Responses to CASAC Questions on the Ozone ISA from Consultant Dr. Duncan Thomas

Questions from Dr. Tony Cox

1. *Can valid determinations of manipulative or interventional causation … be made based on observed associations of the types analyzed in the ISA?*

Response: I discussed this question extensively in my earlier comments on the draft $PM_{2.5}$ PA. Briefly, there exist methods of causal inference designed to reanalyze observational epidemiology data as if it were from a randomized trial with the goal of estimating the Average Causal Effect of a hypothesized intervention. There have been a handful of such publications in the air pollution literature (see my previous citations), as well as a few "Accountability" studies based on real world "natural experiments" (work stoppages, introduction of new regulations, short term interventions around Olympic Games, etc.). However, the vast bulk of air pollution studies have not been designed or analyzed for the purpose of assessing manipulative or interventional causation. Nevertheless, the consistency of the findings from numerous observational studies, the concordance with short term human experimental studies (e.g., chamber or panel studies), and animal experiments, along with other lines of evidence supporting biological plausibility, as outlined in the preface to the ISA, allows a causal interpretation in terms of the likely effect of air pollution on the various health endpoints, if not a quantitative estimate of the predicted magnitude of the effect of a hypothetical intervention. See also my responses to some parts of question 2.

This point has been cogently discussed in a recent commentary by Carone, Dominici & Sheppard (2019), who conclude "In our view, causal inference methods should not be used as another opportunity to weaponize science against itself. Policymakers cannot wait for the data, study designs, and analytic tools that will ensure unarguable causal inferences: stalling until perfect evidence arises is irresponsible and does not protect public health." See also (Goldman and Dominici 2019): "a requirement of manipulative causation fails to recognize the full depth and robustness of existing approaches in epidemiology, statistics, and causal inference and the degree to which they deal with confounding factors."

2. *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. *Is this actually a "formal causal framework"?*

Response: That depends upon the meaning of the term in quotes. The approach used in the ISA does not exploit the emerging framework of "causal inference" that constitutes one type of "formal causal framework." However, the "weight of evidence" machinery (Committee_to_Review_the_IRIS_Process 2014) used here is certainly a well-established and appropriate formal framework for reaching causal judgements combining evidence across scientific disciplines. The machinery of statistical causal inference is not capable of or intended to synthesize evidence across multiple studies from multiple scientific disciplines.

Unlike Dr. Cox, I do not find the definitions in Table II of the Preface (and their application in Tables IS-4 and IS-5) to be "logically incoherent and ambiguous"; rather, the definitions are operational rather than conceptual, in line with those used by IARC and other scientific agencies, based on specific criteria for the types of evidence required to attain each category, e.g., for a determination of a "causal relationship" the following is required:

"… Generally, the determination is based on multiple studies conducted by multiple research groups, and evidence that is considered sufficient to infer a causal relationship is usually obtained from the joint consideration of many lines of evidence that reinforce each other. "

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

Response: see my response to question *h* below.

c. does a "causal relationship" determination imply a manipulative causal relationship?

Response: That depends upon the context in which the term "causal relationship" is used. In the statistical literature on causal inference, yes, the goal is to estimate the effect of an intervention on differences in the expected outcome within an individual under different hypothetical scenarios. In the epidemiological literature — and as used in the ISA — it refers to the existence of a mechanism under which exposure is a contributing factor, which may imply that a change in exposure would be expected to change the outcome, but that is not the primary sense in which the term is used. See also my response to question 2.*h*.

d. Can causal determinations be incorrect?

Response: Yes, of course, any human judgment could be incorrect. That is true of those reached by a large body of experts as well but is much less likely!

e. is it clear how uncertainty about which category is correct should be (or has been) resolved in assigning a final causal determination category?

Response: While the process for deciding upon which category is appropriate is clearly described, I do not see much if any discussion about any disagreements about the choice of category were resolved, except in terms of justification for changes in the categorization since the 2013 ISA (Tables IS-4 and IS-5). These explanations seem cogent to me.

f. is it clear how observations could be used to test and falsify a given causal determination if it is not correct?

Response: The question appears to ask whether "relevant data" not already considered, i.e., new studies, could falsify a conclusion in the ISA. While it is always possible that new data will emerge that leads one to question a previous determination, such speculation would be beyond the scope of the ISA.

g. is the correctness of each causal determination in table ES-1 formally and transparently evaluated in the ISA?

Response: I find that the evidence provided in section IS.4 and the supporting appendices to provide compelling support for the determinations in Table ES-1 and the supporting Appendices (to the extent that I have been able to read parts of it and to the extent of my epidemiologic expertise), and the process for reaching these judgments to be clearly described and transparent.

h. Does a determination that an exposure-response (or concentration-response (C-R)) relationship is a "causal relationship" imply that it is entirely causal,

Response: No, but that depends upon what is meant by "entirely causal." Epidemiologists have long recognized a "complex web of causation (MacMahon and Pugh 1970) meaning that no single factor is ever both necessary and sufficient to cause disease. A "causal relationship" is generally held to mean that a risk factor is a real component of one of the "sufficient component causes" of disease (Rothman 1976).

i. Does a determination that a C-R relationship is a "causal relationship" imply 100% certainty that it is causal?

Response: Obviously that depends upon the confidence with which that judgment has been reached. This seems like a semantic quibble.

j. Does a determination that a C-R relationship is a "causal relationship" imply that it is causal for every member of a population,

Response: not necessarily. More likely the magnitude of the effect will vary across subgroups of the population, but biology being essentially the same across all humans, it is likely that a causal association in the population at large will be true to some extent for any subgroup. Of course, it is possible that some subgroups will have no association at all: e.g., men are not likely to be at risk of ovarian cancer and those who lack a particular genotype that is essential for metabolism of a particular agent may be absolutely immune.

k. Are the five categories mutually exclusive?

Response: yes.

l. Are the five categories collectively exhaustive?

Response: yes.

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%?*

Response: Causal inference methods aim to estimate the "Average Causal Effect", not the probability of causality. The "Probability of Causation" (PC) is an estimate of the probability that a specific individual's disease was caused by some aspect of his exposure history, essentially an individualized version of the epidemiologic concept Population Attributable Risk Fraction. The PC has been frequently used in toxic tort litigation and setting guidelines for compensation policy, although it has come in for criticism, but is irrelevant for judging the causality of an observational association in populations.

3. *… are its conclusions derived by valid inference from true premises? Are the stated conclusions implied by the data and analyses used to support them? Are they consistent with other data and analyses that are at least as good as those selected? Are they appropriately caveated?*

a. Study selection and interpretation:

Response: The various questions for subsection 3(*a):i-vi* and (*b-g*) below require extensive substance matter knowledge of air pollution epidemiology and toxicology that are beyond my expertise. I have to defer to the EPA experts who drafted this ISA and the other consultants to

respond to the specifics about the selection and interpretation of specific studies and any omissions therein. However, I have added a few comments on specific questions about which I have some expertise.

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?*
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?*

Response: I agree that the Moore (2013) should be discussed. These two papers were among the few air pollution publications that used causal inference techniques, as I cited in my $PM_{2.5}$ draft PA comments, so the apparent difference between the conclusions of the two papers merits comment. It appears that the lack of significance from the later paper may be due in part from having to restrict to the 41% of the original 195 geographical grids for which the "experimental treatment assignment (ETA)" assumption was valid (i.e., that the probability of exposure assignment being above or below the 90 ppb threshold being analyzed was between 10% and 90%). The authors go on (following the sentence quoted by Dr. Cox) to state: "the fact that the CMRIER analysis does not provide significant results may be due to the lack of power to detect an effect with inverse weighting estimation. A more efficient estimation approach like TMLE estimation [10] could improve the estimation precision."

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

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)?*
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**?***

Response: Of those studies cited in support of these determinations that I am familiar with, the authors have gone to appropriate lengths to control for such confounders, to the extent possible with the available data. Of course, residual confounding can never be excluded from any observational epidemiology study.

d. *Is it clear that the individual studies cited in support of the ISA's causal determinations of "causal" or "likely to be causal" have adequately controlled for biases due to exposure estimation errors or exposure misclassification errors?*

Response: While the various studies differ in their methods of exposure estimation, few to my knowledge have used formal methods of measurement error correction. However, the bias from measurement error would generally be in the direction of reducing effect sizes and power, not introducing false positives. An exception would be for multi-pollutant analyses where it is possible for some effect of a better measured and causal pollutant to be transferred to the estimate for a worse measured noncausal one, as discussed at length in my response to the same question in my comments on the draft $PM_{2.5}$ PA.

e. *Do you find in the Executive Summary, or elsewhere in the ISA, a clear explanation of the extent to which the key evidence supporting the ISA's causal determinations is sensitive to uncontrolled or incompletely controlled confounding and/or ecological associations?*
f. *More generally, is it clear how criteria for individual study quality were applied to each study used in making causal determinations, and what the results were? (See Table Annex 6-1, cf p. 6-67.) Is it clear how the limitations of each individual study were taken into account in causally interpreting their reported associations and in making causal determinations?*
g. *Does the ISA make clear how its causal determinations would change if evidence from associations caused by confounding, residual confounding, measurement error, or unverified modeling assumptions were excluded?*

Response: This seems rather speculative, absent any evidence that those studies included have failed to adequately address these possible biases.

4. *Is the biological evidence presented in the ISA to support causal determinations correctly stated, correctly interpreted, relevant for predicting effects of changes in the ozone NAAQS, and up-to-date?*

Response: Yes, as explained in my responses to various other questions by Dr. Cox. (By "biological evidence" I assume you mean to include epidemiology, amongst the other lines of evidence, e.g., toxicology, which would be largely beyond my expertise.) See the following response in particular for the question about "*predicting effects of changes in the ozone NAAQS*".

5. *Does the biological evidence presented in the ISA provide well-validated scientific information suitable for predicting the effects on public health of changing NAAQS standard for ozone?*

Response: I believe the evidence presented in the ISA is suitable for reaching a causal interpretation of the effects of air pollution on human health. The ISA does not address the implications of potential changes in the NAAQS; it is my understanding that that will be addressed in the draft Policy Assessment document that I have not seen yet.

6. *Is each of the causal determinations summarized in Table ES-1 (especially those labeled "causal relationship" or "likely to be causal relationship") the only possible causal determination conclusion that is justified by, or consistent, with current scientific evidence? Could different causal determinations be equally well justified (or better justified) by the information presented, or by the totality of current scientific evidence?*

Response: Any judgment of a "causal" or "likely to be causal" relationship is potentially subject to differences of opinion amongst experts. It is my opinion that the various determinations summarized in Table ES-1 are well justified by the totality of the evidence, based on what I have read in sections IS.4, the supporting appendices, and my general background knowledge of the field of air pollution epidemiology. I do not claim to have read more than a portion of the ISA or to have an exhaustive knowledge of the substance matter of air pollution epidemiology, however. That said, I would consider it highly unlikely that any other conclusions would be "equally well justified" or "better justified" than those reached by the authors of the draft ISA.

7. *Are there changes in the design, analysis, selection, or interpretation of individual studies or in the ISA's processes for interpreting and summarizing them that would improve the validity, credibility, and transparency of the ISA's scientific reasoning and conclusions?*

Response: Obviously I would welcome wider application of the techniques of causal inference to observational studies, along the lines of those publications I cited in my response to the draft PM2.5 PA. That said, I believe that the weight of evidence approach used by EPA to evaluate the totality of the evidence, experimental and observational, to be highly appropriate and I have no further suggestions for improvement in that process.

Questions from Dr. Mark Frampton

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

Response: While I find the rationale provided in the ISA for this change in causality determination to be compelling, I do not have the comprehensive knowledge of the air pollution epidemiology literature to address whether there are key studies missing, other new evidence, or inadequately considered weaknesses.

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

Response: Same answer as for the previous question. From my limited experience collaborating with my epidemiologic colleagues on air pollution and obesity, metabolic syndrome, and related conditions, it certainly seems this is a "hot topic" and that "likely" is not unreasonable.

3. ***Change in causality determination for total mortality since the 2013 ISA.*** *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? Also please note that, for effects with causal or likely causal determination, the EPA has restricted consideration of epidemiological studies to those in North America (see PECOS Tool, section 6.1.1.1, page 6-3). That was the case for this determination. Are there epidemiological studies of mortality outside of North America that should be considered?*

Response: I was surprised by the reclassification of the association of short-term ozone exposure and mortality from "likely" to "suggestive," given the enormous body of time-series studies, including large multi-city studies (e.g., NMMAPS and APHEA) showing relationships with cause-specific and total mortality. There is no discussion of the rationale for this change in section ES-4 of the executive decision. Section IS-4.3.1 discusses this evidence only briefly in terms of respiratory mortality only and does not really provide a rationale for downgrading this association. Indeed, Table IS-4 highlights the abundant evidence support the "likely" classification in the 2013 ISA and then says for the 2019 one

> "Recent epidemiologic evidence for respiratory mortality is limited, but there remains evidence of consistent, positive associations, specifically in the summer months, with mean daily 8-h max ozone concentrations between **8.7 and 63 ppb.** When recent evidence is considered in the context of the larger number of studies evaluated in the 2013 Ozone ISA, there remains consistent evidence of an association between short-term ozone exposure and respiratory mortality." [emphasis added]

I have no idea why the EPA would have a policy of relying only on studies from North America. This does not make sense to me, given the availability of high-quality studies from Europe (e.g., the aforementioned APHEA) and elsewhere.

Questions from Dr. Corey Masuca

None of Dr. Masuca's question pertaining to Appendices 1 and 2 are within my area of expertise. I will confine my response to the following one.

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

Response: It is certainly true that exposure is disproportionately distributed, a serious concern known as "environmental justice." This also renders socioeconomic and other factors associated with exposure to be confounders requiring control for epidemiological associations. That ozone affects underprivileged communities disproportionately is worth pointing out in this section but does not alter the need to discuss the appropriateness of the methods used for confounder control. It is my opinion, based on those studies I have been involved in myself or have read in the literature, that the vast majority of the studies relied upon in the ISA have addressed this issue appropriately, to the extent possible with the available data.

Questions from Dr. Sabine Lange

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?*

Response: In addition to bias (of which confounding is one kind), chance can certainly lead to non-causal associations. Assessment of statistical significance is essential to judge the likelihood that an association could be due to chance, so it's incorrect to say that it "is not useful in epidemiology studies." Despite the longstanding and on-going debates about the usefulness specifically of *p*-values for this purpose (Greenland et al. 2016, Wasserstein and Lazar 2016), as opposed to a variety of other approaches (e.g., confidence intervals, Bayes Factors, etc.), they remain the most commonly used method for judging the possibility of chance. I do not see that the EPA has dismissed statistical

significance testing in its evaluation of the evidence, although they correctly do incorporate "trends in data and reproducibility of results" as well as other considerations in their evaluation of the epidemiologic evidence.

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?*

Response: Yes, that is the correct interpretation. An advantage of this design is that by making comparisons with an individual, between-individual confounding is completely eliminated, as are any factors that do not vary over time. While factors other than pollution that do vary over time, like weather, could still be confounders, these can be controlled in the analysis by standard statistical adjustment methods, as in case-control or time-series studies.

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?*

Response: The original case-control design (Maclure 1991) involved a comparison of exposure at the time of the event (or some pre-specified time prior to it to allow for lag effects) to that at some previous comparable ("referent") time. For example, the referent time might involve the same day of the week to control for systematic weekly variation in pollution levels and/or confounders. My colleague, Bill Navidi (1998) pointed out, however, that seasonal variation and especially long-term trends in pollution levels could lead to bias if referent times always preceded event times, even if one or more entire year cycles were included; while there would be no bias if there were no long-term trends and if pollution followed a perfectly symmetric (e.g., sinusoidal) seasonal pattern, departures from such symmetry, as are common for both pollution and meteorology, would lead to bias. Instead, he proposed the "bidirectional case-crossover" design, in which two referent times, one before and one after, equally spaced around the event time, are used. The original Macluer design was intended to study personal time-varying characteristics such as behaviors that could be "triggers" for an event like death or heart attack; in this setting, it would be impossible to observe a behavior that occurred after death! In air pollution studies, however, personal behaviors are not being studied, but ambient exposures are and these can be measured and used meaningfully for comparison. While it is obviously true that pollution *after* the event could not be causally related to the event, the purpose of this design is to get an unbiased estimate of the expected exposure at the time of the event for comparison with the actual exposure at that time and can be interpreted as a sampling-based analog of the standard time-series approach for acute effects (Bateson and Schwartz 1999, Fung et al. 2003, Lu and Zeger 2007). Various versions of this design have subsequently been widely adopted in air pollution studies. Although the original bidirectional design has subsequently been shown to be slightly biased (Lumley and Levy 2000), a modified version involving using fixed time-strata, comparing exposures at event times within each stratum with those at all or selected times (e.g., day-of-week matched times) within the same stratum before and after the event, has been shown to be unbiased (Levy et al. 2001a, Janes et al. 2005a, Janes et al. 2005b), and this design has become the standard in substantive studies (e.g., (Levy et al. 2001b, Di et al. 2017)). As Mittleman (2005) says, "this strategy should be considered the de facto standard approach to the analysis of data arising in studies of the short-term effect of air pollution and weather" (see also references therein for additional studies using this design).

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

Response: If by "dose-concordance" you mean comparability of doses to animals and humans from similar external concentrations and ventilation rates, I would expect that there are so many factors that differ that it would be unreasonable to expect the same dose-response relationships, even if doses could be scaled in comparable units.

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?*

Response: See my response to the previous question. Not being a toxicologist, however, these questions are largely beyond my expertise.

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)?*

Response: Very good question! Directed Acyclic Graphs (DAGs) can be useful tools for visualizing hypothetical relationships among observed and latent variables and for structuring an appropriate analysis strategy (Greenland et al. 1999). Investigators typically have such pictures in mind when conducting an analysis, although they are seldom presented formally in a substance matter publication (they are more commonly included in statistical methods papers). The basic principles that confounders must be controlled using the best available data on known risk factors (or surrogates for unmeasured factors in an attempt to minimize residual confounding), and that intermediate variables on a causal pathway from exposure to disease not be adjusted for, nor for colliders (that are determined by both exposure and disease but are not causal for disease), are well understood. The art is in deciding which variables are or are not appropriate to adjust for. While there are a variety of formal statistical methods for dealing with adjustment uncertainty (Maldonado and Greenland 1993, Greenland 1996, Viallefont et al. 2001, Crainiceanu et al. 2007, Pope and Burnett 2007), it remains a matter for expert judgment, both by the original investigators and by critical readers.

Questions for Non-Member Consultants on the Ozone ISA from Dr. Steven Packham

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?*

Response: Yes. In my opinion, while formal statistical causal inference methods have a useful role, particularly for evaluating the predicted effect of a hypothesized intervention, evaluation of whether a health effect of air pollution is causal requires a synthesis of evidence from multiple types of studies, which goes far beyond what these methods are capable of. I support the general weight of evidence framework used by EPA in this and other ISAs for this purpose.

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?*

Response: I see no mention of PM interactions with ozone in section ES4.1. To the extent that there is compelling evidence the ozone enhances PM clearance and mitigates its adverse effects (literature I am unfamiliar with), then it would seem appropriate to mention that here.

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?*

Response: Not being an expert in toxicology, clinical medicine, or biomedical research, all I can say is that I believe *all* these disciplines are relevant, as is my field of epidemiology, which in my opinion has the most direct relevance to human morbidity and mortality. Whether they deserve "*more* explicit and balanced consideration" implies that they are given inadequate consideration in the present and maybe future ISA, which I am not really qualified to answer. It does seem to me that the present draft ISA has attempted to assess all the relevant information from the various disciplines and incorporate them appropriately in their weight of evidence framework, at least with respect to the causality of the various relationships. I can't comment specifically on the comparability of dose rates between humans and model systems or their implications for pulmonary responses, other than to reiterate that there are many factors that differ among them other than dose that could make such comparisons dubious. Dr. Packham's comment that "ozone-induced FEV1 effects are temporary, reversible, and occur at a lower inhaled dose than a truly adverse health effect" sounds plausible, but beyond my expertise to critique.

Questions from Dr. James Boylan

His questions regarding Appendices 1 and 5 and the first few of those on Appendix 2 are beyond my expertise. I will, however, respond to the following two regarding Appendix 2:

- *Is the discussion on copollutant correlations and potential for confounding (Section 2.5) accurate and complete? If not, what additional information needs to be included?*
- *Is the discussion on interpreting exposure measurement error for use in epidemiology studies (Section 2.6) accurate and complete? If not, what additional information needs to be included?*

Response: I found both of these sections to be very clear, accurate, comprehensive, and well reasoned. I am not aware of any additional information that needs to be included. In particular, I agree with the conclusions in section 2.7 about the likely direction and magnitude of any biases introduced by co-pollutant correlations and measurement error. While section 2.6 could elaborate slightly on the available techniques for correcting for exposure measurement error (Thomas et al. 1993, Carroll et al. 2006), these have been applied only rarely in substantive epidemiology studies, so that would really be necessary.

REFERENCES

Bateson, T. F. and J. Schwartz (1999). "Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures." Epidemiology **10**(5): 539-544.

Carone, M., F. Dominici and L. Sheppard (2019). "In Pursuit of Evidence in Air Pollution Epidemiology: The Role of Causally Driven Data Science." Epidemiology **Publish Ahead of Print**.

Carroll, R. J., D. Ruppert, L. A. Stefanski and C. M. Crainiceanu (2006). Measurement Error in Nonlinear Models: A Modern Perspective (2nd Ed.). London, Chapman and Hall CRC Press.

Committee_to_Review_the_IRIS_Process (2014). Review of EPA's Integrated Risk Information System (IRIS) Process. Washington DC, National Academy of Sciences.

Crainiceanu, C. M., F. Dominici and G. Parmigiani (2007). "Adjustment uncertainty in effect estimation." Biometrika **95**(3): 635-651.

Di, Q., L. Dai, Y. Wang, A. Zanobetti, C. Choirat, J. D. Schwartz and F. Dominici (2017). "Association of Short-term Exposure to Air Pollution With Mortality in Older Adults." JAMA **318**(24): 2446-2456.

Fung, K. Y., D. Krewski, Y. Chen, R. Burnett and S. Cakmak (2003). "Comparison of time series and case-crossover analyses of air pollution and hospital admission data." Int J Epidemiol **32**(6): 1064-1070.

Goldman, G. T. and F. Dominici (2019). "Don't abandon evidence and process on air pollution policy." Science **363**(6434): 1398-1400.

Greenland, S. (1996). "Basic methods for sensitivity analysis of biases." Int J Epidemiol **25**(6): 1107-1116.

Greenland, S., J. Pearl and J. M. Robins (1999). "Causal diagrams for epidemiologic research." Epidemiology **10**(1): 37-48.

Greenland, S., S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman and D. G. Altman (2016). "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations." Eur J Epidemiol **31**(4): 337-350.

Janes, H., L. Sheppard and T. Lumley (2005a). "Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias." Epidemiology **16**(6): 717-726.

Janes, H., L. Sheppard and T. Lumley (2005b). "Overlap bias in the case-crossover design, with application to air pollution exposures." Stat Med **24**(2): 285-300.

Levy, D., T. Lumley, L. Sheppard, J. Kaufman and H. Checkoway (2001a). "Referent selection in case-crossover analyses of acute health effects of air pollution." Epidemiology **12**(2): 186-192.

Levy, D., L. Sheppard, H. Checkoway, J. Kaufman, T. Lumley, J. Koenig and D. Siscovick (2001b). "A case-crossover analysis of particulate matter air pollution and out-of-hospital primary cardiac arrest." Epidemiology **12**(2): 193-199.

Lu, Y. and S. L. Zeger (2007). "On the equivalence of case-crossover and time series methods in environmental epidemiology." Biostatistics **8**(2): 337-344.

Lumley, T. and D. Levy (2000). "Bias in the case – crossover design: implications for studies of air pollution." Environmetrics **11**(6): 689-704.

Maclure, M. (1991). "The case-crossover design: a method for studying transient effects on the risk of acute events." Am J Epidemiol **133**(2): 144-153.

MacMahon, B. and T. F. Pugh (1970). Epidemiology: principles and methods. Boston, Little, Brown and Co.

Maldonado, G. and S. Greenland (1993). "Simulation study of confounder-selection strategies." Am J Epidemiol **138**(11): 923-936.

Mittleman, M. A. (2005). "Optimal referent selection strategies in case-crossover studies: a settled issue." Epidemiology **16**(6): 715-716.

Navidi, W. (1998). "Bidirectional case-crossover designs for exposures with time trends." Biometrics **54**: 596-605.

Pope, C. A., 3rd and R. T. Burnett (2007). "Confounding in air pollution epidemiology: the broader context." Epidemiology **18**(4): 424-426; discussion 427-428.

Rothman, K. J. (1976). "Causes." American Journal of Epidemiology **104**(6): 587-592.

Thomas, D. C., D. Stram and J. Dwyer (1993). "Exposure measurement error: Influence on exposure-disease relationships and methods of correction." Ann Rev Publ Health **14**: 69-93.

Viallefont, V., A. E. Raftery and S. Richardson (2001). "Variable selection and Bayesian model averaging in case-control studies." Stat Med **20**(21): 3215-3230.

Wasserstein, R. L. and N. A. Lazar (2016). "The ASA Statement on p-Values: Context, Process, and Purpose." The American Statistician **70**(2): 129-133.