Preliminary Comments from Members of the CASAC on EPA's *Policy Assessment for Review of the National Ambient Air Quality Standards for Particulate Matter – (External Review Draft – September 2019)*

Received as of 10-21-19

Dr. James Boylan

Chapter 2 – PM Air Quality

To what extent does the CASAC find that the information in Chapter 2 is clearly presented and that it provides useful context for the review?

This Chapter discusses particle size distribution, PM emissions, ambient PM monitoring methods and networks, trends in ambient air concentrations, hybrid $PM_{2.5}$ modeling approaches, and background PM. Overall, the information in this chapter is clearly presented and provides useful context for the review. However, there are a few areas that should be expanded to provide additional context for the review.

Sources of PM Emissions

This chapter presents estimated national values for 2014 NEI emissions (e.g., 5.4 million tons of $PM_{2.5}$ (page 2-4); 13 million tons PM_{10} (page 2-6); 1.5 million tons of particulate OC (page 2-7); 431,000 tons of particulate EC (page 2-8); 4.8 million tons of SO_2 (page 2-9); 14.4 millions tons of NOx (page 2-9); 3.6 million tons NH_3 (page 2-9), and 17 million tons VOCs (page 2-9). However, there is no detailed discussion on the uncertainty associated with each pollutant or source sector. Some pollutants and sectors will be much more or much less uncertain than others. For example, SO_2 emissions from electric generating units (EGUs) have low uncertainty since they are typically captured by hourly CEMs. On the other hand, Figure 2-3 shows that dust accounts for 47% of the national PM_{10} emissions and page 2-6 states "quantification of dust emissions is highly uncertain". Therefore, the national PM_{10} emissions of 13 million tons must also be highly uncertain. The uncertainties in the emissions inventory (magnitude, spatial allocation, temporal allocation, and speciation) should be discussed for each pollutant and source sector. In addition, the impact of emission inventory uncertainty on the air quality modeling results and the risk-based analysis presented in Chapter 3 should be discussed. Finally, it would be helpful to add national maps containing county-level emissions for $PM_{2.5}$, PM_{10}, OC, EC, SO_2, NOx, NH_3, and VOCs to show the variability across the country.

Ambient PM Monitoring

This section discusses FRM, continuous FEM, CSN, and IMPROVE monitors. However, there is no discussion on measurement uncertainty (accuracy, precision, bias, and error) associated with these monitors. In Georgia, FEMs typically show higher $PM_{2.5}$ concentrations compared to FRMs. EPA should compare co-located FEMs/FRMs across the country and summarize the results. This section should discuss how differing $PM_{2.5}$ biases associated with FRM, continuous FEM, CSN, and IMPROVE measurements would impact the evidence-based and risk-based $PM_{2.5}$ assessments in Chapter 3?

The section titled "Additional PM Measurements and Metrics" should include a discussion of the Southeastern Aerosol Research and Characterization Study (SEARCH) network which started in 1999 and included measurements of continuous $PM_{2.5}$, continuous OC, continuous EC, continuous sulfate, continuous nitrate, and continuous ammonium at eight sites in the Southeastern U.S.

- D. Alan Hansen , Eric S. Edgerton , Benjamin E. Hartsell , John J. Jansen , Navaneethakrishnan Kandasamy , George M. Hidy & Charles L. Blanchard (2003) The Southeastern Aerosol Research and Characterization Study: Part 1—Overview, Journal of the Air & Waste Management Association, 53:12, 1460-1471, DOI: 10.1080/10473289.2003.1046631
- Eric S. Edgerton , Benjamin E. Hartsell , Rick D. Saylor , John J. Jansen , D. Alan Hansen & George M. Hidy (2005) The Southeastern Aerosol Research and Characterization Study: Part II. Filter-Based Measurements of Fine and Coarse Particulate Matter Mass and Composition, Journal of the Air & Waste Management Association, 55:10, 1527-1542, DOI: 10.1080/10473289.2005.10464744
- Eric S. Edgerton , Benjamin E. Hartsell , Rick D. Saylor , John J. Jansen , D. Alan Hansen & George M. Hidy (2006) The Southeastern Aerosol Research and Characterization Study, Part 3: Continuous Measurements of Fine Particulate Matter Mass and Composition, Journal of the Air & Waste Management Association, 56:9, 1325-1341, DOI: 10.1080/10473289.2006.10464585
- C.L. Blanchard , G.M. Hidy , S. Tanenbaum , E.S. Edgerton & B.E. Hartsell (2013) The Southeastern Aerosol Research and Characterization (SEARCH) study: Temporal trends in gas and PM concentrations and composition, 1999–2010, Journal of the Air & Waste Management Association, 63:3, 247-259, DOI: 10.1080/10962247.2012.748523

<u>Trends in Ambient Air Concentrations</u>

Page 2-25 states "The regions that cluster outside of the typical annual/daily design value ratio line in Figure 2-11 are the Southeast and Northwest U.S. In the Southeast U.S., the annual design values are high relative to the daily design values due to the lack of seasonality in the concentrations and infrequent impacts of episodic events like wildfire or dust storms". I did not see the typical annual/daily design value ratio line in Figure 2-11. Please add this line. Since the Southeast lacks seasonality in the concentrations and has infrequent impacts of episodic events like wildfire and dust storms, it would seem that the Southeast should represent "typical" annual/daily design value ratios and should not be considered an outlier.

Also, the concept of "urban increment" should be discussed and a couple of examples comparing $PM_{2.5}$ speciation in urban areas to nearby Class I areas should be included. Typically, this analysis will show similar levels of sulfate, but higher levels of OC and EC in the urban areas due to mobile sources.

Background PM

The section on background PM adequately covers annual background concentrations but does not adequately discuss background concentrations for the daily $PM_{2.5}$ standard. This discussion should be added to the chapter. In addition, this section should discuss ^{14}C research to discern fossil-derived carbon from "modern" carbon and implications for background OC.

- Bret A. Schichtel, William C. Malm, Graham Bench, Stewart Fallon, Charles E. McDade, Judith C. Chow, and John G. Watson, Fossil and contemporary fine particulate carbon fractions at 12 rural and urban sites in the United States, Journal of Geophysical Research, 113 (2008), doi:10.1029/2007JD008605
- Roger L. Tanner, William J. Parkhurst, and Ann P. McNichol, Fossil Sources of Ambient Aerosol Carbon Based on 14C Measurements, Aerosol Science and Technology (2010), DOI: 10.1080/02786820390229453

Chapter 3 – Review of the Primary $PM_{2.5}$ Standards

What are the CASAC views on the approaches described in chapter 3 to considering the $PM_{2.5}$ health effects evidence and the risk assessment in order to inform preliminary conclusions on the primary $PM_{2.5}$ standards? What are the CASAC views regarding the rationales supporting the preliminary conclusions on the current and potential alternative primary $PM_{2.5}$ standards?

It is stated in Appendix C (page C-2):

"In selecting specific CR functions for the risk assessment, we focus on health outcomes for which the draft PM ISA determines the evidence supports either a "causal" or a "likely to be causal" relationship with short- or long-term $PM_{2.5}$ exposures (U.S. EPA, 2018a). As discussed in Chapter 3 of this draft PA, these outcomes include the following:

- mortality (resulting from long- and short-term exposure),
- cardiovascular effects (resulting from long- and short-term exposure),
- respiratory effects (resulting from long- and short-term exposure),
- cancer (resulting from long-term exposure), and
- nervous system effects (resulting from long-term exposure) in the draft ISA Table 3-1
- (U.S. EPA, 2018a).

We have focused the analysis on short- and long-term PM exposure-related mortality, reflecting its clear public health importance, the large number of epidemiologic studies available for consideration, and the broad availability of baseline incidence data."

It is unclear why cardiovascular and respiratory effects are ignored in the risk assessment. Including these two additional endpoints would provide useful information to help inform the decision on the level of the $PM_{2.5}$ NAAQS.

The risk assessment will only focus on selected CBSAs. Therefore, the CMAQ model performance should be evaluated at these selected locations. Similar model performance statistics shown in Table C-6 should be developed for each CBSA listed in Table C-3 (number of $PM_{2.5}$ monitoring sites range from 1 to 22 at each CBSA). In addition, NMB bubble plots (small circle over the location of each monitor with the color inside the circle representing a range of NMB values) including each monitor included in the risk assessment should be developed for annual performance and each season of the year. Modeling uncertainty should be quantified and incorporated into the risk assessment. CBSAs with poor model performance should be excluded from the risk assessment.

Given the overlapping confidence bounds, EPA should evalaute if there is a statistically different risk between meeting the current NAAQS and alternative NAAQS.

Chapters 3 to 5

What are the CASAC views regarding the areas for additional research identified in Chapters 3, 4 and 5? Are there additional areas that should be highlighted?

Additional ambient monitors should be required to measure ultrafine particles across the country. In addition, emissions from various sources should be measured to understand the contributions to measured ultrafine particles in the ambient air.

Dr. Tony Cox

Response to Charge Questions for Chapter 1

*"**Chapter 1 – Introduction:** To what extent does the CASAC find that the information in Chapter 1 is clearly presented and that it provides useful context for the review?"*

Preliminary Draft Answer: The history and legislative background are clearly presented and they provide useful context for the review. The stated intentions for the PA presented in Chapter 1 (including "to facilitate advice to the Agency and recommendations to the Administrator" from the CASAC; "to serve as a source of policy-relevant information;" and "to be understandable to a broad audience") remain largely unfulfilled. As discussed next, this is due to limitations arising from a) failure to apply principles of sound science and risk communication needed to obtain valid conclusions and to express them clearly; and (b) failure to thoroughly consider recent and causally relevant scientific evidence in addressing policy-relevant questions about whether and to what extent changing exposure would change public health risks. These points are developed in the following comments.

Comments on Chapter 1

p. 1-2 *"The PA is also intended to facilitate advice to the Agency and recommendations to the Administrator from an independent scientific review committee, the Clean Air Scientific Advisory Committee (CASAC), as provided for in the Clean Air Act (CAA)."*

To be most useful in facilitating the CASAC's advice to the Agency and recommendations to the Administrator, the PA should address the following questions using transparent derivations of conclusions from currently available scientific evidence:

1. *What are the valid implications of recent scientific evidence for whether current NAAQS must be revised to protect human health with an adequate margin of safety?* On the one hand, the current draft PA is limited by its omission of much recent and relevant scientific information, e.g., from intervention and accountability studies (Burns J et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst Rev. 2019 May 20; Henneman LR et al. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. J Air Waste Manag Assoc. 2017 Feb;67(2):144-172). Overall, these studies do not clearly reject the null hypothesis of no change in human health risks caused by reductions in PM exposures. On the other hand, the PA misinterprets significant positive C-R associations (see Table C-1) in studies that do not fully control for confounders (e.g.,

by temperature), coincident historical trends (both exposures and health effects decreasing over time), exposure estimation errors, modeling errors, and other potential non-causal sources of positive C-R association, as showing that reducing exposure would reduce health risks. This causal interpretation is not scientifically or logically valid, since other likely explanations for the positive associations have not been systematically tested or ruled out (Campbell DT and Stanley JC (1963), *Experimental and Quasi-Experimental Designs for Research*). Hence, the PA does not provide valid answers to the crucial policy-relevant question of what implications follow from currently available evidence about whether or how much reducing current exposures would cause changes in public health risks.

2. *How would the simulated results in the PA change if the C-R functions used to estimate health impacts (Appendix C) were revised to remove effects of confounders* (such as same-day and lagged daily high and low temperatures and humidity)? More generally, how would conditioning on appropriate adjustment sets change the estimated C-R functions and resulting risk predictions?
3. *How would correcting for errors in estimated exposures change the estimated C-R functions and resulting simulation results?*
4. *How would the simulated results in the PA change if the C-R functions were revised to reflect more recent literature from outside the work of the authors in Table C-1?* The studies relied on in the PA to simulate health impacts of reducing PM2.5 (Tables 3-4 and C-1) are largely the work of a relatively small cluster of co-authors expressing similar conclusions across multiple studies. These studies share common methodological limitations, such as not being designed or analyzed to support valid causal inferences (e.g., failing to control for confounding by lagged daily high and low temperatures), and not distinguishing between estimated and actual exposures levels. For example, Jerrett, Pope, Turner, and Thurston (the lead authors of the first 4 studies cited) are all co-authors of the first study (Jerrett). Some other recent studies by different authors, such as the intervention and accountability studies reviewed by Burns et al. (2019) and Henneman et al. (2017)) (*op cit.*) reach very different conclusions from those in the studies selected by the EPA (Tables 3-4 and C-1). The draft ISA and PA omit most of these studies. This raises the question: How would including more of the recent, relevant, high-quality scientific literature change the C-R associations and simulation results presented in the draft PA?
5. *What alternative underlying interpretations are consistent with the scientific evidence? To what extent do they support alternative standards?* Testing (and refuting, if possible) alternative explanatory hypotheses is widely considered essential for valid causal inference; see e.g., Campbell DT and Stanley JC (1963), *Experimental and Quasi-Experimental Designs for Research*. Explicitly assessing alternative interpretations of evidence is an important part of sound science. As stated by Feynman, "Details that could throw doubt on your interpretation must be given, if you know them. You must do the best you can—if you know anything at all wrong, or possibly wrong—to explain it. If

you make a theory, for example, and advertise it, or put it out, then you must also put down all the facts that disagree with it, as well as those that agree with it. In summary, the idea is to try to give all the information to help others to judge the value of your contribution, not just the information that leads to judgment in one particular direction or another." (http://calteches.library.caltech.edu/51/2/CargoCult.htm) Applied to PM2.5 health effects, this principle would argue for discussing the results of studies such as those reviewed by Burns J et al.(2019) and Henneman et al. (2017), *op cit*. The PA should discuss what specific alternative hypotheses for explaining observed C-R associations (such as those in Table 3-4 and Table C-1) have been tested, how, and what the results were. For example, has the null hypothesis been formally tested and rejected that the C-R associations in the studies in Table 3-4 and Table C-1 are entirely explained by omitted daily high and low temperatures (same-day and lagged for long enough so that conditional independence tests do not indicate further confounding)? If so, what were the results? What are the implications for alternative standards to protect public health with an adequate margin of safety?

6. *Taking into account the answers to the preceding questions, what does currently available scientific evidence show about whether and how much changes in current exposures would affect public health risks?* This question is not addressed by the causal determination judgments used in the draft PA. It is addressed by the simulations presented in the PA, but these treat C-R associations with uncontrolled confounding (e.g. by daily high and low temperatures) as if they were known to be valid causal predictors. This treatment does not provide valid scientific evidence for addressing how changing PM2.5 exposures would change public health risk. (As stated by Dr. Rhomberg, "one cannot simply change the value of one variable [e.g., PM2.5] and suppose that the model's result [e.g., change in mortality] represents what would be expected in a real setting.") The Draft PA thus leaves unanswered the fundamental policy-relevant question of what a scientifically valid analysis of available evidence would show about how changing current exposures would affect public health risks.

The final PA will be most useful to the CASAC if it addresses the preceding questions.

p. 1-2 *"Beyond informing the Administrator and facilitating the advice and recommendations of the CASAC, the PA is also intended to be a useful reference to all parties interested in the review of the PM NAAQS. In these roles, it is intended to serve as a source of policy-relevant information that informs the Agency's review of the NAAQS for PM, and it is written to be understandable to a broad audience."*

To "serve as a source of policy-relevant information that informs the Agency's review of the NAAQS for PM," the PA should use valid and empirically validated scientific methods to address the question of whether and how much changes in policy would affect public health risks. As just mentioned, the current draft PA is based largely on epidemiological evidence of

positive associations between exposures and health effects in studies that do not fully test and control for confounding, coincident historical trends, and other non-causal sources of associations. These associations (such as the beta coefficients in Table C-1) are then used as if they were known to be valid causal predictors for simulating how changes in exposure would change health risks. This is not sound science. The resulting conclusions and predictions are not scientifically valid and should not be used to guide policies that are to be based on sound science.

To be "understandable to a broad audience," the PA should use clearly defined terms to communicate its findings. Those findings should be derived from the evidence presented using explicit, verifiable reasoning. But the PA instead uses "causal determination" categories to communicate findings expressing judgments that cannot easily be tested or verified (or falsified). The causal determination categories lack clear operational definitions and it is not clear exactly what they mean or even that they are logically coherent (Cox 2019, Improving causal determination, Global Epidemiology). In places they appear to communicate subjective impressions, such as whether associations are deemed "suggestive" of causal conclusions, even if the causal conclusions do not logically follow from the evidence presented. This makes it unclear to a broad audience (or even to specialists, as attested by comments received from external experts) what exactly is being claimed in the PA's causal determinations, and the extent to which what is being claimed follows from the evidence presented.

As explained next, the scientific information and conclusions presented in the draft PA are not clear in meaning, transparent in derivation, scientifically valid, empirically validated, or trustworthy as guides to policy. Throughout the PA, simulations and policy-relevant conclusions are based on the fundamental technical error of misinterpreting association as causation. Specifically, they misinterpret the slope of a regression line ($\Delta y/\Delta x$) as showing how much the dependent variable y (e.g., mortality risk) will change if the predictor x (e.g., PM2.5 concentration) were to be *changed* by one unit via an intervention (Appendix C). This differs from what the slope of the regression line actually shows, which is how the conditional expected *observed* value of y would differ if the *observed* value of x were different by 1 unit (Pearl J 2010, "An Introduction to Causal Inference"p. 2, https://ftp.cs.ucla.edu/pub/stat_ser/r354-corrected-reprint.pdf). Such confusions and fallacies can and should be avoided very easily by applying core principles of sound science, including *clarity* in communicating conclusions (accomplished by using terms with unambiguous operational definitions); *transparency* in reasoning used to derive the conclusions from the evidence presented; *reproducibility* of tests of predictions against data, with results of such empirical validation efforts systematically performed and reported; and *objectivity* in selecting evidence and in presenting and interpreting results. (See e.g., https://www.nap.edu/read/1864/chapter/4#39 and Responsible Science: Ensuring the Integrity of the Research Process, Vol. 1 Committee on Science, Engineering, and Public Policy National Academies, 1992, especially pp. 37-39.) The PA does not follow any of these core principles of sound science:

- ***Clarity:*** **Provide explicit, clear, operational definitions of all key terms used to communicate results.** In the PA, the "causal determination" categories lack clearly stated operational definitions. This facilitates misinterpreting evidence of *association* as evidence of *causation* – a fundamental mistake that the Draft ISA and PA make throughout.
- ***Transparency:*** **Provide explicit, transparent, independently verifiable derivations of conclusions from data.** The PA presents causal conclusions based on expert judgments, without transparent derivations. The PA's conclusions about causation do not follow logically from the data presented. They are drawn from studies and data that were neither designed nor analyzed to permit valid causal conclusions. For example, multiple studies in Table C-1 do not fully control for obvious confounders such as temperature.
- ***Reproducibility of tests of predictions:*** **Perform reproducible tests of predictions against observations (e.g., using formal hypothesis testing) and report the results.** The Draft PA does not report results of formal tests of its causal hypotheses and interpretations. As previously noted, it also ignores a large recent literature showing that observations do not support confident rejection of the null hypothesis that interventions to reduce particulate air pollution have not successfully caused hoped-for (and, in some cases, predicted and claimed) reductions in mortality risks (e.g., Burns J et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst Rev. 2019 May 20; Henneman LR, Liu C, Mulholland JA, Russell AG. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. J Air Waste Manag Assoc. 2017 Feb;67(2):144-172).
- ***Objectivity:*** **State and follow explicit, systematic procedures for selecting and evaluating individual studies and evidence on which to base conclusions.** Report the results in transparent, systematic summaries. Carefully qualify causal interpretations and conclusions to acknowledge remaining ambiguities, uncertainties, or conflicts in evidence, while avoiding over-generalizations or subjective interpretations. As noted above, the PA omits much relevant and recent scientific literature and includes many studies with uncontrolled confounding, no correction of exposure estimation errors, and other methodological flaws.

Many comments received from our external expert consultants emphasize the lack of thoroughness, clarity, transparency, and validity in the PA's analysis. For example, Dr. Fred Lipfert cites many omitted studies and states that "The ultimate test of causality is whether public health has actually improved in response to reduced PM2.5 (by a factor of 3 since ca. 1980), after accounting for coincident trends in spatial patterns of reduced smoking and improved medical care. The extant literature does not support this test." Dr. Sonja Sax notes that, "In particular, the current framework lacks a transparent and systematic approach to evaluation of study quality. Therefore, it is unclear how EPA selects and weights studies for determining and classifying outcomes with regards to causality. In addition, how EPA integrates the evidence

is unclear. Again, the ISA process could be improved by providing greater clarity on how evidence from different scientific lines are integrated to make a final causality assessment. … in the PA the beta coefficient is not well defined. … The PA does not specifically define the beta coefficients or the concentration-response functions and how they are derived from the specific studies that are included in the evaluation. … EPA does not provide a clear definition of the concentration-response (CR) functions other than that they are obtained from the epidemiological studies. The epidemiological studies that underlie the CR functions are not intervention studies, but rather are observational studies that report associations between all-cause or cause-specific mortality and estimated ambient air concentrations. … The PA is not explicit with regards to how the underlying epidemiological studies were chosen, and whether these studies reflect associations or true predictions that could be inferred from intervention studies. Furthermore, EPA does not provide a discussion of how important limitations associated with these epidemiological studies (e.g., exposure measurement error and confounding) impact the interpretation of the risk assessment results. … EPA does not discuss the impact of other potential confounders, including meteorological parameters. This is an important source of uncertainty that needs to be included in the PA. … As with other potential confounders, EPA does not address the impact of potential residual confounding in the epidemiological studies. This should also be discussed in the PA. Importantly, it should be part of a quality and risk of bias evaluation in the PM ISA. … EPA did not include a systematic evaluation of study quality and risk of bias of individual epidemiology studies, included the studies EPA relies on in the PA for the risk assessment and does not discuss issues of internal or external validity associated with these studies. This is an aspect that is missing from the overall evaluation of this evidence. … Overall, the beta values used in the BenMAP analyses are from studies that likely did not effectively control for variables that potentially could confound or entirely explain the relationship between PM2.5 and mortality. …[The PA] needs to be more explicit regarding the assumption of causality. That is, that if the relationship is not causal then the risk estimates are not valid. … There are many aspects of the EPA causal framework that could be improved to provide greater transparency and scientific soundness… As was noted by CASAC in its review of the draft PM ISA, EPA excluded many studies that were informative regarding causal associations between PM and various health effects. … The BenMAP model relies solely on the selected epidemiological studies and these studies are not sufficient to confirm that there is a causal association between PM2.5 at current levels of exposure and mortality because of the inherent limitations of these studies (most importantly confounding and exposure measurement issues).” Other external experts provided many similar comments. For example, Dr. North comments that “EPA’s determinations of causality seem to me seriously flawed by confounding. … I do not accept the claim dating back to EPA’s 2009-11 documents that a ‘causal relationship’ is clearly demonstrated at and below the level of PM2.5 in the present NAAQS. … And by combining data from different areas of the country where PM2.5 comes from different sources, data on possible heterogeneity (different slopes in different locations) is being lost.” Dr. Jaffe states that “I do not see a definition for the beta coefficients in the document and so do suggest that these need further clarification.” Dr. Aliferis considers the causality analysis framework

used by the EPA (i.e., the causal determination framework) reasonable, but notes that "Errors in the choice of confounders translate to errors in causal effect estimates. This is the primary weakness." Mr. Jansen notes that "Studies that fail to account for socioeconomic status are missing a key factor affecting health. … I recommend EPA refrain from blurring the definition of biological plausibility."

The PA also does not apply modern causal analysis and inference frameworks (referred to by Drs. Aliferis and North) developed over the past century that could help to avoid errors and fallacies in judgments about causality. Instead, the PA uses a judgment-based framework and simulations based on associations (specifically, the beta coefficients in Appendix C) to derive conclusions and predictions for public health impacts of changes in PM2.5 concentrations. These conclusions and predictions lack scientific justification. They are not derived by the scientific method. They do not follow the principles of sound science discussed above. Their causal implications have not been validated empirically. They misrepresent C-R associations with uncontrolled confounding as if they were valid causal predictors. They have not applied appropriate technical methods to detect and avoid such errors in causal assumptions and interpretations. Although consistent with the approach taken in previous NAAQS reviews and advocated by the previous CASAC, the approach, predictions, and conclusions presented in the PA lack scientific validity.

Some Limitations of the PA's Judgment-Based Framework and Conclusions

Judgment is not a valid substitute for sound science, despite the emphasis on judgment in the "causal determination" framework currently used in NAAQS reviews. Judgments about causation made without formal causal analysis are notoriously error-prone (e.g., Shanks DR, Medin DL, Holyoak KJ (Eds.) (1996) *Causal Learning.* Academic Press). Exposure concentration-response (C-R) associations arise from many sources, including coincident historical trends (in multiple data sets, both PM2.5 exposures and public health effects such as cardiovascular mortality and morbidity are independently declining over time), modeling errors and biases, and uncontrolled or incompletely controlled confounding (e.g., a cold snap on one or a few days may predict both increased PM2.5 levels and, independently, increased mortality rates days or weeks later). *To interpret such associations, without rigorous, reproducible causal analysis and inference, as providing evidence that reducing PM2.5 standards would reduce human health risks or protect human health is not sound science.* Such associations and judgments provide no valid scientific justification for a conclusion that "revision of the suite of primary PM2.5 standards was necessary in order to provide increased public health protection." Intuitively, the error in logic is analogous to the one in noting a significant, strong, consistent correlation between daily ice cream consumption and daily risk of heat stroke in a population, and then concluding, based on this "strong evidence," that reducing ice cream consumption is required to protect people from heat stroke. In analysis of PM2.5 and mortality, as in the example of ice cream consumption and heat stroke, *it is essential to control for confounders such*

as daily temperature extremes and humidity (both same-day and lagged, possibly up to a month before mortality) before interpreting associations causally. The studies relied on in the PA (Table C-1) do not test and control for such confounders. Hence, the resulting estimates of association (the beta coefficients in Table C-1) are not valid measures of causal effects of PM2.5 exposure on public health. Likewise, the simulation results in the PA based on this interpretation are not valid predictors of the effects that would be caused by reducing PM2.5 concentrations, and policy recommendations based on these simulations are not based on sound science.

More generally, it has been well understood for decades in modern epidemiology that C-R *associations* such as those in Table C-1 are not causal *effects* of exposures on health responses (e.g., Petitti DB. Associations are not effects. Am J Epidemiol. 1991 Jan 15; 133(2):101-2), and that interpreting them as if they were is not scientifically valid. As noted in a recent review, "The field of environmental health has been dominated by modeling associations, especially by regressing an observed outcome on a linear or nonlinear function of observed covariates. Readers interested in advances in policies for improving environmental health are, however, expecting to be informed about health effects resulting from, or more explicitly caused by, environmental exposures. The *quantification of health impacts resulting from the removal of environmental exposures involves causal statements. Therefore, when possible, causal inference frameworks should be considered for analyzing the effects of environmental exposures on health outcomes*." (Bind MA. Causal Modeling in Environmental Health. Annu Rev Public Health. 2019 Apr 1;40:23-43. doi: 10.1146/annurev-publhealth-040218-044048.) (Emphases added.) This published recommendation is consistent with advice received by the current CASAC from multiple outside expert consultants, including Drs. Aliferis, North, and Rhomberg. However, such quantitative causal inference frameworks have not been considered in the PA or in the draft ISA to which it refers.

The fundamental methodological error of confusing association and causation is very common in air pollution epidemiology and has been prominent in previous NAAQS reviews. This is part of a broader pattern: human judgment under uncertainty is prone to well-documented, predictable heuristics and biases, and use of formal analytic methods and sound science are often needed to overcome the plausible-seeming but incorrect conclusions arising from judgments. Scientists, like other people, tend to make predictable errors when they rely on their own judgments rather than on formal analysis of data (e.g., hypothesis testing) to reach conclusions. Among the most common and relevant errors are the following (Kahneman D 2011, *Thinking, Fast and Slow*):

- *Answering an easier question instead of a harder but more relevant question* (and failing to notice the substitution). For example, instead of answering the policy-relevant question "Would further reductions in current PM2.5 concentrations cause further reductions in human health risks?" one might instead answer the easier question "Are past PM2.5 concentrations significantly associated with past human health risks?" – and then act as if the answer to the latter were an answer to the former. The current PA makes this mistake. The current causal determination framework tends to institutionalize it. Applying sound

science and modern causal analysis and inference methods can help to avoid such errors, as discussed further in comments from Dr. Aliferis. Nor should it be thought that applying them creates a difficult burden of proof or a need to apply novel or unproven methods. Basic tests for consistency of observational data with proposed causal interpretations (e.g., testing a null hypothesis that mortality rates are conditionally independent of PM2.5 given recent high and low daily temperatures and other covariates) are very easy to perform using readily available mainstream statistical software, and they have long been used successfully in other areas of applied science (e.g., Shipley B. 2000. *Cause and Correlation in Biology*. Cambridge University Press). Other misconceptions to be avoided include the following:

- Being clear about what question is being addressed (e.g., quantifying past correlations vs. predicting future changes in health effects that can be caused by future changes in exposures), and being careful not to conflate different questions, in no way implies or suggests that existing observational data are inadequate to address causal questions, or that an experimental approach is needed to study causation. This is a straw man argument (e.g., Goldman GT, Dominici F. Don't abandon evidence and process on air pollution policy. Science. 2019 Mar 29;363(6434):1398-1400) that should not distract from the need for clarity about what question is being answered, nor from the fact that methods for using observational data to answer causal questions are well developed and should be applied to draw valid causal conclusions (e.g., Bind MA. Causal Modeling in Environmental Health. Annu Rev Public Health. 2019 Apr 1;40:23-43. doi: 10.1146/annurev-publhealth-040218-044048; ; Campbell DT and Stanley JC (1963), *Experimental and Quasi-Experimental Designs for Research*.)
- The need to apply appropriate methods of causal analysis to draw valid causal conclusions in no way implies or suggests either that data from past epidemiologic studies should not be used, or that policymakers should wait for perfect data, studies, and analyses before taking action to protect public health (Carone M, Dominici F, Sheppard L. In Pursuit of Evidence in Air Pollution Epidemiology: The Role of Causally Driven Data Science. Epidemiology. 2019 Aug 12.) This is another straw man argument. Rather, addressing policy-relevant questions about causation using appropriate causal analysis methods – and applying these methods to data already collected, as well as to future studies – better informs policy makers about the likely consequences of different choices that they might make to protect human health. If some or all of the association between PM2.5 and an adverse health effect is explained by uncontrolled confounding (e.g., by cold weather or low income), for example, then policy makers should be informed about this, and not told that reducing exposure alone will reduce risk of the health effect. Conflating association with causation does not help inform policy makers about what actually works or reveal how to actually protect public health. Answering the right causal question using

appropriate methods does provide this information. Thus, being clear about what questions is being addressed, and answering causal questions by applying causal analysis, does not imply that past data should not be used, but only that it should be used correctly to address the questions that policy makers need answered.

- *Failing to seek and use relevant information*. Readily available information that is relevant for predicting the outcomes of different decisions or policies may not be collected, or may be ignored, especially if it conflicts with prior beliefs. For example, information from intervention studies and accountability studies, assessing what has actually happened to public health in the real world following pollution-reducing interventions, is readily available (e.g., Burns J et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst Rev. 2019 May 20; Letter from Dan Greenbaum to Aaron Yeow dated February 21, 2019; Henneman LR, Liu C, Mulholland JA, Russell AG. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. J Air Waste Manag Assoc. 2017 Feb;67(2):144-172.) This information is highly relevant for empirical evaluation of the health effects of changes in regulations. However, most of it is ignored in the current ISA and PA. Similarly, most individual studies associating PM2.5 with health risks do not systematically test or control for potential confounders such as income (poorer people tend to have both higher exposures and, independently, greater health risks) and weather variables such as high and low daily temperatures in the weeks preceding mortality (e.g., if a cold snap predicts both higher PM2.5 concentrations and also increased mortality rates in the ensuing days and weeks, out to about 1 month for cardiovascular mortality (Huang J, Wang J, Yu W. The lag effects and vulnerabilities of temperature effects on cardiovascular disease mortality in a subtropical climate zone in China. Int J Environ Res Public Health. 2014 Apr 11;11(4):3982-94. doi: 10.3390/ijerph110403982; Li C, Dai Z, Yang L, Ma Z. Spatiotemporal Characteristics of Air Quality across Weifang from 2014-2018. Int J Environ Res Public Health. 2019 Aug 27;16(17). pii: E3122. doi: 10.3390/ijerph16173122)). For example, none of the studies listed in Table C-1 of the PA and used to estimate health risks attributed to PM2.5 tests or corrects for confounding by daily high and low temperatures in the days and weeks prior to an adverse outcome, even though it would have been easy to do so in many cases. (The study of Turner et al. 2016 considers confounding and effect modification by daily high temperatures, but not by daily low temperatures and lagged daily temperature extremes.) Comments received from external expert consultants including Dr. Fred Lipfert provide references to other scientifically relevant and high-quality information omitted from the Draft ISA and PA.
- *Seeking and using irrelevant information to inform decisions*. Studies of historical exposure concentration-response (C-R) associations and regression coefficients are irrelevant for predicting how future policy interventions that change exposure concentrations would affect public health risks, since statistical associations do not answer such causal questions (Pearl J 2010, "An Introduction to Causal Inference"p. 2,

https://ftp.cs.ucla.edu/pub/stat_ser/r354-corrected-reprint.pdf). This is especially clear for studies that do not fully control (or control at all) for obvious confounders such as recent daily high and low temperatures. For such studies, including those relied on by the PA (Table C-1) to make risk predictions, positive statistical associations between PM2.5 and adverse health outcomes have no clear valid causal interpretation. It is therefore completely inappropriate to treat them as if they were valid manipulative causal relationships and to use them to predict effects on public health of changes in PM2.5 concentrations, as the PA does. This conflates correlation with causality, and lacks scientific justification. (Calling these confounded relationships "causal" in a causal determination does not address the problems of uncontrolled confounding and residual confounding, nor make it scientifically legitimate to treat these associations as if they represented manipulative causal relationships.)

- *Overconfidence.* Individuals and groups tend to be over-confident in their own judgments. Such overconfidence is often mistaken for competence. Scientists, like other people, are prone to such biased self-perceptions (e.g., Morgan MG. Use (and abuse) of expert elicitation in support of decision making for public policy. Proc Natl Acad Sci U S A. 2014 May 20;111(20):7176-84; Schwardmann P, van der Weele J. Deception and self-deception. Nat Hum Behav. 2019 Jul 29; Veldkamp CLS et al. "Who believes in the storybook image of the scientist?" Accountability in Research. https://doi.org/10.31234/osf.io/uk6ju MLA; Singh GG et al., Group elicitations yield more consistent, yet more uncertain experts in understanding risks to ecosystem services in New Zealand bays. PLoS One. 2017 Aug 2;12(8):e0182233). For example, p. 3-7 of the PA states that the previous Administrator "*particularly noted that the evidence was sufficient to conclude a causal relationship exists between PM2.5 exposures and mortality and cardiovascular effects (i.e., for both long- and short-term exposures) and that the evidence was sufficient to conclude a causal relationship is 'likely' to exist between PM2.5 exposures and respiratory effects (i.e., for both long- and short-term exposures).*" This expressed confidence that the evidence was "sufficient to conclude" important policy-relevant statements about causality was based on the judgments of selected experts, expressed via the "causal determination" framework, which does not (and did not) prevent confounded statistical C-R associations from being misrepresented to policy makers and the public as causal relationships. By contrast, approaches based on more reproducible analysis of data typically reach far less confident conclusions, such as the following:
    - "Multiple studies in the accountability field have found it difficult to attribute significant improvements in air quality or public health attributable to air quality regulations … This difficulty [is] particularly prevalent in studies that diligently control for multiple confounders across domains (location, time, etc.)" (Henneman et al. 2017)
    - "Two trends are apparent in the accountability assessments above. First, often, one study will assess a regulatory action and determine that the intervention led to

a statistically significant change in the response of interest… Later, using additional data, updated methods, and/or accounting for additional factors, those results are found to be less definitive and potentially invalid" (Henneman et al. 2017)

- "It was difficult to derive overall conclusions regarding the effectiveness of interventions in terms of improved air quality or health. Most included studies observed either no significant association in either direction or an association favouring the intervention, with little evidence that the assessed interventions might be harmful. The evidence base highlights the challenges related to establishing a causal relationship between specific air pollution interventions and outcomes." (Burns et al. 2019)

Consistent with EPA's selectivity in reporting and interpreting available scientific evidence, noted in comments received from several external experts, the PA does not mention these reviews and conclusions, nor discuss most of the studies they describe.

It should be noted that Burns et al. state (consistent with the advice that the current CASAC has received from Drs. Aliferis, North, and Rhomberg) that "results on effectiveness should be interpreted with caution; it is important to emphasize that lack of evidence of an association is not equivalent to evidence of no association." This is correct: at best, the lack of clear evidence from dozens of real-world interventions that reducing PM2.5 reduces health risks can and should be used to put an upper bound on how large any causal effects might plausibly be, but it cannot rule out the possibility of effects that are too small to be detected. Nonetheless, the cautious conclusions from such reviews of real-world data on changes in public health following interventions to reduce PM2.5 contrast with the confidence expressed in EPA's judgment-driven conclusions "that the evidence was sufficient to conclude a causal relationship exists between PM2.5 exposures and mortality and cardiovascular effects (i.e., for both long- and short-term exposures)."

Doing Better

The fallacies of answering the wrong question, disregarding relevant information, seeking and using irrelevant information, and being overconfident that the resulting numbers are "sufficient" to draw important real-world conclusions, undermine the scientific validity of conclusions and risk predictions in the Draft PA. These fallacies can be avoided by applying principles of sound science and modern causal inference frameworks developed over the past century (e.g., Wright, S. (1921). "Correlation and causation," J. Agricultural Research. 20: 557-585; Neyman J (1923), ''Statistical Problems in Agricultural Experiments,'' Journal of the Royal Statistical Society Series B (suppl.) (2):107–80; Yule GU. 1926. "Why do we sometimes get nonsense-correlations between time-series? – A study in sampling and the nature of time series." J R Stat Soc. 89:1–63.Simon HA (1952) "On the definition of the causal relation," Journal of Philosophy 49 (16):517-528; Wiener N. (1956) "The theory of prediction," In E. F. Beckmann, editor, In

Modern Mathematics for the Engineer. McGraw-Hill, New York; Blalock HM (1961) "Correlation and causality: The multivariate case." Social Forces, 39(3)March: 246-251; Campbell DT and Stanley JC (1963), *Experimental and Quasi-Experimental Designs for Research*; Granger, CWJ (1969). "Investigating causal relations by econometric models and cross-spectral methods". Econometrica. 37 (3): 424–438; Heckman JJ (2005) "The scientific model of causality," Sociological Methodology, 35(1), 1–97; Pearl J (2009) "Causal inference in statistics: An overview" Statist. Surv. Volume 3 (2009), 96-146.) (Some previous CASAC members have opined that these methods are too new and untested, or even too idiosyncratic, for practical use in NAAQS reviews, but they have been well developed, refined, widely accepted, and widely applied in other areas of applied science; the above references include works by three Nobel Laureates and a Turing Prize winner.)

As explained by Dr. Aliferis, "what modern causal inference methods do is that they create a causal graph that correctly captures the underlying causality. By using this graph the modeler can then decide which variables to use as covariates in order to obtain valid causal effect estimates." For PM2.5, such a graph could reveal whether potential confounders ignored in studies relied on in the current PA (e.g., the studies in Table C-1), such as daily high and low temperatures in the days and weeks preceding a death, must be conditioned on in order to obtain valid causal effect estimates. The current PA undertakes no such analyses. It simply assumes that associations from the studies in Table C-1 can be used to predict how changes in PM2.5 exposure concentrations would change public health risks. However, Dr. Rhomberg cautions correctly that "one cannot simply change the value of one variable [e.g., PM2.5] and suppose that the model's result [e.g., change in mortality] represents what would be expected in a real setting." Dr. Aliferis explains that, "Valid causal interpretation of the association measures requires that all the right covariates and only those have been used in the model. The right covariates can be found in a number of ways, some of which are stronger, and some weaker." Most of the studies in Table C-1 of the current PA make no attempt to include key covariates, such as temperature, and apply no methods, strong or weak, to assure that "all the right covariates and only those have been used." Thus, the resulting associations (e.g., the beta coefficients in Table C-1) cannot be validly interpreted as predictors of causal impacts on public health risk of changing PM2.5 levels. The analysis and conclusions of the PA based on the beta coefficients are fundamentally unsound because they confuse correlation and causation. Comments received from multiple external experts confirm that the beta coefficients used to make risk predictions in the PA lack clear interpretations and validity as predictors of changes in public health caused by changes in exposure.

As emphasized by several of our external expert consultants, failure to use formal causal analysis methods does not, by itself, imply that the results of the causal determination framework are necessarily incorrect. In principle, they might appear reasonable and could have considerable heuristic value, even if their validity is unknown. However, the estimated beta values in Table C-1 specifically reflect associations quantified without identifying or controlling for the right

covariates (e.g., daily high and low temperatures). They are therefore not valid causal effect estimates. In effect, the PA uses statistical associations with uncontrolled confounding and other errors and biases to predict public health benefits from reducing PM2.5. It presents no scientifically valid analyses suggesting that these predictions reflect what would happen (or has happened) following real-world interventions.

Comments on Chapter 3

p. 3-2 *The Administrator particularly noted the "strong and generally robust body of evidence of serious health effects associated with both long- and short term exposures to PM2.5" (78 FR 3120, January 15, 2013). … The Administrator further observed that such studies were part of an overall pattern across a broad range of studies reporting positive associations, which were frequently statistically significant. Based on her "confidence in the association between exposure to PM2.5 and serious public health effects, combined with evidence of such an association in areas that would meet the current standards" (78 FR 3120, January 15, 2013), the Administrator concluded that revision of the suite of primary PM2.5 standards was necessary in order to provide increased public health protection. Specifically, she concluded that the then-existing suite of primary PM2.5 standards was not sufficient, and thus not requisite, to protect public health with an adequate margin of safety. This decision was consistent with advice received from the CASAC (Samet, 2010a).*

This passage reflects an approach to NAAQS review and policy making that led to revisions in PM2.5 standards based on (1) statistical associations between past PM2.5 exposures and public health effects; and (2) a judgment-based causal determination framework. The current Draft PA follows the same approach. The approach misinterprets statistical concentration-response (C-R) associations as showing that reducing exposure would reduce risk of response. But, since C-R associations such as those in Table C-1 do not fully control for well-documented non-causal sources of association (such as confounding by temperature, errors and biases in modeling assumptions, coincident historical trends in exposure and response, and errors in exposure estimates), it is a logical and statistical fallacy to treat these associations as causal relationships that predict the effects on public health of reducing exposures.

p. 3-15 *"The Administrator's final decisions will draw upon the scientific evidence for PM-related health effects, information from the quantitative assessment of population health risks, information from analyses of air quality, and judgments about how to consider the uncertainties and limitations that are inherent in the evidence and information. To inform the Administrator's public health policy judgments and decisions, the PA considers support for, and the potential implications of, placing more or less weight on various aspects of this evidence, air quality and risk information, and associated uncertainties and limitations.*"

The PA provides no valid scientific information about how changing PM air quality standards would change (or, in the recent past, has changed) public health risks. A scientifically sound analysis would require considering relevant real-world evidence that the PM has ignored (e.g., Burns et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst Rev. 2019; Henneman et al. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. J Air Waste Manag Assoc. 2017 Feb;67(2):144-172); clearly defining and then appropriately calculating beta values (or other formulas for quantifying causal effects on public health of changing PM2.5) while correcting for causally relevant covariates (e.g., high and low daily temperatures and other confounders), exposure estimation errors, and modeling errors and biases; and distinguishing between association and causation. Since the PA does not do these things, it should not be used as if it provided valid scientific information about health risks.

p. 3-16 *"The draft ISA defines these causality determinations as follows (U.S. EPA, 2018, p. p-18):*
• Causal relationship: the pollutant has been shown to result in health effects at relevant exposures based on studies encompassing multiple lines of evidence and chance, confounding, and other biases can be ruled out with reasonable confidence.
• Likely to be a causal relationship: there are studies in which results are not explained by chance, confounding, or other biases, but uncertainties remain in the health effects evidence overall. For example, the influence of co-occurring pollutants is difficult to address, or evidence across scientific disciplines may be limited or inconsistent."

These definitions do not address whether reducing exposure would reduce health risks. Hence they do not support valid scientific conclusions about whether reducing exposure would protect human health by reducing human health risks. The phrase "has been shown to result in" lacks clear operational definition. It seems elsewhere to be used to mean only "is held accountable for, in the judgments or opinions of selected experts." The phrase "with reasonable confidence" lacks clear operational definition. However, as previously discussed, confounding has not been ruled out with reasonable confidence (or at all) in the studies relied on in Table C-1 and in the simulations reported later: for example, studies in Table C-1 do not rule out confounding by daily high and low temperatures. It is not clear that these definitions rule out other non-causal explanations for C-R associations, such as coincident historical trends, modeling errors, and exposure estimation errors. In light of these limitations, the causal determination judgments in Table 3-1 are not verified as being scientifically sound or as being derived (or derivable) by correct reasoning from available scientific knowledge and data.

p. 3-19 "*In the last review, the 2009 PM ISA reported that the evidence was 'sufficient to conclude that the relationship between long-term PM2.5 exposures and mortality is causal' (U.S. 12 EPA, 2009, p. 7-96). The strongest evidence supporting this conclusion was provided by epidemiologic studies, particularly those examining two seminal cohort, the American Cancer*

Society (ACS) and the Harvard Six Cities cohorts. Analyses of the Harvard Six Cities cohort included demonstrations that reductions in ambient PM2.5 concentrations are associated with reduced mortality risk (Laden et al., 2006) …Recent cohort studies, which have become available since the 2009 ISA, continue to provide consistent evidence of positive associations between long-term PM2.5 exposures and mortality."

This passage illustrates that EPA's conflation of association with causation extends back to at least the 2009 PM ISA. The underlying studies cited also make this fundamental error. The ACS and Harvard Six Cities studies and updates explicitly address associations, not causation. Laden et al. (2006), in a section titled "Association of PM2.5 with Mortality" refer to "The effect of each 10μg/m^3 increase in average annual PM2.5 pollution…" without noting that associations are not effects (e.g., Petitti DB. Associations are not effects. Am J Epidemiol. 1991 Jan 15; 133(2):101-2). Laden et al, further state in a section on "Statistical Analysis" that "To adjust for temporal trends in mortality, we included an indicator for period. We then assessed the association of mortality with average city specific PM2.5 for the entire period of follow-up (pollution averaged from 1980–1998) and with the period-specific average PM2.5." This is a clear recipe for residual confounding: an "indicator for period" does not fully control for or "adjust for temporal trends in mortality," but instead allows coincident historical trends of declining PM2.5 and mortality rates to create a positive PM2.5-mortality C-R association within each period. Similarly, in their literature review, Laden et al. (2006) state that "Mortality in Dublin decreased by 8% after a 36-μg/m^3 decrease in average particulate air pollution (black smoke) due to a ban on coal sales." In reality, the 36-μg/m^3 decrease in average particulate air pollution had no detectable impact on total mortality rates. The decrease in average particulate air pollution was associated with decreased mortality rates because both independently were decreasing over time. (Zigler CM, Dominici F. Point: clarifying policy evidence with potential-ou\tcomes thinking--beyond exposure-response estimation in air pollution epidemiology. Am J Epidemiol. 2014 Dec 15;180(12):1133-40.) Finally, Laden et al. control for "potential confounders" such as body mass index, but do not control for far more obviously causally relevant covariates such as temperature. Thus, the "strongest evidence" that EPA refers to as "sufficient to conclude that the relationship between long-term PM2.5 exposures and mortality is causal" does not address causality at all, but consists of associations in the presence of uncontrolled confounding in studies which were neither designed nor analyzed to permit valid causal inferences. There is no valid scientific basis for presenting the associations from such studies as evidence that is "sufficient to conclude that the relationship between long-term PM2.5 exposures and mortality is causal."

Other individual studies cited in Chapter 3 have similar flaws (uncontrolled confounding, residual confounding, uncontrolled exposure estimation errors, etc.), highlighting the importance of stating and systematically applying criteria for individual study quality and then reporting the results for each study. For example, the PA states (p. 3-20) that "*Adding to recent evaluations of the ACS and Six Cities cohorts, studies conducted in other cohorts also demonstrate consistent,*

positive associations between long-term PM2.5 exposure and mortality across various demographic groups (e.g., age, sex, occupation), spatial and temporal extents, exposure assessment metrics, and statistical techniques (U.S. EPA, 2018, sections 11.2.2.2, 11.2.5). This includes some of the largest cohort studies conducted to date, with analyses of the U.S. Medicare cohort that include nearly 61 million enrollees (Di et al., 2017b).” However, the Di et al. study did not control for relevant covariates such as daily high and low temperatures, or include actual measurements of air pollution for any individual. Rather, individual air pollution exposure levels were guessed at using techniques such as the following: “To join monitoring data to each residential ZIP code, we identified the nearest monitoring site within 50 km of the ZIP code (based on centroid point) and assigned air pollutant measurements to that ZIP code.” The resulting guesses were then analyzed as if they were error-free measurements – a clear violation of sound statistical analysis for error-prone exposure estimates. Likewise, the Draft PA states (p. 3-20) that “*A recent series of ‘accountability’ studies has additionally tested the hypothesis that reductions in ambient PM2.5 concentrations would be associated with increased life expectancy or a decreased mortality rate (U.S. EPA, 2018, section 11.2.2.6). In their original study, Pope et al. (2009) used air quality data in a cross-sectional analysis from 51 metropolitan areas across the U.S., beginning in the 1970s through the early 2000s, to demonstrate that a 10 µg/m3 decrease in long-term PM2.5 concentration was associated with a 0.61-year increase in life expectancy.”* The PA does not mention more recent work calling such claims into question (Krstić G. A reanalysis of fine particulate matter air pollution versus life expectancy in the United States. J Air Waste Manag Assoc. 2013 Feb;63(2):133-5), or the authors’ response; nor does it repeat the original authors’ caveat that “the ability to control for additional potential confounders, especially various individual and community risk factors that may have policy drivers in common with environmental regulation, is limited.” The PA states (p. 3-21) that “The draft ISA concludes that, ‘collectively, this body of evidence is sufficient to conclude that a causal relationship exists between long-term PM2.5 exposure and total mortality’.” However, since “this body of evidence” consists primarily of associations in studies that did not fully control for causally relevant covariates (such as daily high and low temperatures) and that were not designed or analyzed to permit valid causal inferences, the conclusion that “this body of evidence is sufficient to conclude that a causal relationship exists between long-term PM2.5 exposure and total mortality” is unwarranted. It is not implied by, or consistent with, the principles of sound science previously discussed.

Comments on Chapter 4

p. 4-5 *“To what extent does the currently available scientific evidence strengthen, or otherwise alter, our conclusions from the last review regarding health effects attributable to long or short-term PM10-2.5 exposures?”*

The question of what health effects are "*attributable to*" exposures is different from the more policy-relevant question of what health effects are *preventable* by reducing exposures. Standard epidemiological calculations of attributable risk (and related measures of association such as relative risk, etiologic fraction, probability of causation, and so forth) allow any health effect to be attributed to any exposure as long as they are positively associated (RR > 1), even if the positive association is fully explained by uncontrolled confounding, residual confounding, or other non-causal sources. What policy makers need to know is how alternative policy decisions would change probabilities of consequences (health risks) and how sure we can be about the answer based on currently available evidence. Chapter 4 does not address these questions.

p. 4-8 *"Based on the overall evidence, the draft ISA concludes that, 'this body of evidence is suggestive, but not sufficient to infer, that a causal relationship exists between short-term PM10-2.5 exposure and total mortality' (U.S. EPA, 2018, p. 11-116)."*

It is not clear how the characterization of the body of evidence as "suggestive" that a causal relationship exists is derived from the materials presented. The "evidence" presented primarily addresses associations, not causation: as stated in the Draft PA four lines previously, "Associations with cause-specific mortality provide some support for associations with total (nonaccidental) mortality, though associations with cause-specific mortality, particularly respiratory mortality, are more uncertain (i.e., wider confidence intervals) and less consistent (U.S. EPA, 2018, section 11.3.7)." Associations, especially when they are not free from confounding, historical trends, errors in exposure estimates, and model uncertainties, do not provide evidence about whether or how much reducing exposure would reduce health risks. Insofar as the epidemiological evidence in Chapter 4 addresses such associations, it does not permit valid conclusions about causation.

These methodological limitations, i.e.,

(a) The PA's derivations of conclusions from evidence presented are not clear, explicit, and independently verifiable/checkable (i.e., they are not transparent);
(b) Attribution of associations does not address risk preventable by reducing exposures (i.e., the chapter does not address the right questions to inform policy decisions); and
(c) Associations do not provide evidence that permits valid causal inferences about how future changes in exposures would change future health risks, especially when uncontrolled confounding and other potential noncausal explanations of the associations are present

apply to the individual sections of Chapter 4. For example, the conclusion on p. 4-11 that "Based on the expanded, though still limited evidence base, the draft ISA concludes that, '[o]verall, the evidence is suggestive of, but not sufficient to infer, a causal relationship between [long]-term PM10−2.5 exposure and metabolic effects' (U.S. EPA, 2018, p. 7-61)" does not appear to be clearly better justified than many other possible conclusions, such as that "overall, the evidence does not imply or suggest a causal relationship between long-term PM10−2.5 exposure and

metabolic effects." However, what is "suggested" by a collection of materials may be in the eye of the beholder. Associations that provide no objective scientific evidence that reducing exposure would reduce (or has reduced) health effects might still be deemed by a suggestible reviewer to be "suggestive of, but not sufficient to infer" any desired conclusion. But such judgments do not constitute sound science.

p. 4-14 *"As in the last review, epidemiologic studies continue to report positive associations with mortality or morbidity in cities across North America, Europe, and Asia, where PM10-2.5 sources and composition are expected to vary widely. Such studies provide an important part of the body of evidence supporting the strengthened causality determinations (and new determinations) for long-term PM10-2.5 exposures and mortality, cardiovascular effects, metabolic effects, nervous system effects and cancer (U.S. EPA, 2018)."*

This conclusion explicitly states that "positive associations with mortality or morbidity" in cities "provide an important part of the body of evidence supporting the strengthened causality determinations (and new determinations)." But positive associations that are not free from confounding, coincident historical trends, and other non-causal explanations, do not provide valid evidence for making or strengthening causal determinations. Using them for this purpose amounts to drawing causal conclusions from non-causal evidence, and is not scientifically valid. This has been well understood for many decades in other areas of science and statistics that deal with observational data; see e.g., Campbell DT and Stanley JC (1963), *Experimental and Quasi-Experimental Designs for Research*. To restore sound science and valid conclusions to the NAAQS review process, it is essential to stop using fallible "causal determination" judgments to go beyond what the data show, and to start using rigorous, reproducible analyses of causally relevant data (e.g., from soundly designed intervention studies) to draw valid causal conclusions, consistent with practices for valid causal analysis and inference developed and applied in other areas of applied science over the past century.

Comments on Chapter 5

p. 5-26 *"While research on PM-related effects on climate has expanded since the last review, there are still significant uncertainties associated with the accurate measurement of PM contributions to the direct and indirect effects of PM on climate."*

p. 5-35 *"Limitations and uncertainties in the evidence make it difficult to quantify the impact of PM on climate and in particular how changes in the level of PM mass in ambient air would result in changes to climate in the U.S. Thus, as in the last review, the data remain insufficient to conduct quantitative analyses for PM effects on climate in the current review."*

The PA should give quantitative estimates, with uncertainty bands, for effects on climate change of changing PM2.5. If cleaner air accelerates warming, for example, how large is the effect? (See e.g., Schiermeier Q (2005), "Clear skies end global dimming: Earth's air is cleaner, but this may worsen the greenhouse effect" doi:10.1038/news050502-8, www.nature.com/news/2005/050502/full/050502-8.html.) While it is reasonable to state qualitatively that there are remaining uncertainties, a more quantitative assessment is needed for well-informed policy-making. Recent research, while acknowledging the difficulties and uncertainties involved, has advanced so that it is no longer true that "the data remain insufficient to conduct quantitative analyses for PM effects on climate". For example, one recent quantitative estimate of PM effects on climate states "that eliminating the human emission of aerosols—tiny, air-polluting particles often released by industrial activities—could result in additional global warming of anywhere from half a degree to 1 degree Celsius. This would virtually ensure that the planet will warm beyond the most stringent climate targets outlined in the Paris climate agreement" (www.scientificamerican.com/article/cleaning-up-air-pollution-may-strengthen-global-warming/). Chapter 5 should make clear whether such quantitative predictions are consistent with current understanding based on the best available evidence. (For more recent quantitative modeling results, see Rosenfield et al. (2019), https://science.sciencemag.org/content/364/6446/eaay4194).

The PA should also address how changes in cold weather and warm weather temperature extremes due to reduced PM2.5 are likely to affect public health risks (e.g., Patel D, Jian L, Xiao J, Jansz J, Yun G, Robertson A. Joint effect of heatwaves and air quality on emergency department attendances for vulnerable population in Perth, Western Australia, 2006 to 2015. Environ Res. 2019 Jul;174:80-87; Xing J, Wang J, Mathur R, Pleim J, Wang S, Hogrefe C, Gan CM, Wong DC, Hao J. Unexpected benefits of reducing aerosol cooling effects. Environ Sci Technol. 2016 Jul 19;50(14):7527-34.

Dr. Mark Frampton

General Comments

CASAC appreciates that, in response to a CASAC request in its April 11 letter, EPA has appointed a panel of twelve expert consultants as a resource in the review of this PM PA, as well as for the ozone review. Those panel members have already provided helpful and insightful responses to specific questions posed by the chartered CASAC members. However, CASAC had recommended in its April 11 letter that the EPA "…reappoint the previous CASAC PM panel (or appoint a panel with similar expertise)…". Those recommendations were not followed. We note the relative absence on the appointed panel of consultants with expertise and experience in air pollution epidemiology research, a scientific discipline that is obviously of key importance in the review of the PM standards, and expertise that was specifically requested by CASAC to assist in its review.

CASAC notes the difficulty and limitation in providing cogent and insightful comments on this PA document, given that the ISA has yet to be finalized, and the CASAC advice for revisions to the ISA, that were made in the CASAC letter to the Administrator (April 11, 2019), have yet to be addressed. Thus CASAC is attempting to review policy assessment and planning that is based on an incomplete scientific review.

Chapter 3 – Review of the Primary PM2.5 Standards: *What are the CASAC views on the approaches described in chapter 3 to considering the $PM_{2.5}$ health effects evidence and the risk assessment in order to inform preliminary conclusions on the primary $PM_{2.5}$ standards? What are the CASAC views regarding the rationales supporting the preliminary conclusions on the current and potential alternative primary $PM_{2.5}$ standards?*

3.1.1 provides a clear summary of the most recent review of and actions on the $PM_{2.5}$ standard with regard to indicator (retained as mass-based), averaging time (retained 24-hr and annual), form (annual standard revised to eliminate spatial averaging provisions), and level (annual level reduced from 15 $\mu g/m^3$ to 12 $\mu g/m^3$.

The approach as described in 3.1.2 of the PM PA is well-reasoned and consistent with previous assessments. The PA appropriately emphasizes health outcomes for which the draft ISA determined that the evidence supports either a "causal" or a "likely to be causal" relationship with $PM_{2.5}$ exposures.

The PA appropriately acknowledges limitations in PM research that have been discussed extensively in CASAC's review of the 2018 ISA and previous PM ISAs. Epidemiology studies form the key source of evidence for $PM_{2.5}$ health effects, and yet epidemiology studies generally examine associations, rather than cause and effect. They also do not establish no-effect thresholds, although C-R relationships in the lower ranges of exposures can help inform this. While human studies may potentially inform thresholds and C-R functions for some subclinical outcome markers in relatively healthy subjects, they cannot address such issues for adverse clinical outcomes including mortality, for the most at-risk groups, or for long-term exposures. Despite these limitations, the totality of the data are convincing for causality because of the sheer number of large, well-conducted epidemiology studies showing remarkably consistent effects using a variety of approaches, locations, and populations. In general, the health effect findings have remained robust with adjustments for co-pollutant exposures and known or suspected confounders, including meteorology factors and socioeconomic status. Notably, no confounder, covariate, or single specific PM chemical component has been identified that explains these observed associations between $PM_{2.5}$ and mortality. The epidemiology findings are supported by human clinical studies and toxicology studies that show plausibility and provide potential mechanisms.

Studies published since the previous review do not call into question the causality findings from the previous review, but further strengthen the evidence linking $PM_{2.5}$ exposure with adverse health effects. They also provide additional data at lower exposure levels.

The question addressed in the risk assessment concerns the adequacy of the current $PM_{2.5}$ standard for the protection of public health, and the appropriate range for an alternative standard if deemed necessary. This is addressed in 3.3 and 3.4. Several studies published since the previous review show increased mortality at mean $PM_{2.5}$ concentrations at or below the current standard. Recognizing the significant limitations in determining thresholds, or the shape of the C-R function at low concentrations, the data from the newer studies are most consistent with, or at least do not refute, a linear C-R curve with no discernable threshold. These issues and the remaining uncertainties involved are reasonably addressed in the PA. Alternative and contingent analyses have appropriately been explored.

Cancer mortality

The 2018 PM ISA concluded that the evidence " is sufficient to conclude that a causal relationship is likely to exist between long-term $PM_{2.5}$ exposure and cancer" (U.S. EPA, 2018, section 10.2.7). CASAC, in its letter to the Administrator of April 11, 2019, disagreed, finding that "…the Draft ISA does not present adequate evidence to conclude that there is likely to be a causal association between long-term $PM_{2.5}$ exposure and … cancer." The available data are not able to adequately distinguish between effects of PM exposure on incident cancer (i.e., $PM_{2.5}$ causing cancer) from cancer-related mortality (i.e., $PM_{2.5}$ hastening death in patients with

cancer). The difficulty in separating these is related to the long latency between cancer initiation and clinical diagnosis. PM exposure may hasten mortality via effects on comorbid conditions, such as chronic obstructive pulmonary disease, cardiovascular disease, or infections. Given the relatively few studies of cancer mortality, and the remaining uncertainties, retaining the previous "suggestive" determination is appropriate.

For consistency with the approach taken in the PA for other outcomes, in which only causal and likely causal relationships are considered, cancer mortality should be removed from the risk analysis and estimates described in 3.2 to 3.4.

3.4 Preliminary Conclusions

The risk assessment estimates that the current primary $PM_{2.5}$ standards could allow a substantial number of deaths in the U.S. (p. 3-97). I agree with the preliminary conclusion of the PA, that the risk assessment "…can reasonably be viewed as calling into question the adequacy of the public health protection afforded by the combination of the current annual and 24-hour primary $PM_{2.5}$ standards." (P. 3-98).

With regard to potential alternative standards, I agree the indicator, averaging times, and forms of the primary standards should be retained. The PA provides a reasonable assessment that reductions in the level of the annual standard would also reduce short-term concentrations, justifying the approach of focusing on the annual standard as the principal means of providing public health protection. I agree with retaining the current level of the 24-hr standard at 35 $\mu g/m^3$. For the annual standard, the PA considers a range of 10 down to 8 $\mu g/m^3$ and this range is reasonable and justified by the evidence and risk assessments. Given the increasing strength of evidence at lower concentrations, and the need to protect the public health with an adequate margin of safety, reducing the level of the annual standard to the lower part of this range, 8 or 9 $\mu g/m^3$, is warranted.

Chapter 4 – Review of the Primary PM10 Standard: *What are the CASAC views on the approach described in chapter 4 to considering the $PM_{10-2.5}$ health effects evidence in order to inform preliminary conclusions on the primary PM_{10} standard? What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current primary PM_{10} standard?*

Chapter 4 of the PM PA addresses the PM_{10} Standard. PM_{10} encompasses all particles smaller than 10 µm and therefore includes $PM_{2.5}$. Therefore the PM_{10} standard is relevant for the health effects of $PM_{10-2.5}$, or so-called thoracic coarse particles, which are not addressed in the $PM_{2.5}$ standard. The 24-hr PM_{10} standard of 150 $\mu g/m^3$ was first established in 1987, replacing a prior standard for total suspended particles, and has been retained at all subsequent reviews, including

the most recent in 2012. An annual PM_{10} standard of 50 $\mu g/m^3$ was established in 1987, but revoked in 2006 because, in the view of the Administrator, the evidence did not support a link between long-term exposure to existing ambient levels of coarse particles and health or welfare. The 2009 PM ISA (U.S. EPA, 2009, section 2.3.3), which formed the basis for the 2012 action, concluded that the evidence was "suggestive of a causal relationship" for cardiovascular effects, respiratory effects, and/or premature mortality following short-term exposures.

Approach:

This chapter of the PM PA provides a helpful summary and background of the approaches taken in the previous reviews of the PM_{10} standard, including the key uncertainties that contributed to the "suggestive but not sufficient to infer" causality determination in the previous ISA and the decision to retain the previous standard (page 4-3).

The PM PA has appropriately placed the greatest emphasis on health effects for which the pollutant in question has been determined to be causal or likely to be causal. In the current ISA, none of the identified health outcomes linked to PM_{10} exposure rose to this level of certainty. Because of this, the approach taken in this chapter was more limited than for $PM_{2.5}$, primarily addressing the remaining uncertainty in the evidence base, and this is reasonable and appropriate.

Preliminary conclusions:

Substantial remaining uncertainties the assessment of $PM_{10-2.5}$ exposure and potential for confounding are discussed.

There are new studies of long-term health effects that justify the change in causality determination from "inadequate" to "suggestive". There are also new data somewhat strengthening the evidence for effects on cancer, and short-term effects on mortality, cardiovascular disease, and respiratory disease. But the key uncertainties remain. In light of these continuing uncertainties, the PM ISA determined that for all of the considered $PM_{10-2.5}$ health effects, the evidence was "suggestive but not sufficient to infer" a causal relationship.

Given these considerations, this draft PA concludes that "…the available evidence does not call into question the adequacy of the public health protection afforded by the current primary PM_{10} standard and that evidence supports consideration of retaining the current standard in this review." I agree with this assessment.

Dr. Sabine Lange

My major comments about this PM policy assessment are summarized below. Details to further support these points can be found after the summary of all the major points.

A reference list can be found at the bottom of this document for those studies that are not referenced in the PM PA.

Chapter 3. Review of the Primary Standards for PM2.5

Charge Question: What are the CASAC views on the approaches described in chapter 3 to considering the PM2.5 health effects evidence and the risk assessment in order to inform preliminary conclusions on the primary PM2.5 standards? What are the CASAC views regarding the rationales supporting the preliminary conclusions on the current and potential alternative primary PM2.5 standards?

Major Point #1: Because the PM proposed rule will be out before this PA is modified, it is not clear how the comments provided by CASAC will have bearing on the decisions being made based on this PA. This is exemplified by the fact that the comments that CASAC recommended for the PM ISA were not incorporated into the PM PA (although there was enough time to do so). The EPA should provide CASAC and the public with some assurance (and evidence) that their hard work in reviewing and commenting on the PM NAAQS documents is being used to inform the forthcoming proposed and final rules.

- It is difficult to comment on this document knowing that the substantial comments from CASAC and the public about the underlying ISA have not been taken into consideration. Many of the comments I raised about the ISA I will again discuss here.
- The EPA should use the comments put forth by the CASAC and the public not just to modify this document, but also to inform the next document (presumably the proposed rule).
- A specific example: *3.1.2. General Approach in Current Review*: For this review, the EPA notes that the focus is on "Causal" or "Likely Causal" determinations from the ISA, and that they focus on information from key epidemiologic and controlled human exposure studies. However, the causality determinations are from the Draft ISA, and do not take into consideration the CASAC's recommendations about those causality determinations. In fact, the ISA is due to be finalized after the public comment period for this document is over – so I ask, how will this PA change (and how can CASAC provide relevant comments on it) if the causality determinations change from the draft to final ISA? This impacts both those issues that consider the overall strength of evidence

showing the causal association between the PM2.5 and health effects, as well as the risk assessment that uses those endpoints designated as causal or likely causal.

Evidence-Based Considerations in Summary of the ISA

Major Point #2: There are actual causal modeling methods available for determining causality using information similar to what is used in epidemiology studies. This type of work and consideration should be used to assess causality in this review. For EPA's current framework, they need to explicitly communicate their judgement about each component of their causality designation to make it clear to readers how they have come to their causality decisions.

- Fundamentally, when assessing the summary of information from the ISA, it is difficult to get around the fact that the EPA is using a qualitative, subjective, heuristic method for determining causation. This method can generate different conclusions based on the judgement of the user. Therefore, it is difficult to comment on the adequacy of the conclusions, because an adequately objective and analytical method has not been applied to the data. This is made clear in comments by Dr. North and Dr. Aliferis.
- In this document the EPA needs to consistently explain why a particular causality designation was chosen. For example, for a *causal* relationship, there is supposed to be evidence from studies where bias, chance, and confounding has been ruled out with reasonable confidence, whereas "uncertainties remain in the evidence overall" for a determination of *likely causal*. Therefore, these stipulations should be specifically addressed in each section to show why the sum of the evidence has warranted a certain causal determination. Dr. Aliferis notes that the EPA's causal designations are highly heuristic and in fact the causality determination method laid out in the ISA Preamble is a heuristic method based on heuristic judgements. Therefore, if the EPA is using their judgement in deciding on a causal designation, they should specifically walk through the arguments they use.
- The EPA has used some of what they label to be accountability studies to justify their causality conclusions in this PA, which is new from the ISA (accountability was not discussed). Because this is new, they should label what they consider to be an accountability study, preferably by referencing one of the several recent reviews on this topic (Henneman et al., 2017; Rich, 2017). This should include a consideration about whether a study that studies PM2.5 changes before PM2.5 regulations come into place (e.g. Pope et al. 2009) can be called accountability studies, and the importance of having control areas for these studies. In addition, because the EPA is considering studies that look at PM2.5 concentrations over long time periods, they should also consider the work of Greven et al., 2011 and Pun et al., 2017. Those papers also consider trends over time but also use sophisticated analyses to separate local from national trends and find no effect of PM2.5 at the local level. The authors of those studies suggest that this pattern demonstrates the presence of uncontrolled confounders in these analyses.

- It is not clear from the presented evidence that chance, bias, and confounding have been ruled-out for the endpoints with *causal* designations.
- The causality determinations should be informed not just by the potential hazard information in experimental studies, but by the concentrations where those effects occur (i.e. not just that inflammation is demonstrated, but that it is only consistently observed at concentrations much higher than ambient).
- Three causality designations were questioned by CASAC in the last review, and those designations need to be either further justified by the EPA in this document (as well as the final ISA), or changed. These included: PM2.5 and cancer; long-term PM2.5 exposures and nervous system effects; and long-term UFP exposures and nervous system effects.

Major Point #3: The EPA must provide a balanced summary of the study results for each health endpoint. Only communicating the positive results, and not the negative results, null results, or uncertainties in the evidence does not accurately explain the evidence to the Administrator and does not help him make an informed decision.

- The EPA should provide, for each endpoint and exposure length, information about **what data supports, what data does not support, what are the uncertainties, at what concentrations do they occur, and what is new**. Only two of these factors are being reliably discussed in this PM PA: what supports, and what is new. This document is supposed to inform the EPA administrator about the data, but it provides an unbalanced viewpoint.
- The EPA states that the data for long-term PM2.5 and mortality has been shown to be robust across exposure assessment methods, statistical models, diverse geographic regions, and temporal periods. This greatly oversimplifies the actual data where there is demonstrated heterogeneity in effect estimates (Di et al. 2017b); a close to 7-fold difference in the slopes of the associations between analyses using different exposure modeling approaches (Jerret et al. 2017); there is regional heterogeneity (Kioumourtzoglou et al. 2015), figure in details section; and the problems with the temporal associations described under Major Point #2.
- For cardiovascular effects, the EPA often summarizes information as being strong and consistent, when in fact the evidence presented in the ISA is not strong and consistent. For example, there are a lot of null findings for short-term and long-term PM2.5 exposure and ischemic heart disease (IHD) and heart failure (HF; one of the figures from the PM ISA is reproduced in the details section for this point). From Figure 6-17 for PM2.5 and ischemic heart disease and myocardial infarction, of the presented effect estimates only 1 out of 11 is statistically significant, and 4 out of the 6 US studies used PM2.5 exposure estimates from before there was a nationwide monitoring network (thereby requiring extrapolation to estimate PM2.5 concentrations). Negative findings for morbidity endpoints do not provide consistent or strong evidence that it is appropriate to model the risk for the associated mortality endpoint (i.e. IHD).

- The lack of respiratory effects observed in controlled human exposure (CHE) studies of PM2.5 exposure needs to be adequately communicated and addressed for the causality determination as well as in the summary sections.
- The summary section 3.2.1.7 needs to adequately communicate that there is variability in the breadth and depth of study findings, instead of over-simplifying with statements such as "Recent epidemiologic studies consistently report positive associations between long-term PM2.5 exposures and a wide range of health outcomes, including total and cause-specific mortality, cardiovascular and respiratory morbidity, lung cancer, and nervous system effects".
- The choice of studies for use in the pseudo-design value analysis included an excellent description of model validity and exposure data quality. This kind of consideration of study quality should be extended to all key studies and discussed similarly.
- The EPA's summary of risk estimates in section 3.4.1 (preliminary conclusions on the current PM2.5 standard) provides only the highest risk estimates with none of the available confidence bounds. This provides inappropriately high and certain results from the risk analysis. The EPA should provide the range and CIs for the risk estimates: For example, instead of: "the risk assessment estimates up to about 50,000 total PM2.5-related deaths, including almost 20,000 ischemic heart disease deaths, in a single year", it should state something like "the risk assessment estimates total PM2.5-related deaths in the range of 13,500 (2,360-24,200) to 52,100 (41,600-62,300), including approximately 15,600 (11,600-19,400) to 16,800 (12,800-20,500) ischemic heart disease deaths."

Major Point #4: The uncertainties identified in the last PM review should be explicitly addressed to determine whether more certain information is available in this review than there was in the past.

- The EPA could explicitly address in each section whether they have diminished the uncertainties noted by the EPA administrator in the last review (pp 3-8 to 3-9): "The Administrator recognized that uncertainties remained in the scientific information. She specifically noted uncertainties related to understanding the relative toxicity of the different components in the fine particle mixture, the role of PM2.5 in the complex ambient mixture, exposure measurement errors in epidemiologic studies, and the nature and magnitude of estimated risks related to relatively low ambient PM2.5 concentrations. Furthermore, the Administrator noted that epidemiologic studies had reported heterogeneity in responses both within and between cities and in geographic regions across the U.S. She recognized that this heterogeneity may be attributed, in part, to differences in fine particle composition in different regions and cities."
- Upon review of the information in the PM PA, it seems that there are still unknowns with copollutants, C-R functions are still plagued by problems with innate variability that makes them difficult to interpret, none of the studies on regional heterogeneity adequately explained the reasons for the city-specific heterogeneity, and it is not clear what components or sources are causing the observed effects. Therefore, it does not seem

that many of the key uncertainties have been reduced in this review. The expert consultant Mr. Jansen also noted this concern.

- Exposure measurement error continues to be a problem, despite using different methods of exposure assessment – it still not clear how much error there are in the estimates, or which estimates are better.
- There needs to be more discussion of relative toxicity, particularly given statistically issues that can arise from combining effect estimates from multiple pollutants (would have a similar effect of linearizing and obscuring effects as errors in exposure estimates do).
- The magnitude of the risks at relatively low PM2.5 concentrations is still an open concern (see major point #5).
- There is still substantial geographic heterogeneity in effect estimates that have not been adequately explained by differences in particle composition or city- or region-specific characteristics.
- To address uncertainties in the attribution of health effects to total PM2.5 mass rather than specific constituents, I asked the expert consultants two questions:
    - Could different magnitudes of error amongst different variables in regression analyses be masking the effect of a speciated constituent of PM2.5?
    - What happens when multiple potential explanatory factors are included in a single variable in an already-complex multiple regression system? Presumably each PM2.5 component has a different C-R relationship with the health effect (even if that relationship is zero), and each is a somewhat better or worse surrogate for the relationship between actual exposure vs measured exposure. What kind of an impact would this inclusion of multiple potential explanatory factors into one variable have on the final C-R function, and how accurate would that C-R function be?
    - Based on their responses, there are still unresolved issues and potentially substantial uncertainties about concluding that the measured health effects are attributable to total PM2.5 mass, as opposed to one of the constituents. These unresolved issues include problems with a paucity of studies investigating the question, problems with measurement error amongst the measured constituents, and availability of methods that allow correlated air pollutants to be disentangled.

Major Point #5: Errors and heterogeneity in epidemiology study variables can alter the proper shape of the C-R function and obscure thresholds. Therefore, epidemiology studies with these known errors should not be used to determine the shape of the association between PM2.5 and health effects.

- Statistical analysis has demonstrated that epidemiology studies that are subject to errors such as exposure measurement error cannot accurately determine the shape of the concentration-response (C-R) function, or determine the presence of a threshold (Brauer

et al., 2002; Cox, 2018; Lipfert and Wyzga, 1996; Rhomberg et al., 2011; Watt et al., 1995; Yoshimura, 1990).

- **Notably, this concern is clearly expressed in the ISA Preamble that the EPA used as their guide for developing the PM ISA (Section 6c, pg 29):**

  "Various sources of variability and uncertainty, such as low data density in the lower concentration range, possible influence of exposure measurement error, and variability among individuals with respect to air pollution health effects, tend to smooth and "linearize" the concentration-response function and thus can obscure the existence of a threshold or nonlinear relationship. Because individual thresholds vary from person-to-person due to individual differences such as genetic differences or pre-existing disease conditions (and even can vary from one time to another for a given person), it can be difficult to demonstrate that a threshold exists in a population study. These sources of variability and uncertainty may explain why the available human data at ambient concentrations for some environmental pollutants (e.g., PM, O3, Pb, environmental tobacco smoke, radiation) do not exhibit population-level thresholds for cancer or noncancer health effects, even though likely mechanisms include nonlinear processes for some key events."

- Given the available statistical analyses, and EPA's own assessment of the ability for population or epidemiology studies to determine the shape of the C-R function and the presence of a threshold, it is unclear why the EPA continues to draw conclusions about the C-R function shape based on this type of data. To be clear, the problem is not whether a threshold in the data may exist, but rather that even if it did, the epidemiology study is not capable of identifying it.
- In general, the conclusion of a linear effect with no threshold is made multiple times in this document (pg 3-7, 3-10, 3-20, 3-21, 3-24, 3-25, 3-33, 3-41, 3-42, 3-50, 3-70, 3-96) but for all of those claims, consideration needs to be made for the problems with epidemiology studies being able to demonstrate this effect.
- In addition, the graphs that are presented in the ISA are not convincing that there is a linear shape for the C-R (figure shown in the details section).

Major Point #6: There are a substantial number of controlled human exposure (CHE) study results available that can be used not just as binary yes/no information for potential biological plausibility of epidemiology studies, but instead can provide information about dose, timing of effects, and potential sensitive populations.

- This information allows possible effect pathways to be narrowed, not just expanded, and a more specific understanding of *likely* mechanisms of action of PM2.5 should be the goal, not just a continued expansion of *possible* pathways.
- **There is one human exposure study (published in Hemmingsen et al. 2015a & 2015b) that observed effects on vascular function and heart rate variability (HRV)**

at relevant low PM concentrations and therefore should be discussed more by the CASAC and by the EPA.

Evidence-Based Considerations for PM2.5 Concentrations in Key Studies

Major Point #7: More consideration should be given to the PM2.5 concentrations at which effects occur in CHE studies, because they provide more definitive evidence of health effects from PM2.5, as well as measured exposure concentrations. **In particular, the results from Hemmingsen et al. (2015a, 2015b) need to be reviewed and the mean concentration compared to measured 5-hour concentrations in the US**.

Major Point #8: Mean PM2.5 concentrations from short-term and long-term studies should not be combined or compared.

- I asked the CASAC expert consultants whether it is appropriate to compare daily PM2.5 concentrations to the annual average? Based on the consultant responses, I have reached the conclusion that PM2.5 concentrations from short-term studies provide information for short-term health effects and are more directly comparable to the 24-hour NAAQS. They do not provide information comparable to the annual average in long-term studies nor are the daily concentrations on which these health effects are based directly comparable to the annual NAAQS.

Major Point #9: The EPA needs to carefully consider what they are measuring and comparing when they derive pseudo-design values. Pseudo-design values don't really represent concentrations or conditions for either short- or long-term studies.

- The EPA's pseudo design value method is quite confusing, and it is not clear that this is the best way to consider the adequacy of the current standard. For example, what does it mean that, for example, 25% of the study area population is in an area that met the standard? Because these conclusions are based on estimates from epidemiology studies that average the effect across multiple locations, there is no way to say that in the particular city that met the standard, that there was actually a significant association between PM2.5 and a health effect.
- I asked two questions of the expert consultants on this point:
    - Is it informative to derive annual average pseudo-design values for study areas in short-term studies (that look at effects of day-to-day PM2.5 concentration changes), in order to determine whether these study areas attained the current annual standard?
    - Answers from the expert consultants confirm my concerns that useful information is not obtained by using annual average pseudo-design values to determine if an area met the annual average standard for an epidemiology study looking at short-term PM2.5 changes.

    - Is the pseudo-design value in a single geographic area particularly informative, when the association between PM2.5 and the health effect is driven by the differences between study areas?
    - This seems like a point that needs to be further addressed by the EPA to justify their use of pseudo-design values to determine if PM2.5-associated health effects were occurring in areas that met the current or alternative NAAQS.
- An alternative to EPA's pseudo-design value analysis would be to assess the design values in particular cities for which city-specific short-term mortality estimates are available (e.g. Franklin 2007, 2008, Dai et al. 2014, Baxter et al. 2017, etc). The results can be separated by those cities that have positive associations between PM2.5 and mortality, and those with negative associations (setting aside the important consideration of statistical significance for the moment). I conducted an example assessment of this kind with the results in my details section.
- The EPA's summary of their pseudo-design value analysis needs to be clarified.

Risk-Based Considerations

Major Concern #10: I have concerns about the methods used to derive the risk estimates.

- More stringent criteria are used to choose epidemiology studies for the pseudo-design value analysis than are used for risk quantification in the risk analysis. In the pseudo-design value calculation, a very thoughtful set of criteria were applied to included studies based on the quality of the exposure assessment. Studies chosen for the risk assessment should have similar study quality criteria applied to them. Of the 8 studies used in the risk assessment, only three met the exposure qualifications in the pseudo-design value analysis: Di et al. 2017b, Zanobetti et al. 2014, and Baxter et al. 2017. There also does not seem to be a consistent application of criteria concerning studies with populations not readily generalizable to the broader US population.
- The EPA makes a big leap from C-R functions derived in epidemiology studies, to the applicability of those functions to risk estimation. For example, to be applicable for risk estimates the C-R functions have to assume causality (discussed above and elsewhere); if the endpoint is not all-cause mortality then there needs to be consideration of competing risks; considerations of effects of measurement error and confounding on the C-R function; the equations used to apply the C-R function to a population estimate. Very little methodological detail is provided in this document for how the risk assessment is done and why particular mathematical choices are made, beyond just referencing the BenMAP manual. This information needs to be expanded and the above concerns addressed to generate realistic risk estimates.
- Hazard ratios and relative risks are conceptually conflated in this document, and in some cases used interchangeably. This is incorrect, as has been well-documented by Sutradhar and Austin, 2017, and others.

- The more concerning conflation, however, is the substitution of beta-coefficients derived from Cox proportional hazards models into time-series-type equations for the purposes of estimating population risk. As discussed in the details section, this replacement is not equivalent, and **to do properly requires population estimates of the instantaneous hazard of the event occurring at a specific point in time in the population where all covariates are set to zero. The EPA uses (population * mortality rate) to estimate this function without justification of why that is an appropriate substitution.** The EPA needs to justify the use of these values in this equation, if they want to use this as the basis of their quantitative risk assessment.
- It seems like there hasn't been a significant reduction in attributable absolute risk from the last standard (set at 15 ug/m3) to this standard. The 2010 PM risk assessment attributed 3-17% of IHD mortality (depending on the location) to PM2.5 concentrations at the 15 ug/m3 standard, and this risk assessment attributes 13-14% of IHD mortality to PM2.5 concentrations at the 12 ug/m3 standard. They use different effect estimates, but they are both from the ACS study (Pope 2015 compared to Krewski 2009), and the Kreswksi effect estimate is actually higher. Some explanation of the changes in risk estimates between assessments would be helpful for understanding whether risk has been decreased, or just generally what kind of variability in estimates can be expected from changes in risk assessment methodology.

Major Concern #11: There is very little quantitative uncertainty analysis provided with this risk assessment.

- It is essential that the EPA begins to capture more of the uncertainty in their risk analysis using quantitative uncertainty methods. There was wide agreement amongst the expert consultants that this type of method both should and could be done.
- Even when the EPA notes the potential causes for some of the uncertainty (e.g. broad 95% CIs), they do not provide information about the significance of this uncertainty or how it should be interpreted.
- The EPA incorrectly concludes that uncertainties in the shape of the C-R function at low PM2.5 concentrations would not differentially affect the risk estimates between the current and alternative standards. However, Figure 3-12 in the PA shows that the concentrations of PM2.5 to which risk is being attributed change with changing standards, and therefore the shape of the C-R function at those concentrations will differentially affect the risk estimates.
- Just taking into account the uncertainty quantified by the 95% CIs of the C-R functions, the risk estimates between the current standard and the alternative standards overlap, showing that there does not seem to be an expectation of a statistically significant decrease in risk with a decrease of the PM2.5 annual standard.

Future Research Recommendations:

- If EPA is going to support future research on determining the mechanisms of action behind epidemiology findings, then they also need to determine *a priori* how they will handle negative findings, as well as which epidemiology associations they will support for investigation. Similarly, future research into shapes of C-R functions should be supported by research to determine how epidemiology studies with errors in the data can determine the underlying shape of the C-R function.
- The EPA should support research that focuses on using causality methods and determining causal pathways for the potential associations between PM2.5 and various health endpoints.
- The EPA should support further development of quantitative uncertainty methods.
- The EPA should support further development of quantitative risk assessment methods.

Conclusion

All that being said, the EPA Administrator is making a decision (and the CASAC is making a recommendation) based on the data and analyses as they stand today. Both the desire for, and difficulties of, using causal analytics models is well-summarized in a recent commentary by Carone et al., 2019. Those authors note that "greater adoption of cutting-edge data science tools and causal inference principles into mainstream air pollution epidemiology as an important step forward", but the authors also state that we cannot wait on obtaining the perfect data or analyses before making policy decisions to protect public health. Indeed, the EPA has not waited, having set their first standards for PM2.5 more than 20 years ago, based on far less data and more rudimentary analyses than we have now. The question faced today is whether the available data and analyses support the current standard or justify a lower standard.

Using the last review as a benchmark, the question is: has the risk assessment changed since the last standard was set? i.e. is there a reason to expect that the risks are greater or more certain at the current standard than was known 7 years ago, and if so, what would be an acceptable level of risk? I am referring to the entire assessment as a risk assessment, not just the quantified portion. The ISA is a hazard assessment and considers dose-response to a limited extent. The REA/PA is a dose-response assessment, exposure assessment, and risk integration/characterization.

Hazard Identification – This has not substantively changed since the last assessment. Most of the causality designations are the same, and the ones that have been upgraded from *suggestive* to *likely* are those that CASAC expressed concerns with. Even if there was more certainty in those new endpoints, they don't provide evidence that risks are occurring at lower concentrations. In the last review the EPA already expressed their greatest degree of certainty in the association between PM2.5 concentrations and mortality and CVD, so the certainty for those key endpoints by definition cannot be greater in this review.

Dose-Response – In the last review the EPA concluded that the dose-response for the major hazards (total mortality, cardiovascular and respiratory morbidity and mortality) were linear with no threshold. The EPA is concluding the same thing in this review, and the steepness of the slopes of the linear estimates is similar between the last review and the current review. Therefore, there is nothing changed in the dose-response.

Exposure – PM2.5 concentrations have continued to decrease nationwide since the last review. Fundamentally, the standard controls health risk by changing the exposure. Most of the exposure data being measured or modeled in the epidemiology studies is from the early 2000s with no data later than 2013. Therefore, the impact of lowering the standard in 2012 hasn't been assessed or captured in these studies.

Risk Integration – I don't consider the quantitative risk estimates to be reliable. There are concerns with the mathematics used to derive risks from hazards (see my detailed comments), about the causal estimands based on associative studies (noted in Carone et al., 2019), the lack of uncertainty estimates, etc. However, because the EPA is using a linear to zero dose-response with essentially the same slope as the last review, it does not matter what the absolute risks are because every standard above zero will be associated with some risk. Equally we cannot use a cut-off for how much absolute risk is acceptable (e.g. 5,000 deaths is ok, but not 10,000 deaths), because that is arbitrary and untenable, as well as being too dependent on the accuracy of the methods used to estimate the risk.

So where does that leave us? With the assessment of mean concentration and pseudo-design values for which there have also been concerns raised? There is no real expectation that any epidemiology study will be conducted that doesn't show an effect at decreasing exposure concentrations because of the way these studies are modeled and because of the errors that linearize the associations and obscure thresholds. The known noise in the data prevents the kind of granular information that is required to discern between standards that are only a few $\mu g/m^3$ different. This same noise, and the lack of methods demonstrating manipulative causality, make it very difficult to predict whether changing the standard will have any impact on public health.

Chapter 4. Review of the Primary Standards for PM10

Charge Question: What are the CASAC views on the approach described in chapter 4 to considering the PM10-2.5 health effects evidence in order to inform preliminary conclusions on the primary PM10 standard? What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current primary PM10 standard?

4.2.1.1 Cardiovascular Effects – Long-Term Exposures

The EPA notes in this section that "The evidence relating long-term PM10-2.5 exposures to cardiovascular mortality remains limited, with no consistent pattern of associations across studies and, as discussed above, uncertainty stemming from the use of various approaches to estimate PM10-2.5 concentrations (U.S. EPA, 2018, Table 6-70)." Although EPA states at the end of this paragraph that there are some high-quality studies that have shown a positive association between PM10-2.5 and cardiovascular endpoints, overall this is not a convincing summary to justify the upgraded "suggestive" causal determination. For example, the EPA cites information in Figures 6-34 and 6-35 from the ISA as showing positive associations with IHD, MI and stroke, but none of the associations in those figures were statistically significant, with several studies being completely null or negative.

Similarly, for long-term metabolic effects, the suggestive designation is based on a single epidemiology study with non-statistically significant effects. This is not convincing of a "suggestive" causal designation. The "suggestive" causal designations for short-term metabolic effects, and nervous system effects are similarly poorly supported and unconvincingly summarized in this PA.

Given the relative paucity of data and causal associations between PM10-2.5 and health endpoints, I support the EPA's rationale for recommending that the PM10 standard does not need to be changed.

Detailed Information about Major Points

Major Point #2 – Causality Determinations

3.2.1.1 Mortality: Long-Term PM2.5 Exposures:

On pg 3-20, the EPA discusses the results from several recent "accountability" studies, conducted by Pope et al. 2009 and Correia et al. 2013. These studies evaluate whether life expectancy is correlated with PM2.5 concentrations over two to three time periods in various areas of the US. Considering these papers to be accountability studies is new for this document, because the term wasn't used in the ISA. Because accountability is a newly introduced topic for this document, the EPA should provide some background and information about what accountability studies are supposed to assess. There are several recent reviews that provide this information (Henneman et al., 2017; Rich, 2017). Those reviews describe accountability studies as "assessments of past environmental policies" (Henneman) that "evaluate the extent to which an air pollution improvement of regulation in a city or region beneficially impacted public health". The strength of these studies is often in their ability to control for confounders that cannot be controlled for in typical epidemiology studies through the use of control areas or populations that were not subject to the environmental policy, but otherwise were similar to the

areas that were regulated. These allow the researcher to move to the second step of the ladder of causality, which addresses interventions and not just associations (associations are step one; Pearl and Mackenzie, 2018). The Pope et al. 2009 paper describes the association between changes in life expectancy and PM2.5 in 1979-1993 and 1999-2000 in 51 metropolitan areas in the US. However, although air pollution was reduced during this timeframe, PM2.5 regulations were not introduced at the federal level until 1997, so this analysis is not an assessment of past environmental policies. It also lacks the feature that makes natural experiments so valuable – there is no control population to demonstrate that changes over time in important considerations like healthcare (e.g. improved diagnostics, preventative care, and disease treatment) are not responsible for the change in life expectancy. Although Correia et al. (2013) does include a timeframe under which PM2.5 regulations would have been enacted (2007), there is still a lack of control population to ensure that the changes in life expectancy are truly related to PM2.5. These studies are better described as time-series studies that use life expectancy as the dependent variable and that use changes over time of PM2.5 concentrations to investigate the association, rather than differences in PM2.5 concentrations across space (as is often done with long-term cohort studies). Because these studies that consider the associations between PM2.5 and mortality over time are included and discussed in this ISA, so too should the studies by Greven et al., 2011 and Pun et al., 2017. Those papers include a sophisticated analysis of these same considerations: that is, the association between the national trend in PM2.5 and mortality over time. In addition, they consider the change in local concentrations of PM2.5 that are different from the national trend, with the hypothesis that whether one is looking at national-level time trends or local-level time trends, the PM2.5-mortality association should be the same. This was not the case – in both studies, the authors found an association between PM2.5 and mortality at the national level (similar to the results of Pope et al. 2009 and Correia et al. 2013), but no association between PM2.5 and mortality at the local level. This suggests an uncontrolled confounder is the actual culprit in the association between national trends in PM2.5 and mortality, which emphasizes the importance of having a proper control population, and calls into question the results from these longevity studies. In addition, in contrast to the results presented by Di et al. 2017 and Shi et al. 2016, when Correia et al. restricted their analyses to counties with year 2000 PM2.5 concentrations of < 10 ug/m3, the association between life expectancy and reductions in PM2.5 became non-significant. For concentrations < 12 ug/m3 there was a non-significant positive association between PM2.5 reduction and life expectancy. This suggests a threshold in the analyses of Correia et al. In addition, this study assumes a linear relationship between life expectancy and PM2.5 concentrations, but the data using cutpoint analyses shows different slopes with different concentration cutpoints, suggesting a non-linear association.

3.2.1.1 Mortality: Long-Term PM2.5 Exposures:

Adequate consideration of bias, chance, and confounding:

- Bias: Likely to be considerable exposure measurement error or misclassification bias; particularly impacts the estimates of the shape of the C-R function as well as the ability to identify a threshold;

- Chance: many of the studies show statistically significant results; no information about p-hacking or publication bias that could impact the judgement of whether chance has been considered;
- Confounding: evidence of unmeasured and important confounding from the work by Greven et al., 2011 and Pun et al., 2017; regional heterogeneity not adequately explained by city-specific characteristics or pollutant sources.

3.2.1.1 Mortality: Short-Term PM2.5 Exposures:
Adequate consideration of bias, chance, and confounding:

- Bias: Likely to be considerable exposure measurement error or misclassification bias as noted by (Avery et al., 2010a, 2010b) particularly impacts the estimates of the shape of the C-R function as well as the ability to identify a threshold;
- Chance: many of the studies show statistically significant results; no information about p-hacking or publication bias that could impact the judgement of whether chance has been considered;
- Confounding: I haven't seen the same kinds of analyses as the Greven work done in these types of studies, but there have been concerns raised by CASAC about adequate control for temperature, and the ecological nature of these studies makes them particularly weak at discerning more than surface correlations, not full causation; Unexplained substantial city-by-city heterogeneity also suggests unmeasured confounding.

3.2.1.2. Cardiovascular Effects:
The EPA notes in section 3.2.3.1 that "controlled human exposure studies provide limited insight into the occurrence of cardiovascular effects following PM2.5 exposures likely to occur in the ambient air in areas meeting the current primary PM2.5 standards and are of limited utility in informing conclusions on the public health protection provided by the current standards." However, the EPA uses the CHE studies to establish biological plausibility for the occurrence of cardiovascular effect following PM2.5 exposures at ambient concentrations. Causality determinations should be informed not only by whether an effect occurs at any concentration in a CHE study, but also by whether that effect occurs at ambient concentrations relevant to setting the standard.

Similarly, in the same section the EPA notes that "there is uncertainty in extrapolating the effects seen in animals, and the PM2.5 exposures and doses that cause those effects, to human populations.". They also state that "Most of the animal toxicology studies assessed in the draft ISA have examined effects following exposures to PM2.5 concentrations well-above the concentrations likely to be allowed by the current PM2.5 standards. Such studies have generally examined short-term exposures to PM2.5 concentrations from 100 to >1,000 µg/m3 and long-term exposures to concentrations from 66 to >400 µg/m3 (e.g., see U.S. EPA, 2018, Table 1-2)." This uncertainty should apply to considerations of biological plausibility in the causality designation, not just to quantifying the standard.

3.2.1.4 Cancer:
CASAC expressed substantial concerns with the "likely to be causal" designation for PM2.5 and cancer. The primary basis for this concern was the exposure assessment in studies of lung cancer mortality: in most of the epidemiology studies, the exposures were measured only slightly before, concurrently, or *after* the cancer mortality. From the CASAC PM ISA comments: "However, the issue of the long lag time that can exist between the inciting exposure and the first clinical signs of cancer is not adequately addressed in the ISA. Most of these studies evaluated PM2.5 exposures a few years before cancer diagnosis or death. Over these time frames, it is likely that most of the lung cancer cases already had the disease, albeit in a pre-clinical state, at the time the exposure was assessed. Thus, the findings in these studies may reflect reduced survival of already incident cancer, rather than true increased lung cancer incidence." There are also no animal studies showing direct effects of PM2.5 on cancer formation, with the only positive animal results coming from a group that pre-initiated the animals with urethane.

3.2.1.5. Nervous System Effects: Long-Term PM2.5 Exposures:
The following is directly from the CASAC letter to EPA about the Draft PM ISA: "The EPA does not provide adequate evidence for the conclusion that there is likely to be a causal association between long-term PM2.5 exposure and nervous system effects. In Table 8-20, the EPA identifies the following as providing high quality or consistent evidence of this relationship: toxicology studies on brain inflammation and reduced cognitive function, and epidemiology studies of reductions in brain volume and reduced cognitive function in adults. For a likely causal conclusion, there would have to be evidence of health effects in studies where results are not explained by chance, confounding, and other biases, but uncertainties remain in the overall evidence. In addition, the determination should be made based on multiple studies by multiple research groups (p. P-12). The toxicology studies have largely been done by a single group. Those animal toxicology studies that were completed by other groups do not provide adequate evidence because the control animals were exposed to gaseous pollutants (Tyler et al., 2016) or were exposed for only two weeks in addition to OVA-sensitization (Campbell et al., 2005). For the brain size epidemiology studies, brain volumes were only measured once in each person and were compared between people. But brain volume can vary up to two-fold between normal people (Reardon et al., 2018), so this seems like an endpoint that could be subject to substantial error. Additionally, the cognitive function epidemiology studies found largely non-statistically significant results (see Figures 8-3, 8-4, and 8-5), including two of the studies that the EPA cited in Table 8-20 (Weuve et al., 2012 and Tonne et al., 2014). Altogether, this data does not provide evidence of health effects that are not explained by chance, confounding, or bias, and that have been done by multiple research groups."

3.2.1.5. Nervous System Effects: Long-Term UFP Exposures:
The following is directly from the CASAC letter to EPA about the Draft PM ISA: "The ISA does not provide adequate evidence to support the conclusion that there is likely to be a causal association between long-term UFP exposure and nervous system effects. There are no

supportive human studies, and the EPA has not considered the appropriate dosimetric adjustments, or rodent-to-human differences in the respiratory tract, that would help extrapolate the animal data to humans. In addition, most of the animal studies that provide coherence were done by a single group in a single location."

Major Point #3 – Balanced and Accurate Reporting of Results

Section 3.2 – Evidence Based Considerations:
At the beginning of this section the EPA notes that: "The draft ISA uses a weight-of-evidence framework for characterizing the strength of the available scientific evidence for health effects attributable to PM exposures (U.S. EPA, 2015, Preamble, Section 5). This framework provides the basis for robust, consistent, and transparent evaluation of the scientific evidence, including its uncertainties, and for drawing conclusions on PM-related health effects."

However, the CASAC provided substantial comments in their review of the PM ISA noting that the EPA's framework did not provide a consistent and transparent evaluation of the scientific evidence, and specifically lacked: inclusion and exclusion information for studies in particular chapters, clear application of study quality assessment when deriving conclusions from included studies (such as consideration of bias, chance, and confounding in epidemiology studies), and transparent methods for weighing contradictory evidence (such as blood pressure findings discussed above). The EPA should more accurately portray the benefits and shortcomings of their system for the ISA in this document.

3.2.1.1 Mortality: Long-Term PM2.5 Exposures:
The EPA notes that "the draft ISA concludes that positive associations between long-term PM2.5 exposures and mortality are robust across recent analyses using various approaches to estimate PM2.5 exposures (e.g., based on monitors, modeling, satellites, or hybrid methods that combine information from multiple sources) (U.S. EPA, 2018, section 11.2.5.1), across statistical models (U.S. EPA, 2018, section 11.2.5.2), across diverse geographic regions and populations, and across a range of temporal periods including the periods of declining PM concentrations". This summary over-generalizes the results and obfuscates the complexity of the data. For example:

- Addressing the robustness of results across recent studies: Di et al (2017b) showed amongst a group of recent studies of PM2.5 assessing mortality, that there was significant heterogeneity in the effect estimates between studies using a random effects meta-analysis method (I^2 = 95.9%; Figure S6).
- Addressing the robustness of associations across exposure measurement methods: One of the key studies used in this risk assessment, Jerrett et al. (2017) [Note: the publication year for this study is mislabeled as 2016 in this PA], provides effect estimates using the same mortality data from the ACS CPSII cohort, but different methods for modeling exposure. They found a close to 7-fold difference in the slopes of the associations

between models (compare Ln(1.02) to Ln(1.14)). While all the associations were positive, they do not seem to be robust to modeling choices.

- Addressing the consistency across diverse geographic regions and populations: likely the EPA means here that the referenced studies covered many geographical locations. However, the EPA did not assess whether there was consistency in associations between long-term PM2.5 concentrations in different regions (because the studies typically represent all the study areas, not separate areas). Zeger et al. 2008 provides effect estimates for 3 different regions, and there are not consistent relationships between PM2.5 and mortality in these regions (slightly negative effect in the Western region, positive in Eastern and Central regions). Similarly, Kioumourtzoglou et al (2016) divided the US into 8 regions, and demonstrated positive effects in the Southeast, South, and Northwest, no significant effect in the Northeast, Central, West, and North Central regions, and a significant negative association in the Southwest (see figure reproduced below). While potential effect modifiers have been offered by in this study, it is not clear how much of the heterogeneity is explained by these modifiers. In another study by the same authors, Kioumourtzoglou et al (2015) investigate effect modification of city-by-city hetereogeneity in long-term PM2.5 mortality associations by considering PM2.5 speciation. Some of the heterogeneity may be explained by PM2.5 species, but again it is difficult to tell how much.

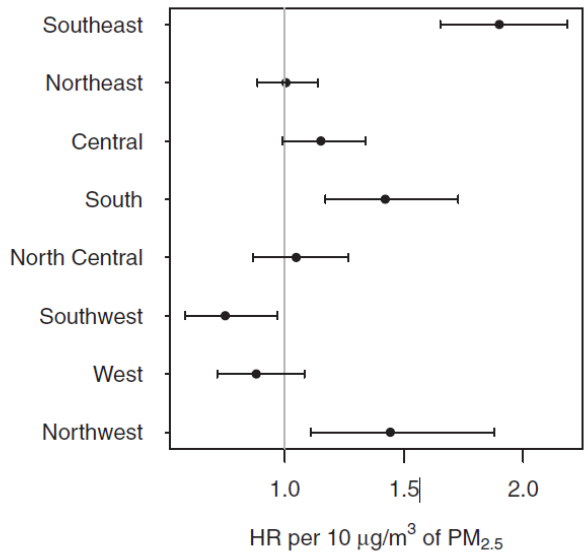


FIGURE 1. By-region $PM_{2.5}$–mortality effect estimates, presented as HRs (95% CI) per 10 µg/m^3 of $PM_{2.5}$.

Kioumourtzoglou et al. (2016), Figure 1 demonstrating regional heterogeneity in long-term PM2.5-mortality associations.

- The statement about consistency over multiple temporal periods does not consider the work of Greven et al. 2011 and Pun et al. 2017 that demonstrate the problems with those types of analyses, or Correia et al. 2013 who showed that the associations with changes in lifespan were not seen in the later period of their study (greater details of these points are provided under Major Point #2).

EPA also notes that "associations persist in analyses restricted to long-term exposures below 12 µg/m3 (Di et al., 2017b) or 10 µg/m3 (Shi et al., 2016) (i.e., indicating that risks are not disproportionately driven by the upper portions of the air quality distribution);". This is not the case with Correia et al. (2013) as noted above, who did not see significant effects of PM2.5 reduction on life expectancy when only using counties with PM2.5 concentrations less than 10 ug/m3, or less than 12 ug/m3 (Table reproduced below).

eTable 3: Summary of selected regression analyses by baseline $PM_{2.5}$ levels for 545 counties (Dataset 1, 2000 – 2007)

Selected counties and analysis	# Counties	β (SE, p) for 10 µg/m^3 $PM_{2.5}$ (full model)
2000 $PM_{2.5}$ < 10µg/m^3	100	-0.28(0.39, 0.482)
2000 $PM_{2.5}$ < 12µg/m^3	186	0.50(0.27, 0.065)
2000 $PM_{2.5}$ < 14µg/m^3	301	0.61(0.21, 0.004)
2000 $PM_{2.5}$ < 16µg/m^3	430	0.36(0.19, 0.064)
2000 $PM_{2.5}$ < 18µg/m^3	511	0.47(0.18, 0.009)
2000 $PM_{2.5}$ > 18µg/m^3	34	0.85(0.82, 0.314)
2000 $PM_{2.5}$ > 16µg/m^3	115	0.87(0.38, 0.023)
2000 $PM_{2.5}$ > 14µg/m^3	244	0.28(0.27, 0.305)
2000 $PM_{2.5}$ > 12µg/m^3	359	0.15(0.21, 0.462)
2000 $PM_{2.5}$ > 10µg/m^3	445	0.27(0.18, 0.126)

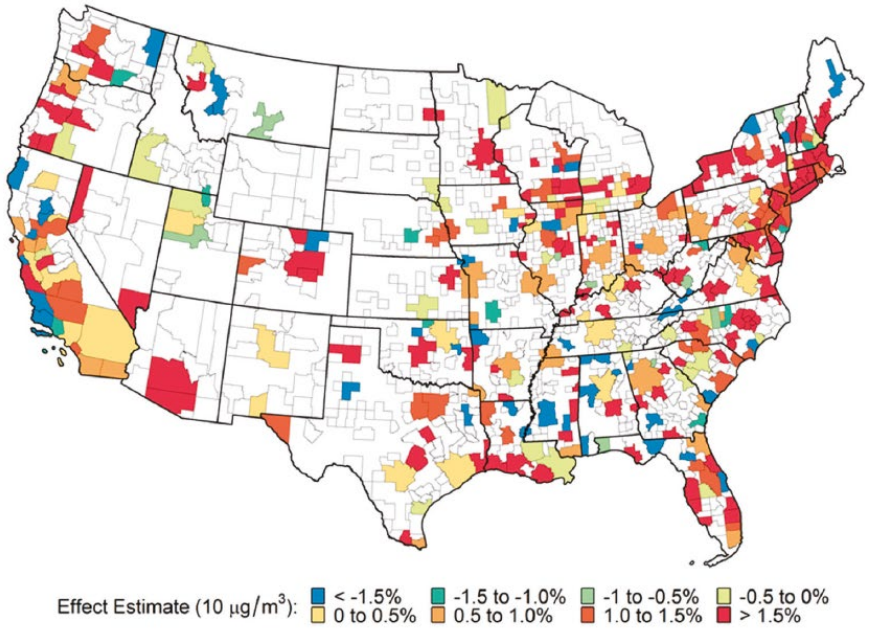
*Corresponds to the covariate pattern in Model 3 of Table 2 (main text) or Model 3 of eTable 2b. Covarites include change in income, change in population, change in high-school graduates, change in proportion of black population, change in proportion of Hispanic population, change in lung cancer mortality rate, change in COPD mortality rate. Analysis used: STATA 11.0, REGRESS clustered by MSA, using the "weight" statement.

Correira et al. (2013) eTable 3 demonstrating a lack of PM2.5 association with longevity in counties with PM2.5 concentrations <10 and <12 ug/m3.

3.2.1.1 Mortality: Short-Term PM2.5 Exposures:

Regional heterogeneity in associations between PM2.5 and mortality is still an open question. Despite substantial analyses that have considered housing characteristics, commuting, household heating type, meteorological features, and poverty, only up to 13% of the variability amongst cities has been explained (Baxter et al. 2018; Figure reproduced below). This was one of the concerns raised by the EPA Administrator when the PM2.5 NAAQS was set in 2012 (pg 3-9) and should be specifically addressed in this document when making policy recommendations.

Fig. 1 Area-specific associations of total non-accidental mortality and fine particulate matter (PM2.5) at lag 1: 312 US core-based statistical areas and metropolitan divisions



Baxter et al. 2018, Figure 1, demonstrating regional heterogeneity in associations between short-term PM2.5 concentrations and mortality.

The EPA notes that there is "strong evidence for ischemic events and heart failure" which supports the conclusion of increased CVD mortality. However, there is not strong evidence of ischemic events – Figure 6-2 and 6-3 in the ISA shows mostly null and non-statistically significant associations between PM2.5 and HA or ED visits for IHD and HF. Similarly, most of the CHF studies showed non-statistically significant results.

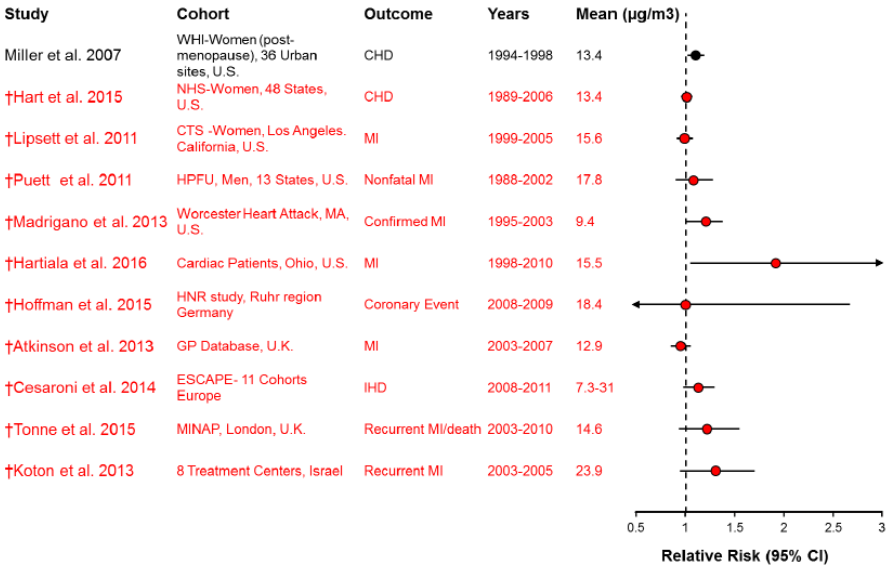
The EPA presents the short-term mortality studies as "primarily positive", when in fact there are some negative studies and some non-statistically significant positive studies. This type of variability needs to be considered when EPA is presenting study results.

3.2.1.2 Cardiovascular Effects: Long-Term PM2.5 Exposures:

The EPA notes in this section that "Positive associations with cardiovascular morbidity (e.g., coronary heart disease, stroke) and atherosclerosis progression are observed in several epidemiologic studies (U.S. EPA, 2018, sections 6.2.2. to 6.2.9)". This is particularly important because one long-term mortality endpoint used in the risk assessment in this PA is ischemic heart disease (IHD). However, there are considerable inconsistencies in some of the cardiovascular morbidity endpoints, such as with associations between long-term exposure to PM2.5 and IHD and myocardial infarction, which are overwhelmingly not statistically significant (Figure 6-17, reproduced below). Of the presented effect estimates only 1 out of 11 is statistically significant,

and 4 out of the 6 US studies used PM2.5 exposure estimates from before there was a nationwide monitoring network (thereby requiring extrapolation from other datasets). Negative findings for morbidity endpoints do not provide consistent or strong evidence that it is appropriate to model the risk for the associated mortality endpoint (i.e. IHD). A similar pattern of non-significant effect estimates is seen with stroke (Figure 6-18), even though the EPA specifically calls out stroke as an endpoint with positive associations.

If the EPA is intentionally focusing on those studies that show positive health effects of PM2.5 exposure, regardless of whether the literature also includes negative studies, then they should state that up front so that it is clear to readers that only positive evidence is being summarized.



†Studies published since the 2009 Integrated Science Assessment for Particulate Matter.

Circles represent point estimates; horizontal lines represent 95% confidence intervals for $PM_{2.5}$. Black text and circles represent evidence included in the 2009 PM ISA; red text and circles represent recent evidence not considered in previous ISAs or AQCDs. Mean concentrations in $\mu g/m^3$. Hazard Ratios are standardized to a 5 $\mu g/m^3$ increase in $PM_{2.5}$ concentrations. Corresponding quantitative results are reported in Supplemental Table 6S-16 (U.S. EPA, 2018). WHI = Women's Health Initiative; CHD = Coronary Heart Disease; MI = Myocardial Infarction; IHD = Ischemic Heart Disease; NHS = Nurses Health Study; CTS = California Teachers Study; HPFU = Health Professionals Follow-up Study; ESCAPE = European Study of Cohorts for Air Pollution; HNR = Heinz Nixdorf Recall study; MINAP = Myocardial Ischemia National Audit Project.

Figure 6-17 Associations between long-term exposure to $PM_{2.5}$ and Ischemic Heart Disease or Myocardial Infarction. Associations are presented per 5 $\mu g/m^3$ increase in pollutant concentration.

2018 PM Draft ISA, Figure 6-17, demonstrating the largely non-significant effect estimates in studies of long-term PM2.5 exposure on ischemic heart disease or myocardial infarction.

Other endpoints whose presentation is misleading are long-term exposure and markers of systemic inflammation, coagulation, and endothelial dysfunction. The EPA presents these as having positive associations, but those are only seen with the longitudinal studies. Cross-

sectional studies for these endpoints are overwhelmingly null, and this discrepancy should be presented. If the EPA thinks that longitudinal studies are better able to determine important differences, then this kind of study quality consideration should be included in the discussion.

3.2.1.3 Respiratory Effects – Short-Term PM2.5 Exposures:
On page 3-44 of this document the EPA notes in a footnote that "In contrast, controlled human exposure studies provide little evidence for respiratory effects following short-term PM2.5 exposures (U.S. EPA, 2018, section 5.1, Table 5-18)." This information is not communicated in the section describing the respiratory effects of short-term PM2.5 exposure. Given the large number of studies that have investigated the respiratory effects of short-term PM2.5 exposures in CHE studies, the lack of effects needs to be discussed in this section and adequately communicated to the Administrator.

3.2.1.7 Summary:
The considerations and uncertainties described above should be included in the summary to adequately communicate both the supportive and the non-supportive evidence for these health endpoints. For example, the statement "Recent epidemiologic studies consistently report positive associations between long-term PM2.5 exposures and a wide range of health outcomes, including total and cause-specific mortality, cardiovascular and respiratory morbidity, lung cancer, and nervous system effects" is misleading about the breadth and depth of the results, including many results that do not report positive associations, and endpoints for which few studies have been done (e.g. nervous system effects). As noted elsewhere in these comments it is certainly not appropriate to represent the results seen in various studies as "consistently positive".

3.2.2 Potential At-Risk Populations:
EPA should include discussion about the strength of the evidence for each different potential at-risk population (i.e. the causality determination for each one, because the causality structure is different than for the health effect endpoints).

3.2.3.2.1 PM2.5 Air Quality Distributions Associated with Mortality or Morbidity in Key Epidemiologic Studies
The EPA states on page 3-61 that "Based on the information in Figure 3-3 to Figure 3-6 and Table 3-3, key epidemiologic studies conducted in the U.S. or Canada indicate generally positive and statistically significant associations between estimated PM2.5 exposures (short- or long-term) and mortality or morbidity across a wide range of ambient PM2.5 concentrations." However, these figures also demonstrate a number of non-positive or non-statistically significant associations between estimated PM2.5 exposures and mortality or morbidity. For example, lung cancer shows few statistically significant associations, as do long-term and short-term PM2.5 exposures with respiratory morbidity.

The choice of studies for use in Figure 3-8 included an excellent description of model validity and exposure data quality. This kind of assessment should be extended to all epidemiology studies that use modeling data for their exposure assessment, not just for this section. Quality assessments of modeled data should also consider that models sometimes adjust for monitored data, but they may end up getting the right answer for the wrong reasons. This problem can lead to models that fit past concentrations but may not accurately predict future concentrations.

3.4.1 Preliminary Conclusions on the Current PM2.5 Standards

The EPA notes in their summary that certain lifestages may be at comparatively higher risk of experiencing PM2.5-related health effects, including older adults. However, in the PM ISA Chapter 12, which investigates the evidence of subpopulations who are more sensitive to PM, the EPA concludes that "Overall, while PM2.5-associated effects are observed in older adults, evidence is inadequate to determine if older adults are at increased risk for effects compared to younger adults." Therefore, the summary in the PM PA mischaracterizes the evidence in the PM ISA about the additional risk to older adults.

In addition, the EPA provides a summary of the risk estimates in this section. However, the estimates provided are the highest of the estimates for the total and IHD mortality calculations, instead of including the range of estimates or the CI of the estimates. The EPA should more accurately summarize these risk estimates by providing ranges, preferably both with information from the CI and the different C-R functions. For example, instead of: "the risk assessment estimates up to about 50,000 total PM2.5-related deaths, including almost 20,000 ischemic heart disease deaths, in a single year", it should state something like "the risk assessment estimates total PM2.5-related deaths in the range of 13,500 (2,360-24,200) to 52,100 (41,600-62,300), including approximately 15,600 (11,600-19,400) to 16,800 (12,800-20,500) ischemic heart disease deaths." This modification should be included in the other summary areas as well. For example, when the EPA states that "potential alternative annual standards with levels from 11.0 down to 9.0 µg/m3 could reduce PM2.5-associated mortality broadly across the U.S., including in most of the 47 urban study areas evaluated.", the quantification should be included so that readers understand the amount of uncertainty in the estimates. For example, the EPA could state that "the absolute risk estimates for total PM2.5-related deaths for an alternate standard of 11 ug/m3 were in the range of 10,700 (1,880-19,300) to 41,000 (32,800-49,100); for an alternate standard of 10 ug/m3 were in the range of 9,710 (1,700-17,500) to 37,800 (30,200-45,300); and for an alternate standard of 9 ug/m3 were in the range of 8,650 (1,510-15,600) to 34,600 (27,600-41,500)." This provides risk estimates as well as some consideration of uncertainty to help the reader judge accuracy.

Major Point #4: Uncertainties

3.2.1.1 Mortality: Long-Term PM2.5 Exposures:
Remaining uncertainties:

- Evidence for PM2.5 effects gets weaker as the health endpoint becomes less severe (less evidence for HAs and ED visits than for mortality; even less evidence for less severe effects such as changes in heart rate, blood pressure, inflammation).
- Unaddressed concerns about bias in the estimates and evidence of unaccounted for confounding.
- Needs to be additional discussion of experimental evidence that is available for chronic animal studies. The lack of mortality observed in these studies should be discussed.

3.2.1.1 Mortality: Short-Term PM2.5 Exposures:
Remaining uncertainties:

- EPA notes in their biological plausibility section that the evidence for how short-term PM2.5 exposure can lead to downstream effects on CVD and respiratory morbidity (and from there to mortality) is "limited". In addition, there needs to be more information about how short-term exposures to PM2.5 at ambient concentrations could be leading to a physiological response that results in death.
- There are unaddressed concerns about bias in the estimates and innate weaknesses in the ecological epidemiology studies that provide the vast majority of the evidence.
- There needs to be additional discussion of experimental evidence that is available for short-term animal studies. The lack of mortality observed in these studies should be discussed. Also, the minor results in human CHE studies.

3.2.1.7 Summary:
The summary section reiterates the original question: "To what extent does the currently available scientific evidence strengthen, or otherwise alter, our conclusions from the last review regarding health effects attributable to long or short-term fine particle exposures? Have previously identified uncertainties been reduced? What important uncertainties remain and have new uncertainties been identified?

The uncertainties previously identified by the EPA Administrator were "related to understanding the relative toxicity of the different components in the fine particle mixture, the role of PM2.5 in the complex ambient mixture, exposure measurement errors in epidemiologic studies, and the nature and magnitude of estimated risks related to relatively low ambient PM2.5 concentrations. Furthermore, the Administrator noted that epidemiologic studies had reported heterogeneity in responses both within and between cities and in geographic regions across the U.S."

The EPA notes that more studies have demonstrated robust effects in copollutant analyses. However, none of the remaining factors (relative toxicity, exposure measurement error, risks at

low PM concentrations, or geographic heterogeneity) are addressed in this section. Because these factors were identified as key and important uncertainties, they should be specifically addressed to communicate whether recent evidence has reduced or changed these uncertainties.

In addition, in the brief discussions about study inconsistencies in this section, there is a list of reasons why studies might produce different results, but no mention of study quality in the evaluation.

3.4.1 Preliminary Conclusions on the Current PM2.5 Standards

In this section the EPA states that "Studies published since the last review have reduced key uncertainties and broadened our understanding of the health effects that can result from exposures to PM2.5." As noted above, the EPA should specifically address the uncertainties identified by the EPA Administrator in the last review. It seems from my review of the data that while more studies have been conducted since the last ISA that consider uncertainties like copollutants, C-R functions, regional heterogeneity, and PM2.5 components and sources, none of them really clarifies any of the underlying uncertainty. There are still unknowns with copollutants, C-R functions are still plagued by problems with innate variability that makes them difficult to interpret, none of the studies on regional heterogeneity adequately explained the reasons for the city-specific heterogeneity, and it is not clear what components or sources are causing the observed effects. Therefore, it does not seem that many of the key uncertainties have been reduced in this review.

3.4.2.1 Potential Alternative Standards – Indicator:

The EPA states here and elsewhere that "many PM2.5 components and sources are associated with health effects, and the evidence does not indicate that any one source or component is consistently more strongly related with health effects than PM2.5 mass". As has been stated by others, this is an illogical conclusion because it stands to reason that some of the sources and components of PM2.5 will be more toxic than others - automobile exhaust more so than road dirt, heavy metals more so than sea salt.

The choice to use total PM mass is primarily based on whether particular PM2.5 species show more consistent associations with health effects than total PM2.5. This kind of determination seems like it would fall prey to a specific problem that can be caused by exposure measurement error: namely, that having different levels of error associated with different explanatory variables in a regression can cause misleading results such that, for example, the variable measured with the least error is identified as the primary "culprit" for the health effect regardless of whether it is causally associated with the health effect (Carrothers and Evans, 2000; Fewell et al., 2007; Lipfert and Wyzga, 1996; USEPA, 2018). Because of monitoring technology and precision, as well as spatial variability in total PM2.5 compared to speciated PM2.5, total PM2.5 could be measured with less error than its constituents. My questions about this topic for the expert consultants were:

- **Could different magnitudes of error amongst different variables in regression analyses be masking the effect of a speciated constituent of PM2.5**?

From Dr. Lipfert:

> "In general, the database for PM constituents is much more sparse than for PM2.5 per se; there are fewer locations, shorter periods of record and limited daily data. I searched for relevant mortality studies from the US, UK, or Canada and found the following papers, none of which were cited in the PA (Specific references in Dr. Lipfert's comments). Most of these studies were from localized areas. None of them used long-term data from the Harvard Six Cities, American Cancer Society, or Medicare cohorts; the Veterans Cohort Study used nationwide data (Lipfert et al., 2006b). Measurement errors relative to PM2.5 are likely since it tends to be regionally distributed while most of its constituents are more local. There may also be uncertainties in the chemical analyses, notably for carbon compounds that appear to be the most important constituents."

From Dr. North:

> "Another emphatic yes. By combining the C-R relations from different areas the effect of the speciated constituents will be masked. Don't mix Salt Lake salt and old deposits of trace metals resuspended by wind from the evaporated portions of the Lake, and California's wood smoke exposures. Get information on each separately!"

From Dr. Sax:

> "The differences in PM composition across regions has been hypothesized to account for the large heterogeneity in effect estimates across regions. However, I do not know of any study that has identified a specific PM constituent that is associated with the observed overall PM effects. It seems reasonable that the exposure measurement error would vary significantly for individual constituents, but it is unclear whether using PM mass masks any effect of individual constituents. Evaluations of individual constituents (such as sulfate) both in experimental and epidemiology studies do not necessarily support the adverse effects observed in studies that evaluate PM2.5 mass."

From Dr. Thomas:

> "Yes, in principle. It has been well known that measurement errors in one variable can bias the estimates of the effect of another variable (Zeger et al. 2000). The magnitude and direction of this bias depends on the correlation of the two variables and of their measurement errors, as well as the strength of the true effects of both variables. While the various pollutants may be fairly highly correlated, their measurement errors are likely to be less so, and two-pollutant models including, say, PM2.5 and a gaseous pollutant are still likely to be relatively robust and better

than ignoring the co-pollutant. PM species are typically more difficult to disentangle and have seldom been incorporated simultaneously in multi-pollutant models."

- **What happens when multiple potential explanatory factors are included in a single variable in an already-complex multiple regression system?** Presumably each PM2.5 component has a different C-R relationship with the health effect (even if that relationship is zero), and each is a somewhat better or worse surrogate for the relationship between actual exposure vs measured exposure. **What kind of an impact would this inclusion of multiple potential explanatory factors into one variable have on the final C-R function, and how accurate would that C-R function be?**

From Dr. Lipfert:

> "A common problem is that of multicollinearity, since many PM constituents are interrelated. A better procedure might be to develop hypotheses from toxicity data and test them against specific causes of death, relying more on physiology than statistics per se."

From Dr. North:

> "Yes, if you mix it all together then you can't tell which ingredients in the resulting stew may be toxic. C-R ought to be done with disaggregation, so one can see the effect of speciation. And it may be that weather and SES are even more important than PM of any species at low levels in predicting health effects. Let's include these factors separately while gathering the data. By separating them we might develop much better information about the impacts on public health, and what strategies might reduce adverse impacts on public health."

From Dr. Sax:

> "This is an area of uncertainty that needs more attention and study. As noted above studies that have tried to evaluate specific constituents of PM have generally reported very inconsistent evidence and there is not clear single component that appears to explain the associations that are observed for PM mass. This is further complicated by the fact that studies have reported statistically significant associations not only between PM and mortality, but also between other criteria air pollutants (nitrogen dioxide, carbon monoxide, sulfur dioxide, and ozone) and mortality (e.g., Stieb et al., 2002), yet all of these air pollutants are rarely included in recent epidemiology studies as potential confounders. And this does not include the myriad of other air pollutant (e.g., benzene, formaldehyde) that correlate with the criteria air pollutants. There is more inconsistency in the literature regarding confounding effects of co-pollutants than EPA generally recognizes in discussing this issue and this should be more fully addressed."

From Dr. Thomas:

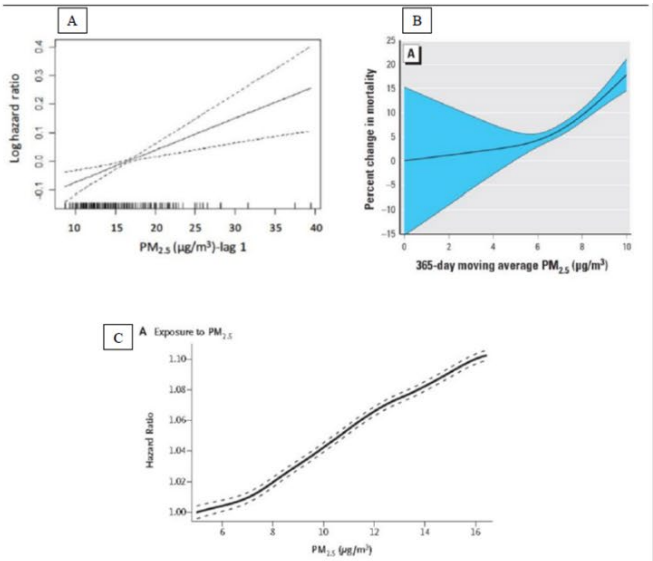
> "This depends upon the degree of multi-colinearity among the variables in the regression equation. The general effect of adding too many highly correlated variables is to inflate the standard errors of the estimates rather than introduce any bias (except in certain conditions, as described in my response to question 5). Adjusting for confounders such as contextual variables (e.g., socioeconomic status in a cohort study of chronic effects) or temporal variables (e.g., weather in a time series study of acute effects) is unlikely to lead to substantial inflation of standard errors and is necessary to control for confounding. On the other hand, multiple pollutants tend to be fairly highly correlated, particularly for constituents of PM. Hence, few epidemiologic studies of either types have attempted to include more than two pollutants in the same model, and then only to assess the extent to which estimates of the effect of one pollutant are affected by inclusion of the other. Novel statistical methods that allow smoothed estimation of multivariable effects such as Bayesian Kernel Machine Regression (KMR) have become popular for analyzing high-dimensional genomic and other data, but have only recently been applied in air pollution epidemiology, e.g., (Bobb et al. 2014)."

Based on the responses from the expert consultants, there are still unresolved issues about concluding that the measured health effects are attributable to total PM2.5 mass, as opposed to one of the constituents. These unresolved issues include problems with a paucity of studies investigating the question, problems with measurement error amongst the measured constituents, and availability of methods that allow correlated air pollutants to be disentangled.

Major Point #5 – Linear No-Threshold Concentration Response

3.2.1.1 Mortality: Long-Term PM2.5 Exposures:

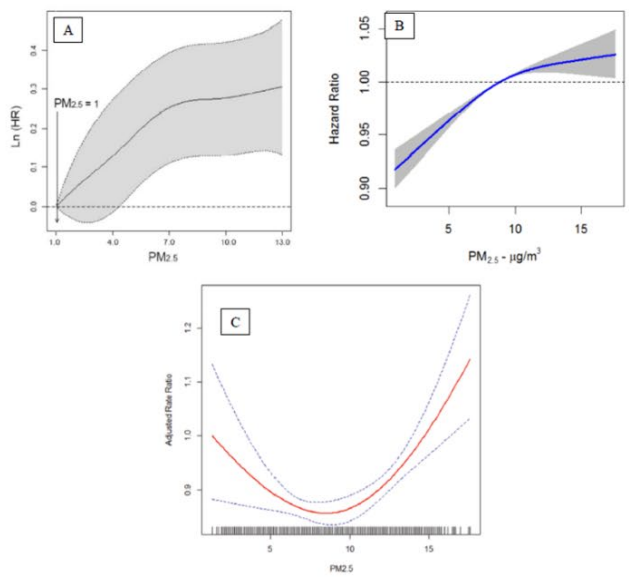
The evidence that has been presented to demonstrate that the association between long-term PM2.5 exposure and mortality are linear is not convincing: for example, in section 11.2.4 of the 2018 Draft PM ISA that addresses this point, 6 curves are presented and only 2 show a linear shape (Figures 11-22 and 11-23 reproduced below). If the EPA is concerned that the data is too uncertain to be confident in these sometimes-odd shapes, then it is equally difficult to be confident in a linear shape.



Note: Shaded areas or dotted lines indicate 95% confidence intervals. The tick marks on the *x*-axis identify the distribution of observations according to $PM_{2.5}$ concentrations.

Source: Permission pending, Panel A Lepeule et al. (2012); Panel B Shi et al. (2015); Panel C Di et al. (2017c)

Figure 11-22 **Examples of concentration-response relationships between long-term $PM_{2.5}$ exposure and total (nonaccidental) or all-cause mortality in (A) the Harvard Six Cities Study using penalized splines (1974–2009); (B) long-term time-series study; (C) the Medicare Cohort using thin-plate spines.**



Note: Shaded areas or dotted lines indicate 95% confidence intervals. The tick marks on the *x*-axis identify the distribution of observations according to $PM_{2.5}$ concentrations.

Source: Permission pending, Panel A Pinault et al. (2016); Panel B Crouse et al. (2015); Panel C Villeneuve et al. (2015).

Figure 11-23 **Examples of concentration-response relationships between long-term $PM_{2.5}$ exposure and total (nonaccidental) mortality in (A) nonparametric estimates; (B) in the CanCHEC cohort study; (C) the Canadian National Breast Screening Study.**

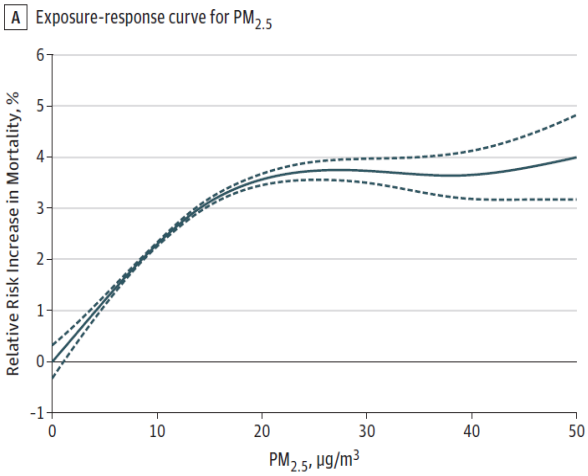
2018 PM Draft ISA, Figures 11-22 and 11-23, demonstrating the variability in shapes of C-R curves derived in studies investigating associations between long-term PM2.5 concentrations and mortality.

3.2.1.1 Mortality: Short-Term PM2.5 Exposures:

As with the long-term mortality studies, the EPA states for short-term mortality studies that "These studies have used various statistical approaches and consistently demonstrate a linear relationship with no evidence of a threshold." However, as discussed above and in the ISA preamble, the presence of known error measurement and misclassification error (Avery et al., 2010a, 2010b) can obscure a threshold. In addition, the EPA have noted in the past that city-to-city heterogeneity can obscure thresholds in the data (O3 PR 2014). The EPA also state on page 3-24 that "studies have not conducted extensive analyses exploring alternatives to linearity when examining the shape of the PM2.5-mortality concentration-response relationship." Therefore, given the known difficulties in determining curve shapes for this type of data, and the fact that extensive analyses have not been conducted, there is not enough information to conclude that the C-R effect between short-term PM2.5 and mortality is linear. As well, the shapes of the C-R curves are not convincing that the observed effects are linear (e.g. Di et al. 2017a, figure reproduced below) – this shows an approximately linear curve at lower concentrations and flattens out at concentrations of ~20 µg/m3, which is not consistent with other studies that have shown associations at higher concentrations. Because these are daily concentrations, levels above

20 µg/m3 would not be that uncommon in this dataset (and therefore the shape is not caused by a paucity of data at concentrations above 20 ug/m3).

Figure 5. Estimated Exposure-Response Curves for Short-term Exposures to Fine Particulate Matter ($PM_{2.5}$) and Ozone



A 2-pollutant analysis with separate penalized splines on $PM_{2.5}$ (A) and ozone (B) was conducted to assess the percentage increase in daily mortality at various pollution levels. Dashed lines indicate 95% CIs. The mean of daily exposure on the same day of death and 1 day prior (lag 01-day) was used as metrics of exposure to $PM_{2.5}$ and ozone. Analysis for ozone was restricted to the warm season (April to September). Ppb indicates parts per billion.

Di et al. 2017a, Figure 5, demonstrating the concentration-response relationship between short-term PM2.5 concentrations and mortality (panel b removed because it addressed ozone, not PM2.5).

Major Point #6 – Controlled Human Exposure Studies

3.1.2. General Approach in Current Review:

In the general approach to the current review, the EPA states that they focus on information from key epidemiologic and controlled human exposure studies. The problem, as ever, is that there are many studies that have been conducted and inevitably these studies do not always find the same results. The fundamental question about what to do with conflicting results is one that CASAC asked EPA to answer in the 2018 Draft PM ISA, and it continues to be an important problem in this document. In fact, it is further complicated by the need here to summarize information from the very substantial PM ISA, which inevitably results in skating over the complexities of conflicting results. However, just because the problem is difficult, does not mean that there should not be an attempt to provide some of the important considerations about study results.

One way to do this is to use summary figures, such as the forest plots that EPA often used in the ISA, which allow many results to be studied together. While this type of summary has been done with the epidemiology study results, CHE study results would also benefit from these sorts of summaries. Interpreting the results from CHE studies of particles has proven to be tricky, with different results being generated by different studies. These results may be dependent on PM generation methods, exposure times and concentrations, the population being studied, or other factors. But even a relatively simple breakdown of these studies would be immensely helpful in deciphering the general impacts of PM on the human body, even in the presence of these methodological complications. Below I have provided what I think is an appropriate summary of the CHE data, which includes figures that provide a simple summary of many of the relevant CHE studies, and helps to determine which endpoints are and are not likely to be relevant to PM exposure, and at what concentrations/doses. I have listed the studies in order of a simple total PM2.5 exposure calculation (described below, essentially time x concentration x total ventilation rate), which helps to visualize the concentration-response of the measured health endpoints.

For the figures below: Total Exposure is calculated via the equation: ((no Exer exp time*5 L/min•m2 / 1000)+(exer exp time * exer EVR / 1000))*PM Conc*2m^2. 2 m^2 refers to average body surface area. EVR for Brauner 2008 was estimated to be 20 L/min•m^2. Blue cells marked as "N" indicate that there was no significant effect of PM2.5 exposure; Red cells marked as "+" indicate that there was a significant effect of PM2.5 exposure in an adverse direction. Green cells marked as "-" indicate that there was a significant effect in the opposite direction of adversity with PM2.5 exposure. Time indicates the time points when the effects were measured after exposure ended. Specifics indicated the endpoints that were affected (for "+" and "-" marked cells). For the "N" cells, only those endpoints are listed that where shown to have changed in other studies. Often multiple papers are published analyzing different aspects of data from the same study. In the figures I try to list all the papers associated with a study. In the endpoint summaries below I try to cite the correct paper for the particular measured endpoint in a study.

I do not include the Lucking et al. 2011 study in this assessment, because it is a diesel exhaust + particle filter study that is not restricted to PM2.5 particles, and there are differences in the gases between the two exposures, so it cannot be definitively shown that effects are due to PM2.5. I also do not include the Vieira studies because they are lacking important experimental details, such as how the exposure was conducted (chamber or facemask?) and the time after exposure when effects were measured.

General Conclusions from My Analysis of PM2.5 Controlled Human Exposure studies

Respiratory symptoms: Only 1 of 4 studies observed respiratory symptoms at lower total PM2.5 exposure and in younger individuals. Respiratory symptoms don't seem to be a consistent effect of PM2.5 exposure.

Respiratory inflammation: 8 studies investigated infiltration of immune cells into lavage fluid (BAL; 2 studies) bronchial biopsy (1 study) or sputum (5 studies) with PM2.5 exposure (**Lange Figure 1**). 5 of these studies also measured soluble inflammatory mediators in the collected respiratory samples. None of the studies showed changes in soluble inflammatory markers. In addition, neither the sputum studies nor the bronchial biopsy study showed infiltration of immune cells. One of the BAL studies (Ghio et al. 2000) saw neutrophil infiltration into lavage fluid at every dose group. The other BAL study (Huang et al 2012) that used lower concentrations did not identify increased neutrophils in lavage fluid with PM2.5 exposure. Therefore, if respiratory inflammation occurs after PM2.5 exposure, it is only measurable using a more sensitive technique such as bronchioalveolar lavage, and with sufficiently high PM2.5 total exposure. Therefore, respiratory inflammation is not likely to be a required precursor event for effects happening at lower concentrations.

Pulmonary function: 10 studies have investigated pulmonary function effects of PM2.5 exposure (**Lange Figure 2**). While a few studies have found decreases in pulmonary function in different populations at varying times after PM2.5 exposure, there is no consistent effect by dose, by time point, or by population. Most studies (7/10) have not observed an adverse effect, and therefore pulmonary function is unlikely to be a consistent effect of PM2.5 exposure.

Pulmonary damage: 7 studies have investigated pulmonary damage caused by PM2.5 exposure (typically measured by changes in total cells, epithelial cells, or total protein in lavage, sputum, or biopsy samples; **Lange Figure 3**). Only a single study demonstrated increased total cells in lavage fluid, and total protein was *decreased* in that study (opposite direction of adversity), and this effect was only significant when all dose groups were combined to compare to the filtered air control (Ghio et al, 2000). All other studies have either shown no change in these damage markers or decreases in damage markers in sputum samples (Gong et al. 2003, Gong et al. 2005). Therefore, PM2.5 exposure is unlikely to cause a consistent effect on pulmonary damage at the concentrations assessed.

Blood counts: 5 studies have investigated changes in blood counts (red blood cells, hemoglobin, hematocrit, etc) with PM2.5 exposure. One study (Gong et al. 2004) demonstrated decreased red blood cells at 22 hours post-exposure in older individuals who were healthy or had COPD. All other studies investigated this endpoint in healthy individuals and found no changes in blood counts. Therefore, it is possible that PM2.5 exposure impacts red blood cell counts at higher PM2.5 total exposures, but there is not enough data for a definitive conclusion.

Blood clotting: 9 studies have investigated changes of factors in the blood clotting cascade caused by PM2.5 exposure (**Lange Figure 4**). 3 studies have shown changes in these factors: Ghio et al. 2000 found increased fibrinogen (no changes in platelets) with PM2.5 exposure when all exposures were grouped together (not in separate concentration quartiles) in healthy younger adults; Ghio et al. 2003 found increased fibrinogen (no changes in platelets, Factor VII, tPA, or

D-dimer) in healthy younger adults under the same exposure conditions; and Mills 2008 found increased platelets with no change in tPA in older adults who were healthy or had coronary heart disease. In contrast, in other studies with both younger and older adults, with and without respiratory disease, at lower or higher total PM2.5 exposures there were no changes in these clotting factors, or opposite changes in clotting factors (Tong et al. 2015, Gong et al. 2003). Therefore, the evidence of PM2.5 exposure causing changes in the clotting cascade is inconclusive.

Systemic inflammation: 11 studies have investigated changes in leukocytes in the blood after PM2.5 exposure, and 11 studies have investigated changes in soluble inflammatory markers (some of these studies overlap, but not all; **Lange Figure 5**). In general, there is inconsistent evidence of changes in blood leukocytes, with Brook et al. 2009 and Behbod et al. 2013 showing increases in total leukocytes and neutrophils in healthy younger adults, but at different time points. Other studies at higher total PM2.5 exposure have shown no changes in these cell numbers, and Ghio et al. 2003 demonstrated decreases in total leukocytes. Other studies show changes in basophils in healthy older adults but not older adults with COPD (Gong et al. 2004) or in monocytes in older adults with coronary heart disease (Mills et al. 2008). Only one study showed any changes in the soluble inflammatory marker IL-6 (Urch et al. 2010), whereas the rest of the studies showed no changes in these markers. Systemic inflammation was also not seen consistently in the same studies as potential cardiovascular functional effects (such as changes in HRV or vascular effects, e.g. Hemmingsen et al. 2015a), making inflammation less likely to mediate these effects.

Vascular function: CHE studies provide evidence of changes in vascular function, although the specific effects are not consistent between studies (**Lange Figure 6**). Hemmingson et al. 2015a shows decreased nitroglycerine (NG)-mediated vascular dilatation with PM2.5 exposure for 5 hours to on average 24 μg/m3 in overweight older adults, whereas Brook et al. 2002 and 2009 show no change in NG-mediated dilatation, but rather decreases in brachial artery diameter or flow-mediated vascular dilatation, respectively. In some of the exposure groups Tong et al. 2015 found decreased flow-mediated dilatation (with no change in brachial artery diameter) in healthy older adults, and Mills 2008 showed no change in various mediator-induced forearm blood flow measures in older adults who were healthy or had coronary heart disease. Some studies investigated vascular adhesion molecules, with increases in ICAM-1 seen in healthy individuals in Gong et al 2003, but not at lower total exposures in older individuals in Tong et al. 2015.

The findings of Hemmingsen et al. 2015a warrant further investigation as a key study in this PA, because they suggest vascular and heart rate variability (discussed below) effects in a potentially sensitive population (overweight older adults) at potentially relevant PM2.5 concentrations.

Blood pressure: EPA has often cited blood pressure changes as a potential consequence of PM2.5 exposure (**Lange Figure 6**). 10 CHE studies have investigated this endpoint in a variety

of populations and at different timepoints after exposure (from 0 to 22 hours), and the evidence is mixed. Siganangabalan et al. (2011) observed increased DBP in younger adults (but not SBP) right after exposure, as did Tong et al. (2015) in older adults. Bellavia et al. (2013) observed increased SBP (but not DBP) during and immediately after PM2.5 exposure in younger adults. However, changes in either DBP or SBP were not observed in the other 7 seven studies, either at higher or lower exposures, in younger or older adults with or without asthma, COPD, or CHD. This includes three studies that did show potential vascular effects (Hemmingsen et al. 2015a, Brook et al. 2002, and Brook et al. 2009). Altogether there is not compelling evidence of PM2.5 exposure causing blood pressure changes in the evaluated studies.

Myocardial vulnerability: Cardiovascular morbidity and mortality have three main contributors: the autonomic nervous system, myocardial vulnerability, and myocardial substrate (Utell et al., 2002; Zareba et al., 2001). Myocardial vulnerability encompasses arrhythmia and ischemic events and is assessed by a number of ECG parameters such as abnormal beats, stroke index, VE, SVE, ST depression or voltage shift, and arrhythmia incidence. Only a few studies have investigated this endpoint, with inconsistent results (**Lange Figure 7**). Langrish et al. 2014 did a comprehensive review of ECGs (more than 12,500 hours of ECG recordings) from CHE studies with exposure to diesel exhaust, ambient air, wood smoke, and carbon UFPs, and found no evidence of cardiac arrythmia with exposure.

Myocardial substrate: This endpoint assesses the myocardium itself using measures such as cardiac output, SaO2, QT intervals and variability, and T-wave amplitude and complexity. There is inconsistent evidence of substrate effects in younger individuals, with some studies showing QT or T-wave alternans increases at lower total PM2.5 exposures (Kusha et al. 2011, Sivagangabalan et al. 2011; **Lange Figure 7**), whereas others have not seen QT effects at higher exposures (Huang et al 2012, Gong et al. 2003). There is some evidence of decreases in SaO2 in older adults at higher total PM2.5 exposures (Gong et al. 2004, Gong et al. 2005). Altogether there is not consistent evidence of effects of PM2.5 on myocardial substrate or vulnerability in the reviewed studies.

Heart rate & heart rate variability: Hear rate and HRV are markers of changes in the autonomic nervous system. Of the 9 studies that investigated changes in heart rate with PM2.5 exposure, none saw increases in heart rate, and one showed decreases (Gong et al. 2003; **Lange Figure 8**). 8 studies have investigated HRV (**Lange Figure 8**). 5 of these studies investigated HRV effects in younger adults and found either no impact of PM2.5 exposure, or changes that were in the opposite direction of adversity (indicating activation of the parasympathetic response, rather than the sympathetic response; Fakhri et al. 2009; Gong et al. 2003). 3 studies investigated HRV effects in older adults, and all but one of the groups (individuals with COPD in Gong et al. 2004) showed some changes in HRV, particularly in markers for the N-N interval, decreases in HF, and/or increases in LF. Hemmingsen et al. (2015), who exposed older, overweight adults to Copenhagen air at 24 ug/m3 PM2.5 for 5 hours seems to be the lowest exposure showing effects,

particularly on HF and LF. This study also shows vascular effects. It is not clear whether these effects reach a level that would be considered "adverse", but they need to be discussed and considered in this PM evaluation.

								Pulmonary Effects				
Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Tissue	Immune Cells	Time	Specifics	Soluble Inflammatory Markers
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	13 (6+7)	64	120	0	0	76.8	Sputum	N	3 & 20 Hrs	Neutrophils	
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	10 (6+4)	140	120	0	0	168.0	Sputum	N	3 & 20 Hrs	Neutrophils	
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (1+9)	47.2	60	60	25	169.9	BAL	N	18 Hrs	Neutrophils	N
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2	BAL	N	18 Hrs	Neutrophils, monocytes	N
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	Sputum	N	22 Hrs	Neutrophils	N
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	Sputum	N	22 Hrs	Neutrophils	N
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	107.4	60	60	25	386.6	BAL	+	18 Hrs	↑ Neutrophils, monocytes	N
Holgate 2003, HEI	Healthy, 27 yo	10 (0+10)	117	60	60	25	421.2	Bronch Biopsy	N	18 Hrs	Neutrophils	
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0	BAL	+	18 Hrs	↑ Neutrophils	N
Ghio 2003, Inh Tox	Healthy, 25 yo	15	120.5	60	60	25	433.8	Sputum	N			N
Gong 2005, Inh Tox	COPD, 73 yo	18 (9+9)	164	60	60	22	531.4	Sputum	N	22 Hrs	Neutrophils, monocytes	
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	Sputum	N	0, 4, & 22 HRs	Neutrophils, monocytes	N
Gong 2005, Inh Tox	Healthy, 68 yo	6 (4+2)	164	60	60	26	610.1	Sputum	N	22 Hrs	Neutrophils, monocytes	
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	Sputum	N	0, 4, & 22 HRs	Neutrophils, monocytes	N
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	206.7	60	60	25	744.1	BAL	+	18 Hrs	↑ Neutrophils	N

Lange Figure 1. Controlled human exposure study results for **pulmonary inflammation** with PM2.5 exposure. BAL = bronchioalveolar lavage. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity.

Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Pulmonary Fxn	Time	Specifics
Brauner 2007, EHP; Brauner 2009 Inh Tox	Healthy, 25 yo	29 (9+20)	9.7	60	90	*20*	40.7	-	*0 Hr*	*↑ TLC*
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	13 (6+7)	64	120	0	0	76.8	N	10 min	TV, BF, FEF, FEV1
Hazucha 2013, Part Fib Tox	Smokers & Ex-smokers, 48 yo	11 (8+3)	108.7	120	0	0	130.4	+	22 Hr	FEV1
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	10 (6+4)	140	120	0	0	168.0	N	10 min	TV, BF, FEF, FEV1
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (1+9)	47.2	60	60	25	169.9	N	0 Hr	FEV1
Sivagangabalan 2011, J Am Coll Cardiol	Healthy, 27 yo	25 (14+11)	154	120	0	0	184.8	N	0 Hr	TV, BF
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2	N	1 & 18 Hrs	FEV1, FEF
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	+	0 Hr	↓TV ↓BF; NC FEV1
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	+	0 Hr	↓TV ↓BF; NC FEV1
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	107.4	60	60	25	386.6	N	0 Hr	FEV1
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0	N	0 Hr	FEV1
Gong 2005, Inh Tox	COPD, 73 yo	18 (9+9)	164	60	60	22	531.4	N	0,4 & 22 Hrs	TV, FEF, FEV1
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	N	0,4 & 22 Hrs	FEV1
Gong 2005, Inh Tox	Healthy, 68 yo	6 (4+2)	164	60	60	26	610.1	+	22 Hrs	↓FEF; NC TV, FEV1
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	N	0,4 & 22 Hrs	FEV1
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	206.7	60	60	25	744.1	N	0 Hr	FEV1

Lange Figure 2. Controlled human exposure study results for **pulmonary function** with PM2.5 exposure. TLC = total lung capacity; TV = tidal volume; BF = breathing frequency; FEF = forced expiratory flow; FEV1 = forced expiratory volume in 1 sec. NC = no change. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity; "-" – statistically significant effect in the opposite direction of adversity.

								Pulmonary Effects			
Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Tissue	Damage Markers	Time	Specifics
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	13 (6+7)	64	120	0	0	76.8	Sputum	N	3 & 20 Hrs	Epith cells
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	10 (6+4)	140	120	0	0	168.0	Sputum	N	3 & 20 Hrs	Epith cells
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (1+9)	47.2	60	60	25	169.9	BAL	N	18 hrs	Total cells, total protein
Behbod 2013, Occup Env Med	Healthy, 27 yo	35 (16+19)	234.7	130.2	0	0	305.6	Sputum	N	22 Hrs	Total cells
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2	BAL	N	18 Hrs	Total cells
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	Sputum	-	*22 Hrs*	*↓total cells*
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	Sputum	-	*22 Hrs*	*↓total cells*
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	107.4	60	60	25	386.6	BAL	N	18 hrs	Total cells, total protein
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0	BAL	+	18 Hrs	↑ total cells;↓total protein
Gong 2005, Inh Tox	COPD, 73 yo	18 (9+9)	164	60	60	22	531.4	Sputum	-	*22 Hrs*	*↓Epith cells*
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	Sputum	N	22 Hrs	Epith cells
Gong 2005, Inh Tox	Healthy, 68 yo	6 (4+2)	164	60	60	26	610.1	Sputum	-	*22 Hrs*	*↓Epith cells*
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	Sputum	N	22 Hrs	Epith cells
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	206.7	60	60	25	744.1	BAL	N	18 hrs	Total cells, total protein

Lange Figure 3. Controlled human exposure study results for **pulmonary damage** with PM2.5 exposure. BAL = bronchioalveolar lavage. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity; "-" – statistically significant effect in the opposite direction of adversity.

Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Blood Clotting	Time	Specifics
Hazucha 2013, Part Fib Tox	Smokers & Ex-smokers, 48 yo	11 (8+3)	108.7	120	0	0	130.4	N	3 & 22 Hrs	Platelets, tPA, D-dimer
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (1+9)	47.2	60	60	25	169.9	N	18 Hrs	Fibrinogen, platelets
Brauner 2008, Part Fib Tox	Healthy, 25 yo	29 (9+20)	10.5	1170	180	*20*	198.5	N	6 & 24 Hrs	Fibrinogen, platelets
Tong, 2015, EHP	Healthy, 58 yo	13 (11+2)	253	120	0	0	303.6	N	0 & 20 Hrs	Fibrinogen, tPA, D-dimer
Tong, 2015, EHP	Healthy, 58 yo + OO	13 (9+4)	253	120	0	0	303.6	-	*0 & 20 Hrs*	*↑tPA; ↓D-dimers; NC fibrinogen*
Tong, 2015, EHP	Healthy, 58 yo + FO	16 (12+4)	253	120	0	0	303.6	N	0 & 20 Hrs	Fibrinogen, tPA, D-dimer
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2	N	1 & 18 Hrs	Fibrinogen, platelets, Factor VII, tPA, D-dimer
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	-	*22 Hrs*	*↓Factor VII; NC Fibrinogen; NC Platelets*
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	-	*22 Hrs*	*↓Factor VII; NC Fibrinogen; NC Platelets*
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	107.4	60	60	25	386.6	N	18 Hrs	Fibrinogen, platelets
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0	+	18 Hrs	↑ Fibrinogen; NC platelets
Ghio 2003, Inh Tox	Healthy, 25 yo	15	120.5	60	60	25	433.8	+	0 & 24 Hrs	↑ Fibrinogen; NC Factor VII; NC platelets; NC tPA; NC D-dimer
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	N	0, 4 & 22 Hrs	Fibrinogen, platelets, Factor VII
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	N	0, 4 & 22 Hrs	Fibrinogen, platelets, Factor VII
Mills 2008, Env Health Perspect	Healthy, 54 yo	12 (0+12)	*176*	60	60	25	633.6	+	2 Hrs	↑ Platelets; NC tPA
Mills 2008, Env Health Perspect	CHD, 59 yo	12 (0+12)	*176*	60	60	25	633.6	+	2 Hrs	↑ Platelets; NC tPA
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	206.7	60	60	25	744.1	N	18 Hrs	Fibrinogen, platelets

Lange Figure 4. Controlled human exposure study results for **blood clotting markers** with PM2.5 exposure. NC = no change; tPA = tissue plasminogen activator. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity; "-" – statistically significant effect in the opposite direction of adversity.

								Systemic Effects					
Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Immune Cells	Time	Specifics	Soluble Inflammatory Markers	Time	Specifics
Hemmingsen 2015, Mutagenesis, Part Fib Tox	Overweight non-smokers, 64 yo	60 (35+25)	24	300	0	0	72.0	N	0 Hrs	Total leukocytes; neutrophils; monocytes	N	0 Hrs	CRP
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	13 (6+7)	64	120	0	0	76.8				N	10 min, 3, & 24 Hrs	IL-6
Hazucha 2013, Part Fib Tox	Smokers & Ex-smokers, 48 yo	11 (8+3)	108.7	120	0	0	130.4	N	3 & 22 Hrs	Total leukocytes; neutrophils; monocytes			
Urch 2010, Inh Tox	Healthy, mild asthmatic, 26 yo	10 (6+4)	140	120	0	0	168.0				+	10 min, 3, & 24 Hrs	↑IL-6 (3 Hrs)
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (1+9)	47.2	60	60	25	169.9	N	18 Hrs	Total leukocytes; neutrophils; monocytes			
Brook 2009, Hypertension, Ramanathan 2016, Part Fib Tox	Healthy, 27 yo	31 (15+16)	148.5	120	0	0	178.2	+	0 & 24 Hrs	↑Total leukocytes (0 Hr); ↑ neutrophils (0 Hr)	N	0 & 24 Hrs	IL-6, CRP
Brauner 2008, Part Fib Tox	Healthy, 25 yo	29 (9+20)	10.5	1170	180	*20*	198.5				N	6 & 24 Hrs	CRP, IL-6
Tong, 2015, EHP	Healthy, 58 yo	13 (11+2)	253	120	0	0	303.6				N	0 & 20 Hrs	CRP, IL-6
Tong, 2015, EHP	Healthy, 58 yo + OO	13 (9+4)	253	120	0	0	303.6				N	0 & 20 Hrs	CRP, IL-6
Tong, 2015, EHP	Healthy, 58 yo + FO	16 (12+4)	253	120	0	0	303.6				N	0 & 20 Hrs	CRP, IL-6
Behbod 2013, Occup Env Med	Healthy, 27 yo	35 (16+19)	234.7	130.2	0	0	305.6	+	1 & 22 Hrs	↑Total leukocytes (22 Hr); ↑ neutrophils (22 Hr)	N	1 & 22 Hrs	CRP, IL-6
Liu 2015 Env Health Perspect, Liu 2017, Env Int	Healthy, 28 yo	55 (29+26)	238	130.2	0	0	310.4				N	1 & 21 Hrs	CRP, IL-6
Bellavia 2013, J AHA	Healthy 27 yp	15 (7+8)	242	130.2	0	0	315.1	N	1 Hr	Total leukocytes; neutrophils; basophils; monocytes			
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2	N	18 Hrs	Total leukocytes; neutrophils; monocytes	N	1 & 18 Hrs	CRP
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	N	4 & 22 Hrs	Total leukocytes; neutrophils; basophils; monocytes	N	4 & 22 Hrs	CRP, IL-6
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	N	4 & 22 Hrs	Total leukocytes; neutrophils; basophils; monocytes	N	4 & 22 Hrs	CRP, IL-6
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	107.4	60	60	25	386.6	N	18 Hrs	Total leukocytes; neutrophils; monocytes			
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0	N	18 Hrs	Total leukocytes; neutrophils; monocytes			
Ghio 2003, Inh Tox	Healthy, 25 yo	15	120.5	60	60	25	433.8	-	*0 & 24 Hrs*	*↓Total leukocytes; NC neutrophils*	N	0 & 24 Hrs	IL-6, CRP
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	N	0, 4, & 22 Hrs	Total leukocytes; neutrophils; basophils; monocytes			
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	+	0, 4, & 22 Hrs	↑Basophils (22 Hrs); NC total leukocytes; NC neutrophils; NC monocytes			
Mills 2008, Env Health Perspect	Healthy, 54 yo	12 (0+12)	*176*	60	60	25	633.6	+	2 & 6-8 Hrs	↑Monocytes (2 Hrs); NC total leukocytes; NC neutrophils	N	2 & 6-8 Hrs	CRP
Mills 2008, Env Health Perspect	CHD, 59 yo	12 (0+12)	*176*	60	60	25	633.6	+	2 & 6-8 Hrs	↑Monocytes (2 Hrs); NC total leukocytes; NC neutrophils	N	2 & 6-8 Hrs	CRP
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	10 (0+10)	206.7	60	60	25	744.1	N	18 Hrs	Total leukocytes; neutrophils; monocytes			

Lange Figure 5. Controlled human exposure study results for **systemic inflammation** with PM2.5 exposure. NC = no change; CRP = C-reactive protein; IL-6 = interleukin 6. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity; "-" – statistically significant effect in the opposite direction of adversity.

Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Vascular Fxn	Time	Specifics	Baroreflex	Time	Specifics
Hemmingsen 2015, Mutagenesis, Part Fib Tox	Overweight non-smokers, 64 yo	60 (35+25)	24	300	0	0	72.0	+	0 Hr	↓NG-Dilatation	N	0 Hr	SBP, DBP
Fakhri 2009, Env Health Perspect (overlap with Brook 2002 & Urch 2005)	Healthy & mild asthmatic, 27 yo	50 (26+24)	121.6	120	0	0	145.9	+	10 min	↓BAD; NC NG-Dilat	N	10 min	SBP, DBP
Brook 2009, Hypertension, Ramanathan 2016, Part Fib Tox	Healthy, 27 yo	31 (15+16)	148.5	120	0	0	178.2	+	0 &24 Hr	↓Flow-Med Dilatation (24 Hr); NC BAD; NC NG-Dilatation	N	0 & 24 HRs	SBP, DBP
Sivagangabalan 2011, J Am Coll Cardiol	Healthy, 27 yo	25 (14+11)	154	120	0	0	184.8				+	0 Hr	↑DBP; NC SBP
Brauner 2008, Part Fib Tox	Healthy, 25 yo	29 (9+20)	10.5	1170	180	*20*	198.5	N	6, 24 HR	RH-PAT score			
Tong, 2015, EHP	Healthy, 58 yo	13 (11+2)	253	120	0	0	303.6	+	0 Hr	↓Flow-Med Dilatation (24 Hr); NC BAD	+	0 Hr	↑DBP; NC SBP
Tong, 2015, EHP	Healthy, 58 yo + OO	13 (9+4)	253	120	0	0	303.6	N	0 & 20 Hr	Flow-Med Dilatation; BAD	+	0 Hr	↑DBP; NC SBP
Tong, 2015, EHP	Healthy, 58 yo + FO	16 (12+4)	253	120	0	0	303.6	+	0 & 20 Hr	↓Flow-Med Dilatation (24 Hr); NC BAD	+	0 Hr	↑DBP; NC SBP
Bellavia 2013, J AHA	Healthy 27 yp	15 (7+8)	242	130.2	0	0	315.1				+	5 Min & 1 Hr	↑SBP (5 min); NC DBP
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7				N	0, 4, & 22 HRs	SBP, DBP
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7				N	0, 4, & 22 HRs	SBP, DBP
Gong 2005, Inh Tox	COPD, 73 yo	18 (9+9)	164	60	60	22	531.4				N	0, 4, & 22 HRs	SBP, DBP
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2				N	0, 4, & 22 HRs	SBP, DBP
Gong 2005, Inh Tox	Healthy, 68 yo	6 (4+2)	164	60	60	26	610.1				N	0, 4, & 22 HRs	SBP, DBP
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2				N	0, 4, & 22 HRs	SBP, DBP
Mills 2008, Env Health Perspect	Healthy, 54 yo	12 (0+12)	*176*	60	60	25	633.6	N	6-8 Hrs	Drug-induced FBF	N	6-8 Hrs	SBP, DBP
Mills 2008, Env Health Perspect	CHD, 59 yo	12 (0+12)	*176*	60	60	25	633.6	N	6-8 Hrs	Drug-induced FBF	N	6-8 Hrs	SBP, DBP

Lange Figure 6. Controlled human exposure study results for **vascular function and baroreflex** with PM2.5 exposure. NC = no change; NG = nitroglycerin; BAD = brachial artery diameter; RH-PAT = reactive hyperemia peripheral arterial tonometry; FBF = forearm blood flow; SBP = systolic blood pressure; DBP = diastolic blood pressure. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity.

Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Myocardial Vulnerability	Time	Specifics	Myocardial Substrate	Time	Specifics
Devlin 2003, Eur Resp J	Healthy, 67 yo	10 (3+7)	120	120	0	0	144.0	N	0 & 24 Hrs	Abnormal beats			
Kusha 2011, EHP	Healthy, 24 yo	17 (9+8)	154	120	0	0	184.8				+	First & Last 5 min	↑T-Wave Alt
Sivagangabalan 2011, J Am Coll Cardiol	Healthy, 27 yo	25 (14+11)	154	120	0	0	184.8				+	0 Hr	↑QT
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2				N	1 & 18 Hr	QT
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	+	0, 4, & 22 Hrs	↓ST-AMD (22 Hrs); NC VE; NC SVE	N	0, 4, & 22 Hrs	SaO2; QT
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	+	0, 4, & 22 Hrs	↓ST-AMD (22 Hrs); NC VE; NC SVE	N	0, 4, & 22 Hrs	SaO2; QT
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0	N	0 & 24 Hrs	Abnormal beats			
Gong 2005, Inh Tox	COPD, 73 yo	18 (9+9)	164	60	60	22	531.4				+	0, 4, & 22 Hrs	↓SaO2 (0 Hr)
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	-	*22 Hrs*	↓VE; ↓SVE; NC ST-AMD	N	0, 4, & 22 Hrs	SaO2
Gong 2005, Inh Tox	Healthy, 68 yo	6 (4+2)	164	60	60	26	610.1				+	0, 4, & 22 Hrs	↓SaO2 (0 Hr)
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	+	22 Hrs	↑VE; ↑SVE; NC ST-AMD	+	0, 4, & 22 Hrs	↓SaO2 (0, 4 Hr)

Lange Figure 7. Controlled human exposure study results for **myocardial vulnerability and substrate** with PM2.5 exposure. NC = no change; QT = QT interval; SaO2 = oxygen saturation. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity; "-" – statistically significant effect in the opposite direction of adversity.

Study	Volunteers	n	PM Conc (ug/m3)	No Exerc Exp Time (min)	Exerc Exp Time (min)	Exerc EVR (L/min m2)	"Total Exp" (ug)*	Heart Rate	Time	Heart Rate Variability	Time	Specifics
Hemmingsen 2015, Mutagenesis, Part Fib Tox	Overweight non-smokers, 64 yo	60 (35+25)	24	300	0	0	72.0			+	0 Hr	↑LFn; ↓HFn; NC SDNN
Devlin 2003, Eur Resp J	Healthy, 67 yo	10 (3+7)	120	120	0	0	144.0			+	0 & 24 Hrs	↓PNN50 (0 Hr); ↓HF (0 Hr); NC SDNN; NC LF
Fakhri 2009, Env Health Perspect (overlap with Brook 2002 & Urch 2005)	Healthy & mild asthmatic, 27 yo	50 (26+24)	121.6	120	0	0	145.9	N	0 Hr	-	*0 Hr*	*↑HF; NC PNN50; NC SDNN; NC LF*
Brook 2009, Hypertension, Ramanathan 2016, Part Fib Tox	Healthy, 27 yo	31 (15+16)	148.5	120	0	0	178.2	N	0 & 24 Hr	N	0 & 24 Hr	SDNN; LF; HF
Kusha 2011, EHP	Healthy, 24 yo	17 (9+8)	154	120	0	0	184.8	N	First & last 5 min			
Sivagangabalan 2011, J Am Coll Cardiol	Healthy, 27 yo	25 (14+11)	154	120	0	0	184.8	N	0 Hr			
Huang 2012, Inh Tox	Healthy, 25 yo	13 (6+7)	89.5	60	60	25	322.2	N	24 Hr	N	1, 18, &24 Hrs	PNN50; SDNN; LF; HF
Gong 2003, Inh Tox	Healthy, 28 yo	12 (6+6)	141	60	60	17.5	380.7	-	***0,** 4, & 22 Hr*	-	*0, 4, & 22 Hrs*	*↑HF (0, 22 Hr); NC LF; NC PNN50*
Gong 2003, Inh Tox	Asthmatic, 34 yo	12 (6+6)	141	60	60	17.5	380.7	-	***0,** 4, & 22 Hr*	-	*0, 4, & 22 Hrs*	*↑HF (0, 22 Hr); NC LF; NC PNN50*
Devlin 2003, Eur Resp J; Ghio 2000, AJRCCM; Harder 2001, Env Health Perspect, Holgate 2003, HEI	Healthy, 26 yo	30 (1+29)	120	60	60	25	432.0			N	0 & 24 hrs	PNN50; SDNN; LF; HF
Gong 2005, Inh Tox	COPD, 73 yo	18 (9+9)	164	60	60	22	531.4	N	0, 4, 22 Hrs			
Gong 2004, Inh Tox	COPD 73 yo	13 (8+5)	167	60	60	24	581.2	N	0, 4, 22 Hrs	N	0, 4, & 22 Hrs	SDNN; PNN50; LF; HF
Gong 2005, Inh Tox	Healthy, 68 yo	6 (4+2)	164	60	60	26	610.1	N	0, 4, 22 Hrs			
Gong 2004, Inh Tox	Healthy 68 yo	6 (4+2)	167	60	60	26	621.2	N	0, 4, 22 Hrs	+	0, 4, 22 Hrs	↓SDNN (22 Hrs); NC PNN50; NC LF; NC HF
Mills 2008, Env Health Perspect	Healthy, 54 yo	12 (0+12)	*176*	60	60	25	633.6	N	6-8 Hrs			
Mills 2008, Env Health Perspect	CHD, 59 yo	12 (0+12)	*176*	60	60	25	633.6	N	6-8 Hrs			

Lange Figure 8. Controlled human exposure study results for **heart rate and heart rate variability (HRV)** with PM2.5 exposure. NC = no change; LF = low frequency; HF = high frequency; PNN50 = NN50 (number of interval differences of successive NN intervals greater than 50 ms) divided by total number of normal-to-normal intervals; SDNN = standard deviation of normal-to-normal interval. "N" – no statistically significant effect; "+" – statistically significant effect in the direction of adversity; "-" – statistically significant effect in the opposite direction of adversity.

Major Point #7 – PM2.5 Effect Concentrations from CHE Studies

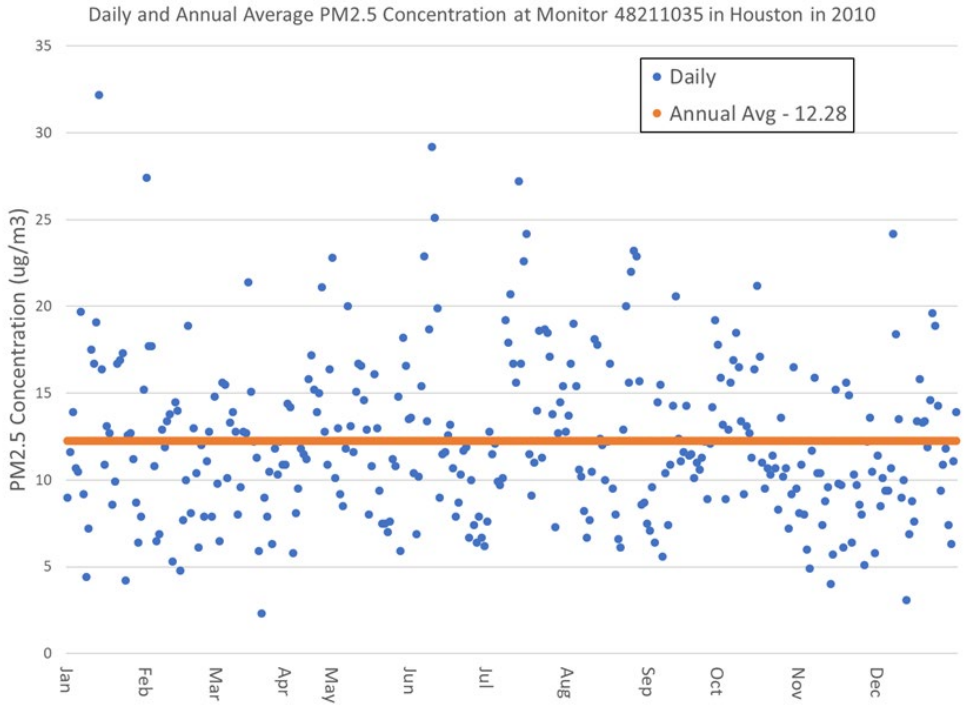
In section 3.2.3.1, the EPA states that "Additional controlled human exposure studies that examine longer exposure periods (e.g., 24-hour as in Bräuner et al. (2008); 5-hour as in Hemmingsen et al. (2015b)), or repeated exposures, to concentrations typical in the ambient air across much of the U.S. may provide additional insight into this issue in future reviews."
It is not clear what the EPA means by this statement. It seems that they are putting the consideration of these studies off until later reviews of the standard, which seems illogical. The Brauner papers essentially saw no effects of 22.5 hr exposure (with exercise) to ~10 µg/m3 PM2.5, but Hemmingsen saw mild effects on vascular function and HRV with 5 hours at 24

µg/m3 (plus gases). EPA presents Hemmingsen as two different studies, but two papers were published from the same data and showed many unaffected markers (metabolic parameters, blood lipids, damage markers, blood pressure, immune cells in the blood, anti-oxidants, soluble inflammatory markers), but they do show the vascular and HRV response. The authors' analysis suggests that this was mediated by neural responses because there were no changes in mechanistic markers such as inflammatory mediators or antioxidants. Because of this study, we should be discussing the adversity of the demonstrated effects, as well as the frequency of a 5-hour measurement of 24 ug/m3 with the current and alternative standards.

Major Point #8 – Comparison of Long-Term and Short-Term PM2.5 Concentration Effects

In section 3.2.3.2 the EPA summarizes the mean ambient PM concentrations measured in various key epidemiology studies (section 3.2.3.2.1); and then for a subset of these studies, the EPA considers whether the paper's study areas would have been in attainment for the current PM standards by deriving pseudo-design values for those study areas (section 3.2.3.2.2). When summarizing mean ambient PM concentrations measured in key epidemiology studies, the EPA includes studies that are investigating effects of both short-term exposure (on the scale of daily to weekly) and long-term exposure (on the scale of one to multiple years). The total mean concentrations of PM2.5 from both study types are then considered in the context of the current annual standard (section 3.2.3.3).

I looked at measured PM2.5 concentrations from one of the monitors in the Houston area (ID # 48211035) for a better general understanding of short-term and long-term PM2.5 concentrations. Included here is a figure that shows the daily and annual average PM2.5 concentrations from 2010 (**Lange Figure 9**). This information helps to inform some of the questions that I ask below.



Lange Figure 9. Daily and annual average PM2.5 concentrations measured at monitor 48211035 in Houston, Texas in 2010.

Typically, this kind of comparison would not be done when deriving a toxicity factor, with short-term exposures being used to derive short-term toxicity factors, and long-term exposures for long-term toxicity factors. The concentration-response functions in the short-term epidemiology studies are derived using day-to-day changes in PM2.5, which doesn't seem like it would be captured in an overall average concentration (as in Lange Figure 9).
I asked the CASAC expert consultants the following question:

- **Is it appropriate to compare daily PM2.5 concentrations to the annual average?**

From Dr. Jaffe:

> "I looked over section 3.2.3.2.1 and especially Figure 3-5. I agree that the results and figure 3-5 are a bit puzzling. It would seem that these studies looking at short term exposures should be compared against the daily PM2.5 standard. In general there is not a good relationship between the annual average PM2.5 and the number of days over the daily standard (35 ug/m3). I also examined data from the Houston site (AQS id 482011035). Figure 1 below shows the annual average, annual 98th percentile and number of days over 35 ug/m3 since 1999 and you can see there is not much relationship with the number of high days and the 98th percentile can be also disconnected with the

annual average. For example in the last two years, there has been a significant increase in the 98th percentile, but not the annual average. Also, see Figure B-8 in PA document."

From Dr. Lipfert:

"This distinction is an artifice of the extant regulatory framework; these are simply two measures of the same typically log-normal frequency distribution of ambient air quality measures. To my knowledge, all C-R functions are essentially linear, including those for health, vegetation damage, and tombstone erosion, but they all require durations of exposure. Short- vs. long-term health effects are discussed below; cumulative exposures should be used for the latter (but seldom have been), just as the number of pack-years is used in smoking epidemiology or working years in occupational epidemiology."

From Dr. Thomas:

"Obviously, there is no reason to expect that acute effects of short-term fluctuations would be similar to the chronic effects of long-term exposure (Kunzli et al. 2001; Thomas 2005). On the other hand, it would not be unreasonable to wonder whether the magnitude of acute effects might differ depending on the long-term average level of pollution. Since the daily variation and long-term average are on different scales, it would be reasonable to use different values to assess them."

From Dr. Rhomberg:

"I will only note that, aside from the purely statistical aspects of the inferences among exposures measured at different durations, there is the toxicological aspect that the dependence of effects on the particular time-course of exposure plays a role, and such dependence is usually treated with asumptions about time-averaging that may be inconsistent with the complexity of underlying dynamics of uptake and clearance as well as damage and repair that ultimately dictate whether and when toxicity may be engendered. That is, the reasons that effects might appear in longer term exposure at levels below those causing impacts from short exposures depend on how the balance of damage and repair processes operate, as well as on whether the longer term risks arise owing to more chances for essentially short-term stochastic events with more chronic consequences once incurred to occur. One needs to ask whether a longer term effect results because of the increased chances that somewhere in the span of averaging time it happened that a set of short-term peaks happened in close enough time-spacing to have some cumulative effect that more widely spaced peaks would not engender, or whether the longer-term effects really depend on inexorable and continuous exposures, even to lower levels. All these questions have to do with averaging time, and any assumptions or calculation methods that entail averaging necessarily invoke some underlying hypothesis about how the effect is attributable to some aspect of the long-term average, despite differences from case to case in the time-varying moment-by-moment exposures."

From Dr. Sax:

> "As presented in Appendix B of the draft PA, it appears that EPA calculates both annual average pseudo-design values and 24-hr maximum pseudo-design values. However, as presented in Figure 3-9, EPA appears to be presenting graphically only the annual average pseudo-design values for both short-term and long-term studies (although this is not clear from the text or from the Figure caption). I agree that it would be most appropriate for EPA to evaluate long-term studies against annual average design values, and short-term studies against the the 24-hr design values. It is unclear why EPA is presenting only the annual average design values when discussing short-term studies."

All together these responses suggest that it is not clear why short-term and long-term exposures are shown together, and if they are, there should be some sort of toxicological justification for why long-term effects may occur at the same concentrations at short-term effects (or vice versa).

3.4.2.1 Potential Alternative Standards – Indicator:

The EPA summarizes their conclusions about potential effects of PM2.5 at concentrations below the standards by emphasizing the mean concentrations measured in epidemiology studies that show associations between PM2.5 and health effects. However, the annual means are from short-term as well as long-term studies. Because the annual means are not informative of the day-to-day changes in PM2.5 to which the health effects are attributed, this type of summary is not helpful for those studies and shouldn't be combined with means from the long-term studies. As well, the means from any of these studies should not be explicitly or implicitly (as in this section) compared to the level of the standard because the mean concentrations may be calculated in ways quite different from the design value.

Major Point #9: Use of Pseudo-Design Values

In section 3.2.3.2.2 the EPA asks the question: are positive associations between PM2.5 and health effects occurring in areas that meet the current or potential alternative PM2.5 NAAQS standards?" To answer this question, they calculate pseudo-design values for each of the study areas in the key papers that have monitoring data available for the DV calculation. With this information the EPA then summarizes the pseudo-design values associated with a proportion of the population in each study. The EPA does this for both the annual and 24-hour standard, and for both short-term and long-term studies. My questions to the expert consultants were:

- **Is it informative to derive annual average pseudo-design values for study areas in short-term studies (that look at effects of day-to-day PM2.5 concentration changes), in order to determine whether these study areas attained the current annual standard?** Although the EPA can technically determine if daily changes in PM2.5 concentration increased health effects in an area meeting the annual standard, **does this really inform the health protectiveness of the annual standard?** It seems that whether

an area showed a positive effect or not could be completely independent of the annual standard and instead dependent on how much the PM2.5 concentrations changed from day-to-day.

From Dr. Jaffe:

"I agree that the short term studies are most relevant to the daily PM2.5 standard. For example one can imagine a case where two studies of the same city but performed at different times of the year could come to rather different conclusions. Eg. Most regions have higher PM2.5 in winter, so a study in winter might shower greater health effects compared to a study down in summer and yet both would have identical annual design values. Based on the discussion in section 3.2.3.2.2 it does not seem like this has been considered."

From Dr. Lipfert:

"As I understand them, design values are only used for regulatory but not scientific purposes and I have no direct comments about them. "Informing health protectiveness" can only be accomplished by comparing health effects measured before and after pollution abatement. This is shown directly in time-series analysis when death counts track pollution peaks, first up and then back down to normal. Such tests have not been successful for long-term pollution abatement, in part because of the time required for cumulative exposures to decrease. This is analogous with the time required to realize health benefits after smoking cessation."

From Dr. North:

"Again, I do not believe I understand the motivation for these pseudo-design values. It seems to me that a short term effect should show up with screening by lag times. I would be concerned more about high levels that persist over multiple days rather than a single high 24-hour value. Annual averages might be appropriate if effects build up over years of exposure, which I think is the case for inhaled tobacco smoke. But this sort of issue is not in my area of expertise. I was trained as a physicist, and then switched to using training in math and probability for risk analysis. What I know of toxicology comes in large part from a few years of serving on the EPA Science Advisory Board's Environment Health Committee. I think you ask a very good question - the pattern of exposure should be important. And the toxicity may vary, depending on chemical composition as well as particle size and exposure pattern."

From Dr. Sax:

"In the PA, EPA acknowledges that the estimated ambient concentrations used in the epidemiological studies that it relies on to evaluate the adequacy of the PM2.5 NAAQS are not the same as the design values that are used to determine compliance with the NAAQS, which is true. To address this, EPA therefore decided to calculate pseudo-

design values to compare with the reported ambient concentrations in the epidemiological studies and determine whether the study areas in the epidemiological studies would have met or violated the current NAAQS. Unfortunately, this does not address the more fundamental question of how the estimated ambient concentrations in the epidemiological studies actually reflect individual exposures to PM, and how the likely exposure measurement error impacts the reported association between PM and various health effects or the shape of the concentration-response function. Instead, EPA appears to use this analysis to conclude that most of the selected studies (18 of 29) – covering both short-term and long-term PM2.5 exposures - that observed positive associations between PM2.5 and mortality and morbidity endpoints had some portion of the population (25-75%) that lived in an area where the air quality met current NAAQS, and for the other 11 studies a majority of the population (> 75%) lived in areas that did not meet the NAAQS. Looking at Figure 3-9, it appears that the few studies had pseudo-design values that were mostly below the NAAQS, the large majority included some values that exceeded the NAAQS, making it difficult to interpret these results in terms of assessing the adequacy of the NAAQS. As noted by EPA, for many of the multi-city studies some locations would likely have met the NAAQS and others would not. It is unclear how useful this analysis is for determining the health protectiveness of the current NAAQS, especially when relying solely on this epidemiological data."

Answers from the expert consultants confirm my concerns that there is not useful information obtained by using annual average pseudo-design values to determine if an area met the annual average standard for an epidemiology study looking at short-term PM2.5 changes.

I also asked the consultants the question:

- In contrast to short-term studies that investigate the effects of day-to-day changes in PM2.5 concentrations within a certain geographic area, long-term cohort studies often look at the association between annual average PM2.5 concentrations and time-to-event data (such as the time from cohort entry to death) over long periods of time. For these studies, it is not uncommon for all study subjects in a single geographic area to have the same (or very similar) exposure assignments (e.g. Jerrett et al., 2017; Thurston et al., 2016), in which case the study is assessing the effects of PM2.5 between geographic areas, instead of within geographic areas. **In this case, is the pseudo-design value in a single geographic area particularly informative, when the association between PM2.5 and the health effect is driven by the differences between study areas?**

From Dr. Jaffe:

"Yes, these studies assign all individuals in a geographic region into the same exposure category but I am not sure what is a better approach. Different communities and neighborhoods within a community have many differences, just as individuals have differences. The association between air pollution and health will depend on the PM2.5

concentrations, but clearly there are complicating and confounding variables (smoking, occupation, age, sex, etc). In general most studies are able to identify and account for the confounding variables. To the extent that a key confounder is missed, this can invalidate the results."

From Dr. Lipfert:

"If we define the "area" as based on the surroundings of an ambient monitoring station, uniformity is assumed in cross-sectional epidemiology and the lack thereof constitutes "measurement error" and biases the slope of the C-R function downward. However, this common scenario is at odds with the real world because each affected individual within that area will have had his/her own personal exposure that tend to be controlled by indoor rather than the outdoor conditions where we typically spend only 10-15% of our time. Indoor air quality is in synch with the outdoors because of infiltration from the outdoors, but each residence has its own long-term offsets from smoking, gas cooking, pets, cleaning agents, fireplaces, etc. I found no long-term spatial correlation between indoor and outdoor PM levels as shown in Figure 1 above (Lipfert, 2015). These indoor-outdoor relationships are consistent with reports from EPA (Baxter et al. 1994, 2007, 2010, 2013, 2014, 2017)."

From Dr. North:

"If long-term exposure underlies the health response, then whether subjects move in and out of the area is important. I learned this from cancer epidemiology. Again, I would like to see case studies on specific areas."

From Dr. Sax:

"As noted above, it is unclear how comparing the pseudo design values for the locations in underlying long-term epidemiology studies (some which are below and others that are above the current NAAQS) informs the adequacy of the current NAAQS. As noted by Dr. Lange, this is further complicated by the fact that the underlying epidemiology study is not assigning exposures using the same criteria as EPA. As noted by EPA, additional uncertainties include the number of monitors that are included in the calculation of the design values that may not reflect the same monitors used in the underlying epidemiology study. EPA should clarify how this analysis is informative for addressing policy questions, given the uncertainties and the clear disconnect between how exposures are estimated in the epidemiology studies and how EPA determines compliance with the NAAQS."

From Dr. Thomas:

"Yes, indeed, it is precisely the differences between study areas that is most important for estimating long-term effects. For reasons discussed above, such association effects are not strictly interpretable as causal effects of an intervention, hence the use of these

> "pseudo-design" values for simulating the expected effects of regulations to change the distribution of exposures."

Altogether, this seems like a point that needs to be further addressed by the EPA to justify their use of pseudo-design values to determine if PM2.5-associated health effects were occurring in areas that met the current or alternative NAAQS.

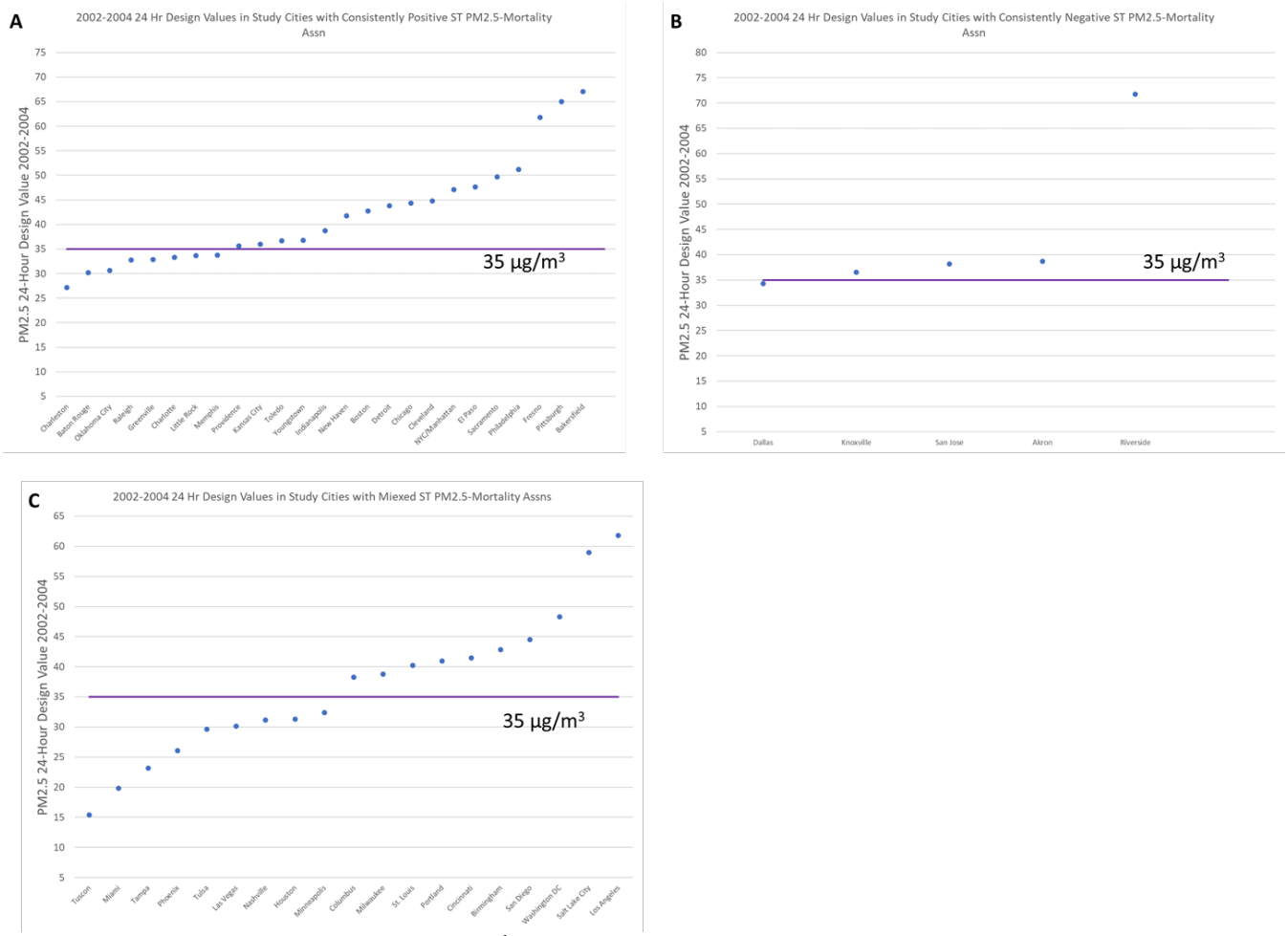
EPA Presentation of Results:

In addition to the unclear interpretation of the pseudo-DV method, the summary of the results presented on page 3-74 is confusing. I am not certain exactly what the EPA is trying to communicate with this summary information. This summary should be revised to be clearer in purpose and message. In addition, some information should be provided about findings for the 24-hr standard pseudo-design vales, in addition to the presentation about annual standard pseudo-design values.

I find it difficult to interpret summary statements such as the following, because of the difficulty in interpreting the results from the pseudo-design value analysis: "For the U.S. studies in this group, annual pseudo-design values from 9.9 to 11.7 μg/m3 correspond to 50th percentiles of study area populations (or health events)."

Alternative Method for Comparing Concentrations in Cities that Meet or Do Not Meet the Standard

An alternative to EPA's pseudo-design value analysis would be to assess the design values in particular cities for which city-specific short-term mortality estimates are available (e.g. Franklin 2007, 2008, Dai et al. 2014, Baxter et al. 2017, etc). The results can be separated by those cities that have positive associations between PM2.5 and mortality, and those with negative associations (setting aside the important consideration of statistical significance for the moment). I conducted an example analysis of this type, evaluating the 2002-2004 24-hr design value in cities that showed consistent positive, consistent negative, or mixed PM2.5-mortality associations in Franklin et al. 2007, Dai et al. 2014, and Baxter et al. 2017 (based on figures in these papers). The results are presented below in Lange Figure 10. This analysis demonstrates that very few of the cities where positive associations were found had a 3-year average 98th percentile 24-hour concentration below 35 μg/m3. More importantly, it showed that there was no discernible pattern in 98th percentile concentrations between studies with consistently positive associations, versus consistently negative, or with mixed associations.



Lange Figure 10. 3-year average 98th percentile 24-hour PM2.5 concentrations (2002-2004) in cities that showed consistently positive (A), consistently negative (B), or mixed associations (C) between PM2.5 and short-term mortality in Franklin et al. 2007, Dai et al. 2014, and Baxter et al. 2017.

Major Point #10: Methods for Risk Assessment

Choice of C-R Function:

For the pseudo-design value analysis the EPA excluded studies "in situations where the study population selection criteria was not random and not likely to be proportional to the underlying population, or the population selection criteria was not clearly specified (e.g., such as in cohort studies like the American Cancer Society cohort (ACS), Nurses' Health Study cohort (NHS), and the Health Professionals Follow-up Study (HPFS))." In the criteria for use of a C-R function from a particular study in the risk assessment, the EPA noted that "For that reason, we favored those epidemiology studies providing effect estimates for populations readily generalizable to the

broader U.S. population (e.g., specific age groups not differentiated by additional socio-economic, or employment attributes)." However, as stated in Appendix B, the ACS study cohort selection criteria was not random and not likely to be proportional to the underlying population – therefore, the three C-R functions from the ACS studies should not meet the EPA's criteria for inclusion in the risk assessment.

In Section C.1.1 the EPA provides a list of criteria that were used for choosing C-R functions for the risk assessment. They note a lack of availability of short-term studies with exposures derived using hybrid approaches. Based on the provided criteria, the EPA should provide reasoning for why Di et al. 2017a effect estimates aren't used – this study meets the EPA's criteria, and they use the hybrid approach for exposure modeling (as with Di et al. 2017b, which is used in the long-term mortality assessment).

Even if the general causal conclusions that the EPA provides are correct, it is still a big leap to go from the provided C-R functions to estimates of risk in the population. This is made worse by EPA's lack of uncertainty assessment. For example, Dr. Aliferis notes the importance of studies that include only the right covariates, not all of them, on the effect estimates. There are also considerations such as the importance of considering cause-specific hazard assessments versus total hazard assessments (the general consideration being that if an investigator is interested in a specific cause of death, a standard Cox PH model will treat a cohort member who dies from an unrelated cause as having been censored, which can inflate the study results; (Austin et al., 2016)).

Hazard Ratios and Relative Risks:

Another aspect of the uncertainty in the risk estimates is the use of C-R functions derived from Cox proportional hazards (PH) models. All the long-term studies provide estimates of the HR (Jerrett 2017 also calculates an HR, but it is mislabeled in the Abstract of the paper, and in EPA's Table C-1 as an RR). Relative risk is the absolute risk of an event happening to a member of group 1 (exposed) divided by the absolute risk of that event happening to a member of group two (unexposed) over a specified time period. The absolute risk is the probability of an event occurring. This is affected by time. For an outcome such as all-cause mortality, eventually all the people in both groups will have the event (probability = 1 for each), and the RR will = 1. The hazard rate can be thought of as the time-specific risk (expected events per unit time) of having the event of interest, provided (i.e., conditioned on the event) that the individual has not already had the event of interest. In a Cox PH model, there is an assumption of a constant ratio between the hazards (i.e. a constant HR) over time, in contrast to the RR that, as noted above, will be time-dependent.

The hazard function is a rate (probability of an event at time $t/\Delta t$), and therefore is not bounded by 0 and 1 as would be the case with a probability (which is what underlies a relative risk). From Sutradhar et al. 2018 the risk definition from hazard: Risk (t) = $S(t,x) = 1- S_0(t)^{\exp(X\beta)}$. Here $S_0(t)$

is the probability of remaining event-free by time t for n individual in the "baseline" or "control" group. This is a probability, and so is bound by zero and one.

Relative Rate of experiencing an event among exposed compared to controls = **HR** = exp(β) = Hazard Function at time t among exposed individuals/Hazard Function at time t among control individuals. By taking the ratio, the HR is now independent of time, and is only dependent on β.

$$HR = \frac{\lambda 0(t)\exp(X1\beta)}{\lambda 0(t)\exp(X2\beta)} = \exp(X1\text{-}X2)\,\beta$$

Relative Risk of experiencing the event by time t = risk function at time t among exposed/risk function at time t among controls: $RR = \frac{1-S0(t)^{\exp(\beta)}}{1-S0(t)}$

These two equations are clearly different from one another. Relative risk is time-dependent, and hazard rate is not. The HR is not an estimate of relative risk. The two estimates are not comparable in magnitude, although they are comparable in direction.

Another note about hazard ratios when EPA is comparing magnitudes of HRs from different studies: the magnitude of the HR will change depending on what covariates are considered in the model (HRs are "non-collapsible"). From Sutradhar 2018: "Authors should refrain from using the magnitude of the HR to describe the magnitude of the relative risk - this is incorrect."

Hazards and Health Impact Functions:

In this risk assessment the EPA references the BenMAP manuals for the methods of estimating health risks from PM2.5 concentration changes – in particular, the health impact functions (https://www.epa.gov/sites/production/files/2015-04/documents/benmap-ce_user_manual_march_2015.pdf). In Appendix C of that manual the EPA derives health impact functions for different functional forms of C-R relationships derived from epidemiology studies, categorized generally as linear, log-linear, and logistic. Different types of studies have different functional forms, with both time-series and Cox PH models being generally log-linear in their form, and case-crossover studies generally logistic. The derivation of the log-linear health impact function is pretty clear from the EPA's explanations on pp C-3 to C-5, and seems pretty straight-forward to apply to effect estimates from a time-series study.

For example, a time-series study is investigating the association between daily mortality count in an area and PM concentration. The equation for that association is:

$$\ln(y) = \alpha + \beta X;$$

where y=daily mortality count; α=intercept+other considered variables; X = PM concentration; and β=the slope of the association (the change in ln(count) per unit change in PM). The health impact function for this aims to answer the question: what is the change in y (Δy) in the population with a particular change in X (ΔX)? The form of the health impact function is then:

$$\Delta y = y_0 - y_c = y_0\left(1 - \frac{1}{\exp(\beta \Delta X)}\right)$$

Where y_0 is the baseline mortality count without a change in PM = mortality rate*total population in a particular area.

In this case, the input for y_0 is the daily mortality count (could also be the annual mortality count, depending on the study), which makes some sense because the original y variable was a mortality count.

However, for the results from Cox proportional hazards models it is trickier. Appendix C of the BenMAP manual devotes a page to this topic (C-11), that essentially states that the association should be treated the same as a log-linear, with notation that the relative risks are equal to the ratio of the hazards, which is equal to exp(β x ΔPM). The equation from pg C-11 is reproduced below:

$$RR = \frac{h(X_0, t)}{h(X_c, t)} = \frac{h_0(t)e^{X_0 \bullet \beta}}{h_0(t)e^{X_c \bullet \beta}} = e^{\Delta PM \bullet \beta},$$

This is not correct. This is the equation for the **hazard ratio (HR)**, not the relative risk. As discussed at length above, these two are not the same. What is not made clear in the BenMAP manual or the PM PA is how the difference in the "left-hand side" variable is dealt with in the EPA's health impact function for Cox PH results (i.e. the difference between hazards and mortality counts). Based on the Cox PH equation below:

$$\ln(hazard, t) = \lambda_0(t) + \beta X;$$

the change in health impact with a change in X (or PM), is the following:

$$\Delta(hazard, t) = (hazard, t)_0 - (hazard, t)_c = (hazard, t)_0\left(1 - \frac{1}{\exp(\beta \Delta X)}\right)$$

Given that the hazard describes the instantaneous rate of an event occurring at a specific point in time, provided that the event has not already occurred, it is not clear what number should be assigned to $(hazard,t)_0$ in this case. The EPA seems to use the same number as for the regular log-linear regression (population*mortality rate). It is also not clear how to interpret the $\Delta(hazard,t)$. In theory $(hazard,t)_0$ could be derived from the original epi study, but all of the other aspects of the equation (e.g. $\lambda_0(t)$) would also have to be available. The EPA needs to justify the use of these values in this equation, if they want to use this as the basis of their quantitative risk assessment.

Major Point #11 – Quantitative Uncertainty Analysis in the Risk Assessment

In section 3.3.2.4 the EPA presents information about the variability and uncertainty in their risk estimates. They capture some quantitative estimates of uncertainty and variability, such as using concentration-response (C-R) functions from different studies, and deriving estimates using the 95% confidence intervals of the C-R functions. These uncertainty measures would be more readily communicated by including them in a table or figure.

EPA assesses many more uncertainties only qualitatively, and several of those uncertainties are considered to have medium to high impacts on the risk estimates, including: simulating air quality to just meet current and alternative standards, representing population-level exposure in 12x12-km grid cells, and the shape of the C-R relationship at low PM concentrations. In addition to these considerations, there are multiple other aspects of the risk assessment that confer uncertainty on the risk estimate. Table 1 below shows very generally some example magnitudes of the potential uncertainties.

Table 1. An incomplete list of possible uncertainties in deriving risk estimates from reductions in PM2.5 concentrations.

Source of Uncertainty	Example Magnitude	Example Reference
Generating the Concentration-Response (C-R) Function in Epidemiology Studies		
Exposure Measurement Error	31-85%	Spatial error + population error (Dionisio et al., 2016)
Model Misspecification Error	50%	Generalized linear vs generalized additive models for PM10 (Sheppard, 2003)
Alternative C-R Functions within One Study	200%	Range of HRs generated using different exposure models (Jerrett et al., 2017)
Causal Relationship	0.35-1	US EPA Expert Elicitation of PM2.5 causality – provided is the range of probabilities that PM2.5 is causing mortality (Mansfield et al., 2009)
Air Quality Monitoring/Modeling		
Air Monitoring	± 10%	Allowable variation in 24-hour PM2.5 monitored concentrations
Air Modeling	± 30%	10-90% range for prediction of change in PM2.5

		concentrations (Mansfield et al., 2009)
Applying the Concentration-Response Function		
Choice of C-R Function	400%	Variability in PM2.5 long-term all-cause mortality estimates presented in Table 3-7 using C-R functions from different studies (PM PA 2019)
Baseline Incidence Rates	± 5%	Influence Analysis 10-90% range for prediction of base mortality rates (Mansfield et al., 2009)
Population Forecasts	± 10%	Influence Analysis 10-90% range for census 2020 population forecasts (Mansfield et al., 2009)
Use of National Estimates	± 1000%	Range of C-R estimates across 77 study cities, compared to the national estimate (Baxter et al., 2017)
Threshold in C-R	6-90%	Change in premature mortality from CPP repeal cutpoint analysis (LML cutpoint on low end of scale; NAAQS on high end of scale) (USEPA, 2017)

The EPA notes that they lack the information to conduct a full probabilistic uncertainty analysis, which is the type of analysis recommended by the World Health Organization (WHO) for this level of complexity of a risk assessment. However, this does not remove the need for a more quantitative assessment of the many sources of uncertainty in the risk estimates. My question to the expert consultants was:

- **Is there a quantitative uncertainty analysis method that the EPA could use for this risk assessment that captures more of the uncertainty and variability of the risk estimates (such as those described in Table 1), in order to better inform CASAC and the EPA Administrator about the impact of these uncertainties**?

From Dr. Jansen:

"Unless I missed it, I saw no discussion of the errors or biases in the measurements data generally nor the FRM/FEM specifically. Such errors and biases certainly exist, can be known, and would have some effect on the assessments. The magnitude of the effect would be study specific and, I would hope, be assessed by the authors of the health effects studies or risk assessors. I know that Dr. Paige Tolbert and the other members of the ARIES team took the issue seriously. As to the PM PA, I do not see that EPA has done a quantitative uncertainty analysis and, thus, has not included this issue. I am certainly a proponent of a quantitative uncertainty analysis being performed. And while I cannot respond with specifics to Dr. Lange's question number 4, I do believe substantial data does exist to derive estimates for many of the items in her Table 1 and EPA should get on with performing that work before the next review. They have been advised to do so in the past."

From Dr. Lipfert:

"Statistical methods are available for within-model random uncertainties, but the questions raised above about appropriateness of specific models have not been addressed and are far more important. There are also uncertainties as to responsible pollutants. PM2.5 is featured in part because of is extensive ambient monitoring network and the use of mathematical models to estimate more detailed locations. However, considering the time periods (decades) and spatial scale (the entire U.S.) it is likely that other pollutants may be involved, especially PM2.5 constituents."

From Dr. North:

"I respond with an emphatic yes to your question. There are lots of possibilities. For example, see the Anne Smith paper of March 2015 and her newly accepted paper."

From Dr. Sax:

"BenMAP uses Monte Carlo analyses from which 95% confidence intervals are derived which incorporates only the statistical uncertainty (the standard error of the beta coefficients) in the risk estimates. EPA could expand this Monte Carlo analysis to include other sources of uncertainty – such as the uncertainty in the variables described in Table 1."

From Dr. Thomas:

"There certainly are methods of uncertainty analysis that can be applied to the analysis of original epidemiologic data, were it available, but this would generally be beyond the capability of EPA staff without access to raw data (see my response to Dr. Cox's question 16). The simulations based on epidemiologic risk estimates used by EPA do address such uncertainties and variability, to the extent possible in meta-analysis."

Based on the information that I provide in Table 1, and the feedback from the expert consultants, there are a number of options for providing more quantitative uncertainty analyses for the PM2.5 risk assessment. The EPA should start using these methods, such as those discussed by Dr. North in his comments.

The EPA notes, rightly, in their limited uncertainty assessment that the 95% CI of the risk estimates provide a wide range on the estimates. They then note that "There are a number of factors potentially responsible for the varying degrees of statistical precision in effect estimates, including sample size, exposure measurement error, degree of control for confounders/effect modifiers, and variability in PM2.5 concentrations." However, this list does not justify the differences in CIs, put them in context, or help them be used to determine the validity or confidence in the risk estimates. Additional discussion of this variability should be incorporated to discuss how this information should be incorporated and interpreted in the risk assessment.

Comments on Qualitative Uncertainty:

In the EPA's qualitative uncertainty assessment, they note that most of the uncertainties are the same between the estimated risks for the current and alternative standards, so they would not impact the relative risk between the two standards. The EPA considers this to be true for the uncertainty about the shape of the C-R function at low PM2.5 concentrations, but this is not the case. Because the alternative standards ascribe risk to increasingly low concentrations of PM2.5, compared to the current standard, then the shape of the C-R function at those lower concentrations (which are disproportionately represented at lower standard levels), could impact the interpretation of the relative risk between the current and alternative standards. This is shown below with a reproduction of Figure 3-12 – the shape of the C-R function in the 6-8 ug/m3 range (for example) would disproportionately affect the risk estimates for the lower standards compared to the annual standard.

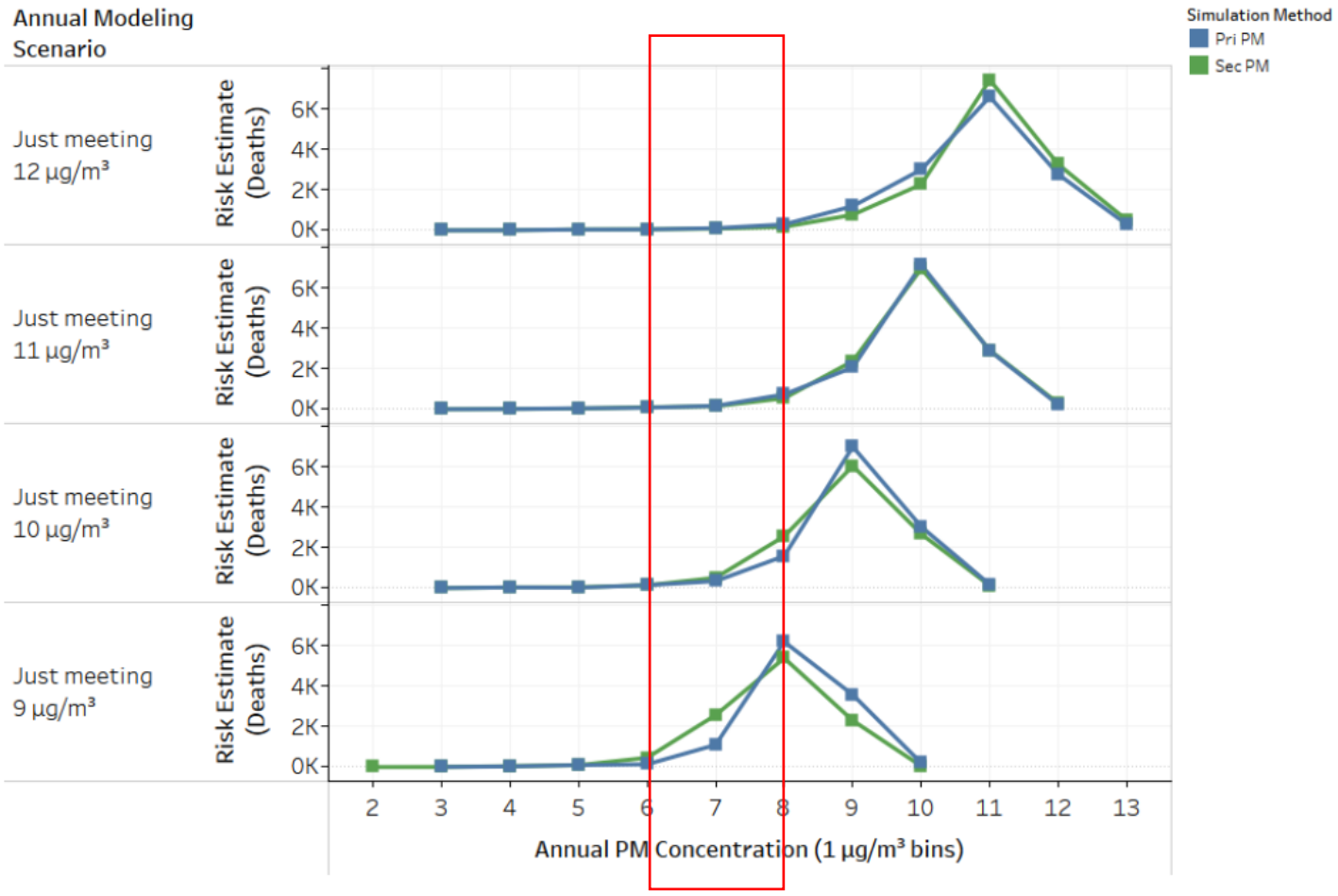


Figure 3-12. Distribution of absolute risk estimates ($PM_{2.5}$-associated mortality) for the current and alternative annual standards for the subset of 30 urban study areas where the annual standard is controlling (blue and green lines represent the Pri-$PM_{2.5}$ and Sec-$PM_{2.5}$ estimates, respectively).[68]

EPA PM PA Figure 3-12 with red box (added by me) to demonstrate the reliance of risk estimates on the shape of the C-R function at lower PM2.5 concentrations.

Dr. Jansen noted multiple times in his responses to questions the importance of carrying the uncertainties in the air quality modeling through the health assessment.

Additional Notes

In Appendix B the EPA states that the pseudo-design value box plots were built using, among other data, US Census population data from 2015. It should be noted that there was not a Census

in 2015. In actuality, the EPA is using Census data from 2010 and extrapolating to 2015 using the methods of Woods & Poole. This should be corrected.

Figure B-6 – this figure is confusingly labeled (only mentions long-term studies, actually short-term and long-term studies shown in figure).

A side-by-side comparison should be offered for the sensitivity analysis shown in Figure B-6, comparing the county population-based method to the actual number of health effects method.

The EPA should clarify whether the "selected betas" in Table C-1 are per 10 ug/m3 PM2.5, or 1 ug/m3 PM2.5, or per some other estimate.

Future Research for PM2.5 NAAQS

Charge Question: What are the CASAC views regarding the areas for additional research identified in Chapters 3, 4 and 5? Are there additional areas that should be highlighted?

In Section 3.5, the EPA identifies an area for future research and data collection as "Further elucidating the physiological pathways through which exposures to the PM2.5 concentrations present in the ambient air across much of the U.S. could be causing mortality and the morbidity effects shown in many epidemiologic studies."
Part of this research should be a plan for what will be done with the data, depending on what kinds of patterns it shows. For example, if low-dose experimental studies show little or no effects (which seems likely, because many higher dose studies have shown little or no effect), then how will this change the conclusions that are being drawn about the epidemiology study results? Or will only positive study results be considered?

In addition, the EPA must put some cap on the amount of time and effort to be spent determining potential physiological pathways of effects demonstrated in epidemiologic studies. Associations with many different endpoints have been demonstrated in studies, including everything from Alzheimer's disease, anemia, baldness, bladder cancer, burglaries, dermatitis, insomnia, obesity, and many more (Cox, 2018 provides a much longer list). What will be the cut-off of associations for which the EPA will look for a mechanism?

The EPA also states a future research goal as "Improving our understanding of the PM2.5 concentration-response relationships near the lower end of the PM2.5 air quality distribution, including the shapes of concentration-response functions and the uncertainties around estimated functions for various health outcomes and populations".
If this is to be done, it must consider the impacts of measurement error and other errors and biases before going too deeply into this research - first determine what kinds of studies can actually determine the shape of the C-R function, and then pursue the research.

The EPA needs to support further causality research on these topics, as described (for example), in Judea Pearl's "Book of Why". The EPA also needs to support research and method development for quantitative uncertainty analyses.

References

Austin, P.C., Lee, D.S., Fine, J.P., 2016. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation 133, 601–609. https://doi.org/10.1161/CIRCULATIONAHA.115.017719

Avery, C.L., Mills, K.T., Williams, R., McGraw, K.A., Poole, C., Smith, R.L., Whitsel, E.A., 2010a. Estimating Error in Using Ambient PM2.5 Concentrations as Proxies for Personal Exposures: A Review. Epidemiology 21, 215–223. https://doi.org/10.1097/EDE.0b013e3181cb41f7

Avery, C.L., Mills, K.T., Williams, R., McGraw, K.A., Poole, C., Smith, R.L., Whitsel, E.A., 2010b. Estimating error in using residential outdoor PM2.5concentrations as proxies for personal exposures: A meta-analysis. Environ. Health Perspect. 118, 673–678. https://doi.org/10.1289/ehp.0901158

Baxter, L.K., Crooks, J.L., Sacks, J.D., 2017. Influence of exposure differences on city-to-city heterogeneity in PM2.5-mortality associations in US cities. Environ. Health 16, 1. https://doi.org/10.1186/s12940-016-0208-y

Brauer, M., Brumm, J., Vedal, S., Petkau, A.J., 2002. Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects. Risk Anal. 22, 1183–93.

Carone, M., Dominici, F., Sheppard, L., 2019. In Pursuit of Evidence in Air Pollution Epidemiology: The Role of Causally Driven Data Science. Epidemiology. https://doi.org/10.1097/EDE.0000000000001090

Carrothers, T.J., Evans, J.S., 2000. Assessing the impact of differential measurement error on estimates of fine particle mortality. J. Air Waste Manag. Assoc. 50, 65–74.

Cox, L.A. (Tony), 2018. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ. Res. 164, 636–646. https://doi.org/10.1016/j.envres.2018.03.038

Dionisio, K.L., Baxter, L.K., Burke, J., Özkaynak, H., 2016. The importance of the exposure metric in air pollution epidemiology studies: When does it matter, and why? Air Qual. Atmos. Heal. 9, 495–502. https://doi.org/10.1007/s11869-015-0356-1

Fewell, Z., Davey Smith, G., Sterne, J.A.C., 2007. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am. J. Epidemiol. 166, 646–55. https://doi.org/10.1093/aje/kwm165

Greven, S., Dominici, F., Zeger, S., 2011. An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information. J. Am. Stat. Assoc. 106, 396–406. https://doi.org/10.1198/jasa.2011.ap09392

Henneman, L.R.F., Liu, C., Mulholland, J.A., Russell, A.G., 2017. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. J. Air Waste Manag. Assoc. 67, 144–172. https://doi.org/10.1080/10962247.2016.1242518

Jerrett, M., Turner, M.C., Beckerman, B.S., Pope, C.A., van Donkelaar, A., Martin, R. V., Serre, M., Crouse, D., Gapstur, S.M., Krewski, D., Diver, W.R., Coogan, P.F., Thurston, G.D., Burnett, R.T., 2017. Comparing the Health Effects of Ambient Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure Estimates. Environ. Health Perspect. 125, 552–559. https://doi.org/10.1289/EHP575

Lipfert, F.W., Wyzga, R., 1996. The effects of exposure error on environmental epidemiology., in: Proceedings of the Second Colloquium on Particulate Air Pollution and Human Mortality and Morbidity. pp. 295–302.

Mansfield, C., Sinha, P., Henrion, M., 2009. Influence Analysis in Support of Characterizing Uncertainty in Human Health Benefits Analysis.

Pearl, J., Mackenzie, D., 2018. The Book of Why: The New Science of Cause and Effect. Penguin Books Limited.

Pun, V.C., Kazemiparkouhi, F., Manjourides, J., Suh, H.H., 2017. Long-Term PM2.5 Exposure and Respiratory, Cancer, and Cardiovascular Mortality in Older US Adults. Am. J. Epidemiol. 1–9. https://doi.org/10.1093/aje/kwx166

Rhomberg, L.R., Chandalia, J.K., Long, C.M., Goodman, J.E., 2011. Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit. Rev. Toxicol. 41, 651–671. https://doi.org/10.3109/10408444.2011.563420

Rich, D.Q., 2017. Accountability studies of air pollution and health effects: lessons learned and recommendations for future natural experiment opportunities. Environ. Int. 100, 62–78. https://doi.org/10.1016/j.envint.2016.12.019

Sheppard, L., 2003. Ambient Air Pollution and Nonelderly Asthma Hospital Admissions in Seattle, Washington, 1987-1994.

Sutradhar, R., Austin, P.C., 2017. Relative rates not relative risks: Addressing a widespread misinterpretation of hazard ratios. Ann. Epidemiol. 28, 54–57. https://doi.org/10.1016/j.annepidem.2017.10.014

Thurston, G.D., Ahn, J., Cromar, K.R., Shao, Y., Reynolds, H.R., Jerrett, M., Lim, C.C., Shanley, R., Park, Y., Hayes, R.B., 2016. Ambient Particulate Matter Air Pollution Exposure and Mortality in the NIH-AARP Diet and Health Cohort. Environ. Health Perspect. 124, 484–90. https://doi.org/10.1289/ehp.1509676

USEPA, 2018. Integrated Science Assessment for Particulate Matter (External Review Draft).

USEPA, 2017. Regulatory Impact Analysis for the Review of the Clean Power Plan: Proposal.

Utell, M.J., Frampton, M.W., Devlin, R.B., Cascio, W.E., 2002. CARDIOVASCULAR EFFECTS ASSOCIATED WITH AIR POLLUTION: POTENTIAL MECHANISMS AND METHODS OF TESTING. Inhal. Toxicol. 8378, 1231–1247. https://doi.org/10.1080/0895837029008488

Watt, M., Godden, D., Cherrie, J., Seaton, A., 1995. Individual exposure to particulate air pollution and its relevance to thresholds for health effects: a study of traffic wardens. Occup. Environ. Med. 52, 790–2.

Yoshimura, I., 1990. The effect of measurement error on the dose-response curve. Environ. Health Perspect. 87, 173–8.

Zareba, W., Nomura, A., Couderc, J.P., 2001. Cardiovascular effects of air pollution: What to measure in ECG? Environ. Health Perspect. 109, 533–538. https://doi.org/sc271_5_1835 [pii]

Dr. Corey Masuca

Comments on Chapter 2

2.2.3 Predicted Ambient PM2.5 Based on Hybrid Modeling Approaches

Need to reference secondary PM2.5 formation modeling (MERPS)

What about personal air monitors, with their shortcomings (i.e., biases of twice the actual concentrations, algorithms, etc.)?

2.4.3 Estimating Background PM with Recent Data

No reference to background transport from interstate transmission.

Comments on Chapter 3

3.1.1.3 Form

More explanation needs for the determination to utilize the 98th percentile, averaged over three years for the 24-hour standard. Does this comport with most epidemiological, toxicological, human studies?

3.1.1.4 Level

Is there a relationship between the short-term standard and the long-term standard for PM2.5?

3.1.2 General Approach in the Current Review

Should pseudo metrics be utilized since they are more based on the levels of air quality designed to meet the standards as opposed to pure or solely-based health data?

3.2.3.2.2 PM2.5 Pseudo-Design Values in Locations of Key Epidemiological Studies

Should pseudo metrics be utilized since they are more based on the levels of air quality designed to meet the standards as opposed to pure or solely-based health data?

3.3 Risk-Based Considerations

Clarification of the importance of a risk assessment

Is a risk assessment needed there is no level of pollutant concentration as which there would be no appreciable risk?

Risk assessments are often wrought with uncertainties.

The lack of definition of an acceptable risk level or range.

3.3.1 Overview of Approach

Regarding health outcomes reviewed, respiratory-related health outcomes are noticeably absent.

Regarding the use of linear interpolation/extrapolation to additional annual standard levels, is the relationship between risk and concentrations linear?

3.4.2.3 Form

Need better explanation of determination of annual mean averaged over three years for annual standard and 98th percentile averaged over three years for the 24-hour standard.

3.4.2.4.1 Alternative Annual Standard Levels

In a discussion of the selection of maintaining the current annual standard versus potential alternative lower standards such as 10 micrograms/cubic meters, only discussions of the pros (benefits) and cons (uncertainties) are discussed. There does not appear to be a sound and solid recommendations to either maintain the current standard or a recommendation for a lower standard such as 10 micrograms/cubic meters.
3.4.2.4.2 Alternative 24-Hour Standard Levels

In a discussion of the selection of maintaining the current 24-hour standard, there is a concise recommendation to maintain the current 24-hour standard. (Page 3-112)

Comments on Chapter 4

4.1.2 Approach in the Current Review

Why the selection of coarse PM as a surrogate for PM!0 when there exist health studies solely based on PM10 and coarse PM and due to the uncertainty of older studies utilized the older "subtraction" methods of determining coarse PM concentrations?

Dr. Steven Packham

Charge Question Chapter 1 – Introduction:

To what extent does the CASAC find that the information in Chapter 1 is clearly presented and that it provides a useful context for the review?

Preliminary Comment:

The information in Chapter 1 is useful. The approach outlined in the Table of Contents is clear. Chapter 1 allows a logical historical analysis of the NAAQS review process.

The INTRODUCTION narrative is written in the present tense; e.g., "*This… document presents the draft policy assessment (PA) for the EPA review of the PM NAAQS*."

The PURPOSE section is also written in the present tense; e.g., *"The PA evaluates the potential policy implications of available scientific evidence…"; "The role of the PA is to help 'bridge the gap' between the Agency's scientific assessments and the judgements required by the Administrator…"*, etc.

Use of present tense in the narrative facilitates the construction of clear and concise sentences. This makes Chapter 1 relatively easy to read and extract critically important bits of information. For example, one can readily extract the stated purposes of the PM PA, as shown later.

Chapter 1 is clear, well written and useful. Staff should be commended.

On the downside, Chapter 1 lacks precision in use of the term, *science*. This is a problem throughout the draft PM PA and all current NAAQS review documents. To demonstrate how a few of the many nuanced meanings and connotations of the word *science* are used - and the challenge this poses to the CASAC whose members represent a variety of scientific disciplines - a list of the PM PA purposes extracted from Chapter 1 is presented. (Hint: Please note purpose 5; subpart a.)

List of PM PA Purposes:

1 Evaluate the policy implications of the available *science*.
2 Help bridge the gap between the Agency's *scientific assessments* and Administrator judgements.
3 Evaluate the adequacy of the current standard.

4 Consider alternatives to the basic standard elements of indicator, averaging time, form, and level.
5 Facilitate advice to the Agency and the Administrator from the CASAC on
   a. Agency assessments of relevant *scientific* information.
   b. Adequacy of the current standard.
   c. Revisions of the current standard as may be appropriate.
6 Be a useful reference to all parties interested in the review of the PM NAAQS.

Forms of the root word science are used liberally as a modifying adjective; e.g., *scientific* evidence, *scientific* knowledge, and relevant *scientific* information.

The document title, Integrated Science Assessment, is particularly interesting. It's probably most often taken to mean an assessment made by integrating evidence, information and knowledge from multiple sciences. But this could be incorrect. Purpose 2 given above in the list of purposes seems to suggest the Agency views itself as performing a pure *Scientific*-Assessment. Which begs the questions: What is the whole point of the scientific assessment? and Is this scientific assessment being conducted in a scientific manner?

The paragraph beginning on Line14, page 1-2 may suggest an answer to the first question. Quote:

> In this draft PA, we take into account the available scientific evidence, as assessed in the external review draft Integrated Science Assessment for Particulate Matter (draft ISA [U.S. EPA, 2018]), and additional policy-relevant analyses of air quality and *risk*. (Emphasis added)

The whole point of the scientific assessment is air quality *risk*. The words risk and risk assessment are not prominently featured in air quality criteria documents prior to 1997.

A review of the PM NAAQS will substantiate the fact that following the promulgation and successful defense of legal challenges of the of PM standards beginning in 1997 (PM PA pp 1-7 to 1-10), the Agency's reliance on risk assessments was markedly increased and words like *scientific information*, and *science assessment* began to refer more to results of risk data and risk assessments.

The EPA publication titled *Review of Process for Setting National Ambient Air Quality Standards* (2006) foreshadowed the NAAQS review process as we know it today. The significance of this publication in the development of an institutional bias toward use and reliance on the inferential science of statistical risk and the confounding, interchangeable use, of the terms *risk-assessment* and *science assessment*, should be mentioned and an acknowledgment

of is impact on the relevance of toxicology and human clinical studies on causal determination in NAAQS ISA's should be included in Chapter 1.

A sentence from Chapter 1, in which the word *scientific* is implied [represented in brackets] but not actually used to modify the words *evidence* and *analyses*, may offer another insight into the Agency's mindset in using the CASAC and public comment as a collective diviner of what *science* is supposed to mean. To quote:

> "Our approach to considering the available [scientific] evidence and analyses in this draft PA has been informed by advice received from the CASA and from public comment.", and "…the final PA will also be informed by advice and recommendations received from the CASAC and… by public comment."

The draft PM PA is replete with confounding terms such as, "the Agency's *scientific (or risk) assessments*," and "the Agency's *assessments of relevant scientific (or risk) information.*" These terms and those listed above in the previous paragraphs refer to a critical body of information in the scientific literature that ultimately needs to get to the Administrator.

The CASAC's independent scientific review of this huge body of information is squarely the committee's statutory responsibility and role in the *need-to-get-to-the-Administrator* step in the NAAQS review process.

Conclusion. It is too late to seriously consider any significant editorial, or scientifically substantive, changes before finalizing the draft PM PA. But, in future PM PA drafts and in the imminent review of the ozone ISA, staff must clarify its definitions and use of terms when referring to *science* in general. It must clarify when, and if, forms of the word *science* are being used in reference to the inferential science of statistical risk as opposed to data from toxicology experiments and controlled human exposure studies.