Public comments to CASAC

PM Policy Assessment Letter and Report

Lianne Sheppard, PhD

December 3, 2019

Oral comments to members of CASAC

Thank you for the opportunity to speak with you today.

I am a Professor of Biostatistics and Environmental and Occupational Health Sciences, and a former member of the chartered CASAC. I am speaking to you today to alert you to the fact that much of the advice to EPA in your draft PM Policy Assessment (PA) consensus response is not fit for purpose, actionable, balanced, or based on well-vetted credible scientific principles. I urge you to substantially revise this draft.

I base my judgment on my review of your draft coupled with the deep dive I have been taking into understanding the methods and data analysis examples advocated by Dr. Cox in his 2017 *Critical Reviews in Toxicology* paper titled "*Do causal concentration–response functions exist? A critical review of associational and causal relations between fine particulate matter and mortality*". First I want to thank Dr. Cox for being very responsive to my questions and requests for data to support my efforts to fully understand and replicate his work. Before I make specific comments, I urge each CASAC member to consider how well you understand the details of the methods the draft is advocating. If you find that your understanding is lacking, I encourage you to refuse to sign onto the advice. The causal inference concepts being advocated in your draft consensus response are technically unsound and if you sign onto this advice your scientific integrity will be forever bound up with them.

My remaining comments give specific examples of my three main points and then conclude with some recommendations.

<u>Not fit for purpose and actionable</u>:

- EPA reviews the scientific evidence from the published peer-reviewed literature. It does not conduct its own analyses of studies, yet much of CASAC's advice is telling EPA how to analyze data.
- The figure comparing C-R functions contains made-up data and is not appropriate.

<u>Not balanced</u>:

- The draft quotes some consultants yet ignores the advice of other consultants. In particular, there are multiple citations to comments by Dr. Aliferis and North, and no citations to comments by any of the other consultants (Drs. Jaffe, Jansen, Lipfert, Parrish, Rhomberg, Sax, or Thomas).
- The draft selectively concerns itself with potential false positive results in the literature (in its discussion of residual confounding, measurement error, model uncertainty, etc.), yet fails to acknowledge or address potential false negative results.

- While omitted confounding is always a potential concern in observational study analyses, the draft consensus response is overstating its presence in the air pollution literature. All responsible environmental epidemiologic analyses address confounding as thoroughly as possible. Published analyses in reputable journals have been further vetted during the peer review process. If there were serious problems with measured variables omitted from the analyses, they would have been detected in this process. In fact the only egregious example I have seen of omitted confounding in a peer-reviewed published paper is in Dr. Cox's analysis of monthly time series data in his 2017 paper.
- The draft consensus response has multiple examples of quote mining, which is a misleading use of the scientific literature by quoting text out of context and distorting the intent of the author(s).

<u>Not based on well-vetted credible scientific principles</u>:

- Appendix A is replete with technically unsound advice. If the causal methods were well vetted and broadly understood for application to air pollution studies, there would be many well-respected peer-reviewed papers published that implement these methods. This appendix should be removed; failing that, CASAC members who don't fully understand all of the technical points being made should be cautious about aligning their own scientific integrity with this text. Here are several key points; see my written comments for further detail:
    - The advice assumes that EPA's weight of evidence causal judgments should rely on inference in individual studies that meet the impossible-to-satisfy and technically unsound causal inference standards advocated by one CASAC member.
    - The advice assumes that causal effects can be learned from analysis of observational data. This is incorrect because it implies that causal inferences can be made without making untestable assumptions. It is impossible to make causal inferences from observational data without making untestable assumptions. See e.g., Carone et al 2019.
    - The advice incorrectly advocates that Bayesian networks can be used to learn causal relationships. This is incorrect. Only correlations can be learned from observational data, not causal relations, yet Bayesian networks are often used to encode causal relations. Furthermore edges in Bayesian networks do not encode conditional independence relations, and Bayesian network graphs are not reproducible because they vary by the algorithm and order of variables.

<u>Recommendations</u>:

- Remove all technically unsound advice, including all of Appendix A.
- Remove all irrelevant advice unrelated to the charge questions, including all text from report pages 2-8 (p 2 line 29 through page 8 line 37).
- Clearly articulate all areas of consensus and lack thereof, and spell these out in the letter.
- Add a statement consistent with CASAC's PM ISA comments that recognizes this CASAC does not have the breadth, depth, or diversity of expertise to make the judgments needed. The addition of consultants has not remedied this situation.

Written comments, Lianne Sheppard, PhD

December 3, 2019

Overarching comment

Much of CASAC's advice to EPA in its November draft letter and consensus response is not fit for purpose, actionable, balanced, or based on well-vetted credible scientific principles. I urge you to substantially revise it.

I urge each CASAC member to consider how well you understand the details of the methods you are advocating in this letter. If you find that your understanding is lacking, I encourage you to refuse to sign onto the advice. The causal inference concepts being advocated in your draft consensus response are technically unsound and if you sign onto this advice your scientific integrity will be forever bound up with them. I believe this is because many key recommendations are driven by the research and publications of Tony Cox. As I have begun to look into his work to assess it critically, I have seen a pattern of repetitive arguments that are not bolstered by sound causal inference theory, rather they are accompanied by unstated (and unverified) assumptions, and illustrated using implausible scenarios. I believe his peer-reviewed published papers have been poorly vetted, given the nature of the journals where they have been published, yet this point of view is dominating CASAC's advice.

Comments on CASAC's consensus comments on Chapter 1

Much of CASAC's advice on this Chapter is irrelevant and/or inappropriate. I provide specifics below.

Much of CASAC's advice is irrelevant to the purpose of Chapter 1

- Most of the discussion of this chapter is irrelevant to the chapter and the charge question. It should be removed. This includes the sections titled "Improving Policy Relevance" (p 2, l 29), "Improving Broad Understandability" (p 5, l 1), and "Facilitating CASAC Advice and Recommendations to the Administrator" (p 6, l 1).
- CASAC's scientific and technical approach does not belong in Chapter 1 of the PA. This is neither legislative background nor history. The lines beginning on p 1 line 31 and ending on p 2 line 9 should be removed.

Much of CASAC's advice on Chapter 1 is inappropriate

- CASAC's claim (d) (line 27, p 1) that it is emphasizing "sound science throughout the review process, including reexamination of long-standing assumptions and frameworks used in previous reviews" fails to understand or recognize the work done by previous CASACs, inappropriately elevates un-vetted scientific concepts as valid and appropriate, and inappropriately appreciates CASAC's proper role and responsibilities in the context of the Clean Air Act mandate.
- The draft consensus response's recommendations for revising the PA (p 2 lines 16-27) are so broad and unspecific that they are impossible to address; they don't belong in the discussion of Chapter 1.
- In advising EPA to include "omitted evidence" (p 2 l 37), CASAC fails to recognize the timeframe (i.e. end date) for the peer-reviewed literature that EPA included in the PM review.

- It is inappropriate for CASAC to argue that causal inference methods must be used by EPA and patently untrue that they are "readily available" for application in this setting (line 33, page 3). Considerable scientific advancement of causal inference methodology is needed in the realm of observational data, which encompasses the bulk of the air pollution health effects literature. (See e.g. Carone et al 2019 for a discussion of the state of causal inference for air pollution epidemiology.)
- It is incorrect for CASAC to claim that "unsound technical methods are used" (p 3 line 41). While it is correct that the risk assessment uses estimates of association to illustrate population risks, this does not make the risk assessment technically unsound. While it is fair for CASAC to request that EPA acknowledge explicitly that estimates of association are the basis for the effect estimates in the risk assessment, it is not appropriate to conclude that the entire exercise is invalid. The risk assessment provides useful perspective to complement EPA's evaluation of the scientific evidence, which is alone more than sufficient to justify that the current standard should be reduced. The draft consensus response falsely claims that "none of the key studies controls for potential confounders …. using appropriate methods", and incorrectly cites Pearl's body of work as providing justification for this statement. Pearl's book (2009) does not address causal inference from observational data, which is the type of data being considered in the risk assessment.
- In claiming the "causal determination conceptual framework [is] not validated for application to PM2.5" (p 4 l 13+), the draft is implying that there is a way to validate a causal framework empirically. This advice effectively assumes that one doesn't need to make untestable assumptions to draw causal conclusions from observational data. This is incorrect. Similarly, the draft consensus response's demand that relationships be "empirically validated" and avoid "unverified modeling assumptions" (p 4, l 27+), is another example where it is effectively assuming that one can draw causal conclusions without making untestable assumptions. Similarly, by demanding "empirically verifiable" and "empirically validated" actions (p 6, points (2) and (3)), the draft consensus response is making the same incorrect assumption that it is possible to draw causal conclusions from observational data without making untestable assumptions.
- The draft consensus response is conflating a weight of evidence causal determination with causal inference methods for application to individual studies (e.g. in its advice beginning p 5 l 8). Furthermore, in contrast to CASAC's perspective, in drawing its weight of evidence causal judgments, EPA is in fact using "causal" in the usual common knowledge sense that reducing PM exposure will indeed reduce health consequences. This determination is based on judgment taking into account multiple lines of evidence from multiple different studies with different designs, strengths and weaknesses. The fact that many of the studies that inform this judgment estimate association and do their best to rule out residual confounding and other biases, does not change the definition or interpretation of EPA's weight of evidence causal determination framework.
- The draft consensus response should refrain from claiming that published results reviewed by EPA can be explained by historical trends (e.g., p 6 l 41). The only peer-reviewed published study I have seen that reported a time series analysis that omitted any control for historical time trends is Dr. Cox's 2017 paper in *Critical Reviews in Toxicology*. (Note: this analysis of 14 years of data from Boston was conducted on the monthly time scale; this is an

additional problem.) Papers with such an omission do not get published in quality journals as they get rejected during the peer review process.

- The draft consensus response's arguments about specific problems such as omitted or incompletely controlled confounding, and model choices are motivated by a time series design (e.g., p 6 l 41, Appendix A). Not only is this a much weaker study design than a cohort design where EPA based most of its weight in its recommendations, but the arguments imply that the solutions to the concerns raised are simple. Many of the solutions, such as incorporating daily minimum temperature or month, make absolutely no sense in the cohort study setting.
- The draft consensus response includes multiple examples of quote mining, i.e. it quotes many reputable papers out of context and presents a different perspective than the point being made by the authors and evident from reading the full paper. Examples include quotes from Fuentes et al (2009) (p 7 l 3), Dominici et al (2014) (p 7 l 6), Carone et al (2019) (p 13 l 10).

Comments on CASAC's Appendix A "Possible Technical Options"

The advice given in this appendix is not viable and should be removed. I urge all CASAC members who don't fully agree with some or all of the advice in this appendix to decline to go along with it. I urge all CASAC members who don't understand some or all of the advice in this appendix to decline to go along with it. Signing onto advice you don't fully agree with and/or understand aligns your personal integrity with this perspective. Here are some reasons to be concerned with the content of this Appendix:

- The advice assumes that EPA does their own analyses of data it reviews from published studies. This is incorrect.
- The advice assumes that EPA's weight of evidence causal judgments should rely on inference in individual studies that meet impossible-to-satisfy and technically unsound causal inference standards advocated by one CASAC member. While this CASAC member's methodological approach has been published, it has not been published in the mainstream causal inference literature and it has not been validated by any thought leaders in causal inference methodology. It is inappropriate for CASAC to be advocating major changes in the approach to the PA based on methods that can be demonstrated to be technically unsound and have not be adequately vetted by causal inference experts.
- The advice assumes that causal effects can be learned from analysis of observational data. This is incorrect because it implies that causal inferences can be made without making untestable assumptions. It is impossible to make causal inferences from observational data without making untestable assumptions. See e.g., Carone et al 2019.
- The advice implicitly assumes that all studies of interest to the PA use the time series design. Multiple arguments and examples presented only address time-varying exposures and outcomes as can be seen on p A-1 lines 16-18, line 25, line 38; p A-2 line 23, line 38. This is incorrect.
- The advice incorrectly advocates that Bayesian networks can be used to learn causal relationships. (p A-1 l 20+) This is incorrect. As Pearl states in his book (2009), only correlations (dependencies or conditional dependencies) can be learned from observational data; causal

relations cannot be learned from observational data, and Bayesian networks are often used to encode causal relations. It is incorrect to equate the edges of a Bayesian network graph with conditional dependencies because edges in Bayesian networks do not encode conditional independence relations. Furthermore, use of Bayesian networks will not produce reproducible causal graphs. Different learning algorithms will produce different graphs and when these graphs are (incorrectly) applied to infer causal relationships between variables, will lead to different conclusions from the same data. Furthermore, another indication that these graphs are not reproducible, different graphs can be obtained by offering any single algorithm the same set of variables but in a different order.

- See the public comments provided by Helen Suh and co-authors for discussion of the inappropriate interpretation of Pun et al (2017) and Eum et al (2018). (p A-2 l 25+)
- The advice makes the argument that exposure measurement error in air pollution epidemiology studies is differential. It provides absolutely no justification for this specious claim. (p A-3 bullet starting l 1)
- The "relatively quick fix" is vague and not actionable. (p A-3 bullet starting l 7)
- The advice is mixing the concepts of model uncertainty and heterogeneity of effect estimates across locations for the same model. (p A -3 line 21 to end of section) In discussion of the Fuentes et al (2009) table included in the document, the draft consensus response fails to appreciate that these authors were demonstrating how, given the same model, different estimation approaches leverage the combined (overall) effect differently in the city-specific estimates. Furthermore, the draft consensus response incorrectly states the effect estimate for Richmond ranges from -2.24 to 2.76 across four modeling approaches. However the latter number is a SE estimate, not an effect estimate; it should be replaced by .72, the effect estimate in the adjusted model. Also, the text fails to acknowledge that 1) for the "local" estimates column all city-specific estimates are consistent with the pooled estimate, and all four estimation methods give essentially identical pooled estimates and SEs. This Fuentes et al (2009) paper does not provide a demonstration of model uncertainty, which is about using distinctly different models for the same set of data, rather this paper addresses how different estimation methods for the same models vary in how they leverage between-city information.
- The entire thrust of Appendix A is to argue from multiple perspectives that the estimates used by EPA should be adjusted so they are attenuated towards the null and their uncertainty estimates increased. This technical advice makes the implicit assumption that the myriad studies reviewed by EPA represent false positives. Furthermore, it gives absolutely no attention to the possibility that false negatives are also possible. To me this suggests a clear lack of balance in CASAC's technical advice.

Comments on CASAC's consensus comments for Chapter 3

While there are many more concerns to address in the Chapter 3 comments, I call out a few below:

- The draft consensus response's criticism of EPA's quantitative risk assessment quotes Carone et al (2019) (p 13 l 10+). This quote fails to recognize one of the key points of that commentary, namely that no causal conclusions can be reached from observational data without making untestable (i.e., unverifiable) causal assumptions. Furthermore, that commentary states "we do not expect the emergence of a methodological silver bullet since many of the challenges of

drawing valid causal inferences about air pollution health effects stem from inherent limitations of the observational nature of the available data", meaning that causal inference tools are not available or straightforward to apply to air pollution studies, as is implied in much of the draft consensus response, and directly stated by Dr. Cox in his individual comments (p B-16 l 9+). The draft consensus response also fails to acknowledge that its misrepresentation of the Carone et al commentary completely contradicts the commentary's conclusion, namely that "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." Finally, the text following the quote (p 13 l 20+) misrepresents Pearl's work since his 2009 book did not address observational data (as I assume is also true in the 2010 journal article cited since it is derived from Pearl's 2009 book), a setting where it is impossible to draw causal conclusions without making unverifiable conclusions.

- The draft consensus response is selectively highlighting a few comments from consultants that some committee members agree with (e.g., p 13, 14), while ignoring the entirety of the consultants' advice. This is not balanced and also exhibits why the use of consultants is counterproductive to Congress' vision for CASAC in the Clean Air Act. I note that in contrast to the references made in the draft consensus response, several consultants made it clear that EPA's causal framework is reasonable and that scientific advice to EPA should not demand approaches that current methods are unable to support. For example:
    - Dr. Aliferis states: "The causality analysis framework currently used by the EPA is, in my assessment, a reasonable framework given the limitations of causal discovery methodologies available until recently, the fact that it is very hard to conduct complex manipulations (policy) involving the environment, and that cause-effect horizons are lengthy."
    - Dr Rhomberg states: "All this is an unavoidable aspect of observational studies. All one can ask for is that the asserted functional structure of the model is reasonable in view of our broader understanding of what kinds of causative processes might operate (or be of interest), that the simplifications of the actual causal nexus are not obviously misleading or ignoring known or plausible important influences, and (to speak directly to Dr. Cox's questions) that we do not over-interpret the functional forms as actual measures of causal impacts without considering the pitfalls and limits to such interpretation." And "It would not be prudent (or consistent with the CAA mandates) to reject a causal effect simply because it cannot be unambiguously measured and characterized. There is no basis to default to a conclusion of no causation unless a rigorous demonstration of its particulars is at hand. We have overall patterns of outcomes in many studies that certainly suggest causative roles for PM in risks of health impacts, and these patterns are repeated with some level of consistency across studies."
    - Dr. Thomas states: "This is not to disparage the bulk of epidemiologic studies, only to emphasize that it would be inappropriate to dismiss them as not addressing causation, given their concordance and the general conformity with the criteria used by epidemiologists for decades to qualitatively evaluate causation." And "In general, I find both the draft PA and the draft ISA to be well written, authoritative, and comprehensive reviews of the literature and thoughtful discussion of the policy implications, including

limitations thereof", and "…the authors have done an excellent job of addressing these uncertainties and their policy implications."

- P 16 lines 9-25: This argument completely overlooks the analyses conducted by Di et al (2017) which were restricted to exposures below 12 ug/m3. Quoting directly from that paper's abstract, "Increases of 10 µg per cubic meter in PM2.5 and of 10 ppb in ozone were associated with increases in all-cause mortality of 7.3% (95% confidence interval [CI], 7.1 to 7.5) and 1.1% (95% CI, 1.0 to 1.2), respectively. When the analysis was restricted to person-years with exposure to PM2.5 of less than 12 µg per cubic meter and ozone of less than 50 ppb, the same increases in PM2.5 and ozone were associated with increases in the risk of death of 13.6% (95% CI, 13.1 to 14.1) and 1.0% (95% CI, 0.9 to 1.1), respectively." Furthermore, *it is not appropriate for the draft consensus response to include made-up data*. One can make up data to show any point; the importance for the Policy Assessment is to address what the evidence say, as reported by Di et al 2017 and other cohort study analyses. The point made in this paragraph of the draft consensus response also ignores the substantial weight of the findings given the huge populations covered in the recent Canadian and US cohort studies.

References

Carone, M., Dominici, F., & Sheppard, L. (2019). In Pursuit of Evidence in Air Pollution Epidemiology. 23 Epidemiology, 1. https://doi.org/10.1097/ede.0000000000001090

Dominici, F., Greenstone, M., & Sunstein, C. R. (2014). Particulate Matter Matters. Science, 344(6181), 34 257–259. https://doi.org/10.1126/science.1247348

Fuentes, M. (2009). Statistical issues in health impact assessment at the state and local levels. Air 11 Quality, Atmosphere & Health, 2(1), 47–55. https://doi.org/10.1007/s11869-009-0033-3

Pearl, J (2009) *Causality: Models, Reasoning, and Inference*, NewYork: Cambridge University Press, second edition.