Preliminary Individual Comments

CASAC Ozone Review Panel

On Health Risk and Exposure Assessment

Updated March 21, 2014

Contents

Mr. George Allen 2
Mr. Ed Avol 7
Dr. Michelle Bell 9
Dr. Joseph Brain 10
Dr. David Chock 13
Dr. Ana Diez Roux 21
Dr. Daniel Jacob 24
Dr. Steve Kleeberger 26
Dr. Fred Miller 27
Dr. Ted Russell 34
Dr. James Ulman 39
Dr. Sverre Vedal 41

Mr. George Allen

These comments focus on Chapters 1-4 of the Health REA.

General Comments.

Overall these chapters, especially chapter 4 (Air Quality Considerations) are substantially improved over the first draft REA, and are generally responsive to CASAC comments on that draft. The implementation of the HDDM rollback method is particularly impressive in its scope.

Charge Questions.

Chapter 1: Introduction

Q #1. To what extent 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?

This brief chapter clearly summarizes the recent history of the O3 NAAQS process, including the legal challenges to the 2008 final O3 NAAQS rule and the "reconsideration" process that EPA unsuccessfully pursued in response to those challenges. My comments on the Policy Assessment 2nd draft discuss the implications of these legal challenges and court rulings in more detail.

Chapter 2: Conceptual Model

Q #2. To what extent does the Panel find that the discussions accurately and clearly reflect the air quality, health effects, exposure and risk considerations relevant for quantitative exposure and risk assessment, building from information contained in the final ISA? What are the views of the Panel on the additional flowchart provided for the overall assessment and the additional information regarding specific elements of the exposure and risk assessments?

This chapter provides a useful and accessible overview of the complex REA process. The flowchart in Figure 2-1 (page 2-2) is helpful in understanding how the various parallel components of both exposure assessment models (APEX) and risk assessment (BENMAP)

support the overall risk characterization. Including the relevant REA chapters for each component in this flowchart is very useful. Each component of the REA process is clearly summarized in a well-structured manner; these summaries provide the reader with a contextual understanding of the more detailed discussions in later chapters. Figure 2-2 (page 2-11) is a very clear summary of the causal determinations of both long and short-term O3 health effects. Page 2-21, lines 24-29 provide a helpful discussion of "attributable risk" and how estimates of total risk remaining after meeting the existing standard form the "reference value" for evaluation of risk reductions from meeting potential lower standards.

Chapter 3: Scope

Q #3. To what extent does the Panel find the scope of the health risk and exposure assessment is clearly communicated? To what extent does the panel find the additional flowcharts for each analytical component to be useful additions?

This chapter covers the scope of the key design elements of the REA in a well-organized manner. The various "Conceptual Diagram" flowcharts are very helpful in helping the reader to understand each of these major components of the REA. Section 3-4, Air Quality Characterization, page 3-12, lines 7-21 provide a good description of how this 2nd draft develops risk estimates referenced to zero O3 concentration and from all sources – a substantial change from how previous risk estimates were developed, and responsive to CASAC comments from the previous draft REA.

Chapter 4: Air Quality Considerations

Q #4. What are the views of the Panel on the appropriateness of the methods used to characterize O3 air quality for the exposure and risk assessment? What are the views of the Panel on the HDDM-based adjustment methodology used to adjust O3 concentrations to just meet the existing O3 standard and alternative standards?

This chapter describes and characterizes the results of the CMAQ-based HDDM adjustment method, as the CASAC recommended in the review of the first draft REA. This is a radical departure from the previous "rollback" mathematical-only simulations in previous O3 NAAQS reviews which while computationally simple were inadequate for this purpose. The HDDM approach takes into account the complex precursor O3 chemistry and how a given emission

reduction scenario may result in both benefits and dis-benefits across the urban to suburban scales.

EPA staff are to be commended for undertaking this major effort and publishing the improvements made to the HDDM rollback method in the peer-reviewed literature (Simon et al. 2013, doi: 10.1021/es303674e), including 129 pages of supporting information. The detailed breakdown of the HDDM method in section 4.3.3.1, pages 4-15 to 4-19, along with the flowchart in figure 4-6, page 4-17 serve as a reasonably detailed description of how the method was implemented for this review.

It is sometimes unclear throughout this chapter if adjustments were based on the equal NOx/VOC emission reduction scenario or a NOx only scenario; see page 4-18 lines 10-27 for one example. The discussion of limitations of the HDDM approach encountered for the substantial emission reduction scenarios needed for New York City and Los Angles (page 4-18 lines 28-35) could be expanded upon – what % reduction of which scenario(s) was required to meet the lower alternative standard levels? Also it seems counter-intuitive that the model's inability to estimate hourly O3 distributions would be encountered at a higher concentration in New York (65 ppb) than in Los Angles (60 ppb), given that Los Angles current O3 levels are substantially higher than for New York City.

Q #5. To what extent does the Panel find that the discussion of uncertainty related to the air quality inputs to the exposure and risk assessment appropriately covers important sources of uncertainty?

The specific limitations of the HDDM method for simulating lower alternative standards in New York City and Los Angles are discussed on page 4-18 lines 28-36 and page 4-19 lines 1-6. The "significantly more uncertain" risk estimates (due to the need to use the 95th percent CI lower bound estimates) is noted; a brief qualitative description of this uncertainty would be useful here.

Section 4.3.3.2, Resulting Air Quality (page 4-19), presents a good overview of how the HDDM adjusted distributions behave in general for both spatial and temporal patterns. A useful comparison of the quadratic rollback and the HDDM adjustment methods is presented in this

section and figures 4-7 and 4-8. It is noted on page 4-20 line 1 that in general hourly O3 distributions are shifted upward with HDDM compared to the quadratic rollback method.

The more general limitation of the HDDM approach for cases where large emission changes (perturbations) are needed to meet lower alternative standard levels (60-65 ppb) is clearly explained in section 4.5, page 4-38 to 4-41. This section quantitatively presents the uncertainty for 50 and 90% NOx reduction conditions, and concludes that the uncertainty of the HDDM method "is small up to 90% emission cuts" (for NOx, page 4-39, lines 9-10). Specific uncertainties are presented for the 15-city case study areas for 75 and 65 ppb cases (page 4-39, lines 27-31). It is unclear if these are NOx-only emission reduction scenarios (as discussed earlier on this page) or not. Table 4-6 (pages 4-42 to 4-51) is a useful summary of qualitative uncertainty for key AQ elements of the risk analysis.

Specific comments.

Page 2-7 lines 24-27: just equal NOx and VOC emission reduction scenarios? or NOx only?

Page 3-12 line 1: "urban cast" typo.

Lines 2-3: NOx only emission scenarios were also considered and should be mentioned here since NOx reductions are likely to be the more effective approach in most scenarios.

Page 3-20 line 7: sentence ends prematurely. "causal relationship with." …

Page 4-2, lines 9-12: it is not "generally agreed" that Castnet O3 data going back to 2006 is of comparable quality to the data reported to AQS. While Castnet QA/QC requirements now (since 2011) meet those required for compliance monitoring by state and local air agencies, the older O3 data did not meet those requirements. See public comments by Alan Leston for more detail. This does not mean that those data cannot or should not be used in the REA; they fill a gap in rural primarily eastern US regions and do have value in this process.

Page 4-5, footnote 6 says composite monitors do not always include the highest design value monitor in every urban area. Page 4-14 lines 12-14 says all monitors were used. This needs clarification.

Page 4-7, Figure 4-4, bottom plot: Sacramento 2008 is oddly elevated by ~ 20 ppb. Is this 4th highest value correct? If yes, a brief explanation in the text or a footnote may be useful.

Mr. Ed Avol

Comments EPA O3 Document, 2nd Draft REA

CHAPTER 4

Fig 4-9 & 4-10, pp4-26 on… I'm curious, and a little concerned, about the HDDM adjustment approach and its admission that downwind surburban sites are excluded from consideration of adjustment impacts. The HDDM adjustment admittedly does not completely account for or consider the downwind rural ozone concentration effects of NOx or VOC adjustments in the upwind urban areas. In several urban locations (like Los Angeles, where downwind reporting sites such as Lake Arrowhead may be excluded from assessment of ozone adjustment impacts evaluations), a potentially misleading picture of decreasing levels in the immediate and downwind vicinity may therefore be created. Since any single monitoring site in a region can place the region in non-compliance, (and locations like Lake Arrowhead have historically been the "high" site in the Los Angeles region), what does exclusion of persistently high downwind sites from the analysis imply for acceptance/endorsement/utility of the HDDM adjustment approach?

CHAPTER 5 MODELING

General Comment – Although I appreciate that one has to select some framework for discussing comparative reductions and portions of the population affected, there is at least one somewhat disquieting aspect to the presentation and interpretation of it for me, and that is the way in which the percentage of populations in the urban case study areas is used. The 15 Urban Case Study Areas represent (Table 4-1, p4-6) populations ranging from 2.7 million to 21 million people (2010 Data), an almost ten-fold range. The ensuing discussions and comparisons present these study areas in terms of percent of children or adults affected at some design value (60, 65,70, or 75 ppb O_3, for example), but 5% of the New York population (at 21+ million) arguably means something quite different, in terms of affected individuals, than 5% of Baltimore's population (at 2.7+ million). Moreover, the percentage perspective does not take into account any adjustment for a "vulnerability" factor (due to race, ethnicity, access to health care, pre-existing conditions etc), so discussion of proportional changes across urban case study areas glosses over some very important indices of public health. I realize that the intent of the presentation is not to compare Atlanta to Sacramento, but still, proportional changes in air quality will likely have disproportional changes in regional public health outcomes…and that observation or acknowledgement does not seem to be considered in any substantive way in the document portions I reviewed (and to my knowledge, is not included in the exposure-response function in the McDonnell-Stewart-Smith model referred to in the HREA).

CHAPTER 8 NATIONAL SCALE MORTALITY BASED ON EPI STUDIES

(nothing to usefully add here; BENMAP is not in my portfolio of expertise)

CHAPTER 9 SUMMARY

General Comment: This ~50 page chapter is a summary? There is a lot of detail and re-visiting of issues that could have arguably been housed in this detail in the other chapters. Additionally, there is no differentiation between key summary conclusions and more general information in this chapter. Surely there must be some key (highlight-able?) summary points that the Reader should be aware of?

Section 9.2 is titled "Key Results", but it rambles on, showing tables and discussing approaches. With all due respect, this is NOT the way to present "Key Results"! This could and should be edited down to a paragraph or two per section, referring to the previous chapters wherein additional supporting details may be found. Key tables or figures should be included to emphasize critical points, not to reproduce them in much the same detail as in their respective home sections. In its current form, much of the important summary conclusions are lost in the morass of re-presentation.

Some Specific Comments

1. P9-1, lines 4 to 6: re-phrase to say, "…short-term O_3 exposures are causally related to both respiratory and cardiovascular effects, and that…"
2. P9-8, Section 9.22 Human Exposure Modeling (Chapter 5)
3. P9-9, lines 2 to 4: delete the second half of this sentence (it doesn't add anything here) and leave sentence to state that "Persons spending a large portion of their time outdoors during afternoon hours experienced the highest 8-hour O_3 exposure concentrations."

Dr. Michelle Bell

Chapter 5. Characterization of Human Exposure to Ozone

The section on APEX would benefit greatly from some discussion of its accuracy. Can a section on evaluation of APEX be included? As written, the section tells us what Apex does, but not how well it does it.

The overall point of the APEX analysis is not made clear in the introductory portions of Chapter 5. As the regulations and most health analyses will be or were based on ambient levels or controlled exposures. Just add more description and detail here for how the results will be used. In general, this chapter could use more discussion of the implications of the results, in addition to the numerical estimates.

The relation of these values, estimated by APEX, in comparison to clinical evidence for benchmarks (page 5-18), needs explanation as the health-based studies are typically based on an overall ozone level, not accounting for activity patterns and so on.

The section on meteorological data used (5.2.4) is a bit vague in places. E.g., “a few to several meteorological stations”. Information on the imputed data (% imputed), etc. would be helpful.

Chapter 8. National scale mortality risk burden based on application of results from epidemiological studies

Overall, this chapter is well written and explained. The methods, interpretation and presentation of methods and results are technically sound and well communicated.

Provide the rationale for the selection for the concentration-response functions used in this section. The choice of these studies needs to be justified. This discussion may belong in Section 7, which gives some information on this (around page 7-18), but it’s still unclear.

The chapter discusses that higher certainty exists for urban areas, so it may be useful to note such areas (the ones used on the original CRFS) on a map or some of the existing maps.

This chapter would benefit from information on why city-specific and national effects may differ. In general, Chapter 8 would benefit from reference to sections in Chapter 7.

Dr. Joseph Brain

Chapter 1: Introduction

1. To what extent 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?

Chapter 1 is brief – six pages long – and yet it does an excellent job in introducing the entire document. The complex history of the ozone standard is well described, including the controversial decisions and legal challenges that occurred during the past decade. The essential facts are there and links to more detailed documents are given. The summary of the goals and approaches is helpful. The final section "Organization of Document" is a brief but useful guide.

What's missing? After informing the reader that the current standard is 75 ppb, based on the annual 4th-highest daily maximum 8-hour average concentration, averaged over three years, there is no comment as to exceedances. An important context early in the document is the status of the current ozone standard. Is it like CO that exceedances are few and far between, or is the current standard problematic?

Chapter 2: Conceptual Model

2. To what extent does the Panel find that the discussions accurately and clearly reflect the air quality, health effects, exposure and risk considerations relevant for quantitative exposure and risk assessment, building from information contained in the final ISA? What are the views of the Panel on the additional flowchart provided for the overall assessment and the additional information regarding specific elements of the exposure and risk assessments?

Chapter 2 has been rewritten and is clearer and more effective than in the previous draft. It clearly shows how this current model builds on information from the now final version of the ISA. I find the flow chart, Figure 2-1, clear and helpful. As indicated in my comments on Chapter 1, why not discuss the topic of exposure and risk assessment in the context of exceedances for the current ozone standard.

One topic that seems to be missing, but was prominent in previous discussions, was the policy-relevant background (PRB) and how that influences our interpretation of the available data. Is it appropriate to compare levels of ozone where health effects are observed to zero ozone, or should they also be compared to the PRB?

In regard to Section 2.2.5, when discussing "exposures of concern," the agency should better defend the three benchmark levels selected, 60, 70, and 80 ppb. One would have thought that one of the benchmark levels would be the current standard, 75 ppb. Why isn't that fourth level included? Don't we really want to know how these other three levels – one above and two below – differ in their consequences from the current standard?

Figure 2.2, "Causal Determinations for O_3 Health Effects," is simple and helpful, but I have several questions about it. Don't we need a brief definition of the three categories: suggestive, likely, and causal? It is true that these "strength of evidence" categories are defined in the ISA, but few readers will have the persistence to look them up. Shouldn't we add a brief definition here of these important categories and what they mean?

Some issues remain unexplored that may be important. For example, what are the practical consequences of adaptation, a well-known phenomenon for ozone exposures? Is a high level of ozone more serious for individuals and communities that rarely encounter elevated levels of ozone compared to more adapted individuals who experience elevated ozone levels much of the time, and perhaps have greater exposures out-of-doors because of climate, and exercising more often outside? Animal data clearly suggest that responses to the same ozone level are influenced by ozone exposure history.

I like the discussion of at risk populations, 2.2.8. In previous documents, we had trouble with this, since at risk could point to increased likelihood of exposure as well as inherent reasons for greater susceptibility, e.g. age and preexisting diseases. A persistent problem is always the limits of at risk populations. To what extent do we base standards on more susceptible populations and to what extent do we want standards to be driven by small numbers of hypersusceptible groups or individuals? For an asthmatic in hospice because of cancer, any increase in ozone may be life threatening.

Overall, Chapter 2 is dramatically improved.

Chapter 3: Scope

3. To what extent does the Panel find the scope of the health risk and exposure assessment is clearly communicated? To what extent does the panel find the additional flowcharts for each analytical component to be useful additions?

I believe that Chapter 3 does a good job of clearly communicating the scope of the health risk and exposure assessment. In particular, the figures and flow charts provide an excellent summary and guide the reader to the more extensive text. Figure 3-1 fits well within the section on air quality characterization and helps the reader. Figure 3-2 accurately summarizes the accompanying text.

Figure 3-3 is useful in terms of linking exposure to uptake of ozone in the lungs. It delineates the multiple factors that affect the link between exposure levels in the ambient environment and the exposure of lungs. I am less happy with the bottom oval, which has all this culminating in changes in the FEV1. Why single out that parameter and ignore other outcomes? Modest changes in FEV1 per se may be of less significance than related inflammatory changes, such as more neutrophils in the BAL. Those changes may be more tightly linked to the pathogenesis of chronic lung disease. Focusing on the FEV1 may lead to minimizing the consequences of ozone exposures on a chronic basis. It also focuses, through the symbol delta, in acute changes. As the text mentions, there is greater interest in how ozone and other air pollutants might affect lung growth as well as the long-term declines in pulmonary function as we age.

Figure 3-4 seems more complicated, and at least for me, was not a very helpful introduction to the text. I'm happy with the final figure, Figure 3-5.

Table 3-1 raises the issue of the empty boxes, the unfilled cells: Why were the categories of respiratory hospitalizations, respiratory ED visits, or respiratory symptoms measured in so few of these cities. Why, for example, were respiratory hospitalizations measured in Detroit, Los Angeles, and New York, but not in the other nine USA urban areas?

I would again point out lack of clarity in terms of background levels of ozone and the related PRB. In the policy assessment draft, especially in Chapter 2, Section 2.4, there is an extensive discussion of nonanthropogenic sources/background ozone. Shouldn't this topic at least be mentioned here?

What do we make of this ozone background? Should health effects at particular ozone levels (when estimating a delta) be related to zero levels of ozone (which never occur) or to our best estimates of background ozone? Moreover, how do we take into account varying background levels in time and space as we model health effects? I also ask why our estimates of exposure and health outcomes don't include the current standard of 75 ppb.

Dr. David Chock

CHAPTER 4: AIR QUALITY CONSIDERATIONS

Charge Questions:

4. What are the views of the Panel on the appropriateness of the methods used to characterize ozone air quality for the exposure and risk assessment? What are the views of the Panel on the HDDM-based adjustment methodology used to adjust ozone concentrations to just meet the existing ozone standard and alternative standards?

5. To what extent does the Panel find that the discussion of uncertainty related to the air quality imputs to the exposure and risk assessment appropriately covers important sources of uncertainty?

This Chapter is a substantial improvement over the first draft version. Replacing the Quadratic Rollback method with the HDDM procedure for estimating the ozone concentration distributions under "just meeting" the standard and alternative standards is an important improvement from the scientific perspective. Using the HDDM approach for adjusting the ozone concentrations to just meeting the existing standards or the alternative standards actually obviates the concern about how the background ozone values should be defined. Another major improvement is the demonstration of the statistical superiority of the Downscaler approach to fuse the modelled and observed ozone concentrations over a large region when both sets of information are available (like in 2007). But the Downscaler approach may not necessarily be readily applicable to the exposure analysis. The Appendices are also very well prepared with a wealth of supporting information.

The key assumptions in this Chapter that allow the incorporation of the HDDM approach in actual applications is the linear relation between the first-order sensitivity coefficients and the ozone concentrations, and between the second-order sensitivity coefficients and the first-order sensitivity coefficients. These assumptions arguably allow the transfer of the statistical relations based on the modeling results for 2007 to the real world by using the observed ozone concentrations at the monitors to replace the modeled ozone concentrations for the grids containing the monitors at the same hours and in the same seasons but for the years different from but not too distant from 2007. These relations of model-run results must be considered strictly empirical. Actually, there is no theoretical basis to justify this regression linking the sensitivities to ozone concentrations. The notion of NOx-limited and VOC-limited conditions may help empirically but far from being reliable across all monitoring sites and all hours of the day in an urban area. The only basis the current approach relies on is the hope that the regression patterns remain relatively unchanged for the corresponding sites and time of day for the period not too distant from the year the regressions are established. Yet this approach may well be the

only sensible way presently to apply the HDDM results to adjust the ozone concentration distributions by precursor emission reductions to meet the current and alternative ozone air quality standards for the different areas under consideration

It is actually a bit surprising that some of the linear fits between the first-order sensitivity coefficients and the ozone concentrations using the model output information for 2007 look pretty good, as illustrated in Appendix 4-D. The empirical relations between the first and second sensitivity coefficients are more tentative, as expected. But at least the generally anticipated negative slopes are observed as we expect the ozone concentrations to eventually decrease and level off as emission controls become more stringent.

For the 50% and 90% emission reductions, the authors also appears to appropriately readjust the sensitivity coefficients by incorporating the range of applicability of the coefficients derived from the base runs and the runs for 50% and 90% emission reductions (See p. 19 of Appendix 4-D.) so that the consistency of the linear relations can be extended to high levels of emission reductions.

There are two other issues that need mentioning, both may be relatively minor at present in comparison to the drastic linear-relation assumptions described above, but still worthy of mention. One is the implicit assumption of local emission reduction. The sensitivity coefficients are built on the assumption that the emission perturbation is described by a uniform, across-the-domain percent US anthropogenic emission reduction. But the resulting sensitivity coefficients are applied only locally for a given urban area without regard to the emission reduction levels of the upwind areas. The resulting errors may be somewhat alleviated if all or most upwind areas also require similar levels of emission reduction to achieve a given ozone air quality standard, or if the high ozone concentrations of an area are dominated primarily by the ozone chemistry of emissions from within the area. The other issue is the predominant use of NOx emission reductions to meet a given standard. Emission reductions by NOx alone is more straightforward but may not be a robust or preferred control strategy in practice in some areas and may lead to an upward shift of the lower portions of the ozone concentration distributions as seen in many metropolitan areas in Figs. 4-9 and 4-10. In these cases, the present or alternative ozone air quality standard may be met, but perhaps at the expense of an increased health hazard due to higher ozone exposures in the more densely populated districts where the VOC-limited conditions tend to prevail, at least initially.

The use of the Downscaler for fusing the observed data and modeling results yields the best performance compared to the eVNA approach. This is not surprising since, in the case of the Downscaler, a large number of parameters are introduced to enable optimization of the fit, whereas eVNA simply invokes a set of straightforward scaling factors for interpolation. Using the estimated parameters for the Downscaler approach to estimate the ozone monitor concentrations when only the modeled ozone concentrations are available may not be readily justified.

In Table 4-6, the authors provide an extensive list of sources of uncertainty in the generation of the air quality output for health risk assessments. The list is very well thought out and thorough.

And their assessments on the directions and magnitudes of the uncertainty impacts of these sources are reasonable. The knowledge-based comments and comments on the influence of uncertainty from these sources on risk estimates are cogent and reasonable. There is one exception: the size of the knowledge-based uncertainty in item F (Applying modeled sensitivities to un-modeled time periods) is actually uncertain in itself. A designation of "low-medium" may be a bit optimistic. It may be more appropriate to designate it as "to be determined" or "potentially low-medium".

In conclusion, I find this chapter to be very well prepared and the authors need to be congratulated for their effort.

Minor Comments:
In the Methods section (4.3.3.1), the text ought to mention the modeling domains and the horizontal grid size used in the model.

For better clarity, the pink and turquoise colors in Figs. 13-27 need to be explained in the text or figure captions in Appendix 4-D.

In Appendix 4-D, it would be useful to mention in the figure caption of Figure 2 that the "75%" mark is used in the illustration instead of the "90%" used in the text.

CHAPTER 5: CHARACTERIZATION OF HUMAN EXPOSURE TO OZONE

Charge Questions:

6. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated and expanded population-based exposure analysis to be technically wound, appropriately balanced, and clearly communicated?

7. Chapter 5 includes several evaluations of key APEX inputs and model outputs, including for example analysis of time-activity data and comparison of actual personal exposures with modeled exposures. What are the views of the Panel on the appropriateness and usefulness of these evaluations and the conclusion drawn from these evaluations?

8. Chapter 5 includes several scenario-based exposure simulations that focus on specific populations or behaviors. What are the views of the Panel on the design, results and interpretation of these additional scenario-based exposure simulations?

9. To what extent does the Panel find that the discussion of uncertainty and variability have covered important sources of uncertainty and variability and appropriately characterized their relationship to the exposure estimates?

This chapter, together with its accompanying appendices, is a result of the EPA's best-effort projection of population exposure under existing and alternative ozone air quality standards. The assessment, interpretation, and presentation of the methods and results are excellent. New material and methodologies have been incorporated, together with a large amount of cross-checkings, verifications, sensitivity tests that significantly enhance the credibility of the conclusions. Noteworthy improvements include the use of the HDDM methodology in place of the quadratic rollback approach, incorporation of new activity data base in CHAD, estimated exposure impact of averting behavior for all school-aged children and for asthmatic school-age children, among others. The attached appendices are quite detailed and comprehensive. Appendix 5F contains the results of a large number of exposure simulations for different population subgroups under different ozone standards. The presentations of the results are clear and concise. Overall, this exposure analysis is fair and credible, and is an impressive achievement. It most definitely is an enormous improvement over the first draft.

Figure 5-1 is an excellent summary of the conceptual framework used in the exposure study.

Section 5.2.3 indicates that the census tract ambient hourly ozone concentrations come from the monitors coupled with the VNA estimation when extrapolations beyond the range of 30 km from a given monitor becomes necessary. However, it should also describe how the air quality modeling results are used to replace the monitored concentrations when cases of achieving the existing or alternative air quality standards are to be modeled. Is the spatial resolution of the monitors retained or replaced by that of the modeling grid? For clarity, a very brief description is in order on how eVNA parameters are used in projecting the concentration distributions for cases that just meet the existing or alternative standards.

The presentation of the exposure assessment results (Section 5.3, pages 5-23 to 5-34) is excellent. The figures are clear, informative and concise. It may be helpful to explain why there are two disconnected lines overlapping 2008 in Figures 5-5 through 5-9. This is a result of two different sets of design value, one for 2006-2008 and one for 2008-2010. The set for 2008-2009

was not used in the analysis. The different levels of design value trigger different levels of emission control, which leads to different levels of exposure.

The chapter provides some discussions on the analysis of time- location-activity data. It is reassuring that the majority of diaries from CHAD comes from surveys conducted in the past decade and that the pre-1990s diaries represent less than 15% of the total diaries available in CHAD. Even in the case where there are some noted differences in activities, like the children's outdoor participation rates of 50% in the 1980's and 35-40% in later decades, the inclusion or exclusion of the 1980's data causes a change in outdoor participation rate of only 3%. (p. 5-37, lies 25 to 36). The authors also indicate an average of 30-minute increase in time spent outdoors in the 1980's data compared to the 2000's data. But when an afternoon outdoor time of at least 2 hours is considered, inclusion of these old data have little adverse influence on the exposure outcomes due to the large variability in the data. Accordingly, it looks reasonable that inclusion of these old diary data would not strongly influence the exposure estimates.

The EPA compared the CHAD diary results with literature on the time spent outdoors and the levels of exertion while outdoors for both asthmatics and non-asthmatics and found little difference between the two groups. This is separately true for school-age children and for adults. However, there seems to be some inconsistency in the literature regarding the *range* in percent of outdoor time spent on strenuous activities by asthmatics. (See p. 5-40, lines 3 to 14.) It would be helpful if this difference could be resolved soon. But this percent *range* difference in the literature is not necessarily grounds to negate the EPA's conclusions, which are based on the analysis of newer available data (See p. 5-39).

It is interesting that there seems to be persistent relative magnitude differences among exposure estimates derived from using ozone concentrations based on the original air monitor values, VNA values and eVNA values. In general, the design of an interpolation scheme ought not to introduce a systematic bias, unless there are some persistent spatial patterns that distinctly favor some biases, like having the highest concentrations constantly occurring in the regions with the highest population density. But these spatial patterns may not be true or persistent in many urban areas. In the case of eVNA, because the modeled concentrations are also involved, some biases may indeed exist, and the difference may in fact reveals the disadvantage of using the monitor values alone to characterize the spatial concentration distributions over the whole urban area.

The explanation for the differences in exposure between the Quadratic Rollback and HDDM is excellent. Indeed, Quadratic Rollback adjusted the high-quantile portion of the concentration distribution with some fixed rules which are not based on science.

Table 5-6 characterizes the major uncertainties encountered and their potential impacts on the APEX exposure assessments. The list is comprehensive and the estimated impacts look

reasonable to the extent of our current understanding. Evidence to challenge the impact assertion is not available. The table also indicates which uncertainties are newly evaluated. This is helpful.

Appendix 5-D contains and extensive list of types or components of exposure variability like simulated individuals, microenvironments, and physiological characterizations, and how APEX incorporates them. Components that may co-vary with the input are also incorporated as necessary. The APEX incorporates the impacts of variability and covariability reasonably and quite thoroughly. One item that is worth mentioning that is also outside the control of APEX is the variability of input meteorological conditions and the resulting ozone concentrations beyond those described in Table 5D-1. Presently, the APEX results on exposure outcome are based on the explicit input of meteorological variables and adjusted ozone concentrations for the period of 2006 – 2010, which may or may not be representative of future scenarios for ozone.

CHAPTER 9: SYNTHESIS

Charge Questions:

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 assessment?

20. To what extent does the Panel find that the discussion of overall uncertainty provides an appropriate context for interpretation of the exposure and risk results?

This chapter is an excellent summary and analysis of findings presented in this Assessment report. The description is comprehensive, accurate and well thought out, even though it is a bit on the lengthy side.

In describing the improvement of the HDDM air quality adjustment methodology over quadratic rollback on p. 9-4, it would be useful to add as part of the first point or as a separate point that the quadratic rollback requires an assumed background ozone concentration which serves as the "floor" for each area. How to set this "floor" can be controversial. The HDDM approach, on the other hand, removes this problem since the background emissions are explicitly included in the model simulations and the predicted ozone concentration changes are directly linked to the US anthropogenic emission reductions alone.

On p. 9-12, line 1, Table 5-6 should be changed to Table 5-7. And on line 25, Figure 9-2 should be changed to Figure 9-3. Also, it would be useful to provide additional clarifications in the figure caption or the footnote of Figure 9-3 that the highest value across the years 2006-2010 in the percent of children with at least one ozone exposure that exceeds 60 ppb for each of the cases of just meeting given air quality standards actually comes from the maximum of each pair of red curves shown in Figure 9-2. A similar linkage is also indicated for Figure 9-5 and Figure 9-4.

The authors highlight two sources of uncertainties in modeling ozone responses to meeting different levels of the standard. (See p. 9-39, lines 15-34) One is the applicability of HDDM sensitivities over large emission perturbations. The other is the variability in data used to create regressions linking the sensitivity coefficients to observed concentrations. The first source of uncertainty can be controlled and is relatively small. But the authors indicate that the second uncertainty is also small. The authors stated that the uncertainty introduced from the *application* of regressions to determine sensitivities were quantified by propagating uncertainties in the sensitivities through to uncertainties in the final predicted ozone concentrations which had standard errors less than 1.4 ppb for all adjustment scenarios (lines 26 to 29). This statement may well be true. But the problem is that the regression steps themselves have high uncertainty because there is no theoretical underpinning to justify the general applicability of these relations. One cannot tell from the level of the ozone concentration alone whether the slope of the first-order sensitivity coefficient relative to the concentration will be positive or negative, let alone the magnitude of the slope. The notion of NOx-limited and VOC-limited conditions may help empirically but far from being reliable across all monitoring sites and all hours of the day in an urban area. It would be helpful if the authors acknowledge that the generality of the regression approach needs further scrutiny.

The authors describe the downward bias in epidemiological risk changes for an urban area associated with meeting different standards due to the tendency that epidemiological models tend to be constructed by using only data concentrated in the urban cores of an area (Section 9.5.3). However, changing the urban area to CBSA without re-adjusting the data base for the model does not rectify this bias. (In fact, it adds more uncertainty to the bias, as pointed out by the authors also.) Neither does normalizing the population size of the risk estimates. The reduced sensitivity of the *existing* epidemiological risk estimates for characterizing the risks in meeting different ozone air quality standards needs to be more explicitly indicated.

Section 9.6 is an excellent conclusion for the overall integrated characterization of health risk associated with different ozone standards based on the enormous tasks undertaken by EPA to date.

Dr. Ana Diez Roux

Chapter 5

6. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated and expanded population-based exposure analysis to be technically sound, appropriately balanced, and clearly communicated?

Overall I found the methods to be clearly presented. Figure 5-1 was especially useful in summarizing the various inputs to the modeling process. The document does a very good job of describing the various sources of data that went into the modeling process and how the data were used.

I also found the description of the model output useful, although some relatively minor editing would improve clarity. For example, the title of Figure 5-2 says "Percent of asthmatic school-age children in all study areas with at least one O3 exposure at of above 60ppb-8 hour while at moderate or greater exertion…" . It is not clear what "one O3 exposure" means in this context. Does it mean that they were engaging in moderate of greater exertion at any time during an 8 hour period with an average of >=60ppb? In order to count as "one exposure" is there a minimum time requirement (for example, must they be engaging in moderate or greater exertion during a least one hour at any time during the 8 hour averaging period?) Perhaps I am misinterpreting the output measure reported here, if so this needs to be clarified.

Figure 5-3 was very helpful as a way to present the results but a bit more clarity in the labeling would help readers better interpret the graphics. For example the labeling of the bottom panel could be "Percent of asthmatic school age children with at least one exposure [see my note above regarding clarifying this metric] at or above 60ppb, 70 ppb and 80 ppb (red, green and blue lines) when air quality was adjusted to just meet standards of 75, 70. 65 and 60 ppb (panels left to right).

Section 5.3.3 provides a very good description of the results. A summary at the end of the section highlighting the key points (especially those that will be of relevance to the PA) would be very helpful.

7. Chapter 5 includes several evaluations of key APEX inputs and model outputs, including for example analysis of time-activity data and comparison of actual personal exposures with modeled exposures. What are the views of the Panel on the appropriateness and usefulness of these evaluations and the conclusions drawn from these evaluations?

I found the evaluations presented in section 5.4.1 useful and well described and the conclusions reasonable. The document has been greatly strengthened through the incorporation of this section. Section 5.4.4 was also useful although the lack of agreement wit the Detroit data in section 5.4.4.1 needs to be explained or at least further discussed with respect to the implications of this for the exposure estimates previously presented.

8. Chapter 5 includes several scenario-based exposure simulations that focus on specific populations or behaviors. What are the views of the Panel on the design, results, and interpretation of these additional scenario-based exposure simulations?

Sections 5.4.3 included very useful information. It would benefit from a concise summary at the end highlighting the key conclusions and their implications from the exposure estimates previously presented.

9. To what extent does the Panel find that the discussion of uncertainty and variability have covered important sources of uncertainty and variability and appropriately characterized their relationship to the exposure estimates?

All important sources of variability and uncertainty are addressed in the text or extensive tables. The document does a very good job of discussion all potential sources of uncertainty and evaluating the extent to which they can be addressed.

Chapter 7: Characterization of Health Risk Based on Epidemiological Studies

13. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated epidemiology-based risk assessment to be technically sound, appropriately balanced, and clearly communicated?

Overall I found the presentation of the methods clear and well justified. The criteria used to select the epidemiologic studies and metrics used in the risk assessment are well described. The limitations of the approach are also adequately noted.

The chapter generally does a good job of describing the results and sensitivity analyses. In general the presentation of results is markedly improved over the prior version. However some additional editing of the language would further improve clarity. The chapter repeatedly refers to “incidence” or “mortality” when what it is referring to (if my interpretation is correct) are actual counts of deaths (epidemiologically incidence and mortality are by definition a proportion or a rate, not a count). In contrast Figure 7-4 dies present true mortality estimates (incidence of death). This language needs to be corrected throughout so that counts of deaths are not referred to as incidence. The titles for figures 7-2 and 7-3 are identical. What is the difference?

The section on short term attributable mortality (pg 7-69) indicates that “the mortality risk metric is generally not responsive to meeting the existing and alternative standard levels”. It is argued that this occurs because of simulated increases on O3 on some days and regions, even when the standard being met is lower. It is noted that this contrasts with clinical study-based risk estimates. Later in the same section it is noted that “the magnitude of the risk reduction increases as lower alternative standards are simulated”. This seems to contradict the previously quoted statement in the same section. These issues need to be clarified. The extent to which the simulated increases of O3 at lower standards is realistic and to be expected in the real world needs to be discussed. How important is this possible increase in considering the health impact of alternative standards?

14. To what extent does the Panel find that the 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 covers the main sources of variability and uncertainty and addresses them appropriately to the extent possible with available data.

The section also appropriately describes sensitivity analyses that have performed to at least partly assess the plausible impact of some of these uncertainties.

15. Adjusting the distributions of O_3 concentrations based on decreasing NOx emissions to just meet the existing and alternative O_3 primary standards resulted, in some cases, in substantial shifts in the spatial and temporal patterns of O_3 across case study urban areas relative to patterns of O_3 that existed for recent air quality, and presumably relative to the patterns present in the study locations of the epidemiology studies from which the concentration response functions were drawn (see section 7.1.1 of the TSD, USEPA, 2012). What are the views of the Panel on the characterization of the degree to which these changes in spatial patterns of O_3 introduce uncertainty in risk estimates when effect estimates based on one spatial/temporal pattern of O_3 (the pattern in the epidemiology study) are applied to a substantially different spatial/temporal pattern of O_3 concentrations?

The spatial pattern may impact the effects estimate if factors correlated with space modify the health impact of O3 (i.e. if there is important heterogeneity over space in the effect estimates within urban areas). In the absence of a clear rationale for this type of effect modification I would argue that the impact of the changing spatial patterns can be ignored. If we believe the effect estimates are capturing the underlying causal effect, then this effect should be approximately generalizable over space.

16. In particular, what are the views on the Panel on the characterization of the level of uncertainty associated with estimates of risk associated with days with relatively lower composite (area-wide average) O_3 concentrations and those with relatively higher composite O_3 concentrations?

I did not see this specifically addressed in the chapter. Perhaps I missed it?

Dr. Daniel Jacob

Chapter 4: Air Quality Considerations

4. What are the views of the Panel on the appropriateness of the methods used to characterize O_3 air quality for the exposure and risk assessment? What are the views of the Panel on the HDDM-based adjustment methodology used to adjust O_3 concentrations to just meet the existing O_3 standard and alternative standards?

I view the methods as appropriate and a major improvement over the quadratic rollback method. A few points seem worth clarifying:

4.1 Pages 4-15,16. It seems that a very important aspect of improving the accuracy of the HDDM analysis for large emissions perturbations is the calculation of sensitivities for three emission levels, thus allowing better representation of the non-linearity. However, I had to wait until the uncertainty analysis in section 4-5 to learn that this calculation at three emission levels was done, and even there it did not tell me what these emission levels were. I recommend that this information be brought up here in the initial description.

4.2. Page 4-16, lines 4-7: how successful are these linear regressions at capturing the variability of the response? The rationale behind a linear fit is not clear.

4.3. Page 4-17: it would be worth clarifying that although nationwide emission decreases were imposed, different levels of emission decreases were used for the different urban areas. I presume that's what was done.

4.4. Page 4-18, paragraph starting on line 28: I did not understand that paragraph at all.

4.5. Figures 4-9 and 4-10: what percentiles correspond to the boxes and whiskers? It would be good to show the design value in those figures. I don't understand why many of the distributions fall far below the design value, even though the emission reductions targeted just meeting the design value. For example, Atlanta in Figure 4-10 seems to show a maximum concentration of only ~50 ppb for a NAAQS of 60 ppb. I'm obviously missing something important here.

4.6 Page 4-31: I don't understand why the national mapping was done only for mean ozone statistics and not for more extreme ozone statistics, in particular the design value. These mean statistics are not well correlated with the design value (Page 4-36, Figure 4-18). I don't get the point of this national mapping.

5. To what extent does the Panel find that the discussion of uncertainty related to the air quality inputs to the exposure and risk assessment appropriately covers important sources of

uncertainty?

5.1 The discussion covers different factors of uncertainty but is lacking in synthesis. The general point of an uncertainty analysis is to quantify the important sources of error and to determine how the errors are expected to add (in quadrature if uncorrelated) to arrive at an overall uncertainty estimate. Without that overall estimate it is not clear how this uncertainty analysis can be propagated to the REA.

5.2 One missing factor of uncertainty that needs some discussion is the ability to quantify the sensitivity of ozone to emission reductions through CMAQ. The standard evaluation of CMAQ with observed ozone concentrations may not help in characterizing that error. There has been some recent literature on comparison of simulated and observed ozone responses to SIP emission reductions (e.g., Dan Cohan's work at Rice) that would be useful to cite.

5.3 The calculation of HDDM sensitivities at three different NOx emission levels is obviously important to better capture the non-linearity in the dependence of ozone on emissions. It must be critical to explain the low errors in applying HDDM to 50% and 90% NOx emission reductions. The text should tell us the three emission levels at which HDDM calculations were done.

Dr. Steve Kleeberger

Chapter 6:

10. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated and expanded lung function risk analysis to be technically sound, appropriately balanced, and clearly communicated?

The methods and results of the expanded lung function risk analyses are sound, balanced, and clearly communicated.

11. What are the views of the Panel on the implementation of the McDonnell-Stewart-Smith model to specify the exposure-response function linking the change in FEV1 to O3 exposure?

The implementation of the McDonnell-Stewart-Smith model is appropriate for exposure/response and change in lung function (FEV1).

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

In my estimation, the discussion of the uncertainty and variability was adequate. Although staff mentions potential causes of variability in response to ozone exposure (p 6-8), it is not clear to me that intrinsic variation (e.g. genetic factors that contribute to wide inter-individual variation) was included in the models. If not, then some discussion is warranted to explain why this is so (e.g. limited reproducibility).

Dr. Fred Miller

Chapter 1: Introduction

1. To what extent 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?

Response: The material on the current approach and the organization of the HREA is clearly communicated and is of the appropriate length.

Chapter 2: Conceptual Model

2. To what extent does the Panel find that the discussions accurately and clearly reflect the air quality, health effects, exposure and risk considerations relevant for quantitative exposure and risk assessment, building from information contained in the final ISA? What are the views of the Panel on the additional flowchart provided for the overall assessment and the additional information regarding specific elements of the exposure and risk assessments?

Response: The flow chart provided in Figure 2-1 is a useful addition that enables the HREAder to see how the different elements of this complicated assessment fit together and how they are covered in the various chapters that comprise the HREA. The reference to Fig. 2-1 at the top of page 2-11 is incorrect as Fig. 2-2 is the relevant figure.

The authors do an excellent job of accurately and clearly discussing the key elements of air quality, health effects, etc that collectively form the risk characterization. They bring key findings in the ISA forward into this chapter to support what their formation of the conceptual model used in the HREA.

Chapter 3: Scope

3. To what extent does the Panel find the scope of the health risk and exposure assessment is clearly communicated? To what extent does the panel find the additional flowcharts for each analytical component to be useful additions?

Response: This chapter is very well written and clearly communicates the scope of the HREA. Elements of the HREA are discussed in a logical order, and the additional flowcharts help the reader understand the key components of each phase (i. e., the characterization of air quality, the exposure assessment, the controlled human exposures, etc.).

Chapter 4: Air Quality Considerations

4. What are the views of the Panel on the appropriateness of the methods used to characterize O3 air quality for the exposure and risk assessment? What are the views of the Panel on the HDDM-based adjustment methodology used to adjust O3 concentrations to just meet the existing O3 standard and alternative standards?

5. To what extent does the Panel find that the discussion of uncertainty related to the air quality inputs to the exposure and risk assessment appropriately covers important sources of uncertainty?

Response: No pre-meeting comments

Chapter 5: Characterization of Human Exposure to Ozone

6. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated and expanded population-based exposure analysis to be technically sound, appropriately balanced, and clearly communicated?

7. Chapter 5 includes several evaluations of key APEX inputs and model outputs, including for example analysis of time-activity data and comparison of actual personal exposures with modeled exposures. What are the views of the Panel on the appropriateness and usefulness of these evaluations and the conclusions drawn from these evaluations?

8. Chapter 5 includes several scenario-based exposure simulations that focus on specific

populations or behaviors. What are the views of the Panel on the design, results, and

interpretation of these additional scenario-based exposure simulations?

9. To what extent does the Panel find that the discussion of uncertainty and variability have covered important sources of uncertainty and variability and appropriately characterized their relationship to the exposure estimates?

Response: This chapter is the first of many successive chapters where the figures are so small that they are of limited value to the reader as one can only deduce overall trends in most of the panels comprising the figures. In addition, the color schemes for 70 and 80 ppb of O_3 are essentially indistinguishable in the figures.

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

10. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated and expanded lung function risk analysis to be technically sound, appropriately balanced, and clearly communicated?

11. What are the views of the Panel on the implementation of the McDonnell-Stewart-Smith model to specify the exposure-response function linking the change in FEV1 to O3 exposure?

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

Response: The updated and expanded lung function risk analysis is technically sound and represents a significant improvement in the approach to this component of the overall O3 risk characterization. The authors clearly describe the main differences between the MSS model for individual responses versus the population model used in this and past assessments. The MSS model is scientifically and biologically defensible, particularly the use of the threshold version of the model even though major differences in risk do not result between the threshold and non-threshold model. The implementation of the MSS model in the HREA is clearly described, and the comparison of the MSS model results to those obtained with the exposure-response model is of tremendous importance. One of the most important statements in the HREA is found at the bottom of page 6-29 where it is stated

> "In most cases, the MSS model gives results about a factor of three higher than the exposure-response function model for school a-aged children. This is expected since, as

> discussed above, the MSS model includes responses for a wider range of exposure protocols (under different levels of exertion, lengths of exposures, and patterns of exposure concentrations) than the exposure-response model of previous reviews".

As noted earlier, the panels comprising the figures are too small and lessen the quality of the chapter. In Fig. 5-10, the colors used for 60 and 70 ppb of O3 are too close to each other and would cause confusion to a reader until they figured out the "stacking" of ppb bars in the different rows.

The description of the additional time activity pattern data recently acquired addresses a concern raised previously by CASAC concerning how activity patterns should be brought up to date.

Discussion of major constituents of uncertainty and variability was well done by the authors. Table 6-16 provides a good summary of the qualitative uncertainties, their likely direction and magnitude, and the extent of the knowledge base underpinning current understanding of the phenomenon being addressed.

Chapter 7: Characterization of Health Risk Based on Epidemiological Studies

13. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated epidemiology-based risk assessment to be technically sound, appropriately balanced, and clearly communicated?

14. To what extent does the Panel find that the 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?

15. Adjusting the distributions of O3 concentrations based on decreasing NOx emissions to just meet the existing and alternative O3 primary standards resulted, in some cases, in substantial shifts in the spatial and temporal patterns of O3 across case study urban areas relative to patterns of O3 that existed for recent air quality, and presumably relative to the patterns present in the study locations of the epidemiology studies from which the concentration response functions were drawn (see section 7.1.1 of the TSD, USEPA, 2012). What are the views of the Panel on the characterization of the degree to which these changes in spatial patterns of O3 introduce uncertainty in risk estimates when effect estimates based on one spatial/temporal pattern of O3

(the pattern in the epidemiology study) are applied to a substantially different spatial/temporal pattern of O3 concentrations?

16. In particular, what are the views on the Panel on the characterization of the level of

uncertainty associated with estimates of risk associated with days with relatively lower

composite (area-wide average) O3 concentrations and those with relatively higher composite O3 concentrations?

Response: The material in Section 7.3 on the selection of model inputs and assumptions is well done and useful in helping the reader understand the overall issues involved with the use of the epidemiological studies in the HREA. While the discussion of uncertainty and variability in Section 7.4 covers the important sources and their relationships to risk estimates, the section could be shortened significantly by a more limited discussion in the text given the same material is essentially provided in Table7-4 in a more succinct manner.

Using the decreasing of NOx emissions as the driver for just meeting the current standard or alternative standards is reasonable and the only really viable approach to lowering O3 levels. While this results in the shifting of spatial and temporal patterns across case study urban areas, the overall effect should cancel out relative to a comparison of the area covered in the epidemiology studies as one would have to invoke that the area not included in these studies are not representative of the broader geographical and socioeconomic area that were included in the epidemiology studies. The uncertainties introduced in risk estimates when effect estimates are based on one spatial/temporal pattern of O3 and are applied to a substantially different spatial/temporal pattern of O3 concentrations are not likely to be any greater than the uncertainties introduced by other factors that are discussed in Chapter 7. Moreover, the central tendency of statistical theory should work to prevent the uncertainties in risk estimates from going only in one direction.

Chapter 8: National Scale Mortality Risk Burden Based on Application of Results from Epidemiological Studies

17. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated national-scale risk analysis to be technically sound, appropriately balanced, and clearly communicated?

18. To what extent does the Panel find the risk and air quality representativeness analyses to be technically sound and clearly communicated?

Response: This chapter is a straightforward and well-written one that covers all of the important aspects of the interpretation and presentation of the methods used. The approach is technically sound. The authors clearly communicated the representativeness of the urban study areas in a national context by examining the major determinants of O3 effect estimates, namely demographics, base-line health conditions, exposure determinants, and climate and air quality. Table 8-6 provides an excellent synopsis of the major findings concerning subcategories of risk attributes and the differences between the urban study areas and the U.S. dataset.

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 assessment?

20. To what extent does the Panel find that the discussion of overall uncertainty provides an appropriate context for interpretation of the exposure and risk results?

Response: The Synthesis chapter is about the right length for condensing the salient points and issues that were dealt with in the HREA. The figures provide a good summary of the major findings for the 4 primary health endpoints that were assessed in a manner that shows the distributional changes among the 15 cities that were assessed.

In Section 9.5, on the overall assessment of confidence in the exposure and risk results, the authors seem to "back off" on the usefulness of the short-term risk based modeling results for the larger study areas using the multi-city times series based estimates compared to what is presented and discussed in Chapter 7. This "backing off" is particularly present in the statement on page 9-42 where the authors write

> "Overall, these sources of uncertainty cause us to have reduced confidence in estimates of short-term risk based on modeling the larger (CBSA-based) study areas using the multi-city time series-based effect estimates. This reduces the utility of the risk assessment in directly informing the decision regarding the level of the standard since we have lower confidence in estimates of absolute risk associated with a given standards level. However, the risk assessment can still be useful in providing estimates of the general magnitude and direction of changes in risk associated with an alternative standard level."

Executive Summary

21. To what extent does the Panel find the Executive Summary to be a useful summary of the data and methods used to estimate human exposures and heath risks and the key results of the assessment?

Response: Overall, the Executive Summary is well written and clearly brings out the most salient points and findings of the HREA. The balance between sections is good. The legend to the figure on E-5 needs to be changed to “top” and “bottom” instead of “left” and “right” if the final version keeps the same publication layout as is currently used. In the section of health based risks for controlled human studies, the authors should eliminate the use of “potentially” when describing the implications of a 20 percent decrease in FEV_1 for persons with existing lung disease – can cause more serious effects needs to be the thrust of this point. On page E-7, the explanation of how to interpret the stacked bars and their colors is excellent and reflects the kind of wording that should be used more frequently in the document.

Dr. Ted Russell

Review of Ozone REA-Health 2nd Draft.

This REA is a marked improvement over the prior Draft, and over the REAs from years past. It shows a very positive evolution in the approach and the presentation. There can still be some improvements, but it has come a long way from the first one I read during the last review. It also represents a tremendous amount of work. The work that has gone in to the air quality characterization and the development of ozone changes in response to emissions controls as described in Appendix 4 is a remarkable amount of effort. While I might have done a few things a bit differently, and likewise interpreted some of the results a bit differently, that could be from my ignorance and not having spent so much time trying to pull off what was done.

From an air quality characterization and analysis standpoint, my view is that the largest step forward in their analysis is their ability to capture ozone responses to emissions controls using an advanced air quality model-based approach, e.g., using CMAQ with HDDM. This has allowed EPA to capture not only the reductions in peak ozone, but also the increases in lower level ozone levels in response to controls. Both of these have been observed, and this gives much greater confidence to their ensuing analyses. One concern was that too much of the important concepts from the Appendix that supports their approach is left in Appendix 4. It is very necessary to read Appendix 4 to have a reasonable view what is being done, and what are the particular strengths and weaknesses (Appendix 4, discussed below, however, is still a bit rough and could use some work). In this REA, they have also added additional ozone response models.

A major general concern was that the figures/figure captions were often not complete and/or clear. It took me a while to see what was being presented, and I often had to go back and forth between the figure and the text to figure out what was being shown. Make each figure/table almost stand alone, i.e., it could be a single slide in a presentation and need little explanation. The more comple figures, in particular, need a more informative caption.

Chapter 2:

Overall, the chapter adequately conveys the first parts of the conceptual framework for conducting a risk and exposure assessment for ozone.

The end of section 2.2.1 should be modified to note that the NOx-limited conditions are found in the summer/high ozone levels. Much of the year, cities can be radical limited due to the lack of sunlight. You may want to characterize these areas as being where “high ozone levels are NOx-limited”.

Chapter 4:

1. Question 4. The use of HDDM-based adjustment is a major step forward. Reading this chapter and the supporting Appendix demonstrate a considerable amount of work, thought and analysis. It also provides results that are much more in line with observations. Kudos to the staff in pulling this off.

There are some concerns, however. First, many important aspects of the method are relegated to Appendix 4, which I can both justify (it gets very technical and is likely of limited interest to many), but also criticize (there are rather important outcomes of using the method and how specific approaches to using the sensitivities). It is appreciated that the staff was able to develop, and have reviewed, a manuscript describing much of their approach, but there are some differences.

One comment that should be made is that, at present, their approach shows some bias in that they primarily utilize a NOx-oriented control approach (e.g., they prefer using a NOx-only set of sensitivities, not the NOx-VOC results, and do not even provide VOC-only approaches). This should be further discussed and defended, e.g., potentially a few analyses showing that a VOC-only approach is largely ineffective in most locations and/or that a VOC-only approach buys little benefit over a NOx-VOC approach. However, I do support the use of a limited number of non-source specific sensitivities as there are a huge number of source-specific analyses that could be done, the choice of which is not apparent at this time. How close would a VOC-only strategy get New York to the 60 ppb level?

A comment between this chapter and Appendix 4 is that I probably would not have chosen the same approach to estimating sensitivities and ozone levels at intermediate control levels, i.e., when control levels are not 50% or 75%. The current approach appears a bit ad hoc, and shows a few major deviations (though limited, and they have an adjustment approach). I might have done the maximum simulations at, say, 85% controls (more towards the extreme end of the controls) and used a cubic spline fit to provide sensitivities at intermediate levels. The spline could provide each of the first and second order sensitivities. Something to think about next time.

The next question I had was exactly how the sensitivities are being applied at each location in the domain. Are they being applied to the CMAQ-simulated value, or (I think) to the VNA/DS-derived value? Figures 3-1 and 4-6 (and Figure 1 in Appendix 4-3) should be edited to make this clear. Fig. 4-6 should explicitely show how VNA or DS is being used. Both might show how/where VNA and DS are being used in the process. Having calculated the location/time specific ozone value, the next question is how to develop the appropriate sensitivity. Should one use the sensitivities calculated directly as described in Appendix 4 specific to the simulated ozone value, or should those sensitivities be adjusted for the difference between the simulated

and observed ozone levels. If the base simulated ozone is 80 ppb (and, thus the sensitivities are consistent with that simulated value), and the observed value is 100 ppb, should the sensitivities be adjusted upwards? One could give reasons both ways, and this should be discussed as well as support for their choice. I would probably adjust, but it is a tough call. (There are other approaches one might consider as well.) I suspect this would make rather little difference, but it should be discussed.

Section 4-5. Section 4-5 is comprehensive and I generally agree with their assessments of the levels of uncertainties, though with a few exceptions. The uncertainty in CMAQ modeling is probably "medium" based upon the model evaluation (which was very extensive). Likewise, the HDDM sensitivities are likewise about medium given that CMAQ results are about medium, and that they also have undergone less extensive review and you cannot directly evaluate the sensitivities using observations. Further, it is not apparent the best way to scale sensitivities when the simulated observation does not match the observations. On page 4-47, it is stated that "… in general we expect that the that the benefits of reducing high ozone concentrations and disbenefits of increasing low ozone would be underestimated." This should be further explained and supported.

Chapter 4 needs a Summary/Key Observations section consistent with the other chapters.

Chapter 6

I am still not wild about how Eq. 6-2 is shown. The assumption is that C and V do not change over the time period, so they are not really a function of t, which is what is shown. Showing that they are varying along with X is inappropriate. One could just as well use t_0 or t_i, indicating that the choice is for time period I, or show that they choose an average over the time period (use a bar over the term).

The uncertainty discussion in this chapter also needs a bit of work. It is noted that the uncertainties in the MSS model parameters are likely larger than 5%, but with little more discussion. Then, Fig. 6-12 uses 5%. Thus, the uncertainties shown are likely greater than shown. This should be noted in the caption. Also, I do not believe that 6-12 should be labeled as elasticities. Elasticity has a specific definition. Those are responses to a 5% increase (as the caption notes). I think an elasticity would be 20 times what is shown (if you still use %, but I might use a fractional elasticity).

Chapter 9

Question 19. Certainly the Synthesis is useful, and the document would suffer without it.

Question 20. The discussion of uncertainty does provide a good context for interpretation of the exposure and risk results. However, I was hoping that the uncertainty discussion would be deeper and more definitive. In particular, it would be useful if the section (9.5) concluded with more direct statements as to how a reader should interpret the overall uncertainties in the risk and exposure assessments for use in standard setting. They could also identify the specific uncertainties that are most key (e.g., contribute the most to their overall confidence in the results) and that should be targeted for further reduction.

With those two answers in mind, while the synthesis is valuable and insightful, it is not without additional problems.

Chapter 9 should deal more with synthesizing the results from the application of various responses over different seasons and different levels. While some analyses dealt with ozone during the warm seasons and only higher exposure levels, others were over the whole range. The discussions that are present are a good start. However, to state "The implications of this is that our estimates of mortality and morbidity risk reductions… are likely to understate…" should be qualified in that the seasonal application can add bias in the other direction.

Page 9-43: l 10-19. This paragraph says things are different and it is important to understand the differences, but does not provide how they are different and what that means. The next paragraph does similarly. It would be good if both of the paragraphs were more informative as to what differently really entails. Use of different metrics will lead to different results, but are they meaning differences? Do they conflict? Are they problematic when the results are interpreted for use in standard setting, e./g., raise concerns about uncertainties, or are they consistent and support the use of the metrics? In general, the synthesis could be more definitive.

9-23, l 2: It is not just one reaction that is of importance. The consumption of radicals also reduces ozone.

9-38, l3-5. Does this statement agree with the analysis found in the ISA? Please link to the ISA.

Appendix 4

Appendix 4 (particularly 4-D) represents a huge amount of work and a major step forward. It also needs a fair amount of work to harmonize the chapters. In particular, the figure numbering and equation numbering should be more specific as to the specific appendix. One might even think about an Introduction to the Appendices overall. Certainly, there needs to be an overall Table of Contents for the Chapter 4 appendices either up front and/or in the HREA Table of Contents.

The evaluation is extensive.

In Appendix 4-D, Section 3.2.3 could be a bit more clear in what is being done to modify observed concentrations. Sections 4-5 and 4-6 could use their own flow diagram specific to that component of the analysis, and with more detail.

I might recommend EPA having a more extensive discussion with the modeling community about how to use sensitivities in adjusting ozone values to meet various air quality metrics. There was not time this time.

Dr. James Ulman

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

- To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated and expanded lung function risk analysis to be technically

sound, appropriately balanced, and clearly communicated?

The staff is to be commended on the both the breadth and depth of the technical analyses, and the effective and concise manner in which they are presented.

- What are the views of the Panel on the implementation of the McDonnell-Stewart-Smith model to specify the exposure-response function linking the change in FEV1 to O3 exposure?

The basis of the MSS model is clearly communicated as is its application to the exposure-response risk estimation. As pointed out in the document, several largely unsubstantiated assumptions had to be made (e.g., extension of young adult age coefficient to children, identical response models for asthmatic and non-asthmatic children). Nevertheless, it is clear that the incorporation of time-dependent inhaled dose and detoxification dynamics as well as inter- and intra-subject variability in the MSS model is, in concept, a substantial improvement over the mean population responses at fixed exertion used in the previous E-R model.

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

The qualitative summary in table 6-16 is effective in defining the key uncertainties as well as indicating their effect on the risk assessment. In addition, a quantitative sensitivity analysis (Fig. 6-12) precisely shows how some (but not all) of the parameters produces uncertainty in that portion of the risk estimate originating from the MSS model. This analysis includes the

uncertainty with respect to intra-subject variation, but does not consider the uncertainty in inter-subject variation. I think that it is important to include a sensitivity analysis with regard to inter-subject variation that might be larger in at-risk populations such as asthmatics than in the general population.

In making the transition from risk assessment to policy, I am most concerned about the lack of sufficient information to include the effect of asthma on the MSS model. As I understand it, the personal exposure pattern for asthmatic and non-asthmatic children are similar, so their quantitatively evaluated risks are indistinguishable. As made clear in the PA document, however, children with asthma may have a greater intrinsic sensitivity to O_3 (i.e., a steeper exposure-response curve) than predicted for children in general.

Of somewhat less but still major concern is the substantial uncertainty due to intra-subject variability. To produce realistic results, it was necessary to modify the continuous distribution that quantifies intra-subject variability in the published MSS model. This was accomplished in the risk predictions by truncating the distribution beyond two standard deviations of its zero mean. As presented in the sensitivity analysis on page 6-41, an increase in the truncation of the intrasubject distribution from ±8% (two standard deviations) to ±20% (stated as the "actual" value on line 8) increases the proportion of children with ΔFEV>10% from 31% to 92%. Although it is not clear to me what is meant by "actual" value, it is obvious that the choice of this has an important effect on the computation of exposure-response distributions.

Dr. Sverre Vedal

Ozone: Health Risk and Exposure Assessment

Ch. 7. Characterization of Health Risk Based on Epi Studies

13. To what extent does the Panel find the assessment, interpretation, and presentation of the methods and results of the updated epidemiology-based risk assessment to be technically sound, appropriately balanced, and clearly communicated?

Random points:

- It's good that risk is based on total risk, not just to the lowest measured level.
- Core-based statistical area (CBSA) used rather than central urban is justified (7-6), although risk estimates are in some instances very sensitive to this choice.
- Substitution of Bell et al. (2004) with Smith et al. (2009) seems OK.
- Exposure based on peak exposure metrics is justified.
- Jerrett et al. (2009) as the only basis for estimating long-term mortality risks is risky, but it's the only game in town. This fact should temper confidence in C-R function.
- It's not surprising that differences in effect estimates drives the cross-city differences in risk reductions (7-74), but it's reassuring to see that.

Points for discussion:

- The estimate of up to approximately 20% of COPD deaths attributed to ozone (7-68) just seems implausible, especially when one considers that the population at risk for dying of COPD is composed of those who are unlikely to exercise and to be outdoors. I know that's what the effect estimate says, but ….
- Use of effect estimates pertaining to the larger populations - in some cases this results in using very a different effect estimate, e.g., NY, 0.0009 vs. for NJ, 0.0001 and 0.0005 (7-28, Table 7-3). Sensitivity to this choice is shown in 7.4.2 and 7.5.3.
- The discussion of variability and uncertainty is general sound and comprehensive. One aspect that is not touched on in discussion of spatial variability in concentrations is the fine-scale spatial variability due to roadway gradients. Near roadway ozone concentrations are considerably lower than city-average values. Also, there are typically no roadside ozone monitors, so concentrations there cannot be captured by most regulatory monitoring networks. Depending on the city, a greater or lesser fraction of the population lives in close proximity to large roadways, and it is not certain, for short-term exposures, that the average city day-to-day concentration pattern is reflected in those living near roadways. This has

implications for population exposure misclassification and isn't reflected in Table 7-4 (7-43). For long-term exposures, the importance of roadside gradients is obvious.

- In Table 7-4 on uncertainty analysis, it isn't clear why simulating ozone concentrations for "attainment of both existing and alternative standards" should be included here among other factors assessed in sensitivity analyses. These are simply different ways of expressing impacts of different regulations that provide different insights.
- I don't understand the conclusion that the mortality metric for short-term exposure is "not responsive to meeting the existing and alternative standard levels" (7-69). Mortality reductions seem to steadily increase with changes to the standard.
- Does not the predicted increase in risk for some study areas (when concentrations are low) when meeting standards (7-70) call into some question the mechanics of the air quality model simulations on how air quality standards are met?
- Observation: A substantial fraction of short-term attributed risk remains after meeting the most stringent alternative standard (7-71).
- As noted, use of regional effect estimates for long-term exposure risk has dramatic impacts on risk (7-79 and Table 7-14), ranging from 0 to 40% of baseline risk, and 27% in Denver – the latter, as others, seems to stretch plausibility – see first bullet in this section.
- Regarding Overall Confidence, in light of the reliance on one study to estimate long-term respiratory mortality effects, and the seemingly large effect estimate, I would have been reluctant to conclude that I had a "reasonable degree of confidence" in these risk estimates (7-86). It also seems inconsistent with the ISA conclusion (ISA 7-31) that there is "limited evidence" for an association between long-term exposure and respiratory mortality, presumably because it is based on only one study.
- In light of the central importance of respiratory (presumably COPD) mortality as an outcome of long-term ozone exposure, consideration should be given to estimating exposures in this group with APEX (if diary profiles are available), not just in asthmatics, the young and the old. Presumably this population might be expected to spend a relatively smaller proportion of time in more exposed settings.

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

- See above for comments on small-scale ozone spatial variability.
- Otherwise, very good.

15. Adjusting the distributions of O3 concentrations based on decreasing NOx emissions to just meet the existing and alternative O3 primary standards resulted, in some cases, in substantial shifts in the spatial and temporal patterns of O3 across case study urban areas relative to patterns of O3 that existed for recent air quality, and presumably relative to the patterns present

in the study locations of the epidemiology studies from which the concentration-response functions were drawn (see section 7.1.1 of the TSD, USEPA, 2012). What are the views of the Panel on the characterization of the degree to which these changes in spatial patterns of O3 introduce uncertainty in risk estimates when effect estimates based on one spatial/temporal pattern of O3 (the pattern in the epidemiology study) are applied to a substantially different spatial/temporal pattern of O3 concentrations?

Well, the Bayes estimates should theoretically make each city C-R function less sensitive to the particular spatio-temporal pattern present during the place and time used for the epi analysis. I don't see any ready alternative now to using the same C-R function, although the assumption of a constant function is a strong one.

16. In particular, what are the views on the Panel on the characterization of the level of uncertainty associated with estimates of risk associated with days with relatively lower composite (area-wide average) O3 concentrations and those with relatively higher composite O3 concentrations?

This characterization is not based on epidemiology (where the confidence interval around the effect smooth widens at both extremes) but rather on findings from human experimental and toxicologic findings, which seem to me to pretty sound in this regard.