and also highlight the strength of the findings from the ISA. This should also be captured in a chapter summary (which should be added).

While the LML is 7.5 ppb, it is not apparent to me that this should necessarily be the highest level used for a cutoff in a sensitivity analysis of assessing risk associated with ozone exposure (though it certainly can be used as one of the levels used).

I am not sure there is great benefit to doing an extensive analysis using the ATUS data base, given its limitations, and it is not apparent how it would materially impact the risk assessment. How would such an analysis be used is not obvious.

1. Question 16: The use of the Bell et al., and Zannobetti and Schwartz CR's is reasonable. They are both highly cited and used in prior studies, which aid in comparisons (which there should be more of in the report).
2. Question 17: Likewise, the Jerrett et al., mortality effect estimate is reasonable. How this is currently presented needs to be improved. In particular, the last sentence in the first paragraph on 8-10, it seems that there are already results, but the rest indicates EPA plans to use the long term effect estimates. The section where using the long term effect estimate is to be used should recapture the ISA conclusion on the causal relationship.
3. Question 18: This section does a good job laying out how the representative analysis was conducted. The section is comprehensive and demonstrates that the urban case studies range over the various air quality, geographic and demographic characteristics experienced by a bulk of the US population. Fig. 1.13 shows that the counties used are more populated (not surprisingly, and also not a bad thing). It might be good to also include something indicating the fraction of population covered, e.g., % of US population (or something like that, but this may have to be adjusted for multi-county MSAs), to show how well you are sampling based upon population (which strikes me as being more important).

The Minor issues:

The Tables and Figures are incorrectly labeled (they all start with 1-)

Table 1-1: It would be informative to note the locations of the four extremes shown.
8-10, l32: What is meant by "high confidence", and such a high confidence is not supported in the manuscript.

8-24, Is the reference to section 7-3 correct?

8-46: In addition to suggesting why Houston and LA might have the lowest effect estimates, please provide similar suggestions for why New York and Detroit have the highest.

Comments on Chapter 9

You can tell this is a work in progress. First, it should start out with a summary of the ISA causal findings and how this motivates the analyses conducted. The way it starts now is not entirely comprehensive in that it does not really cover what was done as part of the national level risk assessment for premature mortality (Chapter 8), which does not look at the just meeting the standard scenario.

I am not sure saying something like "20 to approximately 930 deaths" is that useful without some more context as this is for the various cities, which, of course, will have different numbers of deaths because of size. The % of all-cause mortality provides some of that context. (Also, it should be approximately 20 to 930.) It would also be good to link this back to specific tables in the chapter. Actually, a table, similar to those in Chapter 7, can summarize this information well and provide the desired context. It could contain the range of #deaths per city (but not the specific cities), the total amongst all the cities examined, and the % of all-cause mortality, for both the base and just meeting the standard cases.

The current summary suggests that a future REA may include additional health endpoints. Whether this is the case or not, the Summary should discuss them and the implications of considering them in assessing the adequacy of the current standard and potential alternatives.

The Minor issues:

9-8, l32: You state there is considerable variation. I looked at the results as being remarkably similar.

9-10, l6: You state that 2010 had slightly higher ozone… as compared to what time period? Be explicit here.

Comments on the Welfare Risk and Exposure Assessment

Executive and Integrative Summary

I hope both are coming. Also, parts of the Welfare REA suggest that different, or extended, analyses will be done for the next draft. It would have been nice for these to be mentioned up front. The last paragraph of the synthesis actually does this well… though should have been more detailed in the synthesis chapter.

Comments on Chapter 2

The Conceptual Framework should provide more discussion on how the risks are determined, including discussions about assessing current risk, just meeting the standard, and assessment of the risks from meeting alternative standards. At present, it does not lay out the conceptual part of conducting a risk and exposure assessment (e.g., methods and tools and how the results would be analyzed). This is done, but in too much detail, in the SCOPE.

The Minor issues (similar to Health REA):

2-1,. L4 remove ","

2-2, l 17 (and elsewhere): Use chemical subscripts correctly

2-3, l4 Use of "valleys" here can be misleading (some will think topographically). Change to … local decreases where ozone…"

L6: …, and the NO_2 formed can lead to O_3 formation…" (it can form HNO3 as well)

L8: Likewise, don't use "valleys" in this context.

2-3, l24 replace relatively insensitive with less sensitive., and "both" with "either"

2-5, l7: Do you mean intrusions?

2-6 need a space in "Below ground" in the figure.

2-8 "e"? (At the end of the sentence.)

2-14, l13: You bring up in the conclusions that "air quality models … show that important ecosystems…" If this is part of the conclusion, those should be discussed somewhat more in the chapter along with what has been shown.

Comments on Chapter 3

3-8, l6 Maybe replace extrapolate with estimate?

3-8, l23-25: What alternatives? Where are they described?

Comments on Chapter 4

For what is there, Chapter 4 provides a reasonably clear discussion of the air quality considerations involved in the Welfare REA (WREA). However, it ends abruptly, and how you would deal with meeting alternative standards is not addressed (if that might ever be done… if not, you might actually explain why the difference with the Health REA). If the plan is to evaluate W126-based standards, doing a national analysis where the ozone is adjusted to meet a specific W126 level would have been instructive. The chapter does not do a good job of describing the types of future analyses being contemplated, including what alternative types of rollback approaches are planned or which types of standards to be evaluated are envisioned.

1. Question 4: Appropriateness of years of air quality: Fine choice. You want to avoid 2009 as an abnormally low year in parts of the country.
2. Question 5: Approach used to develop W126 surface: Fusing CMAQ results with observations is good. Looking at the three year average, and the individual years, is good. You should more clearly explain how this works in to the "just meeting the standard" scenario.
3. Question 6. Methods for simulating just meeting the standard

a. Quadratic rollback approach: A problem with the quadratic roll-back (or linear or Weibull, I think) methods are that they do not allow for lower levels to actually increase in response to controls, and a few other issues are of potential consideration as well. Indeed, the health REA proposes using an air quality model-based method (HDDM) to quantify how levels will go down. Further, only having the monitors experience exceedences go down is also problematic, with a specific example captured in the text (if two sites are close, and one is exceeding, only that monitor is reduced: page 4-7).

b. Alternative approaches: Given that a more comprehensive approach (CMAQ with HDDM) is being used in the health REA, its use should be examined for application here. You could use the HDDM sensitivities to reduce ozone at the specific monitors, which are then applied in MATS.

The Minor Comments

4-4, l13 "… sum the <u>weighted</u> hourly…"

4-5 l3: You should explain, at least to some extent, what the Kohut analysis is, and how it will be used, before saying "For the specific application of the Kohut analysis…" Further down the page it is then mentioned the Kohut analysis will be described "in more detail" in chapter 5. Up to this point, there has been no detail. I would give a few sentences at the beginning of this paragraph, or possibly even on page

4-3, about the Kohut analysis.

Comments on Chapter 7

Not a bad start at all. One comment from above: a more detailed discussion of the different analyses to be done as part of the second draft would have been useful.

Dr. Helen Suh

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 5 - Characterization of Population Exposure

7. To what extent does the Panel find the methods used to conduct the exposure analysis technically sound? Does the Panel have any recommendations on the methods used?

The overall methods used to conduct the exposure assessment are sound but too vague in either the description or discussion the approaches that will be used. Since the sensitivity analyses, which will be conducted in the second draft REA, is a key component of the REA, the methods that will be used for these analyses should be discussed in more detail. As is written currently, only a cursory description of the uncertainty analyses is provided. Also, important, but not discussed in sufficient detail or referenced, is model performance, particularly in light of recent improvements made to the model. How these improvements affect model performance will be important to the credibility of findings from exposure and health assessments. This discussion could be added as a new section between the current Sections 5.3.1 and 5.3.2. Similarly, the methods that will be used to estimate multiple repeated exposures and to correct current underestimation of the findings were not discussed. Given the relative paucity of longitudinal activity pattern data, it is not clear how this will be performed.

Additional areas that warrant clarification or further elaboration are listed briefly below:

- A figure that describes the modules, inputs, outputs, years and cities of the APEX model should be added to Section 5.3, as this figure would help the reader (especially those not familiar with APEX or probabilistic exposure modeling) understand the process by which ozone exposures are calculated.
- The age and spatial representativeness of several of the input databases are of concern, particularly as they are related to the estimation of exposures for children and asthmatic children. The datasets that are relevant to children are dated and are from limited geographical areas. As a result, their relevance to current conditions and the 16 examined urban areas is not clear, may vary with geographical area, and will likely be a major source of uncertainty in the APEX model results. A prime example of this is the CHAD database, for which the majority of data were collected more than a decade and even two decades ago. This is particularly true for data collected for children. Further, it is not clear what fraction of the data were collected during the high ozone season or for the 16 urban areas, as would be relevant to this assessment. The impact of the dataset age and temporal and spatial representativeness on the APEX findings should be discussed in more detail. Even if the APEX uses the most current available data, it is possible that these data are too old to be relevant to current scenarios. The extent to which this is true should be discussed.
- Page 5-13 “Averting Behavior and Exposure” section: This section should be moved to Section 5.5 “Variability and Uncertainty”, as it does not contain methods to correct exposure estimates for averting behavior. As a curiosity, when were ozone alerts first issued? Does this time period overlap with the activity pattern data collection? If not, this should be mentioned in the discussion.
- Page 5-15 (corrected version), lines 1-16 (beginning with “Therefore,…”): This discussion is

speculative and convoluted. I would delete and replace with a sentence such as the following "Thus, ozone exposures, particularly for asthmatic children, may be underestimated on high ozone days."

- Page 5-16: It seems possible that other factors, such as gender, would affect variability in EVRs in addition to BSA and exercise level. The discussion should list these factors and if possible consider these factors in the estimation of EVRs or their contribution to uncertainty.
- It will be important to expand the analysis to older adults and to outdoor workers, as is currently planned. This expansion should be of relatively high priority.

8. To what extent does the Panel find the assessment, interpretation, and presentation of the results of the exposure analysis as presented technically sound, appropriately balanced, and clearly communicated?

- The discussion of the results and their interpretation is interesting, but its clarity could be improved with greater organization and detail.
  - Page 5-24, lines 26-31: The overview of the set of figures and tables with results included in the results is confusing. It was very hard to discern to what figures references were being made.
  - A discussion summarizing the key main results should be provided, as no discussion was currently provided beyond the figures and tables. Some of this summary can be found in first two paragraphs of Section 5.6.2 and in Chapter 9 (Synthesis). The summary found in Chapter 9 can likely be removed, while the Section 5.6.2 paragraphs can be moved forward into the presentation of results and comprise the start of the results presentation. As noted below, Section 5.6.2 can subsequently be modified to focus on explanations of the findings and integrating the main results with those from Section 5.6.1.
  - The presentation of the findings would be improved with summary tables or figures (and corresponding discussion) of the distribution of the exposure estimates for each city, and not just the percent of people with exposures above benchmark values.
  - The axes for Figures 5-1 to 5-15 should be labeled more completely and clearly to help the reader. As is currently, the labeling of the figures and tables is confusing.
- Section 5.6.1 was well-written and clear, with interesting findings. The discussion of these findings should be expanded to discuss whether and how these findings will be incorporated into the planned uncertainty or health analyses.
- Section 5.6.2: The discussion of the APEX model results could be enhanced with several modifications:
  - As noted above, the discussion could be refocused to (1) integrate the main findings with the analysis of factors affecting results, (2) discuss the impact or import of the findings to the subsequent health analyses
  - Figures "4-1 to 4-15" should be relabeled to "Figures 5-1 to 5-15".
  - The last paragraph on Page 5-52 should be moved, as it is not a discussion but a result. Perhaps this paragraph and figures should be made into its own section 5.6.2, with the current Section 5.6.2 ("Discussion of Exposure Modeling Results", save the last paragraph) made into a new Section 5.7 that is renamed "Discussion".

9. Regarding the characterization of uncertainties and variability:
a) To what extent does the Panel find that the uncertainties associated with the exposure analysis are

clearly and appropriately characterized?
b) To what extent does the Panel find that the uncertainty assessment is technically sound? Are there other important uncertainties which are not covered?

10. What are the views of the Panel on the sensitivity analyses that EPA plans to conduct as part of the second draft REA to evaluate the influence of uncertainties in the exposure analysis?

The plans for the sensitivity analyses are discussed only briefly, but seem appropriate, with analyses planned to examine each of the major uncertainty contributors. Since the methods used to conduct these uncertainty analyses are not discussed, the technical merit of the uncertainty analyses cannot be evaluated. However, for the sensitivity analyses for air exchange rates, it is likely important to restrict air exchange rates to be temperature- as well as city-specific.

Further, the discussion of averting behavior on high ozone days (as noted above) should be included in this section. Analyses examining the impact of averting behavior on exposure estimates should be performed, possibly by relying on data from the RTI Ozone Averting Behavior study to characterize the distribution of time spent outside for the important population sub-groups and high ozone days.

As noted on Page 5-16, fluctuations in children's activity levels are a major source of uncertainty in MET estimates. However, the impact of these fluctuations on exposure estimates will not be examined. If this examination is not possible given available data, it should be mentioned as a major limitation of the uncertainty analysis.

Other:

- Citation regarding composite monitor constructs:
  - Sarnat SE, Klein M, Sarnat JA, Flanders WD, Waller LA, Mulholland JA, Russell AG, Tolbert PE. An examination of exposure measurement error from air pollutant spatial variability in time-series studies. J Exposure Science Environ Epidemiology 2010; 20, 135–146.

(There may be other citations as well.)

Dr. James Ultman

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 6

1. To what extent does the Panel find the methods used to conduct the risk analysis to be technically sound? What are the views of the Panel members on the methods used?

A clear statement or equation of how the endpoint, %ΔFEV, is calculated is absent from the REA. Pages 6-4 and 6-5 of the ISA leads me to believe that %ΔFEV decrements for **a particular individual** are probably computed relative to filtered air (FA) exposure by a formula such as:

$$\%\Delta FEV = -100[\Delta FEV1_{O3} - \Delta FEV1_{FA}]/FEV1_{O3}$$

where

$$\Delta FEV1_{O3} = (FEV1_{post\text{-}O3} - FEV1_{pre\text{-}O3}) \text{ and } \Delta FEV1_{FA} = (FEV1_{post\text{-}FA} - FEV1_{pre\text{-}FA})$$

The argument for subtraction of the FA exposure is that an intrinsic exercise effect (separate from the increase in ventilation by exercise) would otherwise give a misleading value of the ozone-induced FEV1 decrement. For example, if exercise-induced bronchospasm occurred, then $\Delta FEV1_{O3}$, would actually result from a combination of ozone exposure and ozone-induced bronchospasm.

The argument against subtraction of the FA exposure is the possibility of day-to-day changes in the accuracy of the FEV measurements. Such a variation in accuracy would constitute an error in the %ΔFEV computed from the above formula.

As discussed in some detail on pages 6-4 and 6-5 of the ISA, many healthy subjects exhibit a small improvement in FEV1 upon FA exposure while others exhibit a small decrement. For the subjects that show an improvement, this increases their %ΔFEV decrement when computed by the above formula; this effect is particularly important at small ozone exposure levels when FEV1 decrements are small.

The McDonnell-Stewart-Smith paper in 2012 uses a compromised method between computing %ΔFEV for an individual on the basis of their own $\Delta FEV1_{FA}$ and computing %ΔFEV without subtracting the $\Delta FEV1_{FA}$ contribution. They average the FEV decrements in FA for the entire subject population in a given study, and use that value to correct each individual's $\Delta FEV1_{O3}$. This method should be considered in making final computations.

2. To what extent does the Panel find the assessment, interpretation, and presentation of the results of the risk analysis as presented in Chapter 6 to be technically sound, appropriately balanced, and clearly communicated?

There is some lack of clarity that detracts from the presentation and interpretation of the results. For example, the manner in which the minimum and maximum values are selected in table 6-4 (and other similar tables) is not explained until the end of the chapter. In figures 6-20 to 6-30, it appears that only the median values generated by the MSS model are reported. Will a measure of population distribution (i.e., box plot) be shown in future drafts of the ISA?

3. To what extent does the Panel find the focus of the assessment on lung function decrements in the quantitative risk assessment to be appropriate and informative?

Clearly, the wealth of data available on FEV1, makes it an attractive marker to use in a quantitative risk analysis. Additional parameters such as FEF25-75 or FVC are routinely measured along with FEV1, although they have been reported in the literature to a lesser extent. These parameters alone or in combination with FEV1 provide an alternative basis of quantifying clinically-significant lung dysfunction. I don't believe that this is mentioned in the REA. I wonder whether any of these parameters have been considered.

4. What are the views of the Panel on the use of the two different modeling approaches for specifying the exposure-response function linking the change in FEV1 to ozone exposure?

- In describing the MSS model, the biological/physiological rationale for the differential equation in X_{ijk} should be improved (see the descriptions in the McDonnell (2012) and/or Schegle (2012) papers).

- Clearly, the MSS model has the potential of reducing uncertainty by tailoring its predictions to a subject's specific ventilation rate, body surface and age. In principle, it can also accommodate time-varying and repetitive exposures. The first draft REA demonstrates that there can be a substantial inconsistency between the probabilistic population-based model and the MSS model. In particular, the MSS model appears to make substantially larger predictions of affected individuals (table 6-9). Further comparison of the two exposure-response approaches under alternative ozone standards is necessary.

- In order to capitalize on the MSS model, it is important that all the original study data be available to EPA, including those data collected outside of the EPA laboratories by Adams and by Schegle.

5. What are the views of the Panel on the treatment of the relationship between age and ΔFEV1 in the McDonnell-Stewart-Smith model?

In the absence of appropriate data, the approach of freezing the age factor at 35 years is reasonable. Given that FEV1 decrements, might very well show a further decline with age>35, this may overestimate the response of the elderly. It would be useful to comment on this in the text of the REA.

6. To what extent does the Panel find that the qualitative discussion of uncertainty and variability has

covered important sources of uncertainty and variability and has appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates?

I understand that evidence in the literature supports the assumption that the exposure-response function for healthy children is similar to that in adults; in that case, any difference in health risks would be due only to differences in activity patterns. It is unlikely that this is a reasonable assumption for children with asthma, considered to be an important high risk population. Would it be possible to perform a sensitivity analysis of this assumption using the limited information in the literature?.

7. What are the views of the Panel on additional sensitivity analyses or other approaches to addressing uncertainty and variability?

See item 6.

Dr. Sverre Vedal

Comments on the Health Risk and Exposure Assessment

Comments on Chapter 7 - Characterization of Health Risk Based on Epidemiological Studies

11. Regarding the epidemiologic studies used in the analysis:
a) What are the Panel's views on the set of epidemiological studies selected for use in specifying C-R functions and on the set of C-R functions specified for use in the risk assessment?

- The use of multi-city studies for the health endpoints where these are available is justified. Use of Bayesian shrunk estimates is also justified. It is appropriate to express additional uncertainty when using C-R functions based on only one city.

b) To what extent does the Panel find the detailed descriptions of rationales for the selection of the epidemiological studies and the selection of the set of C-R functions specified using those studies to be appropriate and complete?

- Good.

12. To what extent does the Panel find that the qualitative discussion of uncertainty and variability have covered important sources and appropriately characterized the relationship of those sources of uncertainty and variability to the risk estimates?

- The discussion is rigorous and thorough.

- The greatest source of uncertainty in all ozone epidemiological studies, but especially those that involve elderly people who are ill (ie, those susceptible to dying from ozone exposure), is the presumably substantial exposure measurement error when exposure is based on central ambient monitors, and what this implies for the health effect estimates. The epi studies (specifically, short-term exposure mortality studies) make use of the same monitors used to assess compliance with the NAAQS, so in some sense it doesn't really matter that the relationship between true exposure and mortality could be quite different, or could be due instead to a pollutant (presumably indoor) for which ozone is simply a marker, but whose concentrations are reduced when ambient ozone concentrations are reduced. This source of uncertainty is qualitatively different from the sources itemized in the risk assessment. I'm not sure how this uncertainty should be expressed, or whether it needs to be considered in this context.

- Exposure measurement error could also be a source of variability in effect estimates.

- Table 7-6, C. Comments column. Ozone concentrations do not really vary spatially as a function

of sources, apart from ozone quenching by NO from roadway emissions.

13. Regarding the results of the risk analysis:
a) What are the views of the Panel on the presentation and discussion of risk estimates, including the key observations presented in section 7.6.2?

- Table 7-15. "Pathway-specific" is not very descriptive of cardiovascular vs. respiratory.

- Table 7-24. These are very large effects, likely determined by the effect size in the Boston study. Nevertheless, does ~15% of all asthma exacerbations being due to ozone exposure pass the credibility test?

- Section 7.5.2. This section on just meeting the current standards is very repetitious of the previous section on recent conditions (section 7.5.1). It would be more useful to say something like the assessment is identical to that for recent conditions, and then focus on some differences, if there are any.

- Many of the key points in Section 7.6.2 merely reiterate (rather than summarize) those points detailed previously.

b) What are the views of the Panel on the presentation of the distribution of O_3-related mortality across daily O_3 levels for each city as "heat maps"?

- Nice

- Note the dramatic difference between the Zanobetti (Table 7-7) vs. Bell (Table 7-9) estimates for total O_3-related deaths for Los Angeles.

14. To what extent does the Panel agree with the characterization of overall confidence, including the degree to which the conclusions reached regarding overall confidence are supported by available information?

- Agree with the approach and plan for the next version of the REA.

15. What are the views of the Panel on EPA's discussion of potential refinements to the REA for the second draft, including the plans for quantitative sensitivity analyses, additional refinements to the core risk estimates, and plans for assessment of long-term mortality and morbidity (i.e., plans to model risk for mortality and the decision not to model risk for morbidity endpoints, given data limitations)?

- I do not support performing an exposure and risk assessment for long-term exposure and respiratory morbidity.

- The Lin (2008) study is described as "the strongest of the long-term O_3 exposure and respiratory morbidity studies" (p.7-80, line 21), although in the ISA-3, this honor is bestowed on the CHS studies on new-onset asthma. Note that the long-term exposure study is 2008b and the short-term study is 2008a (line 37) in the references.

Miscellaneous Comments

p. 7-23, lines 17-21. This is a standard point regarding confounding, but has little relevance in the setting of short-term exposure epi studies, which is the focus in the risk analysis, and time series studies in particular, where non-time varying factors do not confound.

p. 7-28, line 29. Generating Bayesian-adjusted city-specific estimates is not a study design. It's merely a way of using study findings to express effect estimates.

p. 7-78, line 21. Should be 50% "higher"?

Dr. Kathleen Weathers

Comments on Chapter 1: Introduction

1. Does the Panel find the introductory and background material, including that pertaining to previous reviews of the O_3 standards and the current review, to be clearly communicated and appropriately characterized?

Yes.

Comments on Chapter 2: Conceptual Model

What's the significance of the colors in Figure 2.1? If none, don't use them. If there is significance, identify in figure legend.

2. To what extent does the Panel find that the materials accurately reflect and clearly communicate the currently available welfare effects evidence, and the relevance of that evidence for quantitative exposure and risk assessment, as characterized in the 3rd Draft ISA?

It does.

Seems that biodiversity could/should be included Figure 2.1.

The statement, "Rural areas, such as national parks, national forests, and state parks, tend to be less directly affected by anthropogenic pollution sources than urban sites." is not necessarily true. Consider sensitive sites, downwind of long(ish)-range transport and the fact that diurnal ozone cycles are dampened at high elevations.

Conclusions:

It would be useful to add a sentence on what is known about the relative importance of natural vs anthropogenic sources of tropospheric ozone (by region, if necessary, e.g., 2-14).

I suggest that the fact that responses are species-specific and that there are data for few dominant tree (shrub or herb) species are available be included here.

This chapter needs some copy editing. Among other things, misspellings and partial sentences exist.

Comments on Chapter 3: Scope

3. To what extent does the Panel find the scope of the welfare risk and exposure assessment to be clearly communicated?

The scope is well communicated. I especially appreciated that the REA started where the last one ended.

Somewhere, perhaps in the last chapter, I would like to see a list of data and synthesis needs that would result in a more robust analysis five years from now.

I suggest that some indication of the gap filling and "fusion" approaches used in this REA be mentioned in the scope chapter (rather than just mentioning that "new" approaches were used).

The use of spatial data and GIS mapping is promising, but, as always, the devil is in the details—the details of extrapolation methods, data to underpin that extrapolation, different spatial resolution of data/data layers. The methods used and especially how data were fused were not entirely clear to me (throughout the document).

If history repeats itself, the uncertainty and variability section (not included in this draft) will be critically important to CASAC's ability to provide constructive comments.

Comments on Chapter 5: Ecological Effects

A few overarching comments.

I think the EPA has done a commendable job trying to extrapolate and make relevant the RBL, visible injury and veg. susceptibility data across the country. It is a good start. That said, the chapter needs significant editing to make clear both the complicated and sometimes convoluted analyses as well as the interpretation of these analyses. Parts of this chapter are quite tedious to wade through and unclear. In the next version of the REA, I suggest not only reporting the (for example) RBL results of analyses, but adding a phrase of interpretation of what they mean (in clear language).

I suggest modifying "These results strengthen our understanding of O3 effects on forests and demonstrate the relevance of the knowledge gained from trees grown in OTC studies: to the "field" relevance."

Section 5.2: I think it important to define what is meant here by threshold—when it is first used.

Page 5.5: The grid cell size should be noted when CMAQ is invoked. It is only later in this section that 12k grid cell sizes are identified. It seems to me a mismatch between 12km grid cell sizes and species-specific seedling responses as well. Over what percentage of any grid cell is any one of the species (and seedlings!?) mapped represented?

Actually, I am not sure how or why mapping seedlings makes sense for the big picture.

The figure and table legends should be more clear and descriptive so that they are interpretable even without the text (e.g., "rolled back to meet the current 8-hour standard" …of xxx; Importance Value—add a phrase about what that, and the legend, mean).

Also, please note the geographic location of Wilderness areas in Table 5.4

I was surprised that there was no acknowledgment in this chapter of the importance of elevation and/or landscape heterogeneity in affecting ozone exposure or synthesis/extrapolation results.

7. Regarding the assessment of relative biomass loss (RBL) for individual species:
a) What are the views of the Panel on the use of the linear model forced through the origin to assess the proportional relationship between the relative biomass loss (RBL) values for each species comparing the RBL at recent ambient conditions to the RBL under the scenario modeling O3 just meeting the current standard?

Section 5.2.1.2: It's a little unclear to me what was done here. The mapped RBLs were then compared using a linear regression?

A linear model is unlikely to be mechanistically explainable. However, if it is true that at 0 W126RBL would also be 0, it seems reasonable to force the lines through the origin.

b) To what extent does the Panel find that this an appropriate analysis to compare the proportional changes in RBL? Does the Panel have suggestions for alternative approaches that provides a comparable result and maintains the cell-by-cell approach to help control for environmental variability?

Proportional changes are useful.

c) To what extent does the panel agree with the approach used to combine the 11 tree species into one analysis?

This is a difficult question and it depends upon to what end the analysis is used. Trees respond differently to ozone, of course and there are sensitive and insensitive species. To what extent does policy need to take into account the response of the most sensitive species? Those questions notwithstanding, my initial response was that the analysis is necessary and useful to get a sense of what the average is.

This (and many others) paragraph is completely unclear to me:
"Using this approach provides two advantages. First, it will in part correct for variability in O3 exposures in different regions. For example, one source of variability is the difference between O3 concentrations measured at the height of ambient monitors and those occurring at the height of the actual tree canopy. In the 2007 Staff Paper (U.S. EPA, 2007a) this difference was addressed by applying a 10%

reduction in hourly O3 values in each grid cell. That methodology introduced uncertainty, but was a useful in comparing the effects of uncertainty in the O3 exposure values."

8. Regarding the assessment of RBL for combined species:
a) To what extent does the Panel support the use of the Importance Values from the U.S. Forest Service to weight the RBL values in extrapolating from individual trees to larger ecosystem level effects?

It seems to me that there is little other choice. It's an imperfect solution, but one that has available data to underpin it. It is important to be able to extrapolate from the trees to the forest (and I would phrase it this way rather than "larger ecosystem level effects," which really don't have any meaning).

b) What are the views of the Panel on the use of the summed-RBL as a metric to use for assessing effects at the larger ecosystem scale?

This needs more explanation and discussion. It's a bit of "fusion" mish-mash.

c) Does the panel have any recommendations for methods to include a wider range of tree species (beyond the 8 species included)?

Not without data to underpin them. Hopefully the FIA/FHM data will help.

9. What are the views of the Panel on the use of federally designated Class I and Critical Habitat areas as endpoints for this analysis? Does the Panel have any suggestions for additional parks beyond Rocky Mountain National Park and Sequoia/Kings Canyon National Park that should be analyzed? Does the panel have recommendations for additional or alternative geographic analysis areas that could be used?

I agree with this approach. These Federally designated areas are sensitive, have a mandate to protect the biological resources, and often have finer spatial scale data to be used in the extrapolation. Since coastal sites along the eastern seaboard are subject to high ozone, I suggest including Acadia or Cape Cod National Parks.

10. To what extent does the Panel find that the vegetation mapping data from USGS and NPS used to generate a scaled-RBL surface in Great Smokey Mountain National Park is appropriate?

For these case studies, it is appropriate and useful to have finer spatial scale data for extrapolation and synthesis. The spatial scale of the data should be noted, however (e.g., page 5-29).

This paragraph was completely unclear to me:
"Whenever possible, only plots from within the park were used. In some cases, no plots were available from within the park and in those cases plots from the same vegetation community in nearby areas were used. In a few cases there were no plots available, and those communities were excluded."

11. Regarding the screening level foliar injury risk assessment:

a) To what extent does the Panel find the updated assessment of foliar injury risk in national parks originally performed in Kohut (2007) to be an appropriate screening level risk assessment?

I find it of use because there is a prior study with which to compare the current analysis. However, as was identified in the text, I do have questions about whether these are the correct controlling variables. I was unsurprised that an analysis of data suggested an inverse relationship between ozone and soil moisture; one would expect exposure to be higher in dry conditions.

b) What are the views of the Panel regarding the potential methods for estimating O_3 exposure at additional parks?

Use of local monitoring data and empirical relationships built in the locations that have spatially explicit data is one approach.

c) What are the Panel's views regarding the appropriateness of requiring that two criteria must be satisfied (i.e., based on both W126 and N100) in order to receive higher risk ratings?

It is a conservative, but potentially more confusing approach. I suggest using just one criterion—the proposed form: W126.

d) Is the Panel aware of any assessments of foliar injury in national parks conducted between 2006 and 2010 that could potentially be used to validate the updated risk ratings?

Check with NPS natural resource managers and FIA/FHM monitoring results.

12. Regarding the assessment of cover of O_3 sensitive species:

a) To what extent does the Panel find the preliminary analysis of sensitive species cover to be an appropriate and useful approach to highlight areas of potentially higher risk due to the presence of sensitive species?

I think that it is a useful approach. I also think it important to take into account the landscape features (e.g., high elevation).

b) To what extent does the Panel find the vegetation mapping data appropriate to assess the cover of O_3 sensitive species in GSMNP?

It's a useful tool.

c) What are the views of the Panel on the decision to not distinguish between vegetation strata (i.e. herb, shrub, tree)? To what extent does the Panel agree with this methodology relative to analyzing the strata individually?

Since sensitivity is likely to vary (as is ozone exposure, by strata), I think it important to look at strata individually as well, but I do not think it's realistic to extrapolate this information across the US. The data are way too sparse.

d) What are the views of the panel on using benchmarks, similar to those used in the Kohut analysis of foliar injury risk, to allow estimates of change between exposure scenarios?

Reasonable.

Comments on Chapter 6: Ecosystem Services

General comments:
I am very pleased that the EPA has chosen to use the existing MEA framework/definitions for ecosystem services. It's important not to reinvent the wheel.

6.2.1.1: Be sure to distinguish when and if ANPP vs NPP is what is meant in these studies.

13. To what extent does the Panel find that EPA has adequately characterized the range of ecosystem services that are potentially adversely affected by O3?

They seem adequately characterized, especially given the inability to quantify effects.

14. To what extent does the Panel agree with EPA's ecosystem services framework, connecting O3 exposure, through ecological effects to ecosystem services?

I agree that it's a reasonable approach.

15. Does the panel agree with EPA's use of combined O3 exposure data with other data sources (e.g. fire data, bark beetle maps, trail maps) to link areas of concern or interest with areas of higher vegetative risk due to O3? Does the Panel have any recommendations for additional datasets and ecosystem services that could add to or improve these analyses?

I Agree. No, sadly.

16. Regarding the analysis of forest yield impacts:
a) To what extent does the Panel agree that the Forest and Agricultural Sector Optimization Model (FASOM) model is appropriate to assess timber and crop yield changes and the effects of those changes on additional ecosystem services?

I don't have experience with this model, however, to the extent that is has been vetted and used, it seems reasonable.

b) What are the views of the Panel on the extrapolation of concentration-response functions across similar species?

In the absence of any other data or information, it may have to suffice, but it is not optimal. Assumptions and details of extrapolation should be clearly articulated.

[The appendix detailing the FASOM analysis will be submitted for review in August]

17. Regarding the analysis of urban forest impacts:
a) To what extent does the Panel feel that the i-Tree model is appropriate for assessing changes to urban forest ecosystem services based on O3 exposure?

Again, I do not have direct experience with this model, however, as a qualitative, comparative approach, it is useful.

b) In order to increase the number of tree species covered by the iTree model, does the Panel have recommendations for additional species that could be included, based on estimates from similar species?

[The appendix detailing the i-Tree analysis will be submitted for review in August]

18. Regarding the use of PnET-CN:
a) What are the views of the Panel on the potential use of the PnET-CN model in the 2nd draft to assess impacts on larger scale ecosystem services (e.g. hydrologic changes, c sequestration)?

I suggest using this model and perhaps another that was parameterized for western systems, initially, such as Day-Cent-Chem (Century derivative), to compare. [Note: I do not know whether a model such as Day-Cent-Chem could be modified or parameterized for this use, however. The developers would need to be contacted.]

b) Does the Panel have recommendations of other models that are accessible to EPA that could be used instead of PnET-CN?

See above.

19. Regarding ecosystem services related to foliar injury:
a) To what extent does the Panel agree that potential visible foliar injury is appropriate to use as a metric to assess potential loss of cultural services associated with recreation in national parks?

I agree.

b) Does the Panel feel that there are O3 benchmarks that could be used to assess changes in foliar injury potential between exposure scenarios similar to those used by Kohut (2007)?

I am not aware of any.

Comments on Chapter 7: Synthesis

20. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall welfare exposure and risk analysis?

This is useful and mostly clear but is really not a synthesis, perse. Rather it is a summary. I think that parts of this section should appear before the other sections. Perhaps as a type of executive summary.

Is the reference point of no ozone reasonable and defensible?

Dr. Peter Woodbury

Comments on the Welfare Risk and Exposure Assessment

Summary Comments – Major Topics

I appreciate the effort that has gone into developing the WREA. EPA staff have made a productive and creative effort to model multiple types of effects of ozone on different kinds of vegetation in different environments. The use of ecosystem services as a conceptual framework is helpful and timely.

Also, there are some efficiencies to be gained by eliciting our comments on a partial draft. But I do think it should be kept in mind that we are reviewing a partial draft, not a complete first draft. For this reason, you may get more comments than you'd like when we review a complete draft.

I found that the charge questions were overly narrow, and were formulated to elicit a "yes/no" answer. There were several major issues with the document that were not addressed by the charge questions. Thus while I have responded to the charge questions, I also provide below some overarching comments. I have labeled each issue below so that I can refer to them in my subsequent responses to charge questions.

(Issue -- crops) Much more attention and analysis is required for crop yield loss. As summarized in the ISA, there is very strong evidence going back for decades that ambient ozone exposure in many locations and years is causing yield loss of sensitive annual crops and trees. In the jargon of ecosystem services, food and feed production is a critically important provisioning service. Effects on sensitive crops such as soybean should be analyzed at the least at the same level of detail as is currently provided for individual tree species. Specifically, the WREA should provide crop yield loss estimates for individual crops separately for different regions for current ambient ozone conditions for individual years, as well as projected ozone scenarios. Lastly, at the end of this expanded section, it should be made clear that decades of evidence, as well as new confirming evidence clearly show that major crop and tree species experience growth and yield loss due to ambient ozone in many regions and years.

Also, it is very important to include yield loss due to ozone as an adverse effect for farmers in regions with high ozone exposure. Such analysis would provide a firm basis for estimating adverse welfare effects. It is inadequate to analyze only sector-wide economic effects of reduced crop yield due to ozone, as is currently done using FASOM-GHG. The sector-wide approach is inadequate because it does not account for yield and income loss experienced by farmers in areas with elevated ozone. Farmers in low ozone regions will be winners of while producers in high ozone regions will be losers. In other words, ozone is causing a shift of income from farmers in higher ozone regions to those in lower ozone regions. Greater income to farmers in low ozone regions does not any way offset the loss of income to farmers in high ozone regions. Thus analyzing only sector-wide impacts misses the important effects on crop yield and on farmers in high ozone regions. This point is related to another major point that impacts should not be averaged among high ozone regions and lower ozone regions. Additionally, soybeans are a global

commodity, and reduced yields in the US due to ozone decrease the competitiveness of US farmers compared to farmers in other countries with lower ozone levels.

(Issue -- competition) More attention needs to be paid to effects of ozone on competition among species that are highly sensitive to ozone competing with species that are less sensitive to ozone in mixed-species stands. For herbaceous species, there are substantial data from experiments showing differential effects on sensitive species grown in mixtures with other species. For woody species, there are ample data on differential sensitivity among species, and ample evidence from theory and physiologically-based modeling studies about effects on sensitive species grown in competition with less sensitive species. This is an important topic because competition exacerbates effects of ozone on sensitive species of vegetation.

(Issue – sensitive) For the risk analysis, do not average effects on sensitive and less sensitive species. This comment applies to both crop species and tree species.

Crops such as soybean are sensitive to ozone and are widespread. Averaging effects on sensitive crops such as ozone along with less sensitive crops gives a false impression of less effect of ozone.

For tree species, there is the same problem with averaging effects of ozone among sensitive and less sensitive species. Also, there is another problem that in many parts of the USA, tree species grow together in mixed species forests, where they compete for light and other resources. See my other main point about competition.

(Issue – regions) For the risk analysis, do not average effects on high ozone regions and low ozone regions. Instead, such high and low exposure regions should be analyzed and discussed separately. This comment applies to both crop and tree species. In high ozone regions, there may be substantial predicted RBL for tree species and for yield for sensitive crop species. While national averages may be useful, they can obscure important regional effects due averaging in large areas of low ozone exposure with areas of higher ozone exposure.

(Issue – scaling) More attention needs to be paid to scaling from tree seedlings to mature trees. This is a well-known issue and there are substantial data in the literature addressing this issue. Previous EPA documents have discussed this issue in depth, including both experimental results on net photosynthesis and growth, as well as modeling studies investigating differences among seedlings and mature trees.

(Issue – background) The issue of how to determine a background ozone exposure pattern to be used for risk assessment needs to be resolved. For example, page A-42 (WREA Appendix) indicates that yield is compared to a “clean air” background, and that relative yield gains of crops and trees is assumed to be zero at ambient ozone. It is not clear to me what this means, is it assumed that ambient ozone is not currently affecting crop and tree growth? This is not supported by the evidence.

(Issue – appendices) Lastly, I think some material that is in appendices needs to summarized in the document itself. There are critical assumptions made, for example, in the FASOM modeling that are buried in an appendix.

Comments on Chapter 1 - Introduction

1. Does the Panel find the introductory and background material, including that pertaining to previous reviews of the O_3 standards and the current review, to be clearly communicated and appropriately characterized?

Yes, this section is useful and should be retained.

Comments on Chapter 2 - Conceptual Model

2. To what extent does the Panel find that the materials accurately reflect and clearly communicate the currently available welfare effects evidence, and the relevance of that evidence for quantitative exposure and risk assessment, as characterized in the 3rd Draft ISA?

This section is also useful and should be retained. However, the next draft should distinguish better what is well known from what is not well known. While there is a lot that is not known about ozone effects on vegetation, there is also a lot that is known, and it is critical to distinguish what is well known from what is less well known. Please refer to my detailed comments for the PA for specific examples of both misleadingly vague language and good, clear language.

Comments on Chapter 3 - Scope

3. To what extent does the Panel find the scope of the welfare risk and exposure assessment to be clearly communicated?

This section is also useful and should be retained. However, much more attention should be paid in the WREA (and the PA) to crop yield loss (see issue "crops" above and response to Question 13 below). Also, more attention needs to be paid to (1) effects of ozone on competition among ozone sensitive species and less sensitive species in mixed-species stands (see issue "competition") and (2) scaling from tree seedlings to mature trees (see issue "scaling").

Comments on Chapter 4 - Air Quality Considerations

Comments for specific page numbers:

NOTE: Page and line numbers may not match the most recent draft, because I had already completed my (hand written) notes before the newer version became available.

Page 4-7. The rollback method is useful, but does not include the potential for increased ozone exposure due to increases in precursor emissions that could occur with new development, for example increased industrial development, petroleum extraction and processes, etc.

Comments on Chapter 5 - Ecological Effects

7. Regarding the assessment of relative biomass loss (RBL) for individual species:
a. What are the views of the Panel on the use of the linear model forced through the origin to assess the proportional relationship between the relative biomass loss (RBL) values for each species comparing the RBL at recent ambient conditions to the RBL under the scenario modeling O3 just meeting the current standard?

It seems fine to force the function through the origin; this represents the structure of the data. The main problem with this approach is that regional variation is not addressed. See also detailed comments (by page number) below.

b. To what extent does the Panel find that this an appropriate analysis to compare the proportional changes in RBL? Does the Panel have suggestions for alternative approaches that provides a comparable result and maintains the cell-by-cell approach to help control for environmental variability?

This approach is reasonable, except that regional impacts should be addressed in addition to national impacts as discussed elsewhere in my comments.

c. To what extent does the panel agree with the approach used to combine the 11 tree species into one analysis?

Combining the 11 species gives the false impression of a modest overall effect, while actually there are very large effects for some species (see issue "species"). Furthermore, the text should analyze and discuss effects on different regions, not just an average response across all grid cells (see issue "regions").

8. Regarding the assessment of RBL for combined species:
a. To what extent does the Panel support the use of the Importance Values from the U.S. Forest Service to weight the RBL values in extrapolating from individual trees to larger ecosystem level effects?

There are substantial problems with this approach, see detailed comments for this chapter below.

b. What are the views of the Panel on the use of the summed-RBL as a metric to use for assessing effects at the larger ecosystem scale?

There are substantial problems with this approach, see detailed comments for this chapter below.

c. Does the panel have any recommendations for methods to include a wider range of tree species (beyond the 8 species included)?

9. What are the views of the Panel on the use of federally designated Class I and Critical Habitat areas as endpoints for this analysis? Does the Panel have any suggestions for additional parks beyond Rocky Mountain National Park and Sequoia/Kings Canyon National Park that should be analyzed? Does the panel have recommendations for additional or alternative geographic analysis areas that could be used?

These special areas are appropriate as endpoints, but currently are overemphasized. More emphasis (and analysis) should be focused on regional impacts of ozone on tree seedlings, forests, crops, and other vegetation.

10. To what extent does the Panel find that the vegetation mapping data from USGS and NPS used to generate a scaled-RBL surface in Great Smokey Mountain National Park is appropriate?

The use of the vegetation map data is appropriate, but there are problems with the overall method as discussed elsewhere in my comments.

11. Regarding the screening level foliar injury risk assessment:
a. To what extent does the Panel find the updated assessment of foliar injury risk in national parks originally performed in Kohut (2007) to be an appropriate screening level risk assessment?

In general, this analysis is useful and appropriate, but see detailed comments below.

b. What are the views of the Panel regarding the potential methods for estimating O_3 exposure at additional parks?

c. What are the Panel's views regarding the appropriateness of requiring that two criteria must be satisfied (i.e., based on both W126 and N100) in order to receive higher risk ratings?

The W126 criterion might be adequate on its own.

d. Is the Panel aware of any assessments of foliar injury in national parks conducted between 2006 and 2010 that could potentially be used to validate the updated risk ratings?

No, but the USDA Forest Service analyses may cover regions near the parks.

12. Regarding the assessment of cover of O_3 sensitive species:
a. To what extent does the Panel find the preliminary analysis of sensitive species cover to be an appropriate and useful approach to highlight areas of potentially higher risk due to the presence of sensitive species?

This topic is important and the analysis has some utility, but there are problems with the current analysis as discussed elsewhere in my comments.

b. To what extent does the Panel find the vegetation mapping data appropriate to assess the cover of O_3 sensitive species in GSMNP?

c. What are the views of the Panel on the decision to not distinguish between vegetation strata (i.e. herb, shrub, tree)? To what extent does the Panel agree with this methodology relative to analyzing the strata individually?

I don't understand the rationale for combining the strata.

d. What are the views of the panel on using benchmarks, similar to those used in the Kohut analysis of foliar injury risk, to allow estimates of change between exposure scenarios?

Benchmarks should be based on quantitative data about foliar injury for sensitive species.

Comments for specific page and line numbers:

NOTE: Page and line numbers may not match the most recent draft, because I had already completed my (hand written) notes before the newer version became available.

Page 5-1, Table 5-1. For the last row of this table, analysis of alteration of terrestrial community composition should be address in the REA. For herbaceous species, there are ample data from experiments showing differential effects on sensitive species grown in mixtures with other species. For woody species, there are ample data on differential sensitivity among species, and ample evidence from theory and physiologically-based modeling studies about effects on sensitive species grown in competition with less sensitive species. This is an important topic because it is quite likely that competition exacerbates effects of ozone on sensitive species of vegetation (see issue "competition").

Page 5-5, Figure 5-1 and text. The figure legend should clarify that these C-R functions are based on data from seedlings, not from mature trees. There should also be more discussion of the issue of scaling from seedlings and saplings to mature trees. Previous versions of the ISA (or rather, criteria documents) covered this issue in some detail.

Page 5-8 and 5-9. For each species, it would be helpful to add a map of the change in biomass loss with the rollback to current and proposed standards.

Page 5-12, Table 5-3 (and elsewhere). There are problems with interpreting the proportion of RBL at the current standard. It would be clearer to present the change in RBL due to meeting the current and alternate proposed standards.

Page 5-13, Figure 5-7. Individual species responses should be presented separately on this figure, not combined, and letters should be used to indicate the individual species. Combining the 11 species gives the false impression of a modest overall effect, while actually there are very large effects for some species (see issue "sensitive"). Furthermore, the text should analyze and discuss effects on different regions, not just an average response across all grid cells (see issue "regions").

Page 5-16 and 5-17, Figures 5-9 and 5-10. I don't know what this figure is supposed to represent, although I understand how the calculation was done (importance value * RBL). It does indicate areas with high importance values and high RBL values, but to what purpose? Elimination of sensitive species where they are not common (thus have a low importance value) might have very important ecological impacts. If the purpose is to combine for all (or for the most common) species to estimate stand-level RBL, then the approach is not valid. As shown in several modeling studies for forests (and experimental studies for herbaceous species), competition in mixed species stand means that the total stand RBL will be much less than the RBL for sensitive species. At the same time, effects on sensitive species are likely to be greater due to increased competition from less sensitive species (see issue "competition").

Page 5-17. This analysis is biased, because it implicitly assumes that all species other than the 11 with C-R functions have no response to ozone, which is not a valid assumption. This bias varies among cells in the map. This bias is stated in the text, but that does not eliminate the bias, nor solve the issue that the bias varies across the map.

Page 5-17. What is the rationale for excluding the most sensitive species (Eastern Cottonwood)?

Page 5-22, Figure 5-15. This figure is not well designed, the Class 1 areas are difficult to see and the exposure values are not clear. I suggest using black for the Class 1 areas, using light grey for the state boundaries, and using inset colored circles to show the W126 exposure values.

Page 5-14, line 14. For this analysis, rather than an arbitrary value of W126 of 10, use cutoff values that are relevant to the proposed alternate standards.

Page 5-23. Table 5-4. In the left-most column of this table and elsewhere in this and other tables, align numbers on the decimal point.

Page 5-31, Figure 5-19 (and subsequent similar figures). Remove title from top of figure. Add some "base map" information for orientation, such as State boundaries, major highways, and major towns near the Park. Spell out abbreviations in all figure titles (such as GSMNP).

Page 5-33. See previous comments on methodological problems with this "scaled" approach.

Page 5-39, Table 5-7. Why was the "moderate" risk class used by Kohut not included here? It is discussed in the text, but should be included in the table. Also, please clarify how and why the ratings differ from those by Kohut.

There needs to be more discussion of how to interpret these risk classes. Actually, it would be more useful if possible to develop more quantitative risk characterization, such as predicted percentage of species with foliar damage and degree of foliar injury.

Page 5-44. This document should review the results of field studies (key older ones with more attention to newer ones) and results of USDA Forest Service foliar injury assessments.

Comments on Chapter 6 - Ecosystem Services

13. To what extent does the Panel find that EPA has adequately characterized the range of ecosystem services that are potentially adversely affected by O3?

Much more attention should be paid in the WREA (and the PA) to crop yield loss. There is strong evidence that common crops have substantial yield loss. This is a critically important provisioning service. Effects on sensitive crops such as soybean should be analyzed in detail in the WREA at the same level of detail as currently provided for individual tree species (see issue "crops).

Also, it is very important to include yield loss due to ozone as an adverse effect for farmers in regions of high ozone exposure (see issue "regions"). Such analysis would provide a firm basis for estimating adverse welfare effects. It is inadequate to analyze only sector-wide economic effects of reduced crop yield due to ozone, as is currently done using FASOM-GHG. The sector-wide approach is inadequate because it does not account for yield and income loss experienced by farmers in areas with elevated ozone. Such effects are real and cannot be "averaged out" by counting potential increases in the price of soybean, because such increases will go primarily to farmers in low-ozone areas. In other words, even if there is only a modest "net" effect of ozone on net producer income in the USA overall, there is still an adverse effect on many farmers in areas with elevated ozone. Ozone is causing a shift of income from farmers in higher ozone regions to those in lower ozone regions. Additionally, soybeans are a global commodity, and reduced yields in the US due to ozone decrease the competitiveness of US farmers compared to farmers in other countries with lower ozone levels.

14. To what extent does the Panel agree with EPA's ecosystem services framework, connecting O3 exposure, through ecological effects to ecosystem services?

15. Does the panel agree with EPA's use of combined O3 exposure data with other data sources (e.g. fire data, bark beetle maps, trail maps) to link areas of concern or interest with areas of higher vegetative risk due to O3? Does the Panel have any recommendations for additional datasets and ecosystem services that could add to or improve these analyses?

The approach is conceptually sound. However, because effects of ozone on so few services can be analyzed quantitatively, these sections should be much shorter, as discussed elsewhere in my comments.

16. Regarding the analysis of forest yield impacts:

a. To what extent does the Panel agree that the Forest and Agricultural Sector Optimization Model (FASOM) model is appropriate to assess timber and crop yield changes and the effects of those changes on additional ecosystem services?

There are serious shortcomings with the endpoints used in this analysis, and in the inability of this approach to account for competition among species and to account for scaling from seedlings to mature trees. Please see my detailed comments elsewhere.

b. What are the views of the Panel on the extrapolation of concentration-response functions across similar species?

In general, such an approach is reasonable, but it must be based on physiological and other biophysical similarities among species, not just economic groupings. Please see detailed comments elsewhere.

[The appendix detailing the FASOM analysis will be submitted for review in August]

While much of this chapter is useful, well written, and appropriate, there are serious deficiencies due to the narrow focus on Class 1 areas and on selected tree species. Correcting these deficiencies will require modest expansion of the WREA to include analysis of effects on the yields of sensitive crops in order to appropriately address adverse welfare effects (see issue "crops"). Such expansion can be done based on methods and analyses already completed.

Additional analysis of effects on crop yield must be included, as discussed in my comments elsewhere including for the PA. Discussion of FASOM-GHG and i-Tree analyses should only be included if they are appropriate and robust, and should be reviewed by this panel. Page A-41 (WREA Appendix) indicates that ozone exposure-response functions for tree seedlings were used to calculate RYLs over their whole life span. Scaling effects from seedlings to mature trees is an important topic that has received substantial attention in the literature as summarized in previous EPA documents for previous ozone standard reviews (see issue "scaling"). Similarly, competition among tree species that differ in ozone sensitivity in mixed-species stands is expected to greatly alter how ozone affects the growth of an individual species (see issue "competition").

Page A-42 (WREA Appendix) indicates that yield is compared to a "clean air" background, and that relative yield gains of crops and trees is assumed to be zero at ambient ozone. It is not clear to me what this means, is it assumed that ambient ozone is not currently affecting crop and tree growth? See issue "background". This is not supported by the evidence. Furthermore apparently only sectoral net results are calculated for consumers and producers. As discussed further in my comments to the PA (Chapter 7), such analysis does not account for the individual producers some of whom will be winners and some of whom will be losers. The evidence strongly suggests that ambient ozone exposure in many locations and years is causing yield loss of sensitive crops and trees. In such locations, producers (and perhaps consumers) will be losing income due to ozone, and that is an adverse effect on them, even if producers in other regions are gaining income (see issues "crops" and "regions". The results of the FASOM

modeling may be useful, but effects on sensitive crop and tree species in higher ozone regions should also be quantified in addition to sector-wide economic impacts.

17. Regarding the analysis of urban forest impacts:
a. To what extent does the Panel feel that the i-Tree model is appropriate for assessing changes to urban forest ecosystem services based on O3 exposure?

There are problems with this analysis, including scaling from seedlings to mature trees (see issue "scaling"), and the need for clarification about how robust the predictions of i-Tree are for specific endpoints. Please see additional comments elsewhere.

b. In order to increase the number of tree species covered by the iTree model, does the Panel have recommendations for additional species that could be included, based on estimates from similar species?

[The appendix detailing the i-Tree analysis will be submitted for review in August]

18. Regarding the use of PnET-CN:
a. What are the views of the Panel on the potential use of the PnET-CN model in the 2nd draft to assess impacts on larger scale ecosystem services (e.g. hydrologic changes, c sequestration)?

As far as I am aware, this model uses a "big leaf" approach to modeling canopy processes. This approach cannot account for the competitive interactions among ozone sensitive and less ozone sensitive species in mixed species forests, which are very common in the Eastern USA. Without inclusion of such processes, and without data to parameterize the responses of mature trees and forest stands to ozone, it is not clear how robust predictions from this model will be for hydrological processes nor for carbon sequestration. See issues "scaling" and "competition".

b. Does the Panel have recommendations of other models that are accessible to EPA that could be used instead of PnET-CN?

The ISA does cite some literature using combinations of whole-tree models and stand-level models to address these issues. For example, the individual tree model TREGRO and the forest stand model ZELIG have been used previously to scale effects of ozone on net photosynthesis to mixed-species forest stands. Such models could be used, or at least more attention could be paid to publications that have already performed such analyses for some regions. I have provided such citations in the past, and they have also been cited in previous EPA analyses, but I am glad to provide such references again upon request.

19. Regarding ecosystem services related to foliar injury:
a. To what extent does the Panel agree that potential visible foliar injury is appropriate to use as a metric to assess potential loss of cultural services associated with recreation in national parks?

This metric is appropriate.

b. Does the Panel feel that there are O3 benchmarks that could be used to assess changes in foliar injury potential between exposure scenarios similar to those used by Kohut (2007)?

Comments for specific page and line numbers:

NOTE: Page and line numbers may not match the most recent draft, because I had already completed my (hand written) notes before the newer version became available.

Page 6-3, Table 6-1. Change "ruffled" to "ruffed".

Page 6-6, line 20-21. Clarify that effects on food and timber production can be quantified via measurements and modeling. This is a major deficiency of this document that should be changed.

NOTE: Page and line numbers below are based on the most recent draft.

Page 6-8, para 1. The numbers in the text don't seem to match those in Table 6-2.

Page 6-8, Table 6-2. What is the basis for the "Forest Management" values in this table? See previous comments about the need to model competition among ozone sensitive species and less sensitive species in mixed-species stands.

Page 6-9. Line 6. Much more important sources of uncertainty are due to not accounting for (1) effects of ozone on competition among ozone sensitive species and less sensitive species in mixed-species stands (see issue "competition") and (2) scaling from tree seedlings to mature trees (see issue "scaling").

Page 6-12, Figure 6-1. I can't see the W126 values on this map, and suggest using a separate panel to do so.

Page 6-13, Effects on crops must be included as part of this assessment! See comments above under Question 13.

Page 6-14, Table 6-3. The FASOMGHG regions are not very meaningful without a map.

Page 6-16, line 10. Change "form' to "from".

Page 6-18, first para. By this logic, increased ozone concentrations create a benefit. Although I understand the rationale, this does not make sense; see comments for Question 13 above.

Page 6-21, Figure 6-4. Add legend to figure.

Page 6-26, line 13. "751 m" – what does this mean?

Page 6-34, last para. The WREA should allocate more effort and length to aspects/endpoints that can be quantified. A short section (a few pages) can mention other aspects that cannot be quantified, but should not be so long.

Page 6-36, line 8. Change "Smokey" to Smoky" throughout all documents.

6-47. See comments elsewhere about the need to account of ozone effects on competition among ozone-sensitive and less sensitive species. See also comments elsewhere about FASOM analyses.

Page 6-48, Table 6-25. Add column header for Column 1 (perhaps it is rank by basal area, or count, or ??).

Page 6-49, Table 6-26. Does the base case mean zero effect of ozone on carbon sequestration? How are growth rates modeled, is the effect of ozone implicitly included in such modeling, if it is based on statistical measures of growth?

6-50. How robust are these predictions of tons of air pollutants removed? Which air pollutants? In general, for the i-Tree, FASOM, and other modeling analyses, more analysis of how robust the predictions are would be helpful.

Comments for Appendix 6-A (15Aug12 version)

Page 6A-39, line 29. Clarify which years were used, etc.

Page 6A-40, lines 8 and 19. How do these C-R functions relate to those used elsewhere in the WREA? If they are the same, why is a 2007 document cited? If they are different, why?

Page A-40, Table 14. I don't know what rationale was used to assume that sugar beets respond to ozone similarly to sorghum and hay. Similarly, why is canola assumed to be similar to soybean? Perhaps these are groupings for economic reasons, because they don't seem to be groupings based on the taxonomy or physiology of the species. Data on the ozone sensitivity of related crops should be used for such groupings.

Page 6A-41. Ozone exposure-response functions for tree seedlings were used to calculate RYLs over their whole life span. Scaling effects from seedlings to mature trees is an important topic that has received substantial attention in the literature as summarized in previous EPA documents for previous ozone standard reviews. Similarly, competition among tree species that differ in ozone sensitivity in mixed-species stands is expected to greatly alter how ozone affects the growth of an individual species (see issue "competition").

Page 6A-42. See comments above under Question 16.

Page 6A56. See previous comments herein and for the PA about the inadequacy of examining only net changes in sectoral income for producers and consumers.

Page 6A56, Table 23. What assumptions are made regarding international trade? If ozone causes reduced yields in the US, would that reduce the ability to compete with Brazilian (or other) soybean production that might have lower ozone levels and thus higher production?

Comments on Chapter 7 - Synthesis

20. To what extent does the Panel find the synthesis to be a useful integration and summarization of key results and insights regarding the overall welfare exposure and risk analysis?

The synthesis should include discussion and quantification of crop yield loss due to ozone, as discussed in comments above. This is an important topic that must be addressed thoroughly in the WREA and carried into the PA.

Comments for specific page and line numbers:

Page 2, line 26. A few sentences should be added to put these results in context.

Page 3, line 12. A few sentences should be added to put these results in context.

Page 3, last para. Effects on crop yield should be included here, see previous comments.

Page 4, line 6-9. Misleading, see previous comments. Effects on individual species are also important, not just all 11 species together. Effects on mature trees, as opposed to seedlings are important. Effects on competition among ozone-sensitive and less sensitive species are very likely to be important.

Page 4, line 24. A few sentences should be added to put these results in context.

Page 7, lines 5-7. The USDA Forest Service Forest Health Monitoring data are very important and should definitely be included.

Appendix B

CASAC Ozone Review Panel Assignment of Priorities (Red Text) of Additional Analyses under Consideration by EPA for Inclusion in the Second Draft Ozone Health Risk and Exposure Assessment

Chapter 2 – Conceptual Framework

1. Addition of conceptual framework diagram and additional text laying out conceptual framework in more clear terms, e.g. major steps in analysis, key inputs, analytical outcomes, risk characterization. - **High**

Chapter 4 – Air Quality Considerations

1. Continued evaluation of CMAQ-HDDM methodology for adjusting ozone air quality distributions to simulate just meeting the current and alternative standards. Conditional on positive evaluation, considering implementing CMAQ-HDDM adjustments for the 2nd draft REA, including the following updates to the CMAQ-HDDM model:
   a. Update to the 2007 modeling platform - **A Given**
   b. Include a longer modeling period (we have not determined if we will model the whole year or just May-Sep, Apr-Oct . . . this will depend on what is feasible time-wise and what is deemed necessary) - **Move to whole year, as resources allow**
   c. Update to CMAQv5.0 if it is available (HDDM is not yet available in CMAQv5.0 and there is no guarantee that it will be ready in time but if it is we'd like to use it) - **Low**
   d. If time permits, explore other approaches for grouping modeled and observed days into bins and calculating average sensitivities. - **Low**
   e. Expand to 16 cities (from the 2 cities evaluated in the Simon et al memo) - **Must**
   f. Considering expanding to two DV periods (we only analyzed 2006-2008 for the memo but we plan to add 2008-2010 for the 2nd draft) - **Must**
   g. Evaluating possibility of using combinations of NOx/VOC reductions. - **Do only for a few cities**

An additional item we are now considering:

2. Evaluation of feasibility of estimating background ozone both using the CMAQ-HDDM model configuration and other available models, as well as considering multiple estimates of background based on alternative definitions - **Document in ISA use of HDDM**

Chapter 5 and Chapter 6

1. Addition of 12 additional urban areas for a total of 16 urban areas - **A Given**
2. Addition of exposure estimates for outdoor workers and people > 65 years old - **High**

3. Addition of exposures estimated for air quality just meeting alternative standards and change in exposure between scenarios just meeting current and alternative standards. - **High**

Chapter 5 (Exposure Assessment)

1. Improvement in the depth and clarity of documentation, presentation of results, and discussion of results; present distributions of exposures - **High**
2. Addition of discussion of representativeness of the 16 urban areas for exposure assessment – **High**
3. Addition of sensitivity analyses (effects on exposure distributions and lung function risk):
a. Considering evaluation of exposures above background concentrations - **Potentially High**
b. Considering evaluation of representativeness of CHAD activity patterns (geographic location, SES factors, occupation, changes over time for adults and for children, high/medium/low ozone days, and other factors to be determined) - **High**
c. Considering developing a method for spatial interpolation of ambient ozone concentrations - **Low**
d. Considering evaluation of exertion levels associated with activities (especially outdoor activities) (model performance validation) - **High**
4. Analysis of repeated exposures for cohorts of school children (i) during the school year, (ii) during the summer (vacation) - **High**
5. Additional clarification of types of uncertainty characterization (qualitative, sensitivity analyses, influence assessment, etc.) - **High**

Chapter 6 – Lung Function Risk Assessment

1. Conditional on availability of data from study authors, update the two lung function risk models with data from additional clinical studies (including Kim et al. (2011) and Schelegle et al. (2009)). - **High**
2. Conditional on availability of data from study authors, and inclusion in the final ISA, evaluate the potential for using the McDonnell et al. (2012) model in estimating lung function decrements. To implement the model in APEX, EPA needs the data used to fit the model to evaluate the uncertainty of the model. - **High**

Chapter 7 – Urban Area Case Studies

1. Addition of assessment of risks remaining after just meeting alternative standards and incremental risk reductions for alternative standards - **High**
2. Addition of assessment of respiratory mortality risks from long term ozone exposures - including potential sensitivity analysis based on modeling risk using regional effect estimates - **High**
3. Addition of confidence intervals for estimates of risk reductions between recent conditions and just meeting current standards and between current standards and alternative standards - **High**
4. Conditional on availability of data from study authors, consideration of adding sensitivity analyses for alternative model specifications (e.g., lag , co-pollutant, regional vs national priors) for

short-term exposure-related mortality and morbidity - **Low except for co-pollutant (high priority); if there is literature relating to specific city on these factors, discuss**

5. Consideration of analysis of composite monitor constructs, e.g. MSA averages, used in source epidemiological studies to determine if we should more closely match those structures in our risk modeling – **Low, but beef up citations**

6. Addition of further analysis of patterns of interest in risk assessment in order to provide better evaluation of results (e.g., explanation of substantially larger effect estimates for short-term exposure related mortality in some cities) - **Nice to do**

7. Conditional on availability of data from study authors, use of actual LML values from epidemiological studies (alternatively, if those data are not available, we are considering constructing LML values based on AQS data for the time periods matching those used in the studies) - **Low**

8. Consideration of sensitivity analysis using interpolation of missing ozone data in deriving composite monitor values (similar to exposure analysis approach) - **Low**

Chapter 8 – National Scale Recent Conditions Risk Analysis

1. Addition of estimates of long-term exposure respiratory mortality - **High**

2. Correction of error in May-September average 8-hr daily maximum concentrations - **High**

3. Application of May-September city-specific concentration-response factors from Bell et al. (2004) if available - **High**

4. Conditional on availability of NMMAPS LML values from study authors, application of those LML values to Bell et al. (2004) mortality estimates - **Low**

5. Conditional on availability of data from study authors, consideration of using city-specific LMLs for cities included in NMMAPS - **Low**

Chapter 8 – Assessment of Representativeness of Urban Case Study Areas

1. Consideration of using American Time Use Survey (ATUS) data to compare national representativeness of selected urban areas for time spent outdoors - **Low**