Responses to CASAC Questions on the Ozone ISA from Consultant Mr. John J. Jansen

Given the broad nature of the questions posed by the CASAC members, I read a good deal of the draft Ozone Integrated Science Assessment (ES, IS, Appendices 1, 2, and 10, and portions of Appendices 3 through 7). I only scanned appendix 8 since there were no questions posed. I have familiarity with welfare effects and could address it if desired, but not until after November 8 as I have a prior commitment. As with my comments on the PM PA, I respond to many, but not all, CASAC member questions and also offer some general thoughts and other comments at the end.

Questions from Dr. Masuca

As to **precursor sources** page 1-7 notes that the focus of the chapter is on sources of USB ozone so a focus on CO, CH4, **global CH4, and international emissions** is appropriate. There is summary discussion of all ozone precursors and sources and it appears adequate. It would be useful to see some information on the "trend" in biogenic VOC emissions in order to understand its year to year (and season to season) variability. To my knowledge I have never seen such a trend plot and recognize it would be difficult given methodological changes over time but it is in the NEI and might put other pollutant trends into perspective (see figure 1-3c on page 1-11). On page 1-12, I question the statement that mobile sources are the **primary** driver to NOx declines. EGU and industrial sources have also declined significantly.

I am not sure I understand the question on **biogenic VOCs**. Assuming the question's focus is on the CTM's ability to accurately assign responsibility for ozone between biogenic and mobile source VOC, the uncertainties of accurately estimating the specie, amount, location, and timing of both biogenic and mobile VOC are significant. I would submit that although a lot of work has gone into estimating mobile VOC, the limitations are similar to what is summarized in the final bullet for biogenic VOC on page 1-21.

As to **ozone photochemistry**, to the extent that new data from PAMS and near road monitoring is leading to updated CTMs or better CTM performance (or even using the data in model performance evaluation), I would suggest adding such a discussion here or in the modeling discussion (section 1.6). Since the near road data, in particular, is recent, it may be too soon to expect such work and results.

Topography is certainly important locally and needs to be handled in the meteorological models. Humidity has an effect and is handled in the chemical mechanisms of CTMs.

I was disappointed that there was no discussion in section 1.7 on the shifting nature of ozone peak concentrations. With the successful lowering of peak ozone in the summer, the likelihood of the peaks occurring in spring or fall is increasing (see below for references). These findings have implications for monitoring and SIP modeling demonstrations. It may also have implications for source attribution and control strategies if what's needed to reduce peaks in the spring are different from the summer.

- Blanchard, C. L. and G. M. Hidy, "Ozone response to emission reductions in the southeastern United States," Atmos. Chem. And Phys. 18: 8183–8202 (2018).

- Blanchard, C. L., et.al., "Emission influences on air pollutant concentrations in New York State: I. Ozone," Atmospheric Environment: X 3 (2019) 100033

Remote sensing data is discussed relative to USB ozone in section 1.6.1.2. Such data is also used in hybrid modeling for exposure estimation on page 2-12.

I did not understand the remaining questions.

Questions from Dr. Frampton

The three questions basically ask whether specific causality classifications or change in classifications is justified. While this area is not in my area of expertise, I do have some comments about the framework for conducting systematic reviews and making causal determinations (see also responses to Dr. Cox below).

While I commend EPA staff and do not envy them the task of searching, reviewing, summarizing, and evaluating the literature, I have always been frustrated by what I perceive as a lack of clear criteria and transparency in the descriptions of what leads to a particular causality classification. I can read several descriptions of evidence and am unable to identify what makes one "causal" and another "likely" or "suggestive" and as a result have a difficult time deciding whether I agree or not. Clear criteria are needed for study inclusion/exclusion, study quality, and causality classification. It is also not always clear which evidence is being given more weight than other evidence.

These issues has been commented on in many past NAAQS reviews (most recently by Dr. Julie Goodman of Gradient on the Particulate Matter Policy Assessment Document (https://yosemite.epa.gov/sab/sabproduct.nsf//1D9FD74E638BDBAE852584950014F5B5/$File/Goodman+Comments+on+Draft+PM+Policy+Assessment.pdf). The problems persist in this document.

You ask whether additional studies should be included. While I do not know of any, I hope others will offer suggestions. No system is perfect and the review process fills the gap. **In the past**, there appeared to be a tendency to initially include studies with significant positive studies and exclude significant negative as well as null studies. And, therefore, suggestions for missed studies were made. This is not as apparent in this ISA as evidenced by the many figures such as Figure 4-2 on page 4-11. Hopefully any missing studies will get identified in this review.

Again, in the past, when significant negative and null studies were included there was a tendency for the narrative to critique (dismiss?) them and simply accept the significant positive studies. This suggests a bias in the review of the evidence. Identified missing studies also ran the risk of post-rationalization. EPA needs to work harder to critique all studies, weight them appropriately, and avoid post-rationalization of additional studies identified in the review process. Unfortunately, this bias is continuing. See for example page IS-27:

> "While there is coherence between epidemiologic and experimental evidence of ozone-induced lung function decrements and pulmonary inflammation, respiratory symptoms were not

associated with ozone exposure in a limited number of epidemiologic studies. However, these studies generally relied on parent-reported outcomes that may have resulted in under- or over-reporting of respiratory symptoms."

As to your first question, EPA's conclusion that - the weakness in the epidemiologic evidence remains despite some new studies showing significant positive and null results - rest on the lack of significant positive results on for HF, IHD, and MI, arrhythmia and cardiac arrest, or thromboembolic disease. I cannot comment on how critical having such evidence is but it appears to prevent the overall epidemiologic evidence from strengthening. However, looking across all the figures (like 4-2) with relatively few significant positive effects seems to be consistent with "suggestive."

As to question 2, I fall back to the lack of criteria for judging classification. While I am of the school that quality human and animal experimental studies at relevant exposures need to be weighted over suggestive epidemiological (associational) studies to establish causality, I do not know whether there are sufficient suggestive epidemiological studies here. Also, there are no summary figures like figure 4-2 to help me.

As to question 3, figure 6-2 seems to show weak evidence for cardiovascular and respiratory mortality begging the question what is driving the total mortality. The quote from page 6-20 relies on the controlled human exposure studies. In a quick scan, I could not find the controlled human exposure discussion or a figure on those studies.

Questions from Dr. Lange

Question #1: I was not aware of this. Imposing "significance" as a criterion on one type of study and not others seems wrong. On the other hand, does it explain the use of the phrase "positive associations" in many places (e.g., see page 4-2) rather than "significant positive associations?" I find such phrasing troublesome as it could be implying more rigor than exists and lacks clarity. If the result is null, it's null. Significant results should be shown distinctly, not diluted by "positive" null results, and weighted more heavily.

Question #2: I believe that is correct. I also seem to recall use of the ozone on, say, another Tuesday in a month, assuming the event happened on a Tuesday.

Question #3: I believe your concern is not limited to mortality. I would expect using the day after for a hospital admission is affected by medical treatment and being confined indoors. If this is a fatal flaw, the studies should be excluded. If not, then there is clearly an uncertainty and the study results should be down weighted.

Question #6: Clearly, the issue should be a key criteria used in the selection of and evaluation of studies. As I stated above, EPA needs clear criteria for study inclusion/exclusion, study quality, and causality classification. It is also not always clear which evidence is being given more weight than other evidence.

Questions from Dr. Packham

Question #2: The issue of "beneficial" effects should certainly be included. I cannot judge whether a total re-write is needed but balancing adverse versus beneficial effects is always a challenge (e.g., see EPA documents on the SOx, NOx, PM NAAQS secondary effects for nitrogen deposition). Stepping back, part of the problem is choosing which metrics to assess and deciding if the responses are, or lead to, a serious adverse outcome. In the welfare effects literature, it is common to declare any observed response or change as an adverse effect which is problematic. In addition to the issue of beneficial effects, there is the issue of recovery or reversibility. I did a search for the terms in the ISA and found some references to them in discussing experimental studies, which is appropriate. However, I did not see how it affected weighting nor causality classification. In other words, if a metric was responsive but recovered, how is that evidence weighted and used in terms of causality classification?

Question #3: Yes. Clearly Schelegle played a key role last time and there was much debate among the panel, CASAC, and public commenters. And I suspect figure ES-3 is destined for the same fate. Your analysis is helpful and EPA should be encouraged to consider it. As I stated above, I am a proponent of weighting quality human and animal experimental studies at relevant exposures over suggestive epidemiological studies to establish causality.

Questions from Dr. Boylan

Overall, Appendices 1 and 2 do a reasonable job covering the issues. There are a few areas that should be added/addressed namely a) adding trends in biogenic VOC's, b) correcting the description on drivers for NOx trends, c) adding a discussion of near road and PAMS data, and d) adding a discussion on the broadening of when peak ozone is occurring. See my comments on Dr. Masuca's question for more detail on these. Also, there should be discussion on how the various approaches to estimating exposure affect the health analyses. In addition, consistent with my recommendation regarding quantitative uncertainty analysis, these sections should provide a summary of what the quantitative uncertainties are for each.

I am not familiar enough with the literature on ozone effects on climate to render an opinion on the appendix's accuracy and completeness. However, I wonder how relevant the appendix is in setting an ozone NAAQS. There is precious little discussion of US ozone nor US consequences. Since the ultimate purpose of the NAAQS is to protect US public health and welfare, one cannot do so without a US context. How does the information provided allow the Administrator to set a level of US ozone concertation that is protective of US health and welfare related to climate effects? The only "level" information provided is between pre-industrial levels and today's levels. Further, the document explicitly avoids any discussion of feedbacks from ozone precursors and their effects on radiative forcing (see pages 9-4, 9-5, and 9-10). As to effects on temperature, how does one deal with the fact that temperature affects ozone formation and ozone affects temperature. It is also disappointing that no work has been done on the issue of tropospheric ozone effects on UV-B shielding. Finally, there is the issue of change vs. adverse effect. Similar to my concerns with much of the ecological effects literature

(especially critical loads literature), only changes are referred to with no perspective. Is the inference that any change is adverse (like in the critical loads literature)? Without such context how does one set a level that protects?

Questions from Dr. Cox

On the issues of the causality framework and causality classification, below are my thoughts:

1. EPA has updated and refined its approach to causality determinations over the years but it has always been qualitative in nature. Despite suggestions to alter and/or add to its approach (especially to add criteria and clarity), EPA has continued to justify the status quo as consistent with past practices endorsed by CASAC.
2. Clear criteria are needed for study inclusion/exclusion, study quality, and causality classification. It is also not always clear which evidence is being given more weight than other evidence.
3. I am of the school that a) associations are not and cannot be causal and b) quality human and animal experimental studies at relevant exposures need to be weighted over suggestive epidemiological (associational) studies to establish causality. In fact, I might argue that robust associational evidence can, at most, infer causality not demonstrate it. The methods advocated by Dr. Cox, once conducted, may or may not alter this statement. To be clear, I am not saying to ignore the associational evidence, only limit how far one interprets it towards causality.
4. In evaluating the strength of the evidence, EPA looks across scientific disciplines (i.e., including epidemiology, controlled human exposure studies, and animal toxicology). This evaluation can lead to a causal classification assuming quality experimental studies are weighted over suggestive epidemiological studies. When EPA evaluates studies across statistical disciplines (i.e., panel, case-crossover, time-series, case-control, and cross-sectional studies), in my opinion, it cannot lead to a causal classification but it can lead to a strengthening of the suggestive associational evidence.
5. It is not always clear which evidence is being given more weight than other evidence. I can read several descriptions of evidence and am unable to identify what makes one "causal" and another "likely" or "suggestive" and as a result have a difficult time deciding whether I agree or not.
6. The assessment, especially of the experimental studies, must deal with beneficial effects, reversibility, and recovery. Stepping back, part of the problem is choosing which metrics to assess and deciding if the responses are, or lead to, a serious adverse outcome. In the welfare effects literature, especially in critical loads research, it is common to declare any observed response or change as an adverse effect which is problematic. In addition to the issue of beneficial effects, there is the issue of recovery or reversibility. I did a search for the terms in the ISA and found some references to them in discussing experimental studies, which is appropriate. However, I did not see how it affected weighting nor causality classification. **In other words, if a metric was responsive but recovered, how is that evidence weighted and used in terms of causality classification?**
7. Finally, I have always been concerned about the term causal vs. the term contribute. We know death certificates are problematic with primary, secondary, and even tertiary causes listed. And I know from personal experience, they are not necessarily correct. How does this factor into the analysis and messaging?

Since I am not a statistician, I cannot respond in detail to your questions. I can only refer to the above. For example, numbers 1 & 2 above would be responsive to your question 2a. As I read through them, some thoughts came to mind. They are listed below:

1. For question 2b, the answer is no simply because the framework does not accommodate those terms. In addition, however, see number 7 above.
2. For question 2h see number 7 above.
3. For question 2l, we need clarity and criteria for the 5 we have, not more. Your list demonstrates the unclear nature and lack of criteria that exists. While it would be nice to have a not causal, I suspect the argument against it is that you cannot prove a negative.
4. For question 3i, your examples show the need for adding studies to the evaluation. No system is perfect and the review process fills the gap as illustrated by your list.
5. For question 3iv, I would suggest that some criteria for minimums on things like confounding would make sense. However, whether a hard line can be established is problematic. Rather, the lack of or minimal confounding analysis should down weight a study.

General Observations

As I stated in my comments on the PM PA I am a proponent of a quantitative uncertainty analysis being performed. And I would argue that the ISA should include a section **in each chapter** on the literature that evaluates the uncertainty in the various components that will make up the risk assessment. One could look to the outline provided in Dr. Lange's question number 4 (questions to consultants on the PM PA) for the beginnings of an outline, adapted of course for ozone. I do believe substantial data and hopefully studies exist to derive estimates for many of the items in her Table 1 and EPA should get on with performing that work. They have been advised to do so in the past. The approaches recommended by Dr. Anne Smith (see below taken from PM PA comments of Dr. North) should also be tried and information embedded in the various studies needed for such analysis should be summarized.

- Smith, A.E. and Gans, W., "Enhancing the Characterization of Epistemic Uncertainties in PM2.5 Risk Analyses, Risk Analysis, 35:361-378 (2015).
- Smith, A.E., "Response to Commentary by Fann et al. on 'Enhancing the Characterization of Epistemic Uncertainties in PM2.5 Risk Analyses,'" Risk Analysis 35:381-384 (2015).
- Smith, A.E., "Inconsistencies in Risk Analyses for Ambient Air Pollutant Regulations," Risk Analysis 36:1737-1744 (2016).
- Smith, A.E., "Response: Author Synthesis and Response," Risk Analysis 36:1780-1792688-1692 (2016).
- Smith, A.E., "Using Uncertainty Analysis to Improve Consistency in Regulatory Assessments of Criteria Pollutant Standards," Perspective, accepted for publication in Risk Analysis (2019). (To become available as soon as typesetting and proofing are done on Early View, https://onlinelibrary.wiley.com/toc/15396924/0/0).

Other Comments

Figure ES-2 on page ES-6 should change the "*" to an up or down arrow to show upgraded and downgraded classification.

EPA should use consistent units when describing a body of evidence. For example, it is frustrating to read, for example, the discussion on pages IS-36 and IS-38 with both ppb and ppm included. It is also misleading.