Responses to CASAC Questions on the Ozone ISA from Consultant Dr. Sonja Sax

Questions from Dr. Tony Cox

Overarching Questions (*in italics*)

1. *Is the scientific information provided by the ISA clear?*

    a. *Is it clear how the ISA's causal determination conclusions can be tested, and either verified or refuted (or left undecided), by observations?*

    In my opinion, the ISA could be clearer in many aspects, including how studies are selected for inclusion (or exclusion), and how the evidence is weighed and specifically used to reach causal conclusions. I generally agree with many of Dr. Cox's suggestions for improving the process to make it more objective. Given the same information, I believe, that different people (or different groups) could come to very different conclusions and causal determinations based on current classification descriptions.

    In the ozone ISA, this is exemplified by the "down grading" of the classification for two specific endpoints – cardiovascular effects and mortality from ozone exposures. Given similar evidence in the prior ozone ISA, EPA concluded that both of these endpoints were "likely causal," whereas now EPA has concluded that the evidence is "suggestive." Several papers published by me and my colleagues, which questioned the original EPA classification for cardiovascular effects are more in-line with the current conclusions by EPA. On the other hand, EPA has concluded that evidence supports a "likely causal" classification for metabolic effects, for which EPA did not provide any classification previously. There are many issues with this new category of health outcomes, and with the way in which EPA evaluated the evidence and made its causal conclusion determinations. For example, being a relatively new health outcome category, which greatly overlap with other health outcomes (i.e., cardiovascular effects) it is unclear how the evidence should be evaluated, including whether studies have sufficiently accounted for the numerous confounders that could contribute to the findings. Furthermore, it is less clear how the animal findings (which appear to be the evidence that EPA finds most compelling) should be interpreted. That is, whether the various downstream effects observed at high levels of exposure would be applicable to development of metabolic disease in humans.

    In a 2013, my colleagues and I published a paper that presented recommendation for improvements to the NAAQS causality framework, including many areas that could benefit from greater clarity. In particular, we use the prior ozone ISA as a case study to highlight areas where EPA could improve the overall evaluation of health effects. This paper is relevant to the current ISA as many of the limitations associated with the prior ISA are still present.

[See Goodman JE, Prueitt RL, Sax SN, Bailey LA, Rhomberg LR. **Evaluation of the causal framework used for setting national ambient air quality standards.** Critical Reviews in Toxicology 2013;43(10):829-49.]

For example, the scientific evidence provided in the ISA generally represents a summary of the findings from selected literature that EPA deems to be most relevant. In general, I understand that the literature that EPA is tasked to review is quite vast, however, the ISA could benefit from being clearer on how the literature is selected (i.e., more specifics on search criteria used, screening on literature, literature included and excluded, and reasons for inclusion and exclusion). This would provide greater transparency related to this process.

One critique in our paper that EPA has partially addressed in the current ISA, is that there previously was a lack of guidance regarding how study quality in considered. In the current ISA, a new addition is some guidance on what constitutes a high-quality study in each of the study domains (epidemiology, toxicology, etc.), which is essentially repeated as an Annex to each Appendix for each health outcome category. However, it is still unclear how the guidance is applied to each individual study, as the study summaries presented little, if any, information related to the study quality or even the strengths and limitations of the studies. EPA could be clearer on whether the study quality guidelines were used to include/exclude studies that fit/or did not fit these study quality criteria. That is, are all the studies that EPA presents considered to be "high" quality studies? I have not verified whether the studies summarized comply with the study quality criteria that EPA presents. While this shows progress in the right direction, EPA needs to take it one step further and present information on how the guidelines were used in the evaluation of individual studies.

In addition, the ISA causality framework would also benefit from some clarity regarding what each causal classification represents. Specifically, is the classification meant to identify the strength of the evidence for a given level of exposure (e.g., below the current NAAQS) or is it simply to identify whether there is evidence enough of expected harm at any level of exposure? That is, does the ISA causality framework represent simply a hazard identification rather than identification of harm at a certain level of exposure. EPA has made some progress on this point by identifying relevant literature to be evaluated using PECOS, which "defines the parameters and provides a framework to help identify the relevant literature to inform the draft 2019 Ozone ISA statements." However, in the PECOS statements EPA specifies an exposure range to be considered. i.e., "0.4 ppm or below for humans, 2 ppm or below for other mammals" only for experimental studies, and for epidemiological studies only specifies "ambient concentrations of ozone," presumably including concentrations above and below the NAAQS. In general, EPA could be more explicit in detailing the distinction between hazard assessment and risk assessment (for example, in the actual definitions for the causal classifications).

b. *Do the concepts and terms used to express key scientific conclusions in the ISA, especially the causal determination categories, have clear scientific meanings (e.g., unambiguous operational definitions)?*

The operational definitions of the causal classifications leave a lot of room for subjective interpretation. For example, for a causal conclusion EPA notes that "the pollutant has been shown to result in health effects in studies in which chance, confounding, and other biases could be ruled out with reasonable confidence." In this case it is unclear what EPA means by "reasonable confidence." What is reasonable to one person may not be sufficient for another.

c. *Is it clear and generally understood and agreed what the key conclusions mean? Specifically, are the causal determination categories used to communicate key conclusions unambiguous and well defined?*

See answer above

d. *Do those who read the ISA have a shared, unambiguous understanding of what its key scientific conclusions (i.e., causality determinations) imply about how or whether changes in ozone air pollution would change public health outcomes?*

This is where further clarity is needed. Are the causal classification conclusions suppose represent a hazard identification only? Or should they also inform whether harm occurs at a certain "relevant" level of exposure? This is unclear in the ISA, where for example, PECOS statements define a level of exposure for experimental studies, but not for epidemiological studies. Because the NAAQS process involved the development of a level of exposure.

2. *Is the scientific information provided by the ISA sound?*

a. *Are its conclusions logically implied by the data and analyses on which they are based?*

Most of the ISA sections that I have reviewed (albeit not in very much detail), including the respiratory section (Appendix 3), the metabolic effects section (Appendix 5), and the cardiovascular section (Appendix 4) as well as the Integrated Synthesis, appear to summarize results from recent studies and provide very general "integration" summaries. There is little with regard to presenting information on study quality or how studies are weighed by EPA when making inferences or drawing conclusions. In general, it is difficult to follow the rationale as presented in the ISA with regards to causal conclusion determinations and there are appear to be inconsistencies in how the evidence is deemed sufficient to select one classification vs. another, such as in the cardiovascular vs. the metabolic outcomes. This highlights the need to have clear protocols and definitions regarding how these conclusions are derived.

b. *Are its conclusions correctly stated and caveated?*

No, it is hard to follow the rationale for the conclusions. For example, although the cardiovascular effects section and metabolic effects section are related, EPA comes to very different conclusions regarding these two endpoints. EPA seems to provide much more discussion regarding the uncertainty and limitations of the cardiovascular evidence, in particular the incongruency between the subclinical results in animal studies and the epidemiological findings. In addition, there appears to be many more studies on the cardiovascular effects, than on metabolic effects and I saw little if any discussion of strengths and limitations of any of the studies in the metabolic effects section compared to the cardiovascular section. I think it is essential that EPA include this information in the text or the tables. For example, if animal studies only included a single exposure dose these studies provide much more limited evidence of an effect because you cannot evaluate a dose-response relationship. For studies that do provide multiple exposures, it is important to consider dose-response, and EPA does not seem to consider this. As noted above, adversity of effect is also an important consideration as transient, reversible effects are less likely to be adverse.

c. *Is it clear than conclusions do not reflect selection bias in the choice of studies relied on? Are its conclusions consistent with other relevant data and studies not included in the ISA?*

I have not conducted a separate literature search or analysis of the data to determine if there is selection bias. I do note that several studies (some of which I am an author of and references provided throughout) are missing from the ISA. More transparency is needed on how EPA selected studies for inclusion/exclusion

d. *Is it clear how studies were selected for inclusion in the ISA, and why individual studies were included or excluded?*

No, I did not see a discussion of how studies were selected or why studies were included or excluded, except for reference to the PECOS statements. I think that more clarity is needed in this regard.

Specific Questions

1. *Question: Can valid determinations of manipulative or interventional causation – that is, how and whether changing exposure would change health risks – be made based on observed associations of the types analyzed in the ISA? I emphasize manipulative and interventional causation (rather than predictive (Granger) causation, but-for causation, epidemiological (attributive) causation, mechanistic causation, etc.) because it is most relevant for policy makers.*

The short answer to this question I think is no. I suspect that Dr. Cox is referring to so called accountability studies, such as his paper [Cox & Popken. 2015. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? Annals of Epidemiology, Volume 25 (3), pgs. 162-173], which explicitly address whether reductions in PM and ozone have contributed to reductions in mortality rates. As noted above, EPA does need to make it clear whether the ISA is simply a hazard evaluation or if it is meant to address the adversity of the effects at a given exposure level.

2. *Questions: The following questions are intended to help assess the conceptual clarity and meaning of the causal determination categories, and of key conclusions expressed using them, such as those in Table ES-1 (p. E-5) of the Draft ISA.*

   a. *A preliminary question: Is this actually a "formal causal framework"?*

      I am unsure of the answer/ or answer is not clear. I generally agree that the current definitions for each classification leave a lot of room for subjective judgement and in general it is not always clear (to me) how EPA weighs the evidence and comes to its final causal conclusions. This could use some revision and clarification.

   b. *Does the ISA's causal determination framework clearly distinguish between necessary and sufficient causation?*

      The short answer to this is no. The ISA does not clearly make a distinction between necessary and sufficient causation as specified by Dr. Cox. This relates to my discussion above regarding clarity on whether the ISA is simply a hazard identification or whether the ISA is meant to identify a level of exposure (as tasked for developing a NAAQS) at which adverse effects are expected to occur. The current ISA and overall NAAQS framework is not clear on this point.

   c. *Does a determination that exposure has a "causal relationship" with a health effect in a population imply that reducing exposure would reduce risk of the health effect in the population, other factors being held fixed?*

      I think again the answer is no. I don't think the data are sufficient to establish this with any certainty. There are too many other factors that influence the development of disease or that contribute to deaths and the processes are so complex that it would be difficult, if not impossible, to assess this with any degree of certainty, particularly at low exposures. The observed associations do not imply causation and it is only by looking critically at all of the data, across all lines of evidence and weighing the strengths and weaknesses of these lines of evidence that we can at least establish plausibility for causation. However, even if causation could be established, given the multiple risk factors for any given disease (respiratory, cardiovascular, etc.) it would be difficult to determine how much reduction in risk from air pollution exposures would influence reductions in disease with any certainty. This is difficult for any risk factor that is evaluated, some which are known modifiable life style choices (smoking, drinking, diet, and exercise) and other which are not (e.g., genetics).

   d. *Can causal determinations be incorrect? (Or, to the contrary, are they performative utterances?)*

      I think that the causal classifications can be interpreted differently based on the current definitions and how EPA presents and interprets the data. That is, they are open to interpretation, particularly for determinations of a "causal relationship." That may be the only classification that could be considered to be "incorrect" based on different interpretations and weighing of the available data because I don't think that causality can be established with absolute certainty. Other classification determinations are more open to interpretation as to whether they are correct or not.

e. *If causal determinations can be mistaken, then is it clear how uncertainty about which category is correct should be (or has been) resolved in assigning a final causal determination category, as in Table ES-1 p. ES-5) of the ISA?*

No, I don't think it is clear. As noted above, EPA should provide more clarity and perhaps caveat the classifications based on the amount of uncertainty in the underlying evidence.

f. *If causal determinations can be incorrect, then is it clear how observations could be used to test and falsify a given causal determination if it is not correct? For example, is it completely clear how someone can use relevant data to show that a determination of "causal relationship" or "likely to be causal" in the ISA is incorrect, if indeed that is the case?*

No, I don't think that as the framework is laid out by EPA, it could be applied consistently and that the same causal determinations would be necessarily developed by different groups. As noted above, this is exemplified by the change in classification for two key health outcomes (CV and mortality) in the current ISA.

g. *If causal determinations can be incorrect, then is the correctness of each causal determination in table ES-1 formally and transparently evaluated in the ISA? In other words, have formal rules for determining the correctness of the causal determinations in Table ES-1 (p. ES-5) from the data and evidence presented been explicitly stated, applied systematically, and the results documented? (If so, where?)*

No, see my comments regarding the NAAQS framework above, as well as Goodman et al. (2013) reference (above).

h. *Does a determination that an exposure-response (or concentration-response (C-R)) relationship is a "causal relationship" imply that it is entirely causal, with no contribution from incompletely controlled confounding, modeling errors and biases, or other non-causal sources? If not, is there a clearly defined lower bound on how much of the relationship (e.g., how much of the slope of a C-R regression line) must be causal in order for the whole relationship to be classified as causal? (If so, what is it?)*

This is an interesting question, and I think that the way that EPA uses the C-R functions in the risk assessment, assumes that there is a causal relationship between the air pollutant (in this case ozone) and the various health effects, under the conditions specified in the underlying epidemiology model and then EPA often applies this to other populations. However, the epidemiological studies are flawed in that they do not account for uncontrolled or incompletely controlled confounders, and have errors and biases that are not always discussed or caveated. Therefore, it is more likely that there does exist some lower bound of the attributable fraction and it is unclear how much of it is necessary for consideration of a true causal effect. I think this is somewhat akin to determining a threshold below which effect are unlikely to occur, something that is very likely for ozone based on the lungs protective mechanisms against oxidative damage.

i. *Does a determination that a C-R relationship is a "causal relationship" imply 100% certainty that it is causal? If not, is there a clearly specified lower bound on how probable it*

must be that the relationship is causal in order for it to be classified as causal? (If so, what is it?)

See previous answer. I think as EPA uses it, yes.

j. Does a determination that a C-R relationship is a "causal relationship" imply that it is causal for every member of a population, or might it be deemed "causal" if it is causal for a sensitive subpopulation only? In the latter case, is there a clearly specified lower bound on the fraction of the exposed population for which the relationship must be causal, in order for the whole relationship to be classified as causal? (If so, what is it?)

This is also an interesting question. EPA currently uses C-R functions in its risk assessment based on selected epidemiological studies. The epidemiological studies are based on different population groups (sometimes children, sometimes the elderly, sometimes all ages etc.), therefore I think it depends on the epidemiological study and the population group that the study includes.

k. Are the five categories mutually exclusive?

I don't think they are necessarily mutually exclusive. Again, as defined I think the scientific evidence could be evaluated differently by different people or groups such that they would arrive at different conclusions (or different categories) given the same evidence. See for example our assessment of the cardiovascular effects of ozone and EPA's conclusions in the prior ozone ISA.

Prueitt, RL; Lynch, HN; Zu, K; Sax, SN; Venditti, FJ; Goodman, JE. 2014. "Weight-of-evidence Evaluation of Long-term Ozone Exposure and Cardiovascular Effects." *Crit. Rev. Toxicol.* 44(9):791-822.

Goodman, JE; Prueitt, RL; Sax, SN; Lynch, HN; Zu, K; Lemay, JC; King, JM; Venditti, FJ. 2014. "Weight-of-evidence Evaluation of Short-term Ozone Exposure and Cardiovascular Effects." *Crit. Rev. Toxicol.* 44(9):725-790.

l. Are the five categories collectively exhaustive?

Not necessarily, as there is no counter classification to each of the levels as noted by Dr. Cox, except if EPA selects the last category "not likely to be a causal relationship" to be the counter point to all other classifications – including "causal" and "suggestive" as well as "likely causal."

m. Can a body of evidence be categorized as "likely to be causal" if the probability of causality based on the evidence is less than 50%?

Based on EPA's definitions, it is hard to say what percent probability of causality needs to be for the evidence to be considered "likely to be causal."

3.

a. Questions: Based on these spot checks, I have the following questions:

i. *Is it clear that the ISA's study selection process has successfully provided a comprehensive, trustworthy, and unbiased selection of the best available science on ozone and health effects?*

As noted above, the ISA would benefit greatly from more transparency regarding the literature search strategy, the selection process for studies included for evaluation, and a summary of why certain studies were included or excluded from consideration.

ii. *Is it clear why results from Moore (2008) are included and cited as "key evidence" but contrary results from Moore (2013) are excluded? More generally, is it clear that study inclusion and exclusion criteria were applied systematically and neutrally to identify and select the best and most up-to-date studies to inform the ISA's conclusions?*

This is an area that does require further evaluation. I agree that EPA should include both positive and negative findings in a balance manner so a not to appear like it is cherry picking.

iii. *Are there other studies that are omitted from the ISA that should be included?*

Although I have not done an independent literature search, it appears that Dr. Cox identified studies that were not included (and EPA did not provide and explanation for exclusion). I note above, and also in comments elsewhere, several publications that my colleagues and I have published that could be considered for inclusion in the ISA.

iv. *Are there studies included in the ISA that should be omitted (e.g., because of uncontrolled confounding, obsolete or incorrect modeling assumptions, conclusions dependent on unverified assumptions, ecological fallacy, lack of causally relevant information, lack of design that can support valid causal inferences, or other methodological problems?)*

Other than general study quality criteria that EPA has identified in an Annex to each of the Appendices, it is still unclear how EPA weighs studies based on study quality. This is an area that EPA still needs to work on and add to it's ISA evaluations. After an assessment of study quality, which would help to identify methodological issues or biases in certain studies, EPA could develop criteria for excluding studies or at the very least for giving certain studies less weight in the overall conclusions and causal determinations.

v. *Is it clear that the process followed in selecting and summarizing scientific studies in the ISA was sufficient to assure accurate, unbiased, up-to-date, and trustworthy summaries of the relevant scientific literature to inform causal determination judgments?*

vi. *Do you find in the Executive Summary a clear explanation of the extent to which the key evidence supporting the ISA's causal determinations consists of, is sensitive to, or is derived from unverified modeling assumptions, or from modeling assumptions that more recent literature has found to be incorrect or inadequate? Have you found information in the ISA on sensitivity of causal determination conclusions to untested, uncertain, or incorrect assumptions? (If so, where? See Table Annex 6-1, cf p. 6-67 for a discussion of what should be done. Has it be done, and is it clear what the results were?)*

b. *Were the epidemiological studies used to support the causal determinations summarized in Table ES-1 (p. ES-5) and Figure ES-2 (p. ES-6) appropriately designed and analyzed to provide valid scientific information and valid causal conclusions about effects of possible future interventions (rather than just conclusions about historical statistical associations. For these observational studies, were criteria for valid study design and analysis for causal inference (specifically for interventional causation) explicitly stated, systematically applied, and the results transparently presented? (If so, where?)*

c. *Is it clear that the individual studies cited in support of the ISA's causal determinations of "causal" or "likely to be causal" adequately controlled for potential confounding and residual confounding by variables such as income and weather variables?*

I did not have the time to conduct an in-depth evaluation of whether EPA applied the criteria for study quality and causal determinations (e.g., Table Annex 6-1) appropriately or correctly. Generally, I found that study summaries for most of the health outcomes I reviewed lacked a clear discussion of the study strengths and weaknesses and there did not appear to be any discussion regarding study quality or how the EPA criteria were applied to individual studies, even for studies that were identified as being "key evidence." This is an area that could be significantly improved. Overall, I think that this limits the transparency of how EPA arrived at the causal conclusions. There are clearly many limitations for all the various studies evaluated and included in the ISA, and these limitations need to be more explicitly discusses and weighed.

I did not have the time or opportunity to review and answer additional questions posed by Dr. Cox, however, I think that some of the answers provided above, partially answer many of these additional queries. Overall, I think that EPA is going in the right direction with developing PECOS statements and providing guidelines for assessment of study quality. It falls short, however, in the implementation and in clearly and transparently showing how these criteria are applied to individual studies.

Questions from Dr. Mark Frampton

1. Change in causality determination for short-term cardiovascular effects since the 2013 ISA.

Question 1: Please comment on the strengths and weaknesses of the epidemiology literature with regard to CV effects of short-term ozone exposure. Are there key studies that are missing? Are the remaining weaknesses, along with the other new evidence, sufficient to justify the change in causality determination?

I believe that the change in causality classification is justified. I refer Dr. Frampton to several studies that were published by me and my colleagues, and the findings from these evaluations are more consistent with the change in the causality determination for CV disease. None of these studies were mentioned or included in the ozone ISA:

Prueitt, RL; Lynch, HN; Zu, K; Sax, SN; Venditti, FJ; Goodman, JE. 2014. "Weight-of-evidence Evaluation of Long-term Ozone Exposure and Cardiovascular Effects." *Crit. Rev. Toxicol.* 44(9):791-822.

Goodman, JE; Prueitt, RL; Sax, SN; Lynch, HN; Zu, K; Lemay, JC; King, JM; Venditti, FJ. 2014. "Weight-of-evidence Evaluation of Short-term Ozone Exposure and Cardiovascular Effects." *Crit. Rev. Toxicol.* 44(9):725-790.

Goodman, JE; Prueitt, RL; Sax, SN; Pizzurro, DM; Lynch, HN; Zu, K; Venditti, FJ. 2015. "Ozone Exposure and Systemic Biomarkers: Evaluation of Evidence for Adverse Cardiovascular Health Impacts." *Crit. Rev. Toxicol.* 45(5):412-452.

Petito Boyce, C; Goodman, JE; Sax, SN; Loftus, CT. 2015. "Providing Perspective for Interpreting Cardiovascular Mortality Risks Associated with Ozone Exposures." *Reg. Tox. Pharmacol.* 72(1):107-116.

2. Metabolic effects, new determination of "likely" for both short- and long-term exposure.

Question 2: Is there sufficient epidemiological evidence of metabolic effects to justify the "likely" determination for both short- and long-term exposures? Are there additional studies that should be considered?

I found the data to be quite limited for this large group of health outcomes, with a lot of overlap between many of these effects and effects that could potentially contribute to cardiovascular disease or even respiratory disease. I think that the large leap that EPA took to conclude that these effects are "likely to be causal" is very premature and the evidence (as presented by EPA) does not appear to justify the classification.

3. Change in causality determination for total mortality since the 2013 ISA.

Question 3: Please comment on the strengths and weaknesses of the epidemiology literature with regard to short-term ozone exposure and total mortality. Are there key studies that are missing? Does the available evidence justify the change in causality determination for total mortality?

EPA seems to be appropriately acknowledging many of the important limitations in the epidemiology literature for ozone and mortality, including potential confounding by other air pollutants, weather and temporal trends, large amounts of unexplained heterogeneity in ozone-mortality effect estimates, and potential for exposure measurement errors or modeling errors. Although total mortality appears to have consistent findings across studies, the cause-specific mortality results are inconsistent, largely null and very imprecise (Figure 6-2).

Questions from Dr. Sabine Lange

Epidemiology Study Questions

1. *It has been established that associations found in an epidemiology study can be due to: causation, bias, chance, and/or confounding. If the concept of statistical significance is not useful in epidemiology studies, then how do the study authors/EPA rule out that chance has caused the observed association?*

   Epidemiological studies cannot be the sole basis for establishing a causal relationship because of the inherent limitations and because, in the case of observational epidemiological studies, you cannot rule out bias, chance and/or confounding with reasonable confidence. The issue is even more difficult when the observed effects are very small and not statistically significant or marginally significant. Because of these limitations it is essential to evaluate all lines of scientific evidence, including experimental evidence (human chamber studies, animal studies, mechanistic studies). By evaluating the consistency and coherence within and across the various scientific lines of evidence, one can obtain a better picture of whether a causal association is more or less likely. More importantly, the evidence may be able to also elucidate levels of exposure at which effects are more likely.

2. *Am I correct in understanding that the intention of ozone case-crossover studies is to compare the ozone concentrations on a day when a health effect occurred for a person, to the ozone concentrations on a day when that health effect did not occur for that person?*

   I believe that is the correct interpretation.

3. *If so, then it would be important that some other factor (not related to ozone) did not prevent the health event from occurring on a control day. These studies often use days before and after the health event as control days, but for mortality studies (such as Di et al., 2017), how can a day after death be used as a control day? It doesn't matter what the ozone concentrations are after a person's death, that person would not be able to respond to that concentration. How should we interpret case-crossover studies that use control days after the event (particularly mortality) occurred?*

   I think that this is a valid question, as you described by selecting days post-health effect this would violate an important epidemiological tenant for assessing a causal relationship – that is, that the exposure must precede the effect.

Experimental Study and Dose Concordance Questions

4. *What is the importance of dose-concordance in establishing the biological likelihood of ozone-mediated effects occurring at relevant exposure concentrations in humans?*

   Particularly in the context of known dose information about ozone: total inhaled dose includes concentration, exposure time, and exercise duration; Hatch et al., (2013) have shown that humans and rats that are exposed to ozone at rest achieve similar alveolar ozone doses, and that humans exercising at 5-times a resting ventilation rate achieved an ~ 5-times higher alveolar

ozone dose; and that ozone concentrations are 2-10 times lower indoors where people spend most of their time.

I think this is a very important issue and one that has not been resolved or evaluated by EPA in weighing the evidence across different studies (i.e., human chamber studies and animal studies).

5. *Is there evidence that the animal models used to assess ozone effects (largely rats, mice, and non-human primates) are more, less, or similarly sensitive to ozone-mediated adverse effects compared to humans, at approximately equal inhaled doses?*

   Again, this is a valid question, and given the evidence as presented in the ISA it is difficult to answer. I think that EPA's assertion that some of the high exposure levels used in the animal studies (based on the Hatch et al., 2013 study) are relevant to ambient exposures in humans is likely to be simplistic at best, and a more detailed analysis to support an answer to this question is warranted.

Causality Question

6. *In the absence of a causality diagram to direct the choice of variables to control in an epidemiological study, how can we judge whether a study has appropriately controlled for confounders, and has not inappropriately controlled for colliders (which can open up pathways between variables that otherwise would not be connected) or mediators (and thereby controlled away the effect)?*

   I don't think this is a new issue and as noted above, this is a particularly important limitation of observational air pollution studies. The study summaries that EPA presents in the ISA fall short of identifying the various limitations in the epidemiological literature and this remains an area of weakness in the overall evaluation of ozone health effects. For determining plausible (but not necessarily absolute) causation, a full integration of all lines of evidence is necessary. As noted previously, relying on only epidemiological evidence is not sufficient.

Questions from Dr. Steven Packham

Question 1. When a causal relationship is conclusive to a high degree of scientific certainty as it is in this case, should this take precedence over causal inference when drafting a NAAQS ISA?

The human chamber studies indicate that for adults exercising for several hours and exposed to a certain concentration of ozone, small (but measurable) FEV1 decrements are observed, generally at levels above about 72 ppb. These effects are reversible and are likely not clinically significant (i.e., people may not even notice the changes in lung function). Some people do report subjective symptoms. As noted in our paper [see Goodman, JE; Prueitt, RL; Chandalia, J; Sax, SN. 2014. "Evaluation of adverse human lung function effects in controlled ozone exposure studies." *J. Appl. Toxicol.* 34(5):516-24], the results from chamber studies need to consider several areas of potential uncertainty. This includes the inter-individual variation in FEV1 measurements that can occur due to factors other than ozone exposures. For example, this intra-individual variation in FEV1 measurements can be up to about 5% and could explain some of the observed statistically significant lung function changes at lower ozone exposure levels. In addition, the level of exertion in these studies is an important consideration as the exercise regimen in most of these (i.e., 40 L min–1 for 6–8 h) is equivalent to work performed during a day of heavy manual labor common of outdoor workers. Therefore, the level of exertion and differential impact of exercise on study participants, in terms of its impact on FEV1 decrements, needs to be considered.

Question 2: Given evidence available from controlled human exposures substantiating causal relationships with a number of physiological responses, including beneficially confounding interactions of ozone on PM clearance, should Sub-section ES4.1 Health Effects in the Draft's Executive Summary, and the entire Integrated Synthesis section of the Draft be rewritten?

The points that Dr. Packham bring up are all valid and need to be incorporated in the discussions in the various relevant sections. While edits may be warranted, I don't think these sections necessarily need to be re-written.

Question 3: Looking ahead, do you think toxicology, clinical human studies, and biomedical research disciplines should be given more explicit and balanced consideration in the development of the present, and future, O3 ISAs with the objective to validate causal relationships and determine hourly inhalation dosage rates for adverse inflammatory responses in pulmonary tissues?

Toxicological, clinical (chamber) human exposure studies and biomedical (mechanistic) studies are included in the evaluation of health impacts in the ISA. In fact, as noted by Dr. Packham in the prior NAAQS assessment of ozone effects, human chamber studies played a key role in identifying the recommended level of exposure to ozone and lead to the revision of the NAAQS – in particular the study by Schelegle et al. (2009). However, there needs to be an important distinction between effects that may be small, transient and reversible (and likely not clinically significant) and effects that may be more serious. It is not possible to determine the level of exposure that would likely result in overt and irreversible harm using human chamber studies, as it would not be ethical. Therefore, the chamber studies must be interpreted within the context of other lines of evidence, including relevant exposures in animal studies and data regarding mode of action in mechanistic studies. For ozone exposures, it is particularly important (as Dr. Packham noted in the background for this question) to understand underlying defense mechanisms in the lungs that would be protective unless these mechanisms are overwhelmed (i.e., at high doses) and can no longer provide the necessary protection against oxidative harm. I also agree with his evaluation of cumulative dose and that EPA's figures summarizing these data

miss the mark with regards to cumulative dose over time. As noted above, these chamber studies involved healthy individuals exercising over several hours and this needs to be accounted for in the interpretation of the findings.

I like the idea of providing information to the public regarding levels that may results in some small reversible effects (that vary across individuals) and setting alerts that people can access on an app to obtain that information. Again, there does need to be a distinction between these benchmarks or guidelines and levels that would likely produce long-term adverse effects and this should also be communicated to the public.

Questions from Dr. James Boylan

Appendix 1 – Atmospheric Source, Chemistry, Meteorology, Trends, and Background

I have not reviewed this section is a lot of detail, but it appears that most of the section is dedicated to defining ozone background and reviewing recent studies and trends in background ozone concentrations. This is an important area of research and warrants detailed discussion. Consideration of background levels of ozone is important in making policy decisions regarding the NAAQS, particularly as the NAAQS is lowered to levels that approach background in many areas (e.g., the western US and areas influenced by wildfires).

I did find some EPA statements to be somewhat contradictory in its summary of the evidence, noting for example, on the one hand that background ozone concentrations can range from 20-50 ppb, and modeling error can be as large as 10 ppb, but then dismissing background ozone by noting that anthropogenic sources contribute to large proportions of measure ozone concentrations (at least on high ozone days). In addition, EPA concludes that the general trend of increasing background ozone from international sources has seen a slow down or reversal more recently. These statements imply that background ozone may not be or should not be given much importance. I would argue that the upper end of the background estimates is quite high relative to actual ambient levels in may areas.

Appendix 2 – Exposure to Ambient Ozone

I briefly reviewed the exposure section. While the section provided a good overview of exposure concepts and models, including new methods that are being use in epidemiological studies to refine exposure measurements, there were some things missing. In particular, I found that the section emphasized too much aspects of exposure measurement error in epidemiological studies and how they can be interpreted, without consideration of the poor correlations between indoor and ambient concentrations and personal and ambient ozone exposures given the time spent indoors. If available, EPA should not only present indoor-outdoor and personal-outdoor ratios, but an estimate of what exposures may actually be vis a vis ambient exposures. It seems to me that the issue of exposure measurement error is very different for ozone than other air pollutants (e.g., particulate matter), because of the impact of ozone scavenging on surfaces. That is, if the exposure is so small because of little penetration of ozone indoors and no indoor sources, can we conclude that ambient measurements – even if refined by modeling – are good surrogates of exposure for people spending most of their time indoors? This is an area that warrants more discussion and consideration than what is provided currently in the ISA. It may be that for large portions of the population, possibly excluding outdoor workers or people that exercise outdoors, ozone exposures may be very low or negligible, but this information is not included in this section and should be.

Appendix 9 – The Role of Tropospheric Ozone in Climate Effects

I am not sufficiently familiar with all of the climate change processes that could be influenced by ozone to appropriately comment on these questions. However, I think this is a new and important area of research and I am glad that EPA is including an assessment of these impacts in the ISA.

Questions from Dr. Corey Masuca

Appendix 1 Atmospheric Source, Chemistry, Meteorology, Trends, and Background Ozone
1.3.1 Precursor Sources

I am not sufficiently familiar with ozone chemistry an all precursors to provide a thorough answer.

1.3.1.2.1 Global Methane

Same as above

1.3.1.2.2 International Emissions of Ozone Precursors

This section focuses on international transport of ozone precursors.

What about local/state/regional transport of ozone precursors?

I agree that this is an equally important issue that warrants consideration, particularly if areas that are exceeding the NAAQS due to transport of precursors from other states.

1.3.1.3.2 Biogenic Volatile Organic Compounds (VOCs)

I am not familiar enough with the models to provide a comment to this question.

1.4 Ozone Photochemistry

With the advent of monitoring for speciated compounds including PAMS and Near-Road Monitoring (NOy), should there be further discussions about the individual chemicals gleaned from the specialized monitoring.

If the speciated chemicals are relevant to ozone formation, they should be included in discussions regarding the impact of these chemicals on ozone.

1.5 Inter-Annual Variability and Longer Term Trends in Meteorological Effects on Anthropogenic and US Background (USB) Ozone

I agree that topography (e.g., in Los Angeles) is an important consideration in the accumulation of air pollution. This should be included with examples. I am unsure about the independent effects due to relative humidity.

Appendix 2 Exposure to Ambient Ozone

2.3 Exposure Assessment Methods

While monitoring, including fixed, ambient monitors and personal and microenvironmental monitors are highlighted, what about remote sensing? Biological sampling in blood or tissue?

This is a good question, although I do not have much to contribute in terms of specific answers.

2.3.2.1 Spatial Interpolation

While attempting to quantity concentrations at locations and areas between concentration points is included under 2.3.2 Modeling, many of these exact same methods (i.e., data averaging, IDW, and kriging) are also utilized for Monitoring data shortcomings.

I agree – these are common methods used when data are limited in general.

2.4.1 Time-Activity Data

Is it possible that ozone exposure through time-activity data may be reduced due to temperature alone, as more people tend to avoid time spent outdoors in the summers during extremely warm/hot/humid, stagnant days which are oftentimes conditions for greater ozone formation?

Yes, I believe this is very possible. I have only seen data on people that may alter their behavior if they have asthma (or kids with asthma) on high pollution days. I have not seen data on activity patterns altered by temperature alone, although these data may be available.

Miscellaneous Question(s)

Due to exposure to ozone being disproportionate for disparate (i.e., lower income, children), should this be emphasis in this section, in lieu of regression analysis confounding/covariate in epidemiological studies for low(er) SES?

Yes – I agree that some discussion on the potential disparate exposures in low income communities or children would be warranted. Although unlike particulate matter, because ozone is formed over time from precursors, it is not always the case that lower income communities might have the highest exposures. EPA does discuss time-activity patterns in children that might contribute to higher exposures (i.e., because of greater time spent outdoors), but this could be made more explicit (e.g., are there any exposure data or actual exposure measurements to confirm this?).