Updated 9-11-12.

Preliminary Individual Comments on Health Risk and Exposure Assessment for Ozone (First External Review Draft, Updated August 2012)

Comments from Mr. G. Allen 2
Comments from Mr. Ed Avol 4
Comments from Dr. Michelle Bell 9
Comments from Dr. Joe Brain 12
Comments from Dr. David Chock 15
Comments from Dr. Ana Diez Roux 18
Comments from Dr. H. Christopher Frey 23
Comments from Dr. W. Michael Foster 34
Comments from Dr. Harkema 37
Comments from Dr. Daniel Jacob 38
Comments from Dr. Steven Kleeberger 40
Comments from Dr. Fred Miller 42
Comments from Dr. Armistead (Ted) Russell 55
Comments from Dr. Helen Suh 59
Comments from Dr. James Ultman 63
Comments from Dr. Sverre Vedal 65

Comments from Mr. G. Allen

September 3, 2012

Preliminary Response to Charge Questions for Health REA, 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 from Mr. Ed Avol

Comments on 1st Draft O3 REA 31Aug2012

General Comments:
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 the California's central valley, it's a shame that the information either wasn't available or that Fresno CA was not selected for study inclusion. However, potential confounding by agricultural exposures, health care access, and demographics in this region would likely have made this a challenging area to consider.

Specific Comments:
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"
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. 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…"
12. 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.
13. 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.
14. 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).
15. Pg2-8, line 3: delete one of the "human" references from the phrase "…*human* controlled human exposure studies…"
16. Pg2-8, line 10: change to read "…more serious *or chronic* health endpoints…"
17. 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."
18. Pg2-8, line 20: add comma after "…controlled human exposure…"

19. Pg2-9, Figure 2-1: change legends to read "Short-term O3 exposures" and "Long-term O3 exposures".
20. Pg2-9, line 16: add comma after "…hospital admissions"
21. 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…!

3.Scope

22. Pg3-2 line 20 – replace "…done as part of…" with "…included in…"
23. Pg3-3 line12 to 13: add parentheses, so sentence reads "…counts of person-occurrences (which accumulate occurrences…over an O3 season)."
24. Pg3-3 line22: run-on sentence needs punctuation; change to read "…moderate or greater exertion. Health effects observed…"
25. 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…)
26. Pg3-4, line 22: add hyphen between "location" and "specific"
27. Pg3-4, lines 27 to 30 : sentence seems redundant.
28. Pg3-5, line 18: change to : "…in the general and susceptible populations, respectively,.."
29. Pg3-5, line26: add comma, so that sentence reads "... be considered, including…"
30. Pg3-7, lines 17 to 21: run-on sentence needs editing, breaking down into shorter sentences with focused statements.
31. Pg3-8, line 23: insert comma to read "…quality, as evaluated…" and remove comma at end of line (following "…(U.S. EPA, 2008a)"
32. Pg3-8, line 24: add commas to read "…This information, along with additional analyses,…"
33. Pg3-8 line27: change sentence to read "…alternative standards, to be presented in the second draft REA…"
34. 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."
35. Pg 3-10, line 6: delete "…, including….", and replace with ":
36. Pg3-10, lines 10 to 14: the sentence as presented does not make sense and needs to be broken apart and re-worded.
37. 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.

38. Pg3-13, line23: Either a word is missing form this sentence, or the order is incorrect; re-word so line reads "Risk measures for lung function risk assessment estimated…"
39. Pg3-14, line 1: age parenthetical comment needs correction.
40. Pg3-16, line 12: change to read "…the 12 *case study* urban areas."
41. Pg3-16, line29: rewrite so line reads "…for which we *presently* only provide…"
42. 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…???
43. Pg3-18, Figure 3-3: for consistency and readability, consider adding "Black identifies…"
44. Pg3-19, line7: change "of just" to "restricted to"
45. Pg3-19, line 8: add "effects" after the phrase "short-term"
46. Pg3-19, line 9: replace "higher confidence" with "greater confidence"
47. Pg3-22, line 26: change "risk" to "risks"
48. Pg3-22, line 27: reword so line reads "…as well as core risk estimates *for* ozone…"
49. Pg3-22,line28: remove "zero" and insert "0 ppm"
50. Pg3-22, line 28: change end of sentence to read "…(LML) *observed* in the epidemiology.studies."
51. 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…"
52. Pg3-23, line6: add comma to read "…ozone is reduced, because…"
53. Pg3-23, line 8; change "higher confidence" to "greater confidence"
54. 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__)."
55. Pg3-25, line 7: is there a reference missing or an extra space?
56. Pg4-1, line 3: change "Chapters 5-7" to "Chapters 5 to 7" (since pages are identified by the chapter-dash-number designation)
57. 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."
58. 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…"
59. Pg4-2, line1: change "must be designed to record" to "must be located to capture"
60. 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,…?
61. Pg8-4, Figure 1.3: …still surprising that the Houston area does not show anything of note here…?

62. 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?
63. 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.
64. 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.

Comments from Dr. Michelle Bell

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.

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.

Updated 9-11-12.

Comments from Dr. Joe Brain

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.

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.

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.

Comments from Dr. David Chock

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. I can appreciate the dilemma here because it is hard to come up with reasonable alternatives. Besides, the real purpose in this exercise is simply to construct a scientifically reasonable ozone concentration distribution an area may have when it just meets the present ozone air quality standard. 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. 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.

Comments from Dr. Ana Diez Roux
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.

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 <u>number</u> of incident cases or total <u>number</u> 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 <u>even if the current standard were met.</u> 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.

Updated 9-11-12.

Comments from Dr. H. Christopher Frey

Submitted September 4, 2012

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.

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 mortaility burden associated with recent ambient O3 leels, 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.

Chapter 5:

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.

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.

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.

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 follow.

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: 2nd and 3rd 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. 2.5th and 97.5th 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 4th highest 8-hr daily maximum concentration. Furthermore, it needs to be crystal clear as to what exactly is being compared here. Is this a 4th highest value in just one year? Is it the three year average of the 4th 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 4th highest values? Or just the 4th 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 4th highest value in one year, or is it the three year average of 4th 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.

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?

Comments from Dr. W. Michael Foster

1st Draft Health Risk and Exposure Assessment for Ozone

<u>Chapter 6: Characterization of Health Risk Based on Controlled Human Exposure Studies</u>

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 <u>health effect</u> included in this portion of the risk assessment is lung function decrement (e.g., ≥ 15% reduction in FEV1) 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 FEV1 that were chosen/evolved (i.e., ≥ 10, 15, and 20%) for comparison. For change in FEV1 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 FEV1 (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 FEV1 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-SmithFEV1 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 dFEV1 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.

Comments from Dr. Harkema

Responses to Charge Questions:

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.

Ouestion 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.

Updated 9-11-12.

Comments from Dr. Daniel Jacob

Response to charge questions 5-6 of Health REA Draft 1

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 (4th 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?

1. 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?

1. 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?

1. 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 from Dr. Steven Kleeberger

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.

Comments from Dr. Fred Miller

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?

   Response – 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.

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?

   Response – 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.

Chapter 3: Scope

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

Response – 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 PRP 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?

Response – 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?

Response – 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.

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?

Response – The set of epidemiological studies selected for use was proceeded by a discussion of the types of data necessary for estimating the change in incidence of a given health effect resulting from a given change in O_3 exposure (pages 7-6 and 7-7). This included the air quality information needed, the C-R functions between the health endpoint of interest and O_3 concentrations, and the availability of baseline health effects incidences and prevalence rates in the population. There was adequate linkage to the ISA document to support the rationale for study selection and for the C-R functions.

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?

Response - Given the above, the authors succinctly describe on page 7-23 the

4 major criteria they used to select the epidemiology studies that would support modeling the health effect endpoints they had identified for analyses on page 7-22. The endpoints listed all pertain to short-term exposure, but the authors noted that consideration is being given to examining some long-term exposure endpoints in the 2nd draft of the HREA. This reviewer would not endorse addressing long-term exposure-related respiratory morbidity at this time due to the limitations in the data available for such analyses.

After selecting the set of epidemiology studies to use, the authors presented an expanded, but relatively succinct, discussion of the factors they considered in identifying the effect estimates and model forms to specify C-R functions for a given endpoint. The 6 factors included: 1) the O_3 exposure metric, 2) single- and multi-pollutant models, 3) single-city vs. multi-city studies, 4) multiple lag models, 5) seasonally-differentiated effects estimates, and 6) the shape of the functional form of the risk model. The last item is particularly important for the affect of O_3 on FEV_1 changes given the recent publication by McDonnell et al. (2012) that clearly shows a threshold for such changes exists.

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

Response – The authors have done an excellent job in Chapter 7 of identifying aspects that either contribute uncertainty or variability to the risk estimates. However, the various additional analyses should be prioritized, as it is unlikely that the Agency will have all of the resources available that are needed to examine all of the variables and topics that have been identified.

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?

Response – The 11 bulleted paragraphs in Section 7.6.2 do a fairly good job of presenting and discussing the key results of the analyses that were conducted for the 1st draft HREA. However, in the 2nd draft, more attention should be given to illustrating the major findings with some simple graphs that convey the findings of "the recent conditions

scenario" versus the scenario where the "current standard" is just met. Readers may well get lost in the magnitude of the extra verbiage used to convey all of the important findings.

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"?

Response – The heat maps are quite useful for the O_3-related mortality data. However, the discussion in the text could be expanded to better convey various examples of the points one can make about the interpretation of the data presented in these maps. For example, just taking the New York data in Tables 7-7 and 7-8, it is not clear why the number of deaths in the 45 to 50 and 50 to 55 ppb categories increase dramatically if the current standard is just met. There are 81 overall fewer deaths but the results for some of the columns seem counter-intuitive and seem to infer that a lower O_3 exposure is worse for a person. Also, presenting the color scale for how many deaths are associated with a color or maybe using a broader color scheme might be useful. For example, currently 47 to 72 excess deaths are shown using the same color.

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?

Response – About 2 pages are devoted to a discussion of the characterization of the overall confidence in the risk assessment and risk estimates derived from the urban case study analyses. The 5 topics discussed on page 7-73 appear reasonable and the authors explain reasonably well their insights about the likely confidence in the outcomes for these topics. However, this reviewer does not understand why the log-linear no-threshold modeling was used for the full range of exposures down to zero O_3 as this belies the real issue that needs to be addressed – that of risks above PRB exposures, particularly since most of the LMLs from the epidemiology studies are below the PRB level in the geographic areas where the epidemiology studies were conducted.

Also, to better understand some aspects relative to characterizing the overall confidence in the various results, the reader has to rely on material presented in Section 7.6.2 as well as in Section 7.6.1. Some integration of the topics covered in these two sections would be helpful in a 2nd draft of the HREA. For example, the authors explain that Bayes-adjusted city-specific exposure-related mortality estimates have increased overall confidence since they combine the elements of the local city-specific signal with a broader national signal. However, in Section 7.7 on refinements being considered for the 2nd draft HREA, they convey that for sensitivity analyses on

short-term exposure-related mortality they would have to obtain Bayes-adjusted city-specific effect estimates for the factors being examined, such as lag structure, and they imply that this can't be done so that the insights gained would be somewhat limited.

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)?

Response – All of the proposed refinements discussed in the HREA are reasonable and would help in identifying the nature and magnitude of public health risks resulting from O_3 exposure. They would also assist in judging the confidence in the point estimates in view of the contributions of variability and uncertainties. However, one of the most important consideration that should be made before working on the 2nd draft of the HREA is to prioritize the potential tasks and sets of analyses that could be conducted. This reviewer is of the opinion that the Agency lacks the resources to conduct all of the various analyses that are discussed for potential inclusion in the next draft.

In addition, this reviewer does not understand why the Agency appears reluctant to identify health risks above PRB levels. The ISA presents modeled base and seasonal data on ozone levels in the absence of anthropogenic sources. This information could be used to identify geographical-specific PRB values for assessing exposure-related health risks above PRB levels. Given that most of the LMLs from the epidemiology studies are below the PRB level in the geographic areas where the epidemiology studies were conducted, the Agency is most likely over estimating controllable risks by not incorporating PRB levels into their risk analyses.

The Higher-Order Decoupled Direct Method capabilities in the Community Multi-Scale Air Quality model are proposed for use in the 2nd draft REA. While using this approach elminates the need to specify values for U. S. background values because they are incorporated directly in the modeling, it is not clear to this reviewer whether the Agency would present risk results for values above these levels in addition to or instead of the no cut-off or the LML cut-off. This is a topic that needs to be discussed at the September 11-13, 2012 meeting.

Chapter 6 : Characterization of Health Risks Based on Clinical Studies

Pre-Meeting General Comments

Chapter 6 is well written and reasonably easy for the reader to follow. The Agency is to be commended for considering using the McDonnell et al. (2010) model, thereby being able to examine the exposure-response function at a population level and not at an individual level. In the opinion of this reviewer, population level risk estimates are more defensible from a risk management viewpoint. Moreover, they are more realistic for use in the NAAQS process because they result in greater risk estimates than individual response functions and more likely capture what is really going on in the population. Importantly, the Agency should be 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 (

Some specific comments are provided below.

Pre-Meeting 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 and the tolerance distribution for individual responses should have a narrow standard deviation, the 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 rations 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 that 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. 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.
6-51, 27	The Graham (2012) citation is not in the Chapter 6 reference list.

Chapter 7. Characterization of Health Risks Based on Epidemiological Studies

Pre-Meeting 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 with no attention to providing risk estimates for effects above PRB levels. Almost all of the LML values in Table 7-5 are below the PRB levels discussed in the ISA. To this reviewer, this casts doubt on the validity of estimating risks down to the LML levels in the various epidemiological studies since the Agency cannot control the level of the PRBs nor can it legislate they be reduced.

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 capabilities in the Community Multi-Scale Air Quality model are proposed for use in the 2nd draft REA. While using this approach elminates the need to specify values for U. S. background values because they are incorporated directly in the modeling, it is not clear to this reviewer whether the Agency would present risk results for values above these levels in addition to or instead of the no cut-off or the LML cut-off. This is a topic that needs to be discussed via an interactive dialogue of the Ozone Panel with EPA staff at the September 11-13, 2012 meeting.

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.

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.
Some specific comments are provided below.

Pre-Meeting 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. Can EPA identify a couple of 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 that the differences in mortality numbers in the "Delta" column from a subtraction of the values in Table 7-8 from Table 7-7 are the result of rounding of the numbers in the separate tables.
7-47, 12	How is the Agency going to show risk via mortality deaths above the Policy Relevant Background levels? Such data would be more useful than using excess deaths above the LML values from epidemiology studies.
7-66, 22	In this paragraph, the numbers of O_3-related deaths across the 12 urban study areas are listed in aggregate and comments are made about Detroit and New York generating higher risk estimates than those for other studies. Would a weighted mean that takes into account population density of the 12 areas be helpful for putting into perspective the overall national significance of the risk estimates?
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. Moreover, unless the epidemiology studies specifically adjusted for the contribution of PM, there is over counting of the number of deaths.
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.
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 as the accuracy of the analyses will be greatly reduced, 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.
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.**
Section 7.7.2	Again, the suggestion is made to prioritize the list of additional refinements to the cores risk estimates since the Agency may not have the resources needed to address all of the items listed in this section.
7-77, 40	Generating CIs for the delta risk estimates would be extremely useful, but how would the authors propose to generate the error term to be used in the CIs? There may be multiple ways to generate unbiased error estimates, yet they may range considerably in their magnitude.
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.

Chapter 8. National-Scale Risk Assessment and Representativeness Analysis

Pre-Meeting 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.

Pre-Meeting 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 be easily 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. The 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).
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. Is 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 very low priority for resource allocation.

Comments from Dr. Armistead (Ted) Russell

Executive and Integrative Summary: I hope both are coming.

Chapter 1:

The Minor issues:

1-6, l30: Remove "any"

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 … <u>predicted</u> Air Quality Index…

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

2-10 Remove a period.

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.

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).

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.

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.

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 from Dr. Helen Suh

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 complicated yet sound. 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, 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.

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

- A figure that describes the calculations (or modules) of the APEX model should be added to Section 5.3.1, 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.
- 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.
- It is not clear to me why the analysis was limited to the general population, children, and asthmatic children, as older adults are also a potential sensitive group. Either justification of this decision is needed or this group should be added to the analysis.

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 included in the results is confusing. It seems that the "first set of figures" (lines 26-27) are missing or perhaps come at the end of the Chapter as Figures 5-22 to 5-25. The "second set of figures for school age children" seems to refer to Figures 5-1 to 5-15, which are again discussed in the following paragraph.
    - The presentation of the main results for school age children is limited, and should be modified to compare results across design values and cities. The presentation of the main results should also be enhanced to present results for asthmatic children and the general population, again with comparisons by design value and by city. Some of this results presentation can be found in the first two paragraphs of Section 5.6.2. These 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.

- 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.

Comments from Dr. James Ultman

Chapter 6

Characterization of Health Risk Based on Controlled Human Exposure Studies

J.S. Ultman

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?

Revisit the concept of correcting FEV1 changes during ozone exposure with those measured during clean air exposure: this is particularly important in exercising individual's who show an increase in FEV1 during clean air exposure. If the effects of ozone exposure and exercise interact (i.e., ozone exposure affects the FEV1 response to exercise), then there is no assurance that taking the difference between ozone exposure and a clean air control will reveal the effect of ozone alone. However, we can say that the change in FEV1 without clean air correction soes represent the effect of ozone exposure plus exercise occurring in exercising individuals. Thus, the comparison of this overall health effect between different exposure scenarios (e.g. just meeting current standard versus just meeting alternative standards) would be a logical basis for the policy assessment.

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?

The origin of the MSS model as described in the current draft of the REA is somewhat obscure. I think that the description given in the Introduction to Schegle's recent article (2012, Inhalation Tox) is easier to follow.

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.

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

In the absence of appropriate data (that might very well show a further decline with age>35), the approach of freezing the age factor at 35 years is a reasonable, conservative approach (that presumably overpredicts the FEV1 response of the elderly).

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. It is unlikely that this is a reasonable assumption for children with asthma, considered to be an important high risk population. Trying to quantify the uncertainty of this assumption should be a priority in the next draft of the REA.

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

See item 6.

Comments from Dr. Sverre Vedal

Vedal 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.

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"?