DATE

EPA-CASAC-20-XXX

The Honorable Andrew R. Wheeler
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: CASAC Review of the EPA's *Policy Assessment for the Review of the National Ambient Air Quality Standards for Particulate Matter (External Review Draft – September 2019)*

Dear Administrator Wheeler:

The Chartered Clean Air Scientific Advisory Committee (CASAC) met on October 22, 2019, and October 24-25, 2019, to peer review the EPA's *Policy Assessment for the Review of the National Ambient Air Quality Standards for Particulate Matter (External Review Draft – September 2019)*, hereafter referred to as the Draft PM PA. The CASAC's consensus responses to the agency's charge questions and individual review comments from members of the CASAC are enclosed. Questions from CASAC members to the pool of non-member consultants and their responses are also enclosed. Major comments and recommendations are highlighted below and detailed in the consensus responses to charge questions.

Overall, the CASAC finds that the Draft PM PA depends on a Draft Particulate Matter (PM) Integrated Science Assessment (ISA) that, as noted in the April 11, 2019, CASAC Report on the Draft PM ISA, does not provide a sufficiently comprehensive, systematic assessment of the available science relevant to understanding the health impacts of exposure to PM, due largely to lack of a comprehensive, systematic review of relevant scientific literature; inadequate evidence and rationale for altered causal determinations; and a need for clearer discussion of causality and causal biological mechanisms and pathways. Given these limitations in the underlying science basis for policy recommendations, and diverse opinions about what quantitative uncertainty analysis and further analysis of all relevant data using the best available scientific methods would show, most CASAC members conclude that the Draft PM PA does not establish that new scientific evidence and data reasonably call into question the public

health protection afforded by the current 2012 $PM_{2.5}$ annual standard. Other members of CASAC conclude that the weight of the evidence, particularly reflecting recent epidemiology studies showing positive associations between $PM_{2.5}$ and health effects at estimated annual average $PM_{2.5}$ concentrations below the current standard, does reasonably call into question the adequacy of the 2012 annual $PM_{2.5}$ National Ambient Air Quality Standards (NAAQS) to protect public health with an adequate margin of safety. The CASAC also finds, in agreement with the EPA, that the available evidence does not reasonably call into question the adequacy of the current 24-hour $PM_{2.5}$ standard, PM_{10} standard, or secondary PM standards and concurs that they should be retained.

Turning to specific comments on chapters in the Draft PM PA, the CASAC finds that Chapter 1 provides a clear, although brief, discussion of legislative background and history that provide useful context for the review. The CASAC recommends that the EPA consider adding a discussion of the exceptional nature of the current CASAC and NAAQS review process, including (a) Administrator Pruitt's "Back to Basics" memorandum; (b) the disbanding of the CASAC PM Review Panel and streamlining of the review process to promote timely advice; (c) the appointment of a pool of non-member consultants to expand the expertise and fields of knowledge used to inform the CASAC's review; and (d) the Administrator's and CASAC's explicit emphasis on sound science throughout the review process, including critically reexamining long-standing practices and assumptions in light of recent data and methods. A brief discussion of the accelerated timeline, which has led to tight schedules, condensed procedures, and a Draft PM PA being produced before the PM ISA was revised/finalized, might also be useful. The CASAC recommends several measures, most already discussed in the CASAC Report on the Draft PM ISA, to more fully realize the Draft PM PA's stated goals of serving as a source of policy-relevant information, being understandable to a broad audience, and facilitating the CASAC's advice to the Agency and recommendations to the Administrator.

The CASAC finds the information in Chapter 2 to be clearly presented and useful as context for the review, but recommends adding discussions of (a) uncertainties in the emissions inventory (magnitude, spatial allocation, temporal allocation, and speciation) for each pollutant or source sector, and how these uncertainties affect the air quality modeling results and risk analysis in Chapter 3; (b) national maps of county-level emissions of pollutants to show variability across the country; (c) measurement uncertainty (accuracy, precision, bias, and error) for Federal Reference Method (FRM), continuous Federal Equivalent Method (FEM), Chemical Speciation Network (CSN), and Interagency Monitoring of Protected Visual Environments (IMPROVE) monitors and results of comparing measurements from co-located FRM/FEMs across the country and (d) how measurement errors and biases for $PM_{2.5}$ affect the evidence-based and risk-based assessments in Chapter 3. Chapter 2 should add discussions of the Southeastern Aerosol Research and Characterization Study (SEARCH) network; air sensor shortcomings and performance issues; the concept of "urban increment," including examples comparing $PM_{2.5}$ speciation in urban areas to nearby Class I areas; and how exceptional events are accounted for in the setting of the standard. Although Chapter 2 adequately covers annual background concentrations, it should add a discussion of background concentrations for the daily $PM_{2.5}$ standard, including ^{14}C research to distinguish between fossil-derived carbon and "modern" carbon.

For Chapter 3, the CASAC urges the EPA to follow the CASAC's advice on the Draft PM ISA by providing a balanced summary of the study results for each health endpoint; updating the causality determinations for $PM_{2.5}$ and cancer and nervous system effects, and for long-term ultrafine particulate exposure and nervous system effects; and describing and quantifying uncertainties (including epistemic uncertainties from measurement errors and model uncertainty) and the extent to which they have changed since the last review. The PM PA should discuss the rationale for comparing long-term and short-term $PM_{2.5}$ concentrations and should clarify the use and interpretation of pseudo-design values. Some members of the CASAC are concerned that the risk assessment approach in Chapter 3 treats regression concentration-response (C-R) functions (that is, functions describing associations between past estimated exposure concentration levels and mortality rates) as if they were causal C-R functions (that is, describing how changing future exposure concentrations would change future mortality rates). Because this is technically unsound, the CASAC recommends that the PM PA explicitly state the implicit assumption that regression coefficients can be used to quantify causality, noting that it is not necessarily a valid assumption, and provide information about whether the assumption has been tested and what the results were. Methodological details should be added to Appendix C of the PM PA, model validation results for BenMAP should be discussed, CMAQ model performance should be evaluated specifically at the study city locations, and study areas with poor model performance should not be used. Future changes in public health risks that might be caused by reducing $PM_{2.5}$ exposures are currently highly uncertain. The CASAC recommends that the PM PA better characterize this uncertainty using quantitative uncertainty analysis. Such an analysis should account for model uncertainty, exposure estimation errors, and both inference (internal validity) and generalization (external validity) uncertainties. As described above and in further detail in the consensus responses, the CASAC members did not come to consensus on whether the new scientific evidence and data reasonably call into question the public health protection afforded by the current 2012 $PM_{2.5}$ annual standard. The CASAC recommends that the final PM PA provide quantitative uncertainty and sensitivity analyses to provide a clearer technical and scientific basis for data interpretation and policy making. The CASAC agrees with the EPA and finds that the available evidence does not call into question the adequacy of public health protection afforded by the current 24-hour $PM_{2.5}$ standard and concurs that it be retained.

Chapter 4 provides a helpful summary and background for approaches taken in previous reviews of the PM_{10} standard. Some CASAC members conclude that new studies since the previous review justify the change in causality determination from "inadequate" to "suggestive" for long-term exposure to $PM_{10\text{-}2.5}$ and mortality and cardiovascular effects, and that the new data strengthen evidence for effects on cancer, and short-term effects on mortality, cardiovascular disease, and respiratory disease. Other CASAC members find the definitions of causal determination categories given in the Draft PM ISA and summarized in the Draft PA PM unclear, and the change to "suggestive" not clearly justified. These members also find that the evidence presented does not imply, or make likely, that $PM_{10\text{-}2.5}$ caused these effects. The CASAC agrees 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."

The CASAC finds much of the information in Chapter 5 on visibility and material effects of $PM_{2.5}$ to be useful, but recommends that the PM PA should more thoroughly address effects of reducing $PM_{2.5}$ on

climate change by providing quantitative estimates and uncertainty bands for effects of changes in $PM_{2.5}$ on global dimming and global warming and their consequences for economic and welfare effects on the United States. At a minimum, estimates of the change in warming caused by a change in $PM_{2.5}$ should be discussed and implications for human welfare in the United States should be evaluated. The CASAC agrees with the EPA that the available evidence does not call into question the protection afforded by the current secondary PM standards and concurs that they should be retained.

The CASAC appreciates the opportunity to provide advice on the Draft PM PA and looks forward to the agency's response.

Sincerely,

Dr. Louis Anthony Cox, Jr., Chair
Clean Air Scientific Advisory Committee

Enclosures

NOTICE

This report has been written as part of the activities of the EPA's Clean Air Scientific Advisory Committee (CASAC), a federal advisory committee independently chartered to provide extramural scientific information and advice to the Administrator and other officials of the EPA. The CASAC provides balanced, expert assessment of scientific matters related to issues and problems facing the agency. This report has not been reviewed for approval by the agency and, hence, the contents of this report do not represent the views and policies of the EPA, nor of other agencies within the Executive Branch of the federal government. In addition, any mention of trade names or commercial products does not constitute a recommendation for use. The CASAC reports are posted on the EPA website at: http://www.epa.gov/casac.

U.S. Environmental Protection Agency
Clean Air Scientific Advisory Committee

CHAIR
Dr. Louis Anthony (Tony) Cox, Jr., President, Cox Associates, Denver, CO

MEMBERS
Dr. James Boylan, Program Manager, Planning & Support Program, Air Protection Branch, Georgia Department of Natural Resources, Atlanta, GA

Dr. Mark W. Frampton, Professor Emeritus of Medicine, Pulmonary and Critical Care, University of Rochester Medical Center, Rushville, NY

Dr. Ronald J. Kendall*, Professor of Environmental Toxicology and Head, Wildlife Toxicology Laboratory, Texas Tech University, Lubbock, TX

Dr. Sabine Lange, Toxicology Section Manager, Toxicology Division, Texas Commission on Environmental Quality, Austin, TX

Dr. Corey M. Masuca, Principal Air Pollution Control Engineer, Air and Radiation Protection Program, Environmental Health Services, Jefferson County Department of Health, Birmingham, AL

Dr. Steven C. Packham, Toxicologist, Division of Air Quality, Utah Department of Environmental Quality, Salt Lake City, UT

SCIENCE ADVISORY BOARD STAFF
Mr. Aaron Yeow, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board, Washington, DC

*Did not participate in review

Consensus Responses to Charge Questions on the EPA's *Policy Assessment for the Review of the National Ambient Air Quality Standards for Particulate Matter (External Review Draft – September 2019)*

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?

The purpose of the Policy Assessment (PA) is to bridge the gap between EPA's scientific assessments and the judgement required by the EPA Administrator when determining whether to retain or revise the National Ambient Air Quality Standards (NAAQS). It is unusual for the CASAC to review a draft PA when the Draft Integrated Science Assessment (ISA) had not been revised/finalized, and to include the Risk and Exposure Assessment (REA) as part of the PA rather than as a stand-alone document that is reviewed and revised/finalized prior to the release of the draft PA. The CASAC stands ready to review additional drafts of the PM ISA, PM REA, and PM PA, should they be provided. This document is limited to providing advice on the Draft PM PA given the documents provided to date.

The discussions of legislative background and history are clearly, although briefly, presented. They provide useful context for the review. The CASAC recommends that the EPA consider adding a discussion of the exceptional nature of the current CASAC and NAAQS review process. Specifically, relevant background on changes in processes and procedures could include: (a) further details of Administrator Pruitt's "Back to Basics" memorandum (adding to the discussion on p. 1-12); (b) the disbanding of the CASAC Particulate Matter (PM) Review Panel and streamlining of the review process to promote timely advice; (c) the appointment of a pool of non-member consultants to expand the expertise and fields of knowledge used to inform the CASAC's review; and (d) the Administrator's and CASAC's explicit emphasis on sound science throughout the review process, including reexamination of long-standing assumptions and frameworks used in previous reviews. A brief discussion of the accelerated timeline, which has led to tight schedules, condensed procedures, and a Draft PM PA being produced before the PM ISA was revised/finalized, might also be useful. Relevant background on methodological changes in the current CASAC's scientific and technical approach in this review cycle could be provided in a separate section. These changes include the following:

(1) Drawing and preserving key conceptual distinctions between:
    a. Association vs. causation
        i. Formal quantitative causal inference vs. judgment for drawing causal conclusions from data;
        ii. Manipulative or interventional causation vs. Bradford Hill or weight-of-evidence (WoE) causation;
        iii. Statistical vs. biological (mechanistic) concepts of causation;
    b. Empirically verified evidence vs. unverified assumptions and models;
    c. Estimated vs. actual individual exposures; and

d. Explicitly derived and independently verifiable scientific conclusions vs. expert judgments.

(2) Emphasis on more effective integration of information from animal toxicology and controlled human exposure studies to:

a. Elucidate and validate potential (i.e., hypothesized) causal biophysical mechanisms underlying epidemiologically suggested health risks; and

b. Better characterize dose-dependent thresholds and causal biological C-R functions for pulmonary inflammation and other physiological responses in other tissues, organs, and systems in response to the physical and chemical characteristics of inhaled PM.

The stated intentions for the Draft PM PA presented in Chapter 1 include "to serve as a source of policy-relevant information;" "to be understandable to a broad audience;" and "to facilitate advice to the Agency and recommendations to the Administrator" from the CASAC. The CASAC recommends that these intentions be more fully realized in the PM PA by undertaking the following measures:

(1) Provide clear operational definitions of the key quantities and terms used to calculate and communicate scientific results;
(2) Apply a more explicit, systematic and transparent process for selecting, evaluating, summarizing, and interpreting studies. Show the results of this process for individual studies relied on in the Draft PM PA (e.g., for the 8 key studies in Appendix C);
(3) Increase transparency and logical soundness in deriving conclusions from data by documenting exactly how conclusions were reached in enough detail so that others can trace and check the logic used;
(4) Distinguish between estimated and actual exposures throughout the PM PA;
(5) Quantify uncertainty and variability in risk predictions; and
(6) Critically discuss the biological realism and results of empirical validation tests of risk predictions and modeling assumptions against observations.

<u>Improving Policy Relevance</u>

To make the information presented in the PM PA most useful and relevant for policy purposes, it is essential to *characterize what is currently known about how interventions that reduce $PM_{2.5}$ exposures affect public health risks,* using the best available scientific evidence and methods. Most members of the CASAC conclude that the current Draft PM ISA and Draft PM PA do not provide this crucial policy-relevant information using sound science due to:

(1) *Omitted evidence.* The CASAC again recommends, as it did for the Draft PM ISA, that the EPA should state, and then systematically apply, specific criteria for selecting, evaluating, summarizing, and synthesizing relevant literature. The Draft PM PA and Draft PM ISA omit dozens of recent and relevant intervention, accountability, and time series studies that provide empirical evidence on how much reducing $PM_{2.5}$ exposures affect public health risks (e.g., Burns et al., 2019; Henneman et al., 2017). Omitting studies that provide empirical evidence on

changes in public health risks following changes in $PM_{2.5}$ exposures makes the Draft PM PA less valuable than it should be as a source of policy-relevant information.

(2) *Focus on associations.* The empirical evidence presented in the Draft PM ISA and Draft PM PA and relied on for quantitative risk predictions in the Draft PM PA focuses on statistical associations between past estimated levels of $PM_{2.5}$ and past mortality and morbidity rates. This information on C-R associations is less relevant for informing decisions than data on whether and how much reducing exposures reduces public health risks. Statistical associations reflect contributions from many sources, such as historical trends, residual confounding, omitted confounders, modeling choices, unmodeled heterogeneity, model specification errors, and exposure estimation errors. For policy makers, only the fraction of a total C-R association that can be controlled by making different decisions is relevant. The current Draft PM PA and Draft PM ISA do not quantify this avoidable fraction of the total C-R association, nor quantify how much of it could be reduced by changing current NAAQS standards. This is the critical information needed to guide scientifically well-informed policy making.

(3) *Insufficient biological validation.* The biological evidence presented in the Draft PM PA does not show that integration and analyses of physiological response data from controlled human exposure studies validate proposed causal mechanisms or modes of action underlying public health risks.

(4) *Appropriate technical methods of causal analysis are not used.* Different technical methods are needed to provide valid answers to questions about effects caused by interventions than to estimate associations (Pearl, 2010). 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, 2019).

As indicated in these references (and discussed further in consultant responses to CASAC questions by Drs. North and Aliferis in Appendix D), appropriate causal analysis, inference, and discovery frameworks developed and widely applied in other areas of science over the past century are now readily available. There has been substantial progress since previous NAAQS reviews in technical and scientific methods for causal analysis, inference, and discovery used in environmental health. Advances include use of causal graph models instead of regression models and use of appropriate adjustment sets to obtain unbiased estimates of causal effects after correcting for measured confounders and other causally relevant variables. These methods and frameworks have not been used in the Draft PM PA. Some members of the CASAC recommend that the PM PA should be updated to reflect progress and the best available methods in quantitative causal analysis, inference, and discovery.

(5) *Unsound technical methods are used, conflating association and causation.* Quantitatively, the Draft PM PA treats regression C-R functions (Appendix C) as if they were causal C-R functions for purposes of simulating how reducing $PM_{2.5}$ exposures would reduce mortality risks. This is

technically unsound. Statistical associations do not in general provide the information needed to determine or simulate how changing exposure would change risk of responses (Pearl, 2010). All of the 8 key studies in Appendix C address C-R associations (e.g., regression functions). None of them quantifies effects on risk caused by changing exposures. None of the key studies controls for potential confounders (e.g., month and daily high and low temperatures in the days and weeks preceding mortality, which explain $PM_{2.5}$-mortality C-R associations in some studies) using appropriate methods (e.g., identifying and conditioning on adjustment sets to obtain unbiased estimates of causal effects). It is not scientifically-sound practice to treat associations with uncontrolled confounding as if they were interventional causal relationships, or to use them to predict or simulate effects on mortality risks of changing $PM_{2.5}$. The PM PA should clearly distinguish between regression and causal relationships throughout, and it should state that it has quantified regression relationships in Appendix C.

(6) *Causal determination conceptual framework not validated for application to $PM_{2.5}$.* Qualitatively, the Draft PM ISA and Draft PM PA rely on a proposed causal determination framework to make qualitative judgments about causality. As far as we have been able to determine, this proposed framework has not been empirically validated for use with $PM_{2.5}$ health effects. It is based largely on the Bradford Hill considerations (addressing associations, not interventional causation) and on assumptions known by experts in causal inference to be unsound, e.g., that "Association, especially association adjusted for potential confounders, is evidence for the causal claim" (IOM, 2008). This assumption has been widely recognized as invalid since at least the early 1960s, when threats to validity of causal conclusions based on associations (such as coincident historical trends, regression to the mean, etc.) were enumerated and discussed as testable non-causal hypotheses by Campbell and Stanley (1963). Some members of the CASAC recommend that the EPA consider updating the PM PA to use conceptually-sound frameworks such as causal graphs, with empirical evidence for implied conditional independence relationship clearly presented and critically discussed.

(7) *Use of unverified modeling assumptions.* The simulation models used in the Draft PM PA to predict public mortality risks caused by $PM_{2.5}$ exposures have not been empirically validated or verified as yielding correct predictions. The predicted effects of changes in $PM_{2.5}$ exposures on changing mortality risks therefore reflect unverified modeling assumptions, rather than empirical conclusions from a scientific process that has rigorously tested predictions against observations. The CASAC recommends that the PM PA clearly identify risk predictions that are not based on empirically validated predictive or causal relationships.

(8) *Omission of climate impacts.* The Draft PM PA does not discuss how reducing particulate pollution warms the climate or reduces land carbon sink capacity (e.g., Harvey, 2018). These effects might be large compared to other effects considered in the Draft PM PA. Well-informed policy making requires that they be addressed quantitatively in the PA using the best currently available information (see consensus responses to Chapter 5 charge questions).

Most members of the CASAC conclude that, for at least these reasons, the Draft PM PA does not fully meet its goals of providing scientifically sound policy-relevant information.

Improving Broad Understandability

Broad understandability of the Draft PM PA's assumptions and conclusions can be improved by clearly defining key quantities and terms used to calculate and communicate scientific results and by clearly distinguishing between several distinct concepts that are conflated in the current Draft PM ISA and Draft PM PA.

Some members of the CASAC recommend that terms such as the causal determination terms "causal" and "likely to be causal" should be given unambiguous operational definitions and that their logical relationships to standard epidemiological concepts of cause (e.g., necessary cause, sufficient cause) and effect (e.g., direct effect, indirect effect, and total effect of exposure on risk) should be made clear. These members are concerned that most readers may not understand that the ISA and PA use "cause" in a specialized statistical sense, to mean something like "an association that is not thought likely to be explained by chance, confounding, or other biases." For example, any two variables with the same historical trends (not explained by chance, confounding, or other biases) would be considered to have a "causal" relationship by this definition, even if changing one would have no effect on the other. Thus, by this definition, increasing college tuition is a "cause" of global warming (or any other unrelated increasing variable), assuming that their increasing trends do not reflect chance, confounding, or other biases. Because this meaning differs from what most people mean and understand by "cause," broad understandability requires emphasizing in the text that "causal" as used in the Draft PM PA refers specifically to associations not thought to be explained by chance, confounding, or other biases; and that it does not necessarily correspond to more usual definitions, or imply that changing one variable would affect the other. Likewise, when Chapter 1 and the rest of the Draft PM PA refer to health "effects" of exposure, it should be made clear that this term is used to refer to correlates of exposure, and that there is no implication that the "effects" would necessarily be eliminated or reduced if exposure were eliminated or reduced.

Broad understandability also requires that the beta coefficients in Appendix C of the Draft PM PA should be explicitly defined. It should be stated in Appendix C of the Draft PM PA and in the text that they are regression coefficients (relating the conditional expected observed value of the dependent variable to the observed values of the predictors, assuming that these are measured without error). They should not be conflated with causal coefficients predicting how the average value of the dependent variable will change if values of the predictors are exogenously set to new values.

Other crucial distinctions that should be drawn and maintained throughout the Draft PM PA (and the PM ISA) include: estimated vs. actual individual exposures; differences between observed average risks for individuals with different exposure levels vs. changes in risks caused by changes in exposure levels; statistical and epidemiological measures of causation in populations vs. biological (mechanistic) causation for individuals; and assumed (or predicted or simulated) changes in mortality risks vs. observed changes in mortality risks following changes in exposures.

Facilitating CASAC Advice and Recommendations to the Administrator

The foundation of the present CASAC's advice and recommendations to the Administrator is sound science, emphasizing:

(1) Dispassionate consideration of the full range of relevant evidence, both positive and negative, about the extent to which reducing exposure reduces public health risks;
(2) Explicit, verifiable, logically sound derivations of conclusions from data and from empirically tested and validated predictive models;
(3) Empirically validated conclusions, with results of reproducible tests of assumptions and conclusions transparently shown and remaining uncertainties and variability quantified;
(4) Clearly communicated findings, expressed using terms with clear operational definitions; and
(5) Candidly discussed alternative interpretations of the data and their possible implications, including assessments of the biological realism of different interpretations.

To facilitate the CASAC's advice and recommendations, therefore, the PM PA should do the following:

(1) *Expand the range of evidence and observational studies considered* to more fully address the full range of scientific evidence (whether positive or negative). The Draft PM PA and Draft PM ISA should systematically apply explicitly stated criteria for selecting evaluating, summarizing, interpreting, and synthesizing studies, and should summarize the results of applying these criteria for each individual study relied on in the PM PA (e.g., those in Appendix C of the Draft PM PA, plus any others that may be added when explicit selection criteria are systematically applied).
    a. Some members of the CASAC and multiple public comments recommend that the Draft PM ISA and Draft PM PA include more of the recent and relevant accountability, intervention, and time series studies showing how public health risks have changed following reductions in $PM_{2.5}$ exposure. Any appearance of bias toward positive studies, as suggested in several public comments, should be absent in the final PM PA.
(2) *Explain exactly how conclusions are derived from observations.* For example, as the CASAC and members of the public have commented, it is not always clear how or whether some of the causal determinations in the Draft PM ISA, repeated in the Draft PM PA, follow from the evidence presented. Likewise, most members of the CASAC and multiple public commentators do not find that the preliminary conclusion in the Draft PM PA (that recent scientific data can reasonably be used to call into question the adequacy of the current primary NAAQS for $PM_{2.5}$ to protect human health) logically follows from the evidence presented, insofar as:
    a. *Confirmatory evidence does not justify revising the current standards.* Additional findings of positive C-R associations confirm what was already assumed in setting the current NAAQS; they do not show that it was incorrect or needs to be revised.
    b. *Alternative interpretations of the evidence have not been thoroughly assessed.* Alternative explanations and interpretations of positive C-R associations, such as that they result from historical trends or incompletely controlled confounding (e.g., by month and daily high and low temperatures in the weeks prior to mortality) have not been tested or refuted in the Draft PM ISA and Draft PM PA.

i. The extent to which positive C-R associations simply reflects particular modeling choices, rather than robust empirical relationships, has not been assessed in the Draft PM PA. Yet, "Different epidemiological models may lead to very different conclusions for the same set of data" (Fuentes, 2009). The same C-R data can often be used to generate either positive or negative associations by making different modeling choices, so that "the associational or regression approach to inferring causal relations— on the basis of adjustment with observable confounders— is unreliable in many settings" (Dominici et al., 2014). Thus, the presence of significant positive associations does not necessarily convey reliable information about causation. The fact that many studies report positive associations does not address the extent to which they result from modeling choices or common confounders.

c. *Omitted evidence.* The Draft PM ISA and Draft PM PA omit some intervention, accountability, and time series studies that, collectively, do not support the hypothesis that reducing $PM_{2.5}$ has clearly reduced mortality risks. Taken together, these studies have not been found to decisively reject the null hypothesis that reductions in $PM_{2.5}$ have not caused greater reductions in risk than would otherwise have occurred. Power calculations in some studies suggest that causal impacts of the sizes assumed in the Draft PM PA should be readily detectable (i.e., the null hypothesis should be confidently rejected), given the large sample sizes and the sampling variability involved; thus, failure to reject the null hypothesis is not explained by small sample sizes or complexity and variability in these studies. Recent reviews of several dozen intervention studies have reported an overall "absence of evidence" that recent interventions to reduce $PM_{2.5}$ levels have caused reduced mortality (although many studies do report positive C-R associations) (e.g., Burns et al., 2019). Acknowledging that absence of evidence is not necessarily evidence of absence (although it is evidence of absence of an effect large enough to be clearly detected), these omitted studies nonetheless do not support the conclusion that further reducing $PM_{2.5}$ is necessary to reduce human health risks. They instead raise questions about why clear effects of interventions are not more readily detected, and whether some long-standing modeling assumptions made in recent NAAQS reviews need to be revisited (especially, that C-R associations, which are readily detected, predict causal impacts of reducing $PM_{2.5}$). This omitted evidence may therefore be important in quantitative uncertainty analyses that address model uncertainty and that assess whether and how much uncertainties about health effects of further reducing $PM_{2.5}$ have changed since the previous review.

d. *Missing quantitative uncertainty and variability analyses*. Findings of positive associations at estimated exposure concentrations below the current standard do not address errors and uncertainties in exposure estimates. In the absence of quantitative uncertainty and variability analysis accounting for model uncertainty, exposure uncertainties, and other epistemic uncertainties, it is unclear what individual exposures were experienced by people with adverse health responses (e.g., how many times greater their individual exposures were than the estimated average exposures for the zip code, square kilometer, county etc. in which they resided). The Draft PM PA (and Draft PM

ISA) present no information on whether uncertainties about causal C-R functions at low concentrations are larger or smaller now than when the current NAAQS were set, but accounting for errors in exposure estimates and for model uncertainties would likely increase total uncertainties compared to those previously considered, since, as noted in multiple public comments, these sources of uncertainty have not yet been quantified.

e. *Missing interventional causation.* Findings of positive C-R associations do not address whether reducing exposure would reduce health risks (e.g., Petitti, 1991), even if they are deemed "causal" in a WoE framework. They therefore cannot logically call into question the adequacy of current standards to protect health, assuming that the current standards should be deemed adequate if changing them would not reduce risks.

f. *Unverified assumptions are not valid evidence of a need to revise standards.* The conclusion in the Draft PM PA that thousands of preventable deaths per year may occur at the current NAAQS levels is based on unverified modeling assumptions, not empirically validated predictive modeling. Unverified assumptions do not constitute valid scientific evidence calling into question the adequacy of current standards.

(3) *Present results of empirical validation tests and quantitative uncertainty and variability analyses.* To facilitate CASAC advice and recommendations to the Administrator based on sound science, risk predictions must be empirically validated, with any remaining uncertainties and variability characterized quantitatively. Empirical validation requires comparing predictions to data. For example, if it is predicted that reducing $PM_{2.5}$ exposure concentrations will reduce mortality rates by more than they would decline otherwise (had $PM_{2.5}$ increased or stayed the same), then this prediction should be compared to data from intervention, accountability, and time series studies with appropriate control groups, to characterize the extent to which it agrees or disagrees with observations. Ranges of effect sizes (i.e., incremental reductions in mortality rate caused by reductions in $PM_{2.5}$) should be quantified. Quantitative descriptions of interindividual variability and heterogeneity and sensitive subpopulations should be provided (e.g., using individual conditional expectation (ICE) plots or other appropriate techniques), and both uncertainty and variability in causal coefficients or causal C-R functions should be characterized.

(4) *Discuss alternative interpretations of available data using plain, clearly defined terms.* This discussion should address whether predicted changes in risk caused by changes in exposures are biologically realistic, and how well they are supported by observed biological changes in people under experimental and real-world conditions. It should address the extent to which observed changes in mortality and morbidity risks following reductions in exposure are consistent with chance, confounding and other biases, modeling choices, historical trends, unmodeled heterogeneity, and other non-causal explanations. Rationales for rejecting non-causal explanations should be explicitly stated and critically evaluated.

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?

Chapter 2 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.

The section on "Sources of PM Emissions" presents estimated national values for 2014 National Emissions Inventory (NEI) emissions. 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. 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. Also, it would be helpful to add national maps containing county-level emissions for $PM_{2.5}$, PM_{10}, organic carbon (OC), elemental carbon (EC), sulfur dioxide (SO_2), oxides of nitrogen (NOx), ammonia (NH_3), and volatile organic compounds (VOCs) to show the variability across the country.

The section on "Ambient PM Monitoring" discusses Federal Reference Method (FRM), continuous Federal Equivalent Method (FEM), Chemical Speciation Network (CSN), and Interagency Monitoring of Protected Visual Environments (IMPROVE) monitors. However, there is no discussion on measurement uncertainty (accuracy, precision, bias, and error) associated with these monitors. For routine monitoring, FRM filters remain in the sampler at or somewhat above ambient temperatures for up to 6 days. FRM filters can lose up to 10% of their non-water mass over 24-96 hours if not removed from the sampler and chilled immediately. Therefore, in field comparisons of co-located FEM and FRM monitors, FEM measurements typically appear to be biased high compared to the FRM, when in reality this is an artifact of field sample handling for the FRM and not an actual limitation of the FEM. The EPA should compare co-located FRM/FEMs 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 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. (Hanson et al., 2003; Edgerton et al., 2005; Edgerton et al., 2006; Blanchard et al., 2013). Also, the discussion on air sensors should include a description of their shortcomings and performance issues.

The section on "Trends in Ambient Air Concentrations" should discuss the concept of "urban increment" and include a couple of examples comparing $PM_{2.5}$ speciation in urban areas to nearby Class I areas. 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." The typical annual/daily design value ratio line is missing and should be added. 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.

EPA's 2016 exceptional events rule allows certain 24-hour PM measurements due to natural events to be excluded from the official design values when compared to the NAAQS. The PA should discuss how exceptional events are accounted for in the setting of the standard.

Hybrid $PM_{2.5}$ modeling approaches have improved the estimates of ambient $PM_{2.5}$ concentrations in areas without monitors. In fact, this is one of the most significant improvements since the last PM NAAQS review. The four approaches presented in the chapter generally agree although the spatial texture of the concentration fields differ among methods.

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 (Schichtel et al., 2008; Tanner et al., 2010).

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?

<u>Consideration of Health Effects and Risk Assessment Evidence</u>

<u>Updates from the draft PM ISA</u>

The CASAC recommends that the EPA follow the CASAC's advice on the PM ISA and update the causality determinations between $PM_{2.5}$ exposure and cancer, long-term $PM_{2.5}$ exposure and nervous system effects, and long-term UFP exposure and nervous system effects. This includes updating the quantitative risk assessment results for $PM_{2.5}$ exposure and cancer.

<u>Balanced and accurate reporting of results</u>

The EPA should provide a balanced summary of the study results for each health endpoint. Adequately communicating available positive, negative, and null results provides useful information for the Administrator. The CASAC suggests that this could be accomplished by providing, for each endpoint and exposure length, what data supports the conclusion, what data does not support the conclusion, what are the uncertainties, at what concentrations do clear effects or associations occur, and what information is new. Similarly, the risk information should be presented with ranges and confidence bounds. As it

stands in this document, in Section 3.4.1 (preliminary conclusions about the current $PM_{2.5}$ standard), the EPA only provides the highest risk estimates with none of the available ranges of results or confidence bounds.

Consideration of uncertainties in the health effects evidence

The CASAC recommends that the EPA more explicitly (and quantitatively, when possible) address the degree to which the uncertainties identified in the last PM review have changed in the current review. Specifically, on pages 3-8 to 3-9 of the Draft PM PA, the EPA provides the following information about the then-EPA Administrator's determination about the uncertainties of the $PM_{2.5}$ review and standard:

> "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 $PM_{2.5}$ in the complex ambient mixture, exposure measurement errors in epidemiologic studies, and the nature and magnitude of estimated risks related to relatively low ambient $PM_{2.5}$ 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." (Emphases added).

More completely addressing the new data that are available to inform these previously specified uncertainties will provide the Administrator with important decision-relevant information about these potentially policy-relevant uncertainties in the health effects evidence.

Linear no-threshold concentration-response (C-R) association

Errors and heterogeneity in epidemiology study variables can affect the apparent shape of the concentration-response (C-R) relationship and can obscure thresholds. Evidence for this has been provided by many peer-reviewed publications (Brauer et al., 2002; Cox, 2018; Lipfert and Wyzga, 1996; Rhomberg et al., 2011; Watt et al., 1995; Yoshimura, 1990) and notably by the EPA in the ISA preamble (US EPA 2015, 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, O_3, 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."

The problem described here is not whether a threshold in the data may exist, but rather that even if it does exist, epidemiology studies may not be capable of definitively identifying the threshold. To address this concern the CASAC recommends that the EPA should explicitly acknowledge in the PM PA that variability and error in the variables can linearize C-R functions and obscure thresholds, and this acknowledgement should be included in those places where the EPA concludes that the relationship between PM and a health effect is linear and has no threshold (as on pages 3-7, 3-10, 3-20, 3-21, 3-24, 3-25, 3-33, 3-41, 3-42, 3-50, 3-70, and 3-96). The CASAC also recommends that the EPA should begin to apply methods (and encourage the epidemiological community to apply methods) to address this particular concern, including errors-in-variables methods. If possible, the EPA should include these types of adjustments when applying the epidemiology C-R functions to their risk assessments.

<u>Comparison of long-term and short-term $PM_{2.5}$ concentration effects</u>

In the Draft PM PA, the EPA should clarify their reasoning for directly comparing long-term and short-term $PM_{2.5}$ concentrations, because this is not usually considered to be a valid comparison. The EPA provided information to the CASAC describing how one of the intentions of the annual $PM_{2.5}$ standard is to control average daily concentrations, with the idea that decreasing annual average concentrations will result in a downwards shift in the distribution of all (or most) daily $PM_{2.5}$ concentrations. This information was very helpful to the CASAC in understanding the comparison of short-term and long-term $PM_{2.5}$ concentrations and it should be added to the PM PA.

<u>Description and conclusions from pseudo-design values</u>

The EPA needs to carefully consider what they are measuring and comparing when they derive pseudo-design values (PDVs). The EPA has not made it clear whether PDVs represent concentrations or conditions for either short- or long-term studies.

Several members of the CASAC and public commentators find that the Draft PM PA does not clearly explain: 1) what information can be gained from determining if an area from a short-term $PM_{2.5}$ study was in attainment of the 3-year annual average $PM_{2.5}$ standard (because the association between $PM_{2.5}$ and the short-term health effect is based on daily changes in $PM_{2.5}$ that may have little to do with the annual average); and 2) what information can be gained from determining if an area from a long-term cohort study attained the annual average standard, because the associations between long-term $PM_{2.5}$ and health effects in these studies is almost always across study areas and so the effect is based on the difference in $PM_{2.5}$ concentrations, not the absolute value of the $PM_{2.5}$ concentrations in any particular study area. The EPA should clarify the use and interpretation of the PDVs.

Methods for quantitative risk assessment

Some members of the CASAC have concerns with how the risk assessment was conducted. The risk assessment approach in Chapter 3 treats regression C-R functions (that is, functions describing associations between past estimated exposure concentration levels and mortality rates) as if they were causal C-R functions (that is, describing how changing future exposure concentrations would change future mortality rates). This is technically unsound, as noted in a recent commentary by Carone et al. (2019) that:

> "…even in the ideal scenario in which all relevant confounders are accounted for and the regression model postulated *a priori* holds true, model-based regression coefficients corresponding to the exposure of interest may still not refer to the causal contrast desired to address the scientific question at hand. This may occur, for example, because the regression coefficients quantify a causal effect on a different scale than desired (e.g., odds ratio versus relative risk) or do not provide the population-level summary desired (e.g., conditional versus marginal interpretation). The limitations of conventional model-based causal inference are exacerbated when modeling is performed on scales less amenable to causal comparisons (e.g., the hazard ratio), or when the exposure of interest occurs over a longer period of time."

Pearl (2010) goes further, noting that association and causal concepts don't mix: associations typically do not provide the information needed to predict correctly how changing one variable would change others. These causal analytics principles are well-established (Dr. Cox's individual comments contain further discussion and references). The Draft PM PA provides no explicit justification for treating regression C-R functions as causal C-R functions. Yet, this cannot validly be assumed without justification, as the two are often very different (Bind, 2019; Please also see responses from consultants Drs. North and Aliferis in Appendix D). Recent reviews and studies have highlighted the importance of the distinction between regression C-R and causal C-R functions (Bind, 2019) and have shown that, for $PM_{2.5}$, regression C-R functions do not necessarily describe the changes in health risks caused by changes in exposures (e.g., Burns et al., 2019 and studies reviewed therein; Henneman et al., 2017). The PM PA should recognize, discuss, and keep clear that there is a real distinction between regression coefficients and causal coefficients. The PM PA should explicitly state the implicit assumption that regression coefficients can be used to quantify causality, note that this is not necessarily a valid assumption (Bind, 2019; Pearl, 2010), and provide information about whether the assumption has been tested and what the results were.

Some members of the CASAC think that the EPA should discuss how the C-R functions used in the risk assessment were selected, and whether the C-R functions have adequately controlled for known confounders of concern for the study type. These members of CASAC think that it may be helpful to give specific examples of types of confounding that should be addressed, including the following:

- Measured confounders omitted from the analyses supporting the Draft PM PA (e.g., the 8 key studies in Table C-1). Examples of omitted confounders would be the month in which a death occurred, or the daily high and low temperatures in the weeks preceding mortality, in studies that

collected these values, but then either did not use them, or categorized, or averaged them over some period (e.g., seasons, or years) before using them.

- Unmeasured (latent) confounders, such as daily high and low temperatures, humidity, and income in studies in Table C-1 that did not collect data on these variables.
- Residual confounding arising from use of broad measurement categories that do not control for confounding due to differences across individuals with the same categorical measurement (e.g., residual confounding by daily high and low temperatures in studies in Table C-1 only adjust for seasonal or annual average temperatures).

In addition, there is very little methodological detail provided in this document about how the risk assessment was done and why particular mathematical choices were made (e.g. see Dr. Lange's individual comments). This information is also not provided in the referenced BenMAP manual, and therefore methodological details should be added to Appendix C of the PM PA. In addition, the EPA should discuss the model validation studies that have been conducted on BenMAP, to demonstrate that the model performance is appropriate.

For the air quality modeling used in the risk assessment, the EPA should evaluate the CMAQ model performance specifically at the study city locations (similar to what is shown in Table C-6 for large regions of the United States) and should not use study areas with low model performance.

Quantitative uncertainty analysis in the risk assessment

The CASAC notes that future changes in public health risks caused by reducing $PM_{2.5}$ exposures are currently highly uncertain. The PM PA needs to capture more of the uncertainty – especially model uncertainty, exposure estimation uncertainty, and other epistemic uncertainties – in their risk analysis using quantitative uncertainty analysis methods. There is wide agreement amongst the public commenters and the CASAC non-member consultants that methods are available to quantify the uncertainty in the EPA's risk analysis. These analyses could specifically address the contributions (if any) to the total associations made by confounding, measurement errors, model uncertainty, and heterogeneity and variability in individual biological (causal) C-R functions, as well as sampling variability. Several methods are suggested (for example, see responses from consultant Dr. North in Appendix D). EPA's 2006 PM NAAQS risk analysis provides an example of a quantitative uncertainty analysis: "These risks were estimated using not only the linear or log-linear concentration-response functions reported in the studies, but also using alternative modified linear functions as surrogates for assumed non-linear functions that would reflect the possibility that thresholds may exist in the reported associations within the range of air quality observed in the studies" (71 FR 61154). A similar approach could be applied in this risk assessment.

Conclusions on $PM_{2.5}$ Standards

The EPA summarized the basis for the Draft PM PA's preliminary conclusion that available scientific information can reasonably be viewed as calling into question the adequacy of the public health

protection afforded by the current primary annual $PM_{2.5}$ standards as follows (Agency Briefing Material, p. 18):

1. Long-standing body of health evidence, strengthened in this review, supporting relationships between short- and long-term $PM_{2.5}$ exposures and various outcomes, including mortality and serious morbidity effects
2. Recent U.S. and Canadian epidemiologic studies reporting positive and statistically significant health effect associations for $PM_{2.5}$ air quality likely to be allowed by the current standards
3. Analyses of pseudo-design values indicating substantial portions of study area health events/populations in locations with air quality likely to have met the current $PM_{2.5}$ standards
4. Risk assessment estimates that the current primary standards could allow thousands of $PM_{2.5}$-associated deaths per year – most at annual average $PM_{2.5}$ concentrations from 10 to 12 µg/m3 (well within the range of overall mean concentrations in key epidemiologic studies).

Some members of the CASAC find this rationale compelling; others CASAC members do not. The CASAC provides the following points and areas of consensus and non-consensus,

Point 1 (Long-standing body of health evidence, strengthened in this review, supporting relationships between short- and long-term $PM_{2.5}$ exposures and various outcomes, including mortality and serious morbidity effects):

All members of the CASAC agree with this statement. However, members of the CASAC differ in their assessments of the causal and policy significance of these associations:

- Some members think that such associations can reasonably be explained in light of uncontrolled confounding and other potential sources of error and bias (discussed below); that associations are not effects (Petitti 1991); and that in intervention studies, reductions of $PM_{2.5}$ concentrations have not clearly reduced mortality risks, especially when confounding was tightly controlled (Henneman et al. 2017; Burns et al. 2019). These members of the CASAC think that, while the data on associations should certainly be carefully considered, this data should not be interpreted more strongly than warranted based on its methodological limitations.
- Other members of the CASAC think that the entire body of evidence for PM health effects justifies the causality determinations made in the Draft PM ISA, that form the basis for the Draft PM PA. This body of evidence, as reviewed in the Draft PM ISA, includes not only associations repeatedly demonstrated in epidemiology studies, but also biological plausibility established by human clinical and toxicology studies. These members find it highly unlikely that the extensive body of evidence on positive C-R associations at low estimated exposures could be fully explained by confounding or by other non-causal explanations.

Point 2 (Recent U.S. and Canadian epidemiologic studies reporting positive and statistically significant health effect associations for $PM_{2.5}$ air quality likely to be allowed by the current standards):

Some members of the CASAC, as well as several public commenters, note that many relevant time series studies, intervention studies, and accountability studies are not included in the literature reviewed in the Draft PM ISA and relied on in the Draft PM PA (e.g., Burns et al., 2019 and studies reviewed therein; Eum et al., 2018; Greven et al., 2011; Henneman et al., 2017; Pun et al., 2017).
Some members of the CASAC think that for point 2, similar to point 1, interpretation of the more recent U.S. and Canadian epidemiology studies should be refined to more fully account for effects of confounding, measurement and estimation errors, model uncertainty, and heterogeneity (see Appendix A for possible technical options); and that evidence from additional negative studies, intervention studies, accountability studies, and time-series studies should also be taken into account. These CASAC members also think that the new studies of positive associations at lower estimated exposure concentrations in Canada and the United States mainly confirmed what had already been anticipated or assumed in setting the 2012 NAAQS. Particularly focusing on the long-term epidemiology studies that supported the 2012 NAAQS and that are cited in the current review, these studies are based on a Cox proportional hazards model that assumes a linear non-threshold relationship between $PM_{2.5}$ and the health effect (often mortality). The long-term $PM_{2.5}$ concentrations in the study areas vary substantially, but the effect of the model is to draw a line through these concentrations to determine the slope of the relationship between the health effect and $PM_{2.5}$ concentrations. Figure 1 illustrates this concept, showing simple linear relationships based on the $PM_{2.5}$ concentration distribution of the study areas and simulated data for the health effect variable. The health effect (Y) variable is simulated based on the estimated slope of the regression line, with some error built in to simulate variability in the data. Each set of data points is based on the concentration distributions and slopes provided in either Krewski et al. (2009) (mean $PM_{2.5}$ concentration of 14 $\mu g/m^3$, slope (beta) of 0.0058 per $\mu g/m^3$), or Di et al. (2017) (mean $PM_{2.5}$ concentration of 11 $\mu g/m^3$, slope (beta) of 0.0081 per $\mu g/m^3$). This figure illustrates that the mean concentration can be lower in one study versus another (Di et al., 2017 versus Krewski et al., 2009), without providing new information about concentrations at which effects occur.

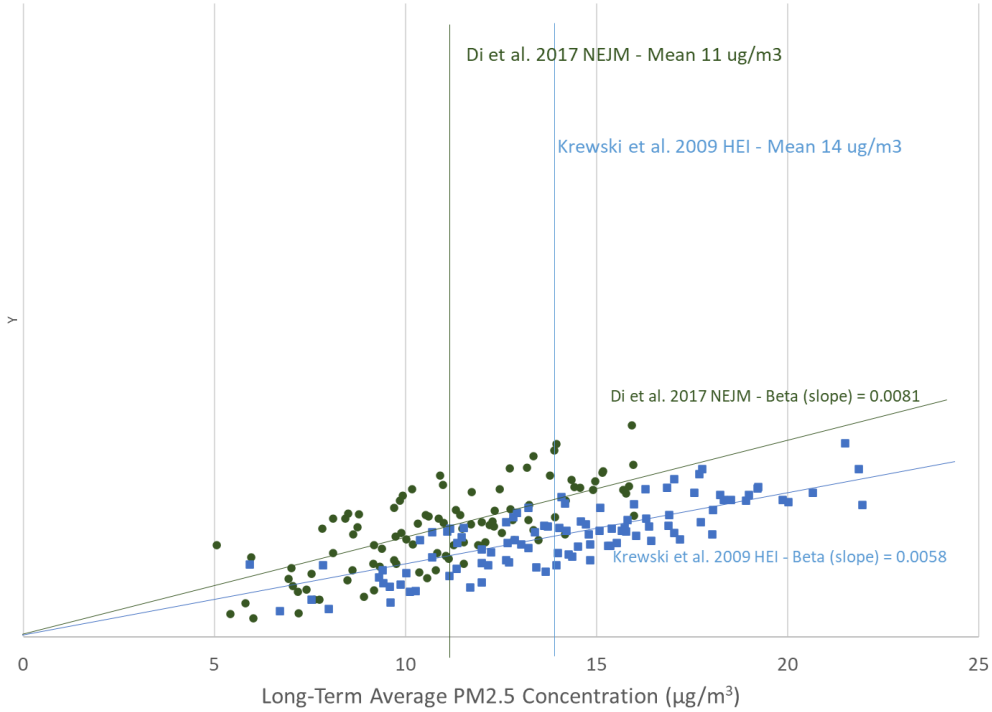


Figure 1. Illustrative simulation of linear relationship between health effect (Y) variable and long-term average $PM_{2.5}$ concentration based on concentration distributions presented in Krewski et al. (2009) and Di et al. (2017).

Other members of the CASAC, as well as the Independent PM Review Panel (members of the disbanded CASAC PM Review Panel), think that the epidemiologic evidence demonstrating the health effects of $PM_{2.5}$ is strong, with biological plausibility provided by human controlled exposure and animal toxicological studies. These CASAC members do not think that this evidence is diminished because of the lack of consistent support from newer intervention and accountability studies. Newer studies demonstrating that the associations between $PM_{2.5}$ and health effects occur at a diversity of locations, in different time periods, with different populations, using different exposure estimation and statistical methods, make this dataset even more robust. Most important are the recent findings that there are associations between $PM_{2.5}$ and health effects in areas with average long-term $PM_{2.5}$ concentrations below the standard, as well as studies that show positive associations even when estimated daily concentrations above 12 $\mu g/m^3$ are excluded. Altogether these members of CASAC find EPA's point 2 to be crucial in their determination that the current scientific evidence calls into question the adequacy of the 2012 $PM_{2.5}$ annual standard to protect public health.

Point 3 (Analyses of pseudo-design values indicating substantial portions of study area health events/populations in locations with air quality likely to have met the current $PM_{2.5}$ standards):

As noted above, the CASAC finds that the pseudo-design value assessment is difficult to interpret.

Point 4 (Risk assessment estimates that the current primary standards could allow thousands of $PM_{2.5}$ -associated deaths per year – most at annual average $PM_{2.5}$ concentrations from 10 to 12 µg/m3 (well within the range of overall mean concentrations in key epidemiologic studies)):

Some CASAC members find that the Draft PM PA's risk assessment conclusion that the current primary standards could allow thousands of $PM_{2.5}$ -associated deaths per year is based on modeling assumptions that are potentially flawed and that have not been empirically verified. Because they are driven by such assumptions, the conclusions from the risk assessment do not comprise valid empirical evidence or grounds for revising the current NAAQS. This is based on concerns noted above that i) the risk assessment treats regression coefficients as causal coefficients with no justification or validation provided for this decision; ii) the estimated regression C-R functions in Appendix C and Chapter 3 of the Draft PM PA have not been adequately adjusted to correct for confounding, errors in exposure estimates and other covariates, model uncertainty, and heterogeneity in individual biological (causal) C-R functions; iii) the estimated C-R functions do not contain quantitative uncertainty bands that reflect model uncertainty or effects of exposure and covariate estimation errors; and iv) no regression diagnostics are provided justifying the use of proportional hazards (PH) and other modeling assumptions. These potential sources of error and bias, and technical methods for addressing them, have been discussed in previous CASAC deliberations and comments on the Draft PM ISA, and in public comments on the Draft PM ISA and Draft PM PA. In response to a request from the EPA, Appendix A discusses some possible technical options for addressing them in the near term.

Other CASAC members find that the approaches used in the Draft PM PA are appropriate and well justified, that the new evidence available since the last review has reduced uncertainties regarding health effects at exposure concentrations below the current standard, and that the evidence strengthens the concern that the current standard is not adequately protective.

As a decision framework for asking to what extent the current scientific literature supports the adequacy (or not) of the 2012 $PM_{2.5}$ standards to protect human health, the CASAC considers how the evidence has changed since the last PM review. One way to view this is by considering the entire PM assessment as a whole to be a risk assessment with the following components: hazard assessment (the ISA), exposure assessment (the PA), dose-response assessment (the PA), and risk characterization (the PA).

<u>Hazard Identification</u> – The purpose of the hazard identification is to determine which health effects could be caused by $PM_{2.5}$ exposure at relevant ambient concentrations (the EPA defines a relevant concentration as being less than 2 mg/m^3). The CASAC agrees that the causal designations have not substantially changed since the last assessment (setting aside the three "likely causal" upgrades, which the CASAC found were not justified in the review of the Draft PM ISA). Although the causal designations have not changed, some CASAC members find that there is greater evidence supporting those causal designations, and at lower exposure concentrations. Other CASAC members did not think that the somewhat lower exposure concentrations in the newer studies are relevant for hazard identification, because this step of the risk assessment process only identifies the hazards, and those identified hazards have not changed in this review.

Exposure Assessment – An exposure assessment determines how much $PM_{2.5}$ the target population is exposed to. $PM_{2.5}$ concentrations have continued to decrease nationwide since the last review. Fundamentally, the NAAQS seek to reduce health risk by reducing exposure. The CASAC had substantial discussions about the studies that are available showing that changes in standards have decreased $PM_{2.5}$ concentrations, and how these concentration decreases may correlate with decreases in health effects attributed to $PM_{2.5}$ (such as mortality). These accountability studies provide potentially crucial information about whether and how much decreasing $PM_{2.5}$ causes decreases in future health effects, which reflects the primary purpose of the NAAQS (to decrease air pollutants to protect the public from experiencing future health effects). This type of interventional causality can be contrasted with epidemiology studies using retrospective information, which determine the association between past $PM_{2.5}$ concentrations and past health effects. Some members of the CASAC identified a number of these types of studies conducted (reviews include Burns et al., 2019 and Henneman et al., 2017) and note that clear evidence of decreases in health effects caused by interventions that decreased $PM_{2.5}$ is lacking, particularly when appropriate controls are included in the analyses. Other members of CASAC do not think that there is enough evidence to make definitive determinations about $PM_{2.5}$ decreases leading to decreases in health effects, and that absence of evidence is not evidence of absence.

Dose-Response Assessment – This aspect of the risk assessment addresses the expected change in the health endpoint caused by a unit change in $PM_{2.5}$. In the last review, the EPA concluded that the dose-response for the major hazards (total mortality, cardiovascular and respiratory morbidity and mortality) were low-dose linear with no threshold. The EPA draws the same conclusion in this review, and the slopes of the linear estimates are similar in the last review and the current review. Some members of the CASAC see this evidence as showing the same dose-response association in the current review as in the last review and note that neither review quantifies effects caused by *changes* in exposure, as opposed to differences in effects associated with *differences* in exposures. Other members of CASAC think that the newer studies provide additional data at lower ambient $PM_{2.5}$ concentrations, strengthening the evidence that health effects are occurring at exposure concentrations below the NAAQS.

Risk Characterization– This aspect of the risk assessment combines the information from the hazard identification, dose-response, and exposure assessments, and sometimes produces a quantitative risk estimate (as is done in this Draft PM PA). However, for the reasons discussed elsewhere in this report and in the individual CASAC member comments, the quantitative risk estimates (particularly in the absence of a quantitative uncertainty analysis) have unknown reliability. Moreover, because the EPA assumes a linear-to-zero dose-response function, every standard above zero will be associated with some risk. Equally, the EPA does not provide information about what a level of acceptable risk above zero would be. Therefore, some CASAC members think that the risk characterization does not provide useful information about whether the current standard is protective or not. Other CASAC members think that the risk characterization does provide a useful attempt to understand the potential impacts of alternate standards on public health risks.

Based on these discussion points, some of the CASAC members conclude that new scientific evidence and data do not reasonably call into question the public health protection afforded by the 2012 $PM_{2.5}$ annual standard. Other members of CASAC find that the weight of the evidence, particularly reflecting

recent epidemiology studies showing positive associations between $PM_{2.5}$ and health effects at estimated annual average $PM_{2.5}$ concentrations below the current standard, reasonably call into question the 2012 annual $PM_{2.5}$ NAAQS as being not sufficiently protective of public health with an adequate margin of safety. The CASAC recommends that the final PM PA provide quantitative uncertainty and sensitivity analyses, considering the foregoing points, to provide a clearer technical and scientific basis for data interpretation and policy making.

The CASAC agrees with the EPA that there is not recent scientific evidence that calls into question the adequacy of the 24-hour $PM_{2.5}$ standard.

Additional Research

What are the CASAC views regarding the areas for additional research identified in Chapter 3? Are there additional areas that should be highlighted?

For Chapter 3, the CASAC agrees with the EPA's listed areas for additional research. The EPA's suggestion about 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. Other areas for additional research that the CASAC recommends are:

- Development of validated, quantitative causal models to describe the causal relationship between changes in $PM_{2.5}$ and changes in health effects of interest, while controlling for levels of confounders and other causally relevant variables.
- Development of quantitative uncertainty methods that include model uncertainty and exposure estimate uncertainties.
- Further development of quantitative causal risk assessment methods.
- Development of better exposure metrics for use in epidemiology studies.
- Development of models to determine the mechanisms of distribution of $PM_{2.5}$ through the body (e.g. PBPK modeling) and biological effects of $PM_{2.5}$ exposures under real-world conditions.
- Development of FRMs for measuring ultrafine particles and elemental carbon. This would help provide the needed ambient monitoring data for ultrafine particles and elemental carbon to support the conduct of observational studies.

Chapter 4 – Review of the Primary PM_{10} 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?

Background

Chapter 4 of the Draft PM PA addresses the PM_{10} Standard. PM_{10} encompasses all particles smaller than 10 µm and so includes $PM_{2.5}$. The PM_{10} standard is specifically intended to protect against the health

and welfare 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 µg/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 µg/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 adverse health or welfare effects. 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. The evidence was considered "inadequate" for long-term mortality and cardiovascular effects, and for cancer.

Approach

This chapter of the Draft 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 for short-term effects in the previous PM ISA, and the decision to retain the previous standard (page 4-3).

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

What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current primary PM_{10} standard?

Some CASAC members find that the new studies available since the previous review justify the change in causality determination from "inadequate" to "suggestive" for long-term exposure to $PM_{10-2.5}$ and mortality and cardiovascular effects. These members also find that the new data somewhat strengthens the evidence for effects on cancer, and short-term effects on mortality, cardiovascular disease, and respiratory disease.

Other CASAC members find that the operational definitions of the causal determination categories given in the Draft PM ISA and summarized in the Draft PA PM to be inadequate, and the scientific, operational meanings of these terms are not self-explanatory. Therefore these members do not know what factual information the EPA seeks to communicate, and therefore do not reach agreement on whether the change to "suggestive" is justified. These members also find that the evidence presented does not imply, or make likely, that $PM_{10-2.5}$ caused these effects. For example, the Draft PM PA interprets evidence of oxidative stress as providing "some evidence of biological plausibility for the findings in a small number of epidemiologic studies" (p. 4-11, line 10), but oxidative stress is a very common biological response to a wide range of conditions, including responses to non-carcinogens.

The CASAC does agree that key uncertainties identified in the last review remain. Given these considerations, this Draft PM 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." The CASAC agrees with this assessment.

What are the CASAC views regarding the areas for additional research identified in Chapter 4? Are there additional areas that should be highlighted?

$PM_{10-2.5}$ health effects research would be strengthened, and some of the remaining uncertainties might be better addressed, with improvements to exposure assessment and quantitative characterizations of remaining errors and uncertainties in exposure estimates. A more extensive network for direct monitoring of the $PM_{10-2.5}$ fraction would enable epidemiological studies to further address the remaining uncertainties.

Additional human clinical and animal toxicology studies of the health effects of the $PM_{10-2.5}$ fraction are needed to better understand biological causal mechanisms and pathways.

Chapter 5 – Review of the Secondary Standards

What are the CASAC views on the approach described in Chapter 5 to considering the evidence for PM-related welfare effects in order to inform preliminary conclusions on the secondary standards? What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current secondary PM standards?

The CASAC finds much of the information in Chapter 5 on visibility and material effects of $PM_{2.5}$ to be useful, while recognizing that uncertainties and controversies remain about the best ways to evaluate these effects. Additionally, the CASAC recommends that the PM PA should more thoroughly address effects of reducing $PM_{2.5}$ on climate change. Specifically, it should provide quantitative estimates and uncertainty bands for effects of changes in $PM_{2.5}$ on global dimming and global warming and their consequences for welfare effects on the U.S. The CASAC agrees with the Draft PM PA's assessment (p. 5-26) that "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." However, the CASAC finds that the discussion does not adequately inform policy makers about current scientific knowledge of effects of $PM_{2.5}$ reductions on global dimming and global warming (e.g., Schiermeier, 2005; Harvey, 2018).

Page 5-35 of the Draft PM PA states that "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 United States. Thus, as in the last review, the data remain insufficient to conduct quantitative analyses for PM effects on climate in the current review." The CASAC believes that it is reasonable to state qualitatively that there are remaining uncertainties, but

recommends that the PM PA should nonetheless provide a quantitative assessment to support well-informed policy-making. Recent research acknowledges the difficulties and uncertainties involved, but has advanced sufficiently 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." (Harvey, 2018) Chapter 5 should explain whether such quantitative predictions are consistent with current understanding based on the best available evidence (see e.g., Rosenfield et al., 2019; Luo et al., 2019). Effects of changes in $PM_{2.5}$ on global dimming and on the land carbon sink should be described (Mercado et al., 2009; USEPA, 2012). The PM PA should also address how changes in cold weather and warm weather temperature extremes due to reduced $PM_{2.5}$ are likely to affect public health risks (e.g., Patel et al., 2019; Xing et al., 2016).

The CASAC concludes that it is appropriate to acknowledge uncertainties in climate change impacts and resulting welfare impacts in the United States of reductions in $PM_{2.5}$ levels, and that it is also appropriate to identify reducing these uncertainties as future research needs. However, in the meantime, the current PM PA should provide policy makers with summaries of current scientific knowledge and quantitative modeling results for effects of reducing $PM_{2.5}$ on each of the following outcomes:

- Warming, dimming, and brightening
- High and low daily temperatures
- Resulting public health impacts related to daily temperatures
- Land carbon sink capacity
- Crop yields, agricultural productivity
- Other weather variables: Precipitation, storms, wind, etc.

At a minimum, estimates of the change in warming caused by a change in $PM_{2.5}$ should be discussed and implications for human welfare in the United States should be evaluated.

The CASAC agrees with the EPA that the available evidence does not call into question the protection afforded by the current secondary PM standards and concurs that they should be retained.

References

Bind, M.-A. (2019). Causal Modeling in Environmental Health. Annual Review of Public Health, 40(1), 23–43. https://doi.org/10.1146/annurev-publhealth-040218-044048

Blanchard, C. L., Hidy, G. M., Tanenbaum, S., Edgerton, E. S., & Hartsell, B. E. (2012). 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. https://doi.org/10.1080/10962247.2012.748523

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 Analysis, 22(6), 1183–1193. https://doi.org/10.1111/1539-6924.00282

Burns, J., Boogaard, H., Polus, S., Pfadenhauer, L. M., Rohwer, A. C., van Erp, A. M., … Rehfuess, E. (2019). Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database of Systematic Reviews. https://doi.org/10.1002/14651858.cd010919.pub2

Campbell, D.T., Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally & Company.

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

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

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., … Schwartz, J. D. (2017). Air Pollution and Mortality in the Medicare Population. New England Journal of Medicine, 376(26), 2513–2522. https://doi.org/10.1056/nejmoa1702747

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

Edgerton, E. S., Hartsell, B. E., Saylor, R. D., Jansen, J. J., Hansen, D. A., & Hidy, G. M. (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. https://doi.org/10.1080/10473289.2005.10464744

Edgerton, E. S., Hartsell, B. E., Saylor, R. D., Jansen, J. J., Hansen, D. A., & Hidy, G. M. (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. https://doi.org/10.1080/10473289.2006.10464585

Eum, K.-D., Suh, H. H., Pun, V. C., & Manjourides, J. (2018). Impact of long-term temporal trends in fine particulate matter (PM2.5) on associations of annual PM2.5 exposure and mortality. Environmental Epidemiology, 2(2), e009. https://doi.org/10.1097/ee9.0000000000000009

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

Greven, S., Dominici, F., & Zeger, S. (2011). An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information. Journal of the American Statistical Association, 106(494), 396–406. https://doi.org/10.1198/jasa.2011.ap09392

Hansen, D. A., Edgerton, E. S., Hartsell, B. E., Jansen, J. J., Kandasamy, N., Hidy, G. M., & Blanchard, C. L. (2003). The Southeastern Aerosol Research and Characterization Study: Part 1—Overview. Journal of the Air & Waste Management Association, 53(12), 1460–1471. https://doi.org/10.1080/10473289.2003.10466318

Harvey, C. (2018). Cleaning Up Air Pollution May Strengthen Global Warming. Scientific American, January 22, 2018. https://www.scientificamerican.com/article/cleaning-up-air-pollution-may-strengthen-global-warming/

Hemmingsen, J. G., Jantzen, K., Møller, P., & Loft, S. (2015). No oxidative stress or DNA damage in peripheral blood mononuclear cells after exposure to particles from urban street air in overweight elderly. Mutagenesis, 30(5), 635–642. https://doi.org/10.1093/mutage/gev027

Hemmingsen, J. G., Rissler, J., Lykkesfeldt, J., Sallsten, G., Kristiansen, J., P, P. M., & Loft, S. (2015). Controlled exposure to particulate matter from urban street air is associated with decreased vasodilation and heart rate variability in overweight and older adults. Particle and Fibre Toxicology, 12(1). https://doi.org/10.1186/s12989-015-0081-9

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. Journal of the Air & Waste Management Association, 67(2), 144–172. https://doi.org/10.1080/10962247.2016.1242518

Institute of Medicine (IOM) (2008). Improving the Presumptive Disability Decision-Making Process for Veterans, Chapter 8. Washington, DC: The National Academies Press. https://doi.org/10.17226/11908

Krewski, D., Jerrett, M., Burnett, R.T., Ma, R., Hughes, E., Shi, Y., Turner, M.C., Pope, C.A., Thurston, G., Calle, E.E., Thun, M.J., Beckerman, B., DeLuca, P., Finkelstein, N., Ito, K., Moore, D.K., Newbold, K.B., Ramsay, T., Ross, Z., Shin, H., Tempalski, B., (2009). Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Res. Rep. Health. Eff. Inst. 5–114; discussion 115-36.

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.

Luo, H., Han, Y., Lu, C., Yang, J., & Wu, Y. (2019). Characteristics of Surface Solar Radiation under Different Air Pollution Conditions over Nanjing, China: Observation and Simulation. Advances in Atmospheric Sciences, 36(10), 1047–1059. https://doi.org/10.1007/s00376-019-9010-4

Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., & Cox, P. M. (2009). Impact of changes in diffuse radiation on the global land carbon sink. Nature, 458(7241), 1014–1017. https://doi.org/10.1038/nature07949

Patel, D., Jian, L., Xiao, J., Jansz, J., Yun, G., & Robertson, A. (2019). Joint effect of heatwaves and air quality on emergency department attendances for vulnerable population in Perth, Western Australia, 2006 to 2015. Environmental Research, 174, 80–87. https://doi.org/10.1016/j.envres.2019.04.013

Pearl, J. (2010). An Introduction to Causal Inference. The International Journal of Biostatistics, 6(2). https://doi.org/10.2202/1557-4679.1203

Petitti, D. B. (1991). Associations Are Not Effects. American Journal of Epidemiology, 133(2), 101–102. https://doi.org/10.1093/oxfordjournals.aje.a115848

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. American Journal of Epidemiology, 186(8), 961–969. 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. Critical Reviews in Toxicology, 41(8), 651–671. https://doi.org/10.3109/10408444.2011.563420

Rosenfeld, D., Zhu, Y., Wang, M., Zheng, Y., Goren, T., & Yu, S. (2019). Erratum for the Research Article "Aerosol-driven droplet concentrations dominate coverage and water of oceanic low-level clouds" by D. Rosenfeld, Y. Zhu, M. Wang, Y. Zheng, T. Goren, S. Yu. Science, 364(6446), eaay4194. https://doi.org/10.1126/science.aay4194

Schichtel, B. A., Malm, W. C., Bench, G., Fallon, S., McDade, C. E., Chow, J. C., & Watson, J. G. (2008). Fossil and contemporary fine particulate carbon fractions at 12 rural and urban sites in the United States. Journal of Geophysical Research, 113(D2). https://doi.org/10.1029/2007jd008605

Schiermeier, Q. (2005). Clear skies end global dimming. Nature. https://doi.org/10.1038/news050502-8

Tanner, R. L., Parkhurst, W. J., & McNichol, A. P. (2004). Fossil Sources of Ambient Aerosol Carbon Based on 14C Measurements Special Issue of Aerosol Science and Technology on Findings from the Fine Particulate Matter Supersites Program. Aerosol Science and Technology, 38(sup1), 133–139. https://doi.org/10.1080/02786820390229453

U.S. EPA. (2009). Integrated Science Assessment for Particulate Matter. EPA/600/R-08/139F. Research Triangle Park, NC. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=216546

U.S. EPA. (2012). Report to Congress on Black Carbon. EPA-450/R-12-001. Research Triangle Park, NC. https://www3.epa.gov/airquality/blackcarbon/2012report/fullreport.pdf

U.S. EPA. (2015). Preamble to the Integrated Science Assessments. EPA/600/R-15/067. Research Triangle Park, NC. http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=526136

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. Occupational and Environmental Medicine, 52(12), 790–792. https://doi.org/10.1136/oem.52.12.790

Xing, J., Wang, J., Mathur, R., Pleim, J., Wang, S., Hogrefe, C., … Hao, J. (2016). Unexpected Benefits of Reducing Aerosol Cooling Effects. Environmental Science & Technology, 50(14), 7527–7534. https://doi.org/10.1021/acs.est.6b00767

Yoshimura, I. (1990). The effect of measurement error on the dose-response curve. Environmental Health Perspectives, 87, 173–178. https://doi.org/10.1289/ehp.9087173

Appendix A

Possible Technical Options

In response to an EPA request for advice on specific practical techniques that could be used in the near term to address these issues, the CASAC notes the following:

- Confounding by measured variables can be addressed by conditioning on their values (e.g., by including them as explanatory variables in structural equation regression models or in the adjustment sets used to estimate the direct and total causal effects of $PM_{2.5}$ on mortality risks). In more detail:
    - A directed acylic graph (DAG) model such as the following can be used to organize the calculation of quantitative adjustment for measured confounders:

      $PM_{2.5}$ ← *month* → *tmin* (daily minimum temperature)
      ↓ ↓ ↓
      mortality risk (associated with $PM_{2.5}$ and other predictors)

    - If there is time to undertake new data analyses, then a package such as bnlearn (http://www.bnlearn.com) can be used to estimate such DAG models from data. (In practice, lagged values of daily minimum and maximum temperatures, as well as humidity and other variables should be included as potential predictors.) The DAG model estimates how the conditional probability of the mortality risk variable depends on the values of the predictors that point into it (in this example, $PM_{2.5}$, month, and tmin). This information is stored in a conditional probability table (CPT) for the mortality risk node in the DAG. This CPT dependence, rather than the total association between $PM_{2.5}$ and mortality risk, is what is needed to create a valid causal simulation of how changing $PM_{2.5}$ alone would change mortality risk probabilities, adjusting for the effects of confounders.
    - To estimate how reducing $PM_{2.5}$ alone (without adjusting month or other variables) would affect mortality risk, condition on a sufficient adjustment set for estimating the direct effect of $PM_{2.5}$ on mortality risk. (These adjustment sets can be calculated using the dagitty package, http://www.dagitty.net/manual-2.x.pdf.) The conditioning and estimation can be done using the partial dependence plot (partialPlot) option in the randomForest package (https://cran.r-project.org/web/packages/randomForest/randomForest.pdf). (Estimate mortality_risk ~ $PM_{2.5}$ + month + … (other variables in adjustment set). The Causal Analytics Toolkit (http://cox-associates.com; http://cox-associates.com:8899/) can be used for all of the preceding calculations.)
    - If no new data analysis is practicable, then as a short term fix, the EPA might consider performing sensitivity analyses in which the simulated effects of reducing $PM_{2.5}$ on reducing mortality risk are adjusted by assuming that 0% (current assumption in Draft

PM PA), 20%, 40%, 60%, 80%, or 100% (empirical estimate in several studies) of the $PM_{2.5}$-mortality association is explained by confounding that is not decreased when $PM_{2.5}$ is decreased.

  - Alternatively, if the EPA concludes that there is a strong, consistent positive association between $PM_{2.5}$ levels and mortality rates in past data, but little or no evidence that interventions that change $PM_{2.5}$ have caused detectable changes in mortality rates, then existing data sets might be used to suggest a plausible upper bound for the largest effect sizes that could be present and yet have escaped confident detection. This approach acknowledges that absence of evidence is not necessarily evidence of absence of a true effect in dozens of intervention studies that have been reviewed (e.g., Burns et al., 2019), while at the same time acknowledging that it is evidence that any effects must be small enough not to have created stronger results than those observed. One possible numerical estimate based on a single study is that a causal effect on the order of 0.1% increase in mortality rate per 10 $\mu g/m^3$ increase in average daily $PM_{2.5}$ exposure concentrations would have been easily detected (Cox et al., 2012). This value was derived by simulating causal effects of different sizes and observing how large they needed to be to be reliably detected (in non-parametric analyses that compared mortality rates on days with high and low $PM_{2.5}$ exposures, matched by city, month of year, and temperature quartile). The much larger body of intervention studies reviewed by Burns et al. (2019) might be used to estimate a tighter upper bound using similar methods.

- *Residual confounding* by measured variables can be addressed by conditioning on the original (higher-resolution) values of the variables, rather than by conditioning on categorized or averaged values (e.g., by conditioning on month rather than on season, or by conditioning on daily high and low temperatures rather than on seasonal or annual averages of their values).
- *Unobserved confounders* can be addressed by using techniques (or results of studies) such as the following:
  - Greven et al. (2011)
  - Pun et al. (2017)
  - Eum et al. (2018)
  - Other techniques for testing and adjusting for unobserved confounders, e.g., using latent variables, are discussed and referenced in individual comments from the CASAC members on the ISA.
  - As a short-term fix, the EPA might consider using an approximate reduction factor of 0.6 (= 0.12/0.20) in estimated excess (above 1) mortality risk ratios (MRRs) to correct for effects of unmeasured confounders that decease over time, based on the ratio 1.20 without adjustment and 1.12 with adjustment in reported by Eum et al. (2018): "We found a 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure to be associated with a 1.20 times (95% confidence interval [CI] = 1.20, 1.21) higher risk of mortality across the 13-year study period, with the magnitude of the association decreasing with shorter study periods. MRRs remained statistically significant but were attenuated when models adjusted for long-term time trends in $PM_{2.5}$. The residual-based, time-adjusted MRR equaled 1.12 (95% CI = 1.11, 1.12) per 10 $\mu g/m^3$ for the 13-year study period and did not change when shorter study periods were examined."

- Errors in exposure estimates and other covariates can be addressed using errors-in-variables techniques, as discussed further in individual comments from the CASAC members on the ISA. A crucial point is that, for the same estimated exposure level in a C-R study, individuals who responded are expected to have had higher actual exposures than otherwise similar individuals who did not respond (assuming that exposure increases response probability); thus, measurement error is not expected to be independent of the outcome (e.g., mortality vs. no mortality).
- As a relatively quick fix, the EPA may wish to consider the specific methods described in the following references:
    - Imai et al. (2011)
    - Blackwell et al. (2015)
- Alternatively, a quicker but rougher approach would be to apply adjustment factors estimated from studies such as Hart et al. (2015). This study noted that "However, to date, no study has attempted to incorporate adjustment for measurement error into a long-term study of the effects of air pollution on mortality" and reported that, in the data set analyzed, "the multivariable adjusted HRs for each 10 $\mu g/m^3$ increase in 12-month average $PM_{2.5}$ from spatio-temporal prediction models were 1.13 (95%CI:1.05, 1.22) and 1.12 (95%CI:1.05, 1.21) for concentrations at the nearest EPA monitoring location. Adjustment for measurement error increased the magnitude of the HRs 4-10% and led to wider CIs (HR = 1.18; 95%CI: 1.02, 1.36 for each 10 $\mu g/m^3$ increase in $PM_{2.5}$ from the spatio-temporal models and HR = 1.22; 95%CI: 1.02, 1.45 from the nearest monitor estimates)."
- *Model uncertainty* can be addressed by using model ensemble methods, such as Bayesian model averaging (BMA) for parametric risk models, or random forest for non-parametric model ensembles (https://cran.r-project.org/web/packages/randomForest/randomForest.pdf). Regression diagnostics should also be clearly presented for quantitative risk models that have been fit to data (e.g., the regression C-R models in Appendix C of the Draft PM PA).
    - A key issue is how much wider the uncertainty intervals around estimated C-R regression functions should become when model uncertainty is included than when it is not. If there is time to do new data analyses, then one option is to use ensemble methods to directly quantify uncertainty bands around partial dependence plots (e.g., using https://www.rdocumentation.org/packages/rfUtilities/versions/2.1-3/topics/rf.partial.ci).
    - If there is time to do new data analyses, then regression diagnostics for Cox proportional hazards (PH) models can be calculated by standard statistical software (e.g., Schoenfeld Residuals Tests, http://www.sthda.com/english/wiki/cox-model-assumptions#testing-proportional-hazards-assumption). These diagnostics should be reported. If the null hypothesis that the PH model describes the data is rejected, then the EPA might consider non-parametric alternatives.
    - Many researchers have recognized that "Different epidemiological models may lead to very different conclusions for the same set of data" (Fuentes et al., 2009) and that estimates of regression relationships often depend heavily on specific modeling choices (e.g., the same data can be used to estimate either positive or negative C-R regression coefficients, depending on what other predictors are included); thus, "There is a growing consensus in economics, political science, statistics, and other fields that the associational or regression approach to inferring causal relations—on the basis of adjustment with

observable confounders—is unreliable in many settings" (Dominici et al., 2014). Non-specialists sometimes mistake consistency of estimated C-R associations as evidence of causality (as in the Bradford Hill considerations), without recognizing the extent to which they may simply reflect modeling choices that enforce such consistency. A useful characterization of model uncertainty should show how widely estimated C-R functions vary as alternative plausible modeling choices are made.

- A very quick, rough adjustment factor might be obtained as follows. The following table for PM_{10} illustrates that C-R risk coefficients estimated by 4 different methods vary widely (e.g., from -2.24 to 2.76 for Richmond, with a pooled estimate of 0.72) (Fuentes et al., 2009). In many cases, the differences among estimates are far greater than would be expected from the estimated SEs alone (e.g., the SE of 0.31 for the "shrunken" method for Richmond), reflecting the importance of model uncertainty and the sensitivity of results to choice of modeling method rather than just to sampling variability in the data. The ratio of the spread in values from different methods to the spread predicted from estimated SEs in tables similar to this one, but for $PM_{2.5}$, might be used to scale up SE-based confidence interval widths to reflect model uncertainty.

1. Table showing how the 4 methods presented in this paper change the resulting estimates of the local effect $\widehat{\beta}^c$. The reported estimates are the health effect times 10^3, $10^3\widehat{\beta}^c$, corresponding to percent increase in mortality per increase in 10 units of PM_{10}.

	Local		Adjusted		Shrunken		Full Bayes	
	$\widehat{\beta}^c$	SE	$\widehat{\beta}^c$	SE	$\widehat{\beta}^c$	SE	$\widehat{\beta}^c$	SE
Syracuse	3.18	1.56	0.72	1.60	0.82	0.31	1.41	0.98
Boston	2.50	1.27	0.72	1.31	0.83	0.31	1.35	0.88
Providence	2.03	1.23	0.72	1.27	0.80	0.31	1.21	0.83
Jersey City	1.25	0.75	0.72	0.81	0.80	0.29	1.03	0.60
Baltimore	0.40	0.62	0.72	0.69	0.66	0.28	0.55	0.52
Newark	0.23	1.05	0.72	1.09	0.68	0.30	0.56	0.72
Philadelphia	0.09	0.66	0.72	0.73	0.61	0.28	0.38	0.56
Washington	0.01	1.53	0.72	1.56	0.69	0.31	0.58	0.87
Kingston	−1.20	2.19	0.72	2.21	0.68	0.31	0.45	1.02
Arlington	−1.60	5.96	0.72	5.97	0.72	0.31	0.71	1.17
Richmond	−2.24	2.74	0.72	2.76	0.68	0.31	0.42	1.09
Pooled Estimate	0.72	0.32	0.72	0.31	0.72	0.32	0.79	0.50

(Source: Fuentes et al., 2009)

- Interindividual variability and heterogeneity in individual biological dose-response functions refers to the frequency distribution of individual biological (causal) C-R functions in an exposed population. Using larger samples does not reduce the variability in C-R functions, although it reduces the width of confidence intervals around the mean values of C-R functions. Heterogeneity in population C-R functions arises from the fact that different clusters of individuals may have significantly different mean C-R functions (and, in more detail, significantly different population distributions of individual C-R functions). Heterogeneity is reflected in the fact that estimated regression C-R functions are significantly different across studies and locations (e.g., cities), e.g., significantly negative in some and positive in others. This information is lost when heterogeneous CR functions are simply averaged, as for the C-R heterogeneity study (Baxter et al., 2017) in Appendix C of the Draft PM PA. Thus, as stated in the 2009 PM ISA,, "the heterogeneity of the shapes across cities makes it difficult to interpret the biological relevance of the shape of the combined curves." (Summary of Expert Opinions on the Existence of a Threshold in the Concentration-Response Function for $PM_{2.5}$-related Mortality).

- If new data analysis is possible, then C-R variability and hetereogeneity can be quantified using individual conditional expectation (ICE) plots (https://cran.r-project.org/web/packages/ICEbox/index.html).
- Otherwise, the distribution of C-R shapes across cites can be used illustrate the extent of realistic heterogeneity in C-R functions.

References

Blackwell, M., Honaker, J., & King, G. (2015). A Unified Approach to Measurement Error and Missing Data: Overview and Applications. Sociological Methods & Research, 46(3), 303–341. https://doi.org/10.1177/0049124115585360

Cox, T., Popken, D., & Ricci, P. F. (2012). Temperature, not Fine Particulate Matter (PM2.5), is Causally Associated with Short-Term Acute Daily Mortality Rates: Results from One Hundred United States Cities. Dose-Response, 11(3), dose-response.1. https://doi.org/10.2203/dose-response.12-034.cox

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

Eum, K.-D., Suh, H. H., Pun, V. C., & Manjourides, J. (2018). Impact of long-term temporal trends in fine particulate matter (PM2.5) on associations of annual PM2.5 exposure and mortality. Environmental Epidemiology, 2(2), e009. https://doi.org/10.1097/ee9.0000000000000009

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

Greven, S., Dominici, F., & Zeger, S. (2011). An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information. Journal of the American Statistical Association, 106(494), 396–406. https://doi.org/10.1198/jasa.2011.ap09392

Hart, J. E., Liao, X., Hong, B., Puett, R. C., Yanosky, J. D., Suh, H., Kioumourtzoglou, M.A., Spiegelman, D., & Laden, F. (2015). The association of long-term exposure to PM2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environmental Health, 14(1). https://doi.org/10.1186/s12940-015-0027-6

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. American Political Science Review, 105(4), 765–789. https://doi.org/10.1017/s0003055411000414

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. American Journal of Epidemiology, 186(8), 961–969. https://doi.org/10.1093/aje/kwx166

Appendix B

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

Dr. James Boylan **B-2**
Dr. Tony Cox **B-9**
Dr. Mark Frampton **B-33**
Dr. Sabine Lange **B-38**
Dr. Corey Masuca **B-100**
Dr. Steven Packham **B-103**

Dr. James Boylan

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?

The CASAC review letter on EPA's Draft ISA submitted to the EPA Administrator on April 11, 2019 recommended the development of a Second Draft ISA for CASAC review and the reappointment of the previous CASAC PM panel (or appoint a panel with similar expertise) in time to review the Second Draft ISA. Instead, EPA has provided CASAC with a pool of consultants that can respond to written questions from the CASAC. Although the pool of consultants has provided additional insight and useful information, they do not serve the same role as the former PM review panel since there are no deliberations and only written answers to specific questions. I feel that the traditional review process (with pollutant specific review panels) is significantly more informative to CASAC's recommendations since it allows verbal discussions and deliberations among experts with differing backgrounds and opinions resulting in a comprehensive examination of controversial topics.

The purpose of the PA is to bridge the gap between EPA's scientific assessments and the judgement required by the EPA Administrator when determining whether to retain or revise the NAAQS. It is unusual to review a draft PA when the ISA had not been finalized. Also, it is unusual to include the REA as part of the PA rather than a stand-alone document that is reviewed and finalized prior to the release of the draft PA. I feel that a final ISA and a second draft of the REA are needed prior to releasing the final PA. Without final versions of EPA's scientific assessments (ISA and REA), it is difficult to provide sound advice on the draft PA.

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. For example, the FRM monitor at Albany, GA (13-095-0007) on March 7, 2019 measured a 24-hour concentration of 30 □g/m^3 (5 □g/m^3 below the level of the 24-hour standard) while the co-located FEM measured 40 □g/m^3 (5 □g/m^3 above the level of the 24-hour standard). EPA should compare co-located FRM/FEMs 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 recently ended). It 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

Trends in Ambient Air Concentrations

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.

EPA's 2016 exceptional events rule allows certain 24-hour PM measurements due to natural events to be excluded from the official design values when compared to the NAAQS. In some cases, identical exceptional events can be treated differently in one location vs. another based on how close the area is to the standard. In both locations, people are impacted by adverse health effects, but the data is removed in one location and not the other. The PA should discuss how exceptional events are accounted for in the setting of the standard.

Hybrid $PM_{2.5}$ Modeling Approaches

Hybrid $PM_{2.5}$ modeling approaches have improved the estimates of ambient PM2.5 concentrations in areas without monitors. In fact, this is one of the most significant improvements since the last PM

NAAQS review. The four approaches presented in the chapter generally agree although the spatial texture of the concentration fields differ among methods.

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?

The study area selection, $PM_{2.5}$ air quality scenarios evaluated, model-based approach to adjusting air quality, and linear interpolation/extrapolation to additional annual standard levels seem appropriate. However, the selection of health outcomes needs additional explanation.

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. If it is due to limited time and resources, that should be stated along with the potential implications on 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 air quality modeling approach to projecting $PM_{2.5}$ concentrations to correspond to just meeting the NAAQS (AQS, CMAQ, Downscaler, SMAT-CE, project monitors to just meet NAAQS, project spatial fields to correspond to just meeting the NAAQS) appears to be scientifically sound. The CMAQ model configuration and input files used in the air quality modeling appear to be appropriate for this application.

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.

According to the PA (page 1-4), "The CAA does not require the Administrator to establish a primary NAAQS at a zero-risk level or at background concentration levels…, but rather at a level that reduces risk sufficiently so as to protect public health with an adequate margin of safety." With a linear, no-threshold CR assumption, any $PM_{2.5}$ level above 0 $\mu g/m^3$ will lead to additional premature deaths. The risk assessment can be used to examine whether the current and alternate standards have sufficiently reduced risk to protect public health with an adequate margin of safety. In order to put the risk assessment results in proper perspective, EPA should compare the current risk assessment results to those deemed acceptable in the previous PM NAAQS review.

The risk assessment presented by EPA gives estimates of $PM_{2.5}$-associated mortality for the current and potential alternative standards. Tables 3-5 and 3-6 estimate $PM_{2.5}$-associated mortality for the current and potential alternative standards for 47 urban study areas, Tables 3-7 and 3-8 estimate $PM_{2.5}$-associated mortality for the current and potential alternative standards in 30 study areas where the annual standard is controlling, and Table 3-9 estimates $PM_{2.5}$-associated mortality for the current and

potential alternative standards in 11 study areas where the 24-hour standard is controlling. In each table containing "CS (12/35) % of baseline", EPA should add another column showing the actual baseline values. In Table 3-5, the baseline number of all-cause deaths from Pope 2015 (pri-PM) would be 51,300/0.071=722,535 deaths, while the baseline number of all-cause deaths from Thurston 2015 (pri-PM) would be 13,500/0.032=421,875 deaths. The baseline number of all-cause deaths from Pope 2015 (pri-PM) is over 300,000 deaths higher compared to the baseline number of all-cause deaths from Thurston 2015 (pri-PM). These large differences should be discussed and the impact on the evaluation of alternative standards should be documented. The changes in absolute risk between the current standard and alternative standards (e.g., 6,600-1,800) are relatively small compared to the differences in baseline number of all-cause deaths (e.g., 300,000).

Although the 95 percent confidence intervals are presented for each point estimate, there is not a quantitative uncertainty analysis. This analysis should look at the impacts of air quality modeling uncertainty on the population-level exposure and risk assessment results. The uncertainty associated with confounders (current analysis assumes 100% of the association is causal) should be evaluated and included in the assessment. Also, the uncertainty related to the shape of the CR relationship for $PM_{2.5}$ exposures and mortality at low ambient PM levels should be included in the analysis. In EPA's 2006 PM NAAQS risk analysis, "These risks were estimated using not only the linear or log-linear concentration-response functions reported in the studies, but also using alternative modified linear functions as surrogates for assumed non-linear functions that would reflect the possibility that thresholds may exist in the reported associations within the range of air quality observed in the studies" (71 FR 61154). A similar approach should be applied in this risk assessment. Finally, given the overlapping confidence bounds in the risk assessment, EPA should evaluate if there is a statistically different risk between meeting the current NAAQS and alternative NAAQS.

In order to recommend whether the current $PM_{2.5}$ standards are adequate or not, I would like to see corroborating supporting evidence from both the evidence-based approach and the risk-based approach. Due to the reasons discusses above, the risk assessment (as presented in this document) is not adequate to justify lowering the $PM_{2.5}$ standard. Therefore, I recommend that the current annual and 24-hour $PM_{2.5}$ standards be retained without revision. In addition, I recommend that an updated risk assessment be developed and shared with CASAC for review prior to the release of the final PA. Review of an updated risk assessment may lead to different policy recommendations on the current annual and 24-hour $PM_{2.5}$ standards.

Chapter 4 – Review of the Primary PM_{10} Standard

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?

Based on the evidence presented in Chapter 4, I support EPA's recommendation to retain the current PM_{10} standard without revision.

Chapter 5 – Review of the Secondary Standards

What are the CASAC views on the approach described in chapter 5 to considering the evidence for PM-related welfare effects in order to inform preliminary conclusions on the secondary standards? What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current secondary PM standards?

Based on the evidence presented in Chapter 5, I support EPA's recommendation to retain the current secondary standards without revision.

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?

EPA should develop Federal Reference Methods for measuring ultrafine particles and elemental carbon and require States to add these monitors to their near-road and NCore monitoring networks. This would help provide the needed ambient monitoring data for ultrafine particles and elemental carbon to support epidemiology studies. 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?"*

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 <u>the primary weakness</u>." 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 ; 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.

The basis for the Draft PA's preliminary conclusion that available scientific information can reasonably be viewed as calling into question the adequacy of the public health protection afforded by the current primary PM2.5 standards has been summarized as follows (Agency Briefing Material, p. 18):

1. *Long-standing body of health evidence, strengthened in this review, supporting relationships between short- and long-term PM2.5 exposures and various outcomes, including mortality and serious morbidity effects*
2. *Recent U.S. and Canadian epidemiologic studies reporting positive and statistically significant health effect associations for PM2.5 air quality likely to be allowed by the current standards*

3. *Analyses of pseudo-design values indicating substantial portions of study area health events/populations in locations with air quality likely to have met the current PM2.5 standards*
4. *Risk assessment estimates that the current primary standards could allow thousands of PM2.5-associated deaths per year – most at annual average PM2.5 concentrations from 10 to 12 μg/m^3 (well within the range of overall mean concentrations in key epidemiologic studies).*

These points do not provide a reasonable scientific basis for questioning that the current primary PM2.5 standards protect human health with an adequate margin of safety, for the following main reasons.

- For point 1, the fact that PM2.5 is associated with mortality and morbidity has been well known for many years, and was already anticipated and assumed in setting the current NAAQS.
- However, *associations* between exposure and response are not the same as *effects* of exposure on response (Petitti DB. Associations are not effects. Am J Epidemiol. 1991 Jan 15; 133(2):101-2). Such associations are expected, and are not necessarily (or usually) preventable by reducing exposure, when confounding, model uncertainty, and other non-causal explanations are present, as they are in the studies relied on by the Draft PA. Specifically, both PM2.5 and mortality and morbidity are positively associated with confounders, such as low income and extreme daily high and low temperatures, that have not been fully accounted for in the studies relied on in the Draft PA, including the 8 studies in Appendix C. Evidence that there are associations when confounders are present is not valid evidence that exposure is causing harm or that reducing exposure would reduce risk to public health.
- My experience analyzing C-R data sets for PM2.5 has been that estimating exposure-response associations without fully controlling for daily temperature extremes, month of year, and other variables does indeed reveal significant positive statistical associations between PM2.5 and mortality, consistent with other findings in the literature; but that these associations disappear once measured confounders such as month and daily high and low temperatures are fully adjusted for (e.g., Cox LAT Jr. Do causal concentration-response functions exist? A critical review of associational and causal relations between fine particulate matter and mortality. Crit Rev Toxicol. 2017 Aug;47(7):603-631. Cox T, Popken D, Ricci PF. Temperature, Not Fine Particulate Matter (PM2.5), is Causally Associated with Short-Term Acute Daily Mortality Rates: Results from One Hundred United States Cities. Dose Response. 2012 Dec 14;11(3):319-43. Cox LA Jr, Popken DA, Ricci PF. Warmer is healthier: effects on mortality rates of changes in average fine particulate matter (PM2.5) concentrations and temperatures in 100 U.S. cities. Regl Toxicol Pharmacol. 2013 Aug;66(3):336-46). Therefore, studies that report such associations, but that do not fully adjust for such confounders, do not provide clear evidence that the reported associations would remain if confounding were more carefully controlled; nor do they necessarily indicate that reducing exposures is necessary or sufficient to reduce health risks.
- The non-robustness of C-R associations between PM2.5 and mortality to careful control for measured (or easily measurable) confounders is broadly consistent with findings from recent reviews of accountability studies and intervention studies: "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.). …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). While these qualitative findings emphasize the difficulty of establishing causal effects, quantitative work shows that it should be easy to identify causal effects (specifically, mortalities preventable by reducing PM2.5) in existing studies if they are of the approximate size (or even 10% of the size) usually assumed based on associations (Cox et al. 2012, op cit.)

- The non-robustness of C-R associations for exposure and mortality and morbidity to changes in modeling choices and assumptions is well recognized in previous literature. "Different epidemiological models may lead to very different conclusions for the same set of data" (Fuentes M. Statistical issues in health impact assessment at the state and local levels. Air Qual Atmos Health. 2009 Mar 1;2(1):47-55). The same C-R data can often be used to generate either positive or negative associations by making different modeling choices, so that "the associational or regression approach [followed in the Draft PA] to inferring causal relations— on the basis of adjustment with observable confounders— is unreliable in many settings"(Dominici F, Greenstone M, Sunstein CR. Particulate Matter Matters Science. 2014 Apr 18; 344(6181): 257–259). Thus, the presence of significant positive associations does not necessarily convey information about causation, but may reflect modeling choices.
- The Draft PA does not present evidence that rules out non-causal explanations for the C-R associations mentioned in point 1, including (a) Confounding by easily measured confounders (such as month and daily high and low temperatures); (b) Confounding by unmeasured confounders and coincident historical trends (e.g., using methods such as those in Greven et al. An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information. J Am =Stat Assoc. 2011;106(494):396-406. doi: 10.1198/jasa.2011.ap09392; Pun VC et al. Long-Term PM2.5 Exposure and Respiratory, Cancer, and Cardiovascular Mortality in Older US Adults. Am J Epidemiol. 2017 Oct 15;186(8):961-969. doi: 10.1093/aje/kwx166; Eum KD, Suh HH, Pun VC, Manjourides J. Impact of long-term temporal trends in fine particulate matter (PM2.5) on associations of annual PM2.5 exposure and mortality: An analysis of over 20 million Medicare beneficiaries. Environmental Epidemiology: June 2018 2(2): p e009.); (c) residual confounding; and (d) Modeling choices and model specification errors (e.g., as revealed by regression diagnostics for the Proportional Hazards model assumed in Appendix C). Since

such potential non-causal explanations for the reported associations have not been tested and convincingly refuted, or rendered unlikely, using data, it is not scientifically valid to conclude that the C-R associations alluded to in point 1 provide reasonable grounds for calling into question the adequacy of the current primary standards to protect human health. To the extent that such associations are expected due to incompletely controlled confounding and modeling error, whether or not further reducing PM2.5 exposures would further reduce human health risks, they do not provide valid evidence that further reductions in PM2.5 are needed to protect human health.

- Suh et al. submitted a written comment on November 5 stating that "We are writing to comment on issues raised in the public comment by Mr. Stewart E. Holm, Chief Scientist for the American Forest & Paper Association, during the Clean Air Scientific Advisory Committee (CASAC) public comment session on the draft EPA Policy Assessment for Fine Particulate Matter (PM2.5) on October 22, 2019, and in individual written comments from Dr. Tony Cox on the Policy Assessment dated October 21, 2019. In discussing the findings from our Pun et al. (2017) [1], Mr. Holm makes factually incorrect statements regarding results from our study and both Mr. Holm and Dr. Cox draw inaccurate conclusions regarding our study findings and their policy relevance." However, I have drawn no conclusions regarding their study findings. I cited them only to illustrate that some, but not all, of reported PM2.5-mortality associations may be explained by temporal trends, which therefore should be controlled for before interpreting the associations causally. Although it seems to me that this is consistent with their own explanation of their findings, if the above references to their work do not support that confounding by unmeasured confounders and coincident historical trends may contribute to C-R associations and should be controlled for, then my references to them may simply be disregarded; the point still stands on methodological grounds, whether or not the works of these authors illustrate it.

- Similarly, point 2, on recent U.S. and Canadian epidemiologic studies reporting positive, statistically significant C-R associations for relatively low concentrations of PM2.5, likewise does not use data to test or refute non-causal interpretations of these associations, including confounding by month, daily high and low temperatures, and other easily measured confounders; residual confounding; unmeasured or unmodeled confounders and trends; and modeling choices and specification errors. In addition, point 2 does not distinguish between estimated exposures for individuals and their actual exposures. In particular, it is left unclear in the Draft PA by how much exposures of individuals with adverse health effects may have differed from estimated exposure concentrations based on areas (e.g., zip codes, square kilometers, or other aggregate areas). Thus, it is not clear from the evidence presented in the Draft PA and ISA whether C-R associations have been observed at individual exposure concentrations below the current NAAQS. Even if such associations were established, it would be important to quantify the extent to which they are due to confounding, modeling choices, or other non-causal explanations. As discussed for point 1, the existence of such associations was

already anticipated when the current NAAQS levels were set. No quantitative uncertainty analysis has been reported showing whether uncertainties about the effects on public health of further reducing PM2.5 are larger or smaller now than when the current NAAQS levels were set. Finally, point 2 ignores the fact that other recent studies have found either no significant or smaller-than-expected positive associations between reductions in PM2.5 and reductions in public health 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 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.) Therefore, point 2 provides no rational scientific basis for concluding that the current NAAQS levels are no longer fully adequate to protect human health with an adequate margin of safety.

- For point 3, the analyses of pseudo-design values again do not distinguish between *actual* individual exposures and *estimated* average exposures, e.g., for broad areas (zip codes, counties, cities, 1 x 1 kilometer areas, etc.) within which individuals live. The pseudo-design values are also assumption-laden and have not been validated; hence, their relevance for real-world health effects of real-world exposures remains unknown.
- For point 4, the risk assessment estimates have not been empirically validated as providing correct (or approximately correct) predictions. The risk assessment amounts to an unverified assumption about how public health risks would respond if PM2.5 exposure were reduced. Such an unverified assumption does not constitute valid scientific evidence for concluding that reducing exposure is necessary to protect human health. The underlying methodology (treating regression C-R functions as if they were causal C-R functions) is not conceptually sound in general (Bind MA. Causal Modeling in Environmental Health. Annu Rev Public Health. 2019 Apr 1;40:23-43; see also comments from consultants Drs. North and Aliferis) and it has not been justified for this specific application to PM2.5 health effects. Therefore, point 4's contention that "that the current primary standards could allow thousands of PM2.5-associated deaths per year" is simply a speculation based on modeling assumptions with neither theoretical validity nor empirical support.
- Taken together, points 1-4 do not refute, or call into question on valid empirical and scientific grounds, the null hypothesis that the current NAAQS levels are fully adequate to protect human health with an adequate margin of safety.

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

Chapter 5 – Review of the Secondary Standards: What are the CASAC views on the approach described in chapter 5 to considering the evidence for PM-related welfare effects in order to inform preliminary conclusions on the secondary standards? What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current secondary PM standards?

Much of the information in Chapter 5 of the Draft PA on visibility and material effects of PM2.5 is useful, but uncertainties and controversies remain about the best ways to evaluate these effects. Additionally, the PA should more thoroughly address effects of reducing PM2.5 on climate change. Specifically, it should provide quantitative estimates and uncertainty bands for effects of changes in PM2.5 on global dimming and global warming and their consequences for welfare effects on the US. The Draft PA states (p. 5-26) that "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." However, the discussion does not adequately inform policy makers about current scientific knowledge of effects of PM2.5 reductions on global dimming and global warming (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; Harvey 2018, "Cleaning Up Air Pollution May Strengthen Global Warming: New research is helping quantify just how big that effect might be" www.scientificamerican.com/article/cleaning-up-air-pollution-may-strengthen-global-warming/).

Page 5-35 of the Draft PA states that "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." It is reasonable to state qualitatively that there are remaining uncertainties, but it is recommended that the PA should nonetheless provide a quantitative assessment to support well-informed policy-making. Recent research acknowledges the difficulties and uncertainties involved, but has advanced sufficiently 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" (Harvey 2018, op. cit.) Chapter 5 should explain whether such quantitative predictions are consistent with current understanding based on the best available evidence (see e.g., Rosenfield et al. (2019), https://science.sciencemag.org/content/364/6446/eaay4194; Luo H, Han Y, Lu C, Yang J, Wu Y, Characteristics of surface solar radiation under different air pollution conditions over Nanjing, China: Observation and Simulation. Advances in Atmospheric Sciences October 2019 36(10):1047–1059). Effects of changes in PM2.5 on global dimming and on the land carbon sink should be described (Mercado LM, Bellouin N, Sitch S, Boucher O, Huntingford C, Wild M, and Cox PM. (2009). Impact of changes in diffuse radiation on the global land carbon sink. Nature, 458, 1014-1017; External Review DRAFT EPA Report to Congress on Black Carbon 3/18/11.) 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).

It is appropriate to acknowledge uncertainties in climate change impacts and resulting welfare impacts in the United States of reductions in PM2.5 levels, and it is appropriate to identify reducing these uncertainties as future research needs. However, the current PA should provide policy makers with

accurate summaries of current scientific knowledge and quantitative modeling results for effects of reducing PM2.5 on each of the following outcomes:

- Warming, dimming, and brightening
- High and low daily temperatures
- Resulting public health impacts related to daily temperatures
- Land carbon sink capacity
- Crop yields, agricultural productivity
- Other weather variables: Precipitation, storms, wind, etc.

At a minimum, estimates of the change in warming caused by a change in PM2.5 should be discussed and implications for human welfare in the United States should be evaluated.

Dr. Mark Frampton

General Comments

In response to a CASAC request in its April 11 letter to the Administrator, 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 stated in its April 11 letter, "Additional expertise is needed for the Clean Air Scientific Advisory Committee (CASAC) to provide a thorough review of the particulate matter (PM) National Ambient Air Quality Standards (NAAQS) documents. The breadth and diversity of evidence to be considered exceeds the expertise of the statutory CASAC members, or indeed of any seven individuals."

CASAC recommended that "…the EPA reappoint the previous CASAC PM panel (or appoint a panel with similar expertise)…". EPA has not done so. The panel of 12 consultants appointed by EPA does not have the breadth or depth of expertise that was represented on the original (dismissed) PM panel, and moreover does not include additional areas of expertise requested. The newly appointed panel of consultants does not include sufficient expertise and experience in air pollution epidemiology research. This is a scientific discipline that is obviously of key importance in the review of the PM standards. None of the current chartered CASAC members are experts in air pollution epidemiology. In addition, the restrictive process for interacting with the newly appointed consultants, which was imposed by EPA without consultation with CASAC, prevents open and frank discussions that are part of the process of achieving consensus. These limitations adversely affect the ability of CASAC to provide the EPA with the best and most relevant advice on the adequacy of the current NAAQS.

We note the difficulty and limitation in providing cogent and insightful advice 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 1. Section 1.4, summarizing the current NAAQS review, is incomplete. It omits a description of the changes to the NAAQS review process, the CASAC, and the PM NAAQS review since 2016. This section of the PA should summarize these changes, and indicate how these changes deviate from the process described in the IRP.

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 provide plausibility and 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 ambient $PM_{2.5}$ concentrations.

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 provide additional evidence that the previously observed associations between estimated ambient $PM_{2.5}$ exposure and mortality extend to concentrations 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 µg/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 µg/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." We 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 <u>understanding the relative toxicity of the different components</u> in the fine particle mixture, <u>the role of PM2.5 in the complex ambient mixture</u>, <u>exposure measurement errors in epidemiologic studies</u>, and the <u>nature and magnitude of estimated risks related to relatively low ambient PM2.5 concentrations</u>. Furthermore, the Administrator noted that epidemiologic studies had reported <u>heterogeneity in responses both within and between cities and in geographic regions across the U.S</u>. 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.**

<u>Evidence-Based Considerations for PM2.5 Concentrations in Key Studies</u>

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.

<u>Risk-Based Considerations</u>

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.

<u>Conclusion</u>

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.

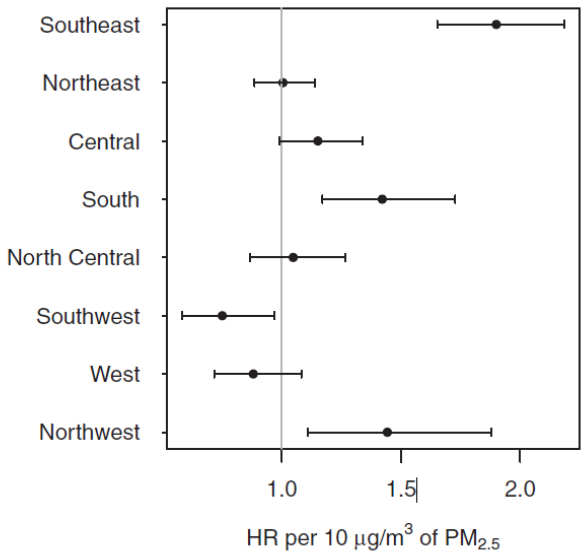


FIGURE 1. By-region $PM_{2.5}$–mortality effect estimates, presented as HRs (95% CI) per 10 μg/m^3 of $PM_{2.5}$.

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/m³ $PM_{2.5}$ (full model)
2000 $PM_{2.5}$ < 10μg/m³	100	-0.28(0.39, 0.482)
2000 $PM_{2.5}$ < 12μg/m³	186	0.50(0.27, 0.065)
2000 $PM_{2.5}$ < 14μg/m³	301	0.61(0.21, 0.004)
2000 $PM_{2.5}$ < 16μg/m³	430	0.36(0.19, 0.064)
2000 $PM_{2.5}$ < 18μg/m³	511	0.47(0.18, 0.009)
2000 $PM_{2.5}$ > 18μg/m³	34	0.85(0.82, 0.314)
2000 $PM_{2.5}$ > 16μg/m³	115	0.87(0.38, 0.023)
2000 $PM_{2.5}$ > 14μg/m³	244	0.28(0.27, 0.305)
2000 $PM_{2.5}$ > 12μg/m³	359	0.15(0.21, 0.462)
2000 $PM_{2.5}$ > 10μg/m³	445	0.27(0.18, 0.126)

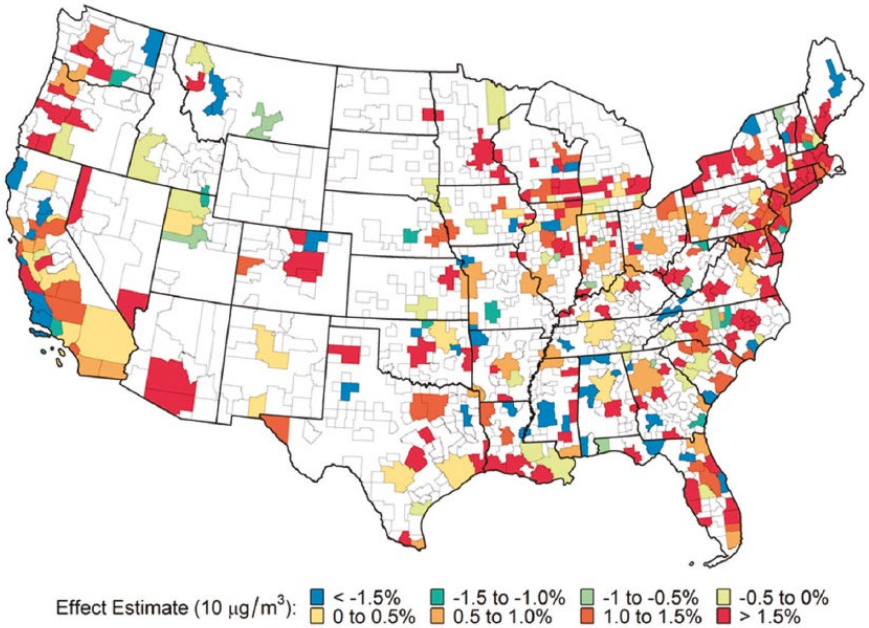
*Corresponds to the covariate pattern in Model 3 of Table 2 (main text) or Model 3 of eTable 2b. Covarites include change in income, change in population, change in high-school graduates, change in proportion of black population, change in proportion of Hispanic population, change in lung cancer mortality 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.

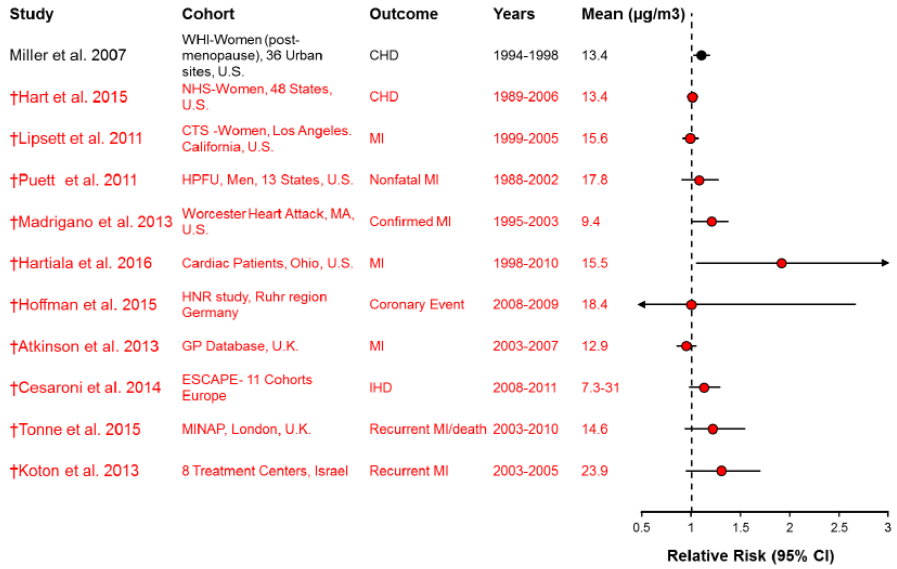
The EPA notes that there is "strong evidence for ischemic events and heart failure" which supports the conclusion of increased CVD mortality. However, there is not strong evidence of ischemic events – Figure 6-2 and 6-3 in the ISA shows mostly null and non-statistically significant associations between 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 µg/m^3. Hazard Ratios are standardized to a 5 µg/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 µg/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:

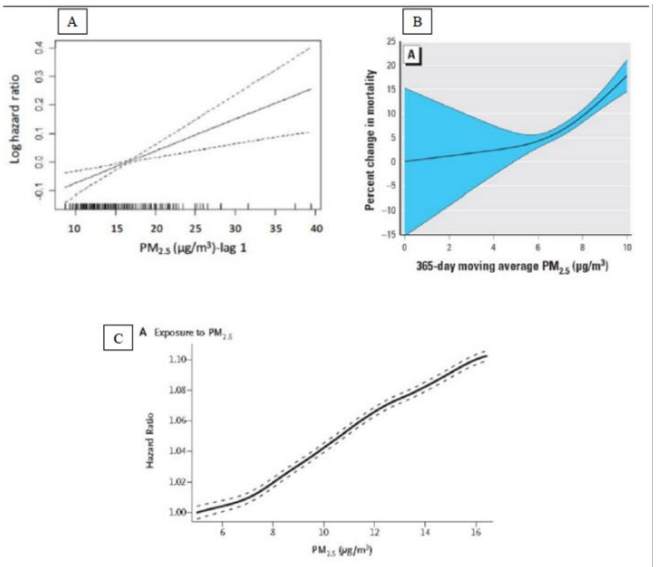
"This depends upon the degree of multi-colinearity among the variables in the regression equation. The general effect of adding too many highly correlated variables is to inflate the standard errors of the estimates rather than introduce any bias (except in certain conditions, as described in my response 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.

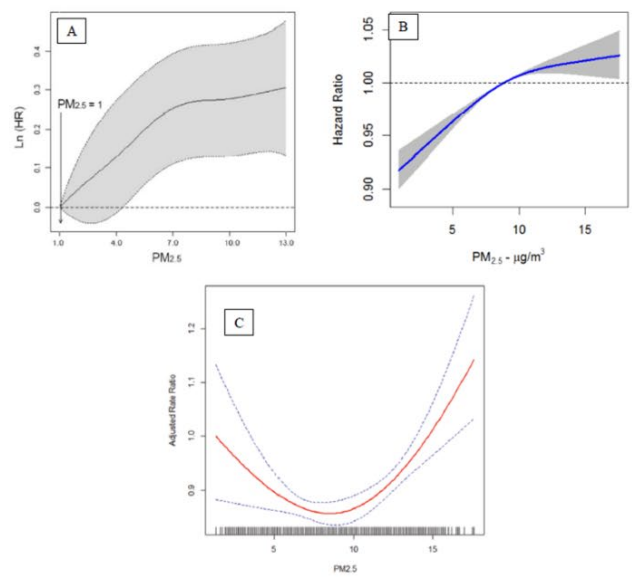
Major Point #5 – Linear No-Threshold Concentration Response

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).

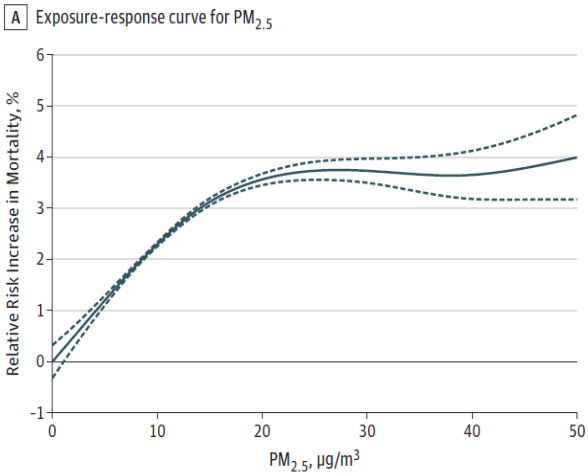
Figure 11-23 Examples of concentration-response relationships between long-term $PM_{2.5}$ exposure and total (nonaccidental) mortality in (A) nonparametric estimates; (B) in the CanCHEC cohort study; (C) the Canadian National Breast Screening Study.

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.

<u>General Conclusions from My Analysis of PM2.5 Controlled Human Exposure studies</u>

<u>Respiratory symptoms</u>: 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.

<u>Respiratory inflammation</u>: 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.

<u>Major Point #7 – PM2.5 Effect Concentrations from CHE Studies</u>

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."

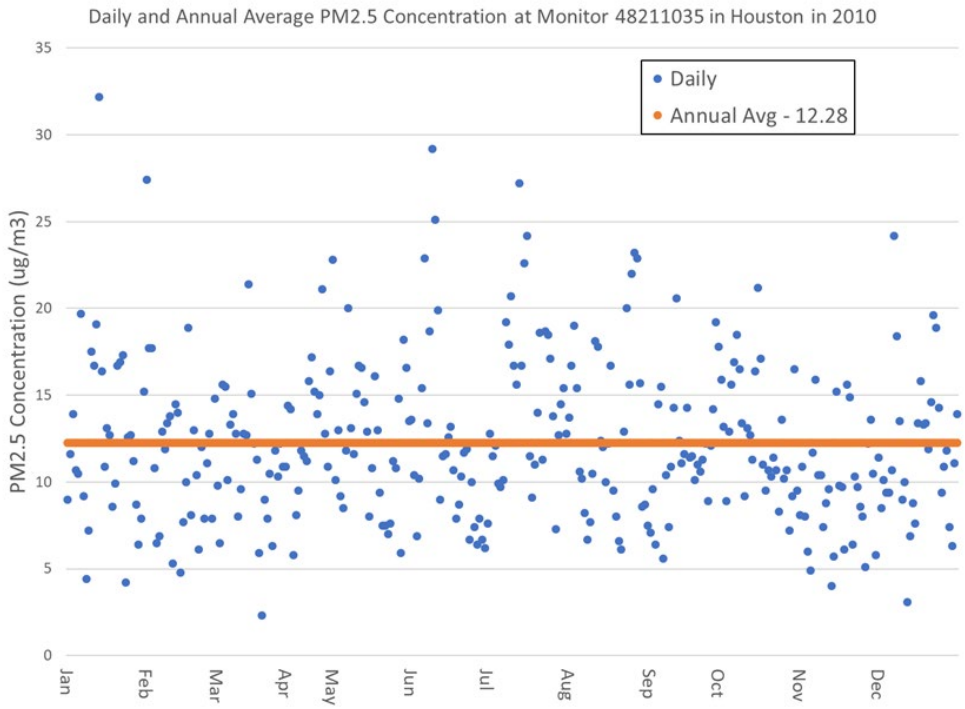
It is not clear what the EPA means by this statement. It seems that they are putting the consideration of these studies off until later reviews of the standard, which seems illogical. The Brauner papers essentially saw no effects of 22.5 hr exposure (with exercise) to ~10 µg/m3 PM2.5, but Hemmingsen 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:

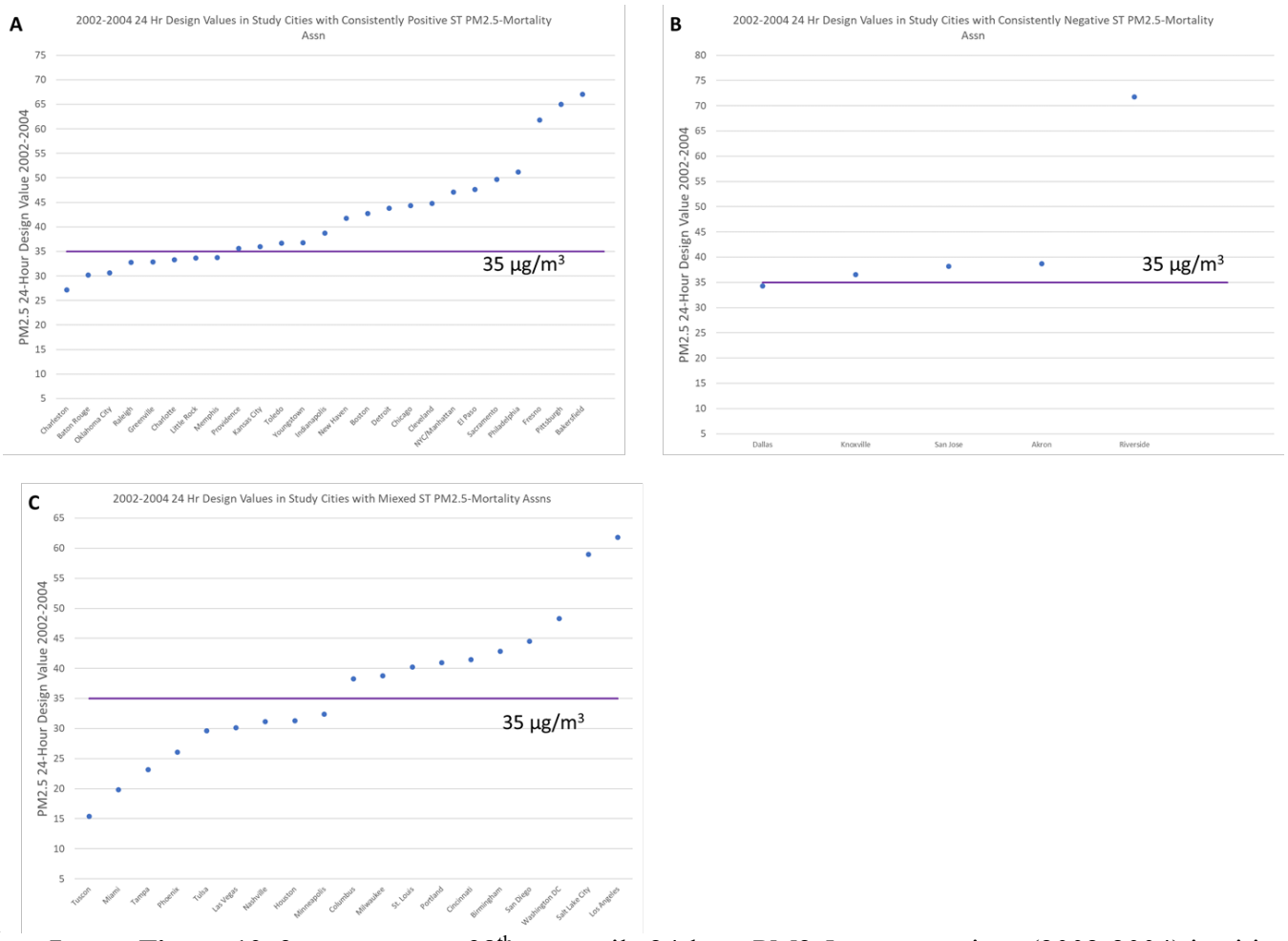
> "Yes, indeed, it is precisely the differences between study areas that is most important for estimating long-term effects. For reasons discussed above, such association effects are not strictly interpretable as causal effects of an intervention, hence the use of these "pseudo-design" values for simulating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

<u>Major Point #10: Methods for Risk Assessment</u>

<u>Choice of C-R Function:</u>

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)).

<u>Hazard Ratios and Relative Risks:</u>
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.

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

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 Δ(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.

<u>Major Point #11 – Quantitative Uncertainty Analysis in the Risk Assessment</u>

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.

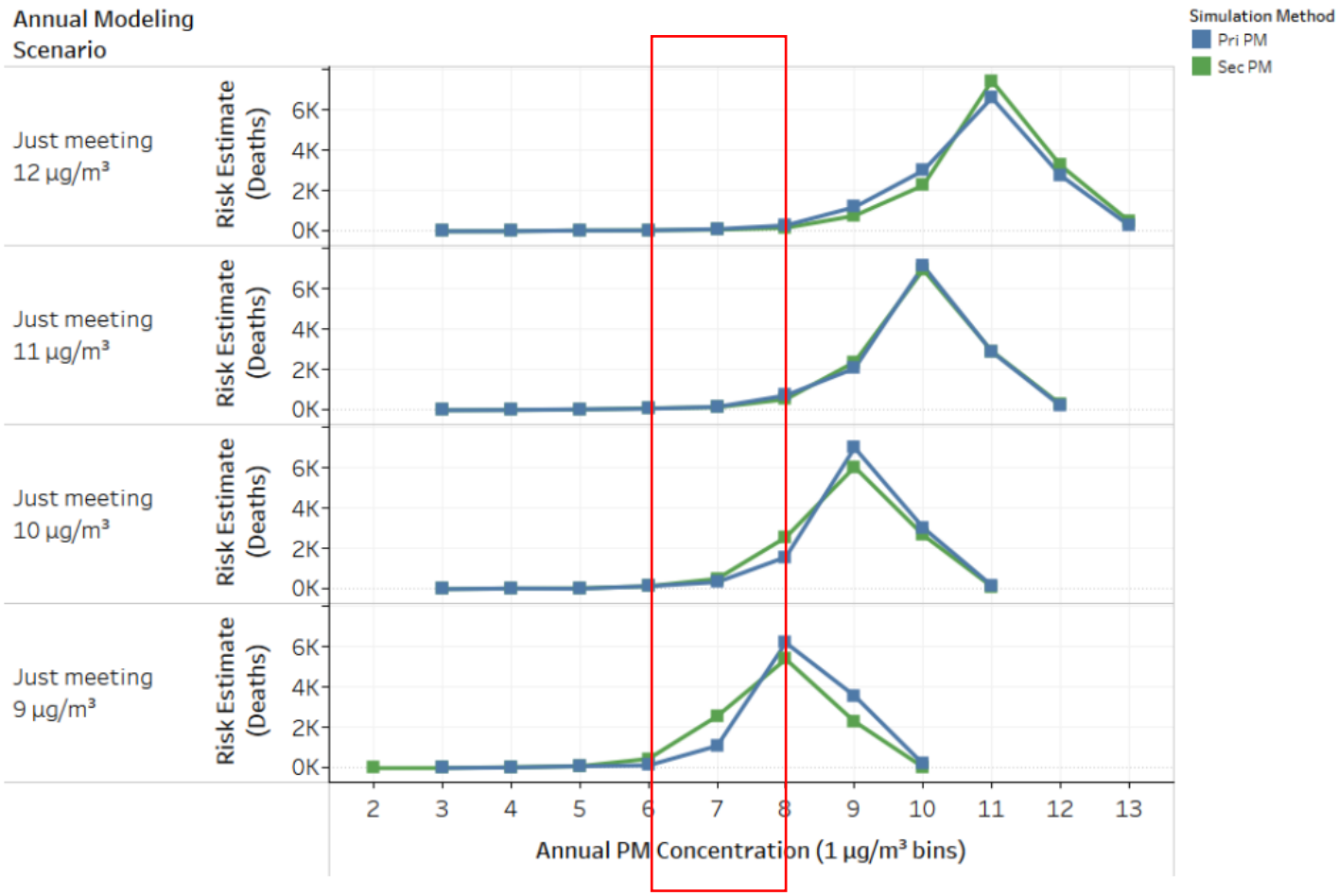


Figure 3-12. Distribution of absolute risk estimates ($PM_{2.5}$-associated mortality) for the current and alternative annual standards for the subset of 30 urban study areas where the annual standard is controlling (blue and green lines represent the Pri-$PM_{2.5}$ and Sec-$PM_{2.5}$ estimates, respectively).[68]

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

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?

3.2.1 Nature of Effects

Table 3.1 - Surprising results of no findings of **causal** instead of **likely to be casual** for respiratory health effects when cardiovascular health effects are **causal**.

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 of Risk Estimates

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 <u>*risk*</u>. (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.

Appendix C

Questions for Non-Member Consultants on the Draft PM PA from CASAC Members

Dr. James Boylan **C-2**
Dr. Tony Cox **C-4**
Dr. Mark Frampton **C-11**
Dr. Sabine Lange **C-13**
Dr. Steven Packham **C-19**

Dr. James Boylan

Chapter 2 – PM Air Quality

- Is the discussion on sources of emissions accurate and complete? If not, what additional information needs to be included?
- Is the discussion on ambient monitoring accurate and complete? If not, what additional information needs to be included?
- Is the discussion on ambient measurement correlations and trends accurate and complete? If not, what additional correlations and trends need to be included?
- To what extent are biases associated with PM_{10}, $PM_{2.5}$, and ultrafine measurements discussed? How would differing $PM_{2.5}$ biases associated with FRM vs. FEM continuous measurements (e.g., FEMs typically show higher $PM_{2.5}$ concentrations compared to FRMs) impact the evidence-based and risk-based $PM_{2.5}$ assessments in Chapter 3?
- Is the discussion on hybrid modeling approaches accurate and complete? If not, what additional information needs to be included?
- Is the discussion on performance methods for evaluating hybrid modeling methods accurate and complete? If not, what additional information needs to be included?
- Is the discussion on background concentrations accurate and complete? If not, what additional information needs to be included?

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

- Is the evidence-based analysis presented in Chapter 3 scientifically sound?
- Is the risk-based analysis presented in Chapter 3 and Appendix C scientifically sound?
- Is the discussion on following topics adequate and complete? If not, what additional information needs to be included?
    - Study area selection,
    - Health outcomes (e.g., decision to focus on mortality and ignore cardiovascular and respiratory effects),
    - Concentration-response functions,
    - $PM_{2.5}$ air quality scenarios evaluated,
    - Model-based approach to adjusting air quality,
    - Linear interpolation/extrapolation to additional annual standard levels, and
    - Characterization of variability and uncertainty in the risk estimates.
- Are the areas for additional research adequate and complete? If not, what additional areas need to be included?

Appendix C – Supplemental Information Related to the Human Health Risk Assessment

- Is the air quality modeling approach to projecting $PM_{2.5}$ concentrations to correspond to just meeting the NAAQS (AQS, CMAQ, Downscaler, SMAT-CE, project monitors to just meet NAAQS, project spatial fields to correspond to just meeting the NAAQS) scientifically sound? If not, what are your concerns and how should they be addressed?
- Is the CMAQ model configuration and input files used in the air quality modeling appropriate for this application? If not, what updates are recommended?
- Are the CMAQ model performance metrics that were evaluated appropriate and adequate for this application? Are there any concerns with the model performance in any of the study areas used in the human health risk assessment? If so, how should these concerns be addressed in the health risk assessment?
- Is the health risk modeling approach using BenMAP-CE appropriate for this application? If not, what are your concerns?
- Do the risk summary tables showing the impact of alternative $PM_{2.5}$ standards and graphical plots showing the distribution of risk across ambient $PM_{2.5}$ levels clearly and accurately summarize the results of the health risk analysis? If not, what additional information should be included?

Dr. Tony Cox

Technical Background for the Questions

Page 3-79 of the Draft Policy Assessment for PM NAAQS, dated 9-05-2019 (henceforth, the "PA") states that "Our consideration of estimated risks focuses on addressing the following policy-relevant questions:

- What are the estimated PM2.5-associated health risks for air quality just meeting the current primary PM2.5 standards?
- To what extent are risks estimated to decline when air quality is adjusted to just meet potential alternative standards with lower levels?
- What are the uncertainties and limitations in these risk estimates?"

The second of these questions – *how much would reducing PM2.5 air pollution levels reduce health risks?* – asks about the effect on health risks of reducing PM2.5 levels. This appears to me to be a question about manipulative or interventionist causality (https://plato.stanford.edu/entries/causation-mani/).

I seek your guidance on how well this causal question is answered by the techniques and models used in the PA. Specifically, the PA uses concentration-response (CR) regression models (Table C-1 of the PA) that describe the health risks associated with different levels of exposure to simulate effects on public health of changing PM2.5 standards. A key technical question is: ***Are the C-R models in Table C-1 appropriate, logically valid, and empirically well-validated, for answering the causal question of how changes in PM2.5 levels would change health risks?*** Both specific and general versions of this question are of interest:

- Are the specific regression models in Table C-1 known to produce valid predictions or simulations of how changing PM2.5 exposure would change population health effects or risks?
- In general, are such regression models appropriate for predicting how an intervention or manipulation that changes the value of a predictor (right-hand side variable) would cause the value (or frequency distribution) of the dependent variable to change? If certain conditions must hold for this to be an appropriate use of regression models, what are they, and has the PA shown that they are satisfied for the C-R regression models in Table C-1?

In more technical detail, Page C-38 of Appendix C states that "BenMAP-CE is an open-source computer program that calculates the number and economic value of air pollution related deaths and illnesses. The software incorporates a database that includes many of the concentration-response relationships, population files, and health and economic data needed to quantify these impacts. BenMAP-CE also

allows the user to import customized datasets for any of the inputs used in modeling risk. For this analysis, CR functions developed specifically for this assessment were imported into BenMAP-CE (section C.1.1). The BenMAP-CE tool estimates the number of health impacts resulting from changes in air quality—specifically, ground-level ozone and fine particles." Page C-1 explains that "our general approach to *estimating PM2.5-associated human health risks* in this review utilizes concentration-response (CR) functions obtained from epidemiology studies to link ambient PM2.5 exposure to risk in the form of incidence (counts) of specific health effects. The derivation and use of this type of CR function in modeling *PM2.5-attributable risk* is well documented both in previous PM NAAQS-related risk assessments (section 3.1.2 of U.S. EPA, 2010) and in Appendix C of the BenMAP-CE User Manual" (emphases added). In turn, the BenMAP-CE User Manual (pages C-1 to C-2, on "Deriving Health Impact Functions") states that "Epidemiological studies have used a variety of functional forms for C-R functions. Some studies have assumed that the relationship between adverse health and pollution is best described by a linear form, where *the relationship between y and PM is estimated by a linear regression* in which y is the dependent variable and PM is one of several independent variables. Log-linear regression and logistic regression are other common forms. … The relationship between Δx and Δy [change in PM2.5 concentration and change in population health response] can be derived from the CR function, as described below, and we refer to this relationship as a health impact function. Many epidemiological studies, however, do not report the C-R function, but instead report some measure of the change in the population health response associated with a specific change in the pollutant concentration. The most common measure reported is the relative risk associated with a given change in the pollutant concentration. *A general relationship between Δx and Δy can, however, be derived from the relative risk*." (Emphases added.) (www.epa.gov/sites/production/files/2015-04/documents/benmap-ce_user_manual_march_2015.pdf) Table E-1 of the BenMAP-CE User Manual ("Core Health Impact Functions for Particulate Matter and Long-Term Mortality") presents various expert estimates for "beta," which is defined and described as follows: "The coefficient for the health impact function. The value of beta (ß) typically represents the percent change in a given adverse health impact per unit of pollution. … Odds Ratios must be converted to beta coefficients to be used in BenMAP-CE." The notes on many of these beta values state "No causality included." Table C-1 of the PA ("Details regarding selection of epidemiology studies and specification of concentration-response functions for the risk assessment") presents the beta values used in the PA simulations, but has no column of notes indicating whether or how any of them addresses causality.

Questions

1. *Are the beta coefficients in Table C-1 of the PA conceptually well defined?* That is, are their intended conceptual meanings and causal interpretations clear and unambiguous? If so, can the definition of beta be expressed using standard epidemiological terms (e.g., controlled direct effects, natural direct effects, total effects, indirect effects, and so forth?) (Petersen ML, Sinisi SE, van der Laan MJ. (2006) Estimation of direct causal effects. Epidemiology. May; 17(3):276-84; Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992, 3:143-155;

Tchetgen Tchetgen EJ, Phiri K. Bounds for pure direct effect. Epidemiology. 2014 Sep;25(5):775-6. doi: 10.1097/EDE.0000000000000154; VanderWeele TJ. Controlled direct and mediated effects: definition, identification and bounds. Scand Stat Theory Appl. 2011 Sep;38(3):551-563; Vansteelandt, Stijn; Bekaert, Maarten; Lange, Theis (2012). Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiologic Methods*. **1** (1, Article 7).)

2. On the same topic of clear definitions, does the discussion of the BenMAP-CE beta coefficients in the PA and underlying documentation (described as typically representing "the percent change in a given adverse health impact per unit of pollution") unambiguously specify which of the following concepts the coefficients represent?
    a. Beta estimates the percent change in the conditional expected *observed* value of the health impact *associated with* a unit change in the *observed* value of the pollution variable. (This might be called the *regression interpretation* of the beta values.)
    b. Beta estimates the percent change in the mean value of the health impact variable *caused by* a manipulation or intervention that changes the value of the pollution variable by 1 unit (e.g., increasing concentrations at all times and locations by 1 microgram per cubic meter above what they otherwise would have been), while *holding the values of all other variables fixed* at the values they had before the intervention. (As concrete examples, the values of lagged daily high and low temperatures, humidity, and co-morbidities such as asthma in the weeks before a death would not be affected by the hypothetical (or counterfactual) change in the pollution variable for a beta coefficient that addresses the health impact on mortality risk of a change in pollution.) (This might be called the *natural direct effect* interpretation of beta.)
    c. Beta estimates the percent change in the mean value of the health impact *caused by* a manipulation or intervention that changes the value of the pollution variable by 1 unit (e.g., increasing concentrations at all times and locations by 1 microgram per cubic meter above what they otherwise would have been), while *allowing the values all other variables to change* in response to the changes in pollution. (As concrete examples, the joint distribution of the values of lagged daily high and low temperatures, humidity, and co-morbidities such as asthma could be changed by conditioning on counterfactual changes in the pollution variable.) (This might be called the *total direct effect* interpretation of beta.)
    d. Beta estimates the percent change in the mean value of the health impact *caused by* a manipulation or intervention that changes the value of the pollution variable by 1 unit (e.g., increasing concentrations at all times and locations by 1 microgram per cubic meter above what they otherwise would have been), while *holding the values of all other variables (e.g., co-exposures, co-morbidities, potential confounders or modifiers such as weather variables) fixed at the values they would be expected to have after the intervention.*
    e. Beta means something different from any of the above (e.g., a controlled direct effect).

Do all of the beta values in Table C-1 refer to the same one of these concepts, or might they refer to different ones (or is the answer not clear)?

3. Similarly, *is the definition of "concentration-response (C-R) relationships" in the PA and its Appendices (cf p. C-38) adequately clear and unambiguous* to support simulation of well-defined causal effects of interventions that change pollution levels? For example, is it clear whether the term "C-R relationship" in the PA refers to natural direct effects of PM2.5 on mortality and other health outcomes, or to total effects, controlled direct effects, effects mediated by PM2.5, or simply slopes of estimated regression lines (associations), or perhaps something else?
4. The PA and BenMAP-CE documentation repeatedly refers to relative risks, odds ratios, attributable risks, and regression coefficients as bases for quantifying causal effects of changes in air pollution on changes in population health responses. Some epidemiology methodologists and experts in causal analysis have distinguished between association and causation (and between "seeing" and "doing") and have warned that relative risks, odds ratios, attributable risks, regression coefficients, and related concepts (e.g., population attributable fractions, probabilities of causation, etiologic fractions) address associations but not causation (including causal effects of interventions), because measures of statistical association do not address how changing one variable would change another (e.g., Pearl J, 2009 Causal inference in statistics: An overview. Statistics Surveys 3: 96-146; Greenland and Robins 2000, Epidemiology, justice, and the probability of causation; Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol. 2002 Apr;31(2):422-9.) *Do the beta coefficients in the PA overcome these methodological objections to using relative risks, regression coefficients, and related measures of association to predict (or simulate) effects of interventions?* If so, how were they overcome? If not, does this imply that the simulations in the PA are not necessarily reliable or valid predictors of the real-world effects on public health of reducing PM2.5? Why or why not?
5. On the same methodological issue, *which, if any, of these measures (relative risks, odds ratios, attributable risks, and regression coefficients) are currently generally accepted in contemporary causal analysis and epidemiology as valid measures of how changing one variable (e.g., exposure) will cause another (e.g., population health responses) to change?* Please provide specific references if available, and identify whether any of these quantities are generally accepted now as valid measures of controlled direct, natural direct, indirect, total, or mediated causal effects.
6. More generally, do the $\Delta y/\Delta x$ values calculated by BenMAP-CE have valid interpretations as *causal impacts* on y of *interventions* that change x? (If x represents daily ice cream consumption in a population and y represents daily cases of heat stroke, would the slope of an estimated regression line or C-R line relating them, β or $\Delta y/\Delta x$, necessarily provide valid predictions or simulations for how an intervention that changes ice cream consumption by 1 unit would change daily cases of heat stroke? Assuming the answer is no, how is the methodology of beta coefficients in Appendix C of the PA essentially different from this example?)

7. *Have the beta coefficients in the PA been empirically validated* as providing approximately correct or usefully accurate predictions or simulations of how interventions that change air pollution would change population health responses? A recent non-EPA review of the empirical evidence of health effects of interventions to reduce ambient PM air pollution did not conclude that there is strong evidence leading to confident rejection of the null hypothesis of no effect: rather, the authors concluded that "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 J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess E. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database of Systematic Reviews 2019, Issue 5. Art. No.: CD010919. DOI: 10.1002/14651858.CD010919.pub2.) To what extent are such findings from actual interventions consistent with the assumptions, confidence intervals, and simulations in the PA?
8. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately describe and discuss the extent (if any) to which their values reflect *potential omitted confounders* of the association between mortality risks and PM2.5 levels (e.g., lagged daily high and low temperatures and humidity in the weeks preceding mortality, if these contribute both to (possibly delayed) mortality and increased energy usage and PM2.5 pollution?)
9. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address the extent (if any) to which their values reflect *residual confounding* of the association between mortality risks and PM2.5 levels (e.g., by daily high and low temperatures and humidity in the weeks preceding mortality in models that only address seasonal, annual, or averaged temperatures)?
10. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address *model uncertainty* (e.g., the possibility that the linear no-threshold model specification is incorrect, e.g., because sufficiently low exposure concentrations do not cause pulmonary inflammation and adverse health effects that occur at higher concentrations)?
11. Does the PA's discussion of beta values, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address *effects on the estimated values of the beta values of exposure uncertainties and estimation errors* (e.g., the possibility that individual exposure concentrations among people with adverse health responses tend to be higher than those among people who did not respond, even when they have the same estimated exposure values)?
12. Does the PA adequately assess *the suitability of the designs of the studies used to estimate beta values for purposes of valid causal inference and simulation? Does the PA's discussion of uncertainty and sensitivity analyses adequately address the internal validity and external validity (generalizability) of the estimated beta values* used to simulate the causal impacts on public health risks of changing PM2.5 levels? (Campbell DT, Stanley JC (1963) *Experimental and Quasi-Experimental Designs For Research*.

Houghton Mifflin. Boston, MA www.sfu.ca/~palys/Campbell&Stanley-1959-Exptl&QuasiExptlDesignsForResearch.pdf)

13. Does the PA's discussion of beta values adequately address *attribution of risk in the presence of joint causes?* For example, if a unit change in PM2.5 levels has different expected effects on mortality risk for a person below the poverty line and during extremely hot or cold weather than it would for an initially similar (exchangeable) person with higher income and no exposure to extreme temperatures, then how much of the statistical "effect" of PM2.5 on mortality risk, as reflected in the beta values in Table C-1, should be attributed to income and weather variables? Conversely, how well does the discussion in the PA make clear how much of the estimated beta value for PM2.5 is actually contributed by other variables (such as temperature extremes and poverty) that would not necessarily be changed by an intervention that reduces PM.5 levels?
14. Overall, does the PA and its underlying documents (e.g., the BenMAP-CE documentation) make a convincing technical case that its simulated health impacts of reductions in PM2.5 are trustworthy and usefully accurate? *How confident can policy analysts and decision makers be in the predictive validity of the simulated results?*
15. Are there other statistical or methodological issues that you would like to comment on that you believe might help the CASAC to assess the validity and soundness of the PA and its simulations for effects on health risks of changing PM2.5 levels, or that might help to improve the technical and scientific quality of the final PA?
16. How can techniques of formal causal modeling and analysis best be applied to improve the clarity of definitions and communication and scientific soundness of simulations, inferences, causal interpretations, generalizations, and policy-relevant conclusions in the PA? Please comment on whether any aspects of the following (or other) causal model formalisms can substantially improve the clarity and scientific soundness of the analyses and simulations in the PA: causal graph and DAG methods, conditional independence tests, intervention and interrupted time series analyses, other quasi-experimental methods, Wiener-Granger causality and transfer entropy, causal dynamic Bayesian networks (DBNs), other information-theoretic and graph methods, Simon-Iwasaki causal ordering , non-parametric structural equations models, mediation analysis.

Supplementary References for Causal Determination Questions from Dr. Tony Cox

Consultants, since several of the CASAC questions are about causal concepts that Dr. Cox has previously referred to in CASAC work, he has asked that the following references be provided:

Cox, L. A., Jr. (2019). Improving causal determination. Global Epidemiology, 1, 100004. https://doi.org/10.1016/j.gloepi.2019.100004

Cox, L. A., Jr. (2019). Communicating more clearly about deaths caused by air pollution. Global Epidemiology, 1, 100003. https://doi.org/10.1016/j.gloepi.2019.100003

Cox, L. A., Jr. (2019). Shapes and definitions of exposure-response curves: A comment on "A matrix for bridging the epidemiology and risk assessment gap." Global Epidemiology, 1, 100006. https://doi.org/10.1016/j.gloepi.2019.100006

Cox, L. A., Jr. (2018). Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. Critical Reviews in Toxicology, 48(8), 682–712. https://doi.org/10.1080/10408444.2018.1518404

Dr. Mark Frampton

1. The EPA concluded in the 2009 PM ISA, and reconfirmed in the 2018 ISA, that the scientific evidence is sufficient to conclude that the relationship between $PM_{2.5}$ exposure and mortality is causal. Previous CASACs have concurred with this conclusion. This finding, among other causality determinations expressed in the ISA, forms a key basis for the risk assessments presented in the current PA. However, the current CASAC was not able to reach agreement on this important issue, as indicated in the letter to the Administrator (April 11, 2019). Some CASAC members are of the opinion that the scientific evidence base is insufficient to support a causal relationship between PM exposure and mortality. Is there evidence that would support a reconsideration of the current and long-held views of this causal relationship as expressed in the PA and ISA?

2. In the long-term epidemiological studies of PM mortality, heterogeneity between and within cities has been cited as a source of uncertainty in drawing conclusions about causality. Please opine on the level of uncertainty that is represented by this heterogeneity, and the impact if any, on the conclusions in the PA.

3. Dr. Cox, Chair of CASAC, has introduced concepts of causality determination that differ from the framework established by the EPA in prior ISAs, used in the current PM ISA, and reiterated in section 3.2 of the current PA. This was a topic of discussion during CASAC's recent review of the PM ISA. Please opine on the adequacy of the causality analysis framework currently used by the EPA, and whether and how the concepts espoused by Dr. Cox should, or should not, be incorporated into the NAAQS causality framework. Also please comment on the implications, of any changes in the causality framework that you would recommend, for the analyses and conclusions in this current PA. For background, see the following documents:

1) Preamble to the Integrated Science Assessments, November 2015, pages 18 to 25 (https://yosemite.epa.gov/sab/sabproduct.nsf/78476291901FF5AE8525835F00634158/$File/ISA_PREAMBLE_FINAL2015.pdf);
2) Follow-up Questions for John Vandenberg from Dr. Cox, letter of 12/17/2018 (https://yosemite.epa.gov/sab/sabproduct.nsf/B28289529495F929852583660057A3EE/$File/Follow-up+questions+for+John+Vandenberg-rev.pdf);
3) Responses from John Vandenberg, letter of 2/20/2019 (https://yosemite.epa.gov/sab/sabproduct.nsf/B48131F413362439852583A7005FDFDA/$File/JVandenberg+response+to+TCox+ltr+of+121718.pdf);
4) CASAC letter to the Administrator, April 11, 2019 - Dr. Cox comments, pages A8 to A27 (https://yosemite.epa.gov/sab/sabproduct.nsf/264cb1227d55e02c85257402007446a4/6CBCBBC3025E13B4852583D90047B352/$File/EPA-CASAC-19-002+.pdf).

4. The CASAC letter to the Administrator (April 11, 2019) states, "There is inadequate evidence for the 'likely to be causal' conclusion for long-term $PM_{2.5}$ exposure and cancer." This is based on

epidemiological studies that do not appear to adequately differentiate incident cancer and cancer-related mortality because the exposure time frames for most of these studies are insufficient to draw conclusions about incident cancer. Do you agree with the CASAC's findings in this matter? Please discuss the evidence (or lack thereof) that supports your opinion.

5. Please comment on the appropriateness and completeness of the approaches used in this PA to assess the risks of exposure, and the assessments of risk reduction of alternate standards, for $PM_{2.5}$ and PM_{10} (sections 3 and 4, respectively).

6. Are there additional key studies that should be considered in sections 3 and 4 of the PA?

Dr. Sabine Lange

Ambient PM2.5 Concentrations in Locations of Epidemiology Studies

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 (Figure 1). This information helps to inform some of the questions that I ask below.

A

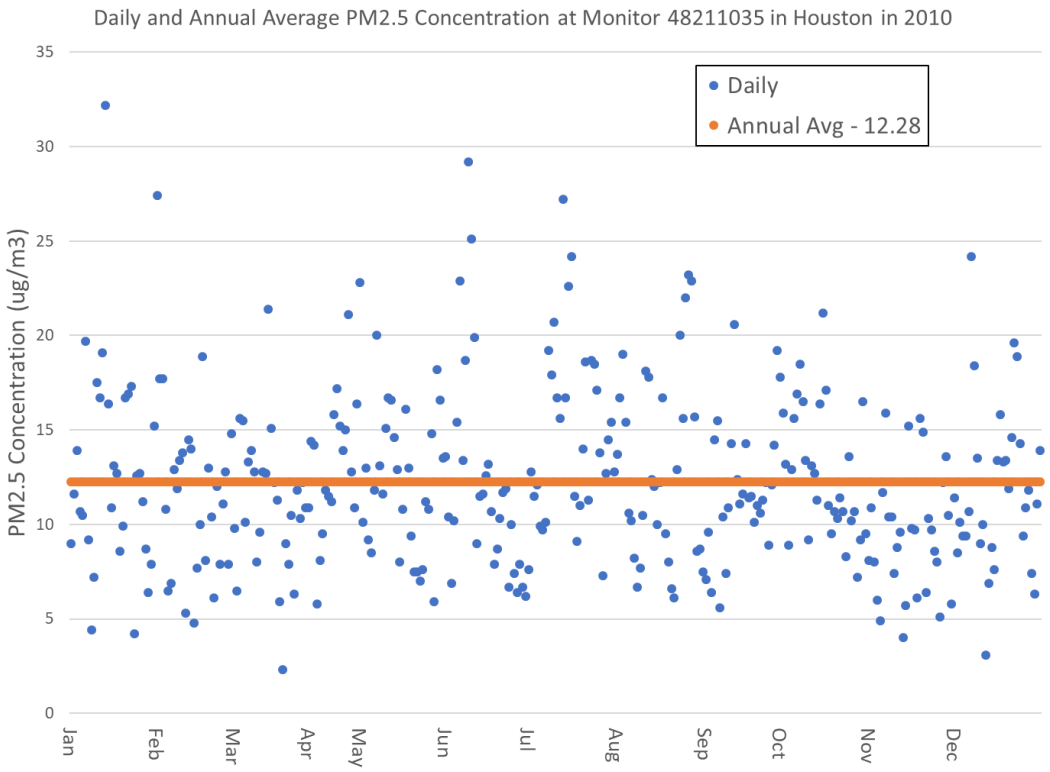


Figure 1. Daily and annual average PM2.5 concentrations measured at monitor 48211035 in Houston, Texas in 2010.

While it seems appropriate to compare the study mean PM2.5 concentration from a long-term exposure to the annual design value (both are on similar time scales), my question is:

1) **Is it appropriate to compare daily PM2.5 concentrations to the annual average?** 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 Figure 1).

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 are:

2) **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.
3) 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?**

Results of Risk Analysis

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. However, many more uncertainties are assessed 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.

4) **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**?

Potential Alternative Standards

In section 3.4.2.1 the EPA discusses whether the available data warrant a different indicator for the PM2.5 standard. This determination 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 are:

5) **Could different magnitudes of error amongst different variables in regression analyses be masking the effect of a speciated constituent of PM2.5**?
6) **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?**

References

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.

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.

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.

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.

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.

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.

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

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.

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.

Dr. Steven Packham

Introduction

The purpose of the NAAQS is to protect public health with an adequate margin of safety. The CASAC is required to review the scientific evidence and knowledge used to establish criteria for pollutants regulated by the NAAQS and to advise the Administrator on whether attainment and maintenance of existing, new and revised NAAQS- allowing [for] an adequate margin of safety - are requisite to protect the public health.

The Overarching context for CASAC Reviews of NAAQS

In the transmittal and charge question memorandum from Director Sasser,[1] reference is made to the two sections of the Clean Air Act governing the establishment and revision of the NAAQS. Director Sasser notes that, since the early 1980s, the independent review function has been performed by the CASAC of the EPA's Science Advisory Board.

A number of other advisory functions are also identified for the committee by section 109(d)(2)(C), which reads:

> Such committee shall also (i) advise the Administrator of areas in which additional knowledge is required to appraise the adequacy and basis of existing, new, or revised national ambient air quality standards, (ii) describe the research efforts necessary to provide the required information, (iii) advise the Administrator on the relative contribution to air pollution concentrations of natural as well as anthropogenic activity, and (iv) advise the Administrator of any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards.

The Historical Context

The NAAQS review process has evolved over time. Important insights into the role of the CASAC can be gained from studying its history. A *Review of the Process for Setting the National Ambient Air Quality Standards*, published in 2006, memorializes the beginning of official and substantive changes in the nature of CASAC reviews.

[1]Erika N. Sasser, Director, Health and Environmental Impacts Division, Office of Air Quality Planning and Standards, United States Environmental Protection Agency

Currently, five documents (the Integrated Review Plan (IRP), the Preamble to the Integrated Science Assessment (ISA Preamble), the ISA (the Integrated Science Assessment document intended to replace former Criteria Documents), and the Human Exposure Risk Assessment (HERA) and lastly the Policy Assessment (PA)) set the scope of scientific knowledge and evidence and the rather prescriptive review process for the present CASAC.

Imagine a flow of hundreds of pre-selected scientific studies moving through the IRP-ISA Preamble-ISA-ERA-PA review pipeline in which the CASAC reviews are essentially limited to responding to Charge Questions such as those given from page 3 of the aforementioned memo.

Specific Charge Questions for Review of the Draft PA

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?
2. **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?
3. **Chapter 3 – Review of the Primary PM2.5 Standards:** 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?
4. **Chapter 4 – Review of the Primary PM10 Standard:** 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?
5. **Chapter 5 – Review of the Secondary Standards:** What are the CASAC views on the approach described in chapter 5 to considering the evidence for PM-related welfare effects in order to inform preliminary conclusions on the secondary standards? What are the CASAC views regarding the rationale supporting the preliminary conclusions on the current secondary PM standards?
6. **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?

Consultants, you will appreciate that the **Draft PM PA** is a draft of the last-and-final document to be reviewed by the CASAC before a final report to the Administrator is prepared on the adequacy of the existing PM NAAQS. With this in mind, please consider offering answers, thoughts or suggestions prompted by some, if not all, of the following questions:

- Are there areas (e.g., specific aspects of biological causation, pulmonary toxicology, or causality in epidemiology) in which additional knowledge is required to appraise the adequacy and basis of existing, new, or revised PM NAAQS to protect public health with an adequate margin of safety?
- Can you suggest additional specific scientific disciplines and areas of biomedical informatics and research (e.g., systems biology methods for clarifying biological causal pathways, mechanisms, modes of action, quantitative causal dose-response relationships) that should be included in future reviews of other criteria pollutants?
- Can you describe the research efforts necessary to provide the required information? To what extent has the needed research already been done, or started? For example, are there crucial experiments or research initiatives that could clarify the shape of the PM2.5-chronic inflammation causal dose-response relationships at relevant exposure concentrations? Are there specific data analyses (e.g., testing for confounding by weather variables over more days prior to mortality) that could clarify the causal interpretation of epidemiological associations relied on in the draft PA to simulate effects of interventions?
- Can you suggest additional areas of scientific literature review on species and individual human organism's capacities of adaptation to inhaled environmental stressors that might help establish margins of safety when exposed to ambient levels of air pollution?
- Could you provide additional information on the relative contributions to air pollution concentrations and resulting health effects of natural and anthropogenic activity? For example, is either one alone, or are both natural and anthropogenic activities together, sufficient to cause the magnitudes of adverse health effects attributed to PM2.5 in the Draft PA?
- Could you provide additional information on any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards?

Again, thank you for your willingness to share your time and expertise with the CASAC.

Appendix D

Responses to CASAC Member Questions on the Draft PM PA from Non-Member Consultants

Dr. Constantin Aliferis **D-2**
Dr. Dan Jaffe **D-14**
Mr. John J. Jansen **D-19**
Dr. Frederick Lipfert **D-24**
Dr. D. Warner North **D-70**
Dr. David Parrish **D-85**
Dr. Lorenz Rhomberg **D-89**
Dr. Sonja Sax **D-95**
Dr. Duncan Thomas **D-119**

Dr. Constantin Aliferis

Questions from Dr. Cox

Question 2*: On the same topic of clear definitions, does the discussion of the BenMAP-CE beta coefficients in the PA and underlying documentation (described as typically representing "the percent change in a given adverse health impact per unit of pollution") unambiguously specify which of the following concepts the coefficients represent?*

a. Beta estimates the percent change in the conditional expected observed value of the health impact associated with a unit change in the observed value of the pollution variable. (This might be called the regression interpretation of the beta values.)

b. Beta estimates the percent change in the mean value of the health impact variable caused by a manipulation or intervention that changes the value of the pollution variable by 1 unit (e.g., increasing concentrations at all times and locations by 1 microgram per cubic meter above what they otherwise would have been), while holding the values of all other variables fixed at the values they had before the intervention. (As concrete examples, the values of lagged daily high and low temperatures, humidity, and co-morbidities such as asthma in the weeks before a death would not be affected by the hypothetical (or counterfactual) change in the pollution variable for a beta coefficient that addresses the health impact on mortality risk of a change in pollution.) (This might be called the natural direct effect interpretation of beta.)

c. Beta estimates the percent change in the mean value of the health impact caused by a manipulation or intervention that changes the value of the pollution variable by 1 unit (e.g., increasing concentrations at all times and locations by 1 microgram per cubic meter above what they otherwise would have been), while allowing the values all other variables to change in response to the changes in pollution. (As concrete examples, the joint distribution of the values of lagged daily high and low temperatures, humidity, and co-morbidities such as asthma could be changed by conditioning on counterfactual changes in the pollution variable.) (This might be called the total direct effect interpretation of beta.)

d. Beta estimates the percent change in the mean value of the health impact caused by a manipulation or intervention that changes the value of the pollution variable by 1 unit (e.g., increasing concentrations at all times and locations by 1 microgram per cubic meter above what they otherwise would have been), while holding the values of all other variables (e.g., co-exposures, co-morbidities, potential confounders or modifiers such as weather variables) fixed at the values they would be expected to have after the intervention.

e. Beta means something different from any of the above (e.g., a controlled direct effect).

Do all of the beta values in Table C-1 refer to the same one of these concepts, or might they refer to different ones (or is the answer not clear)?

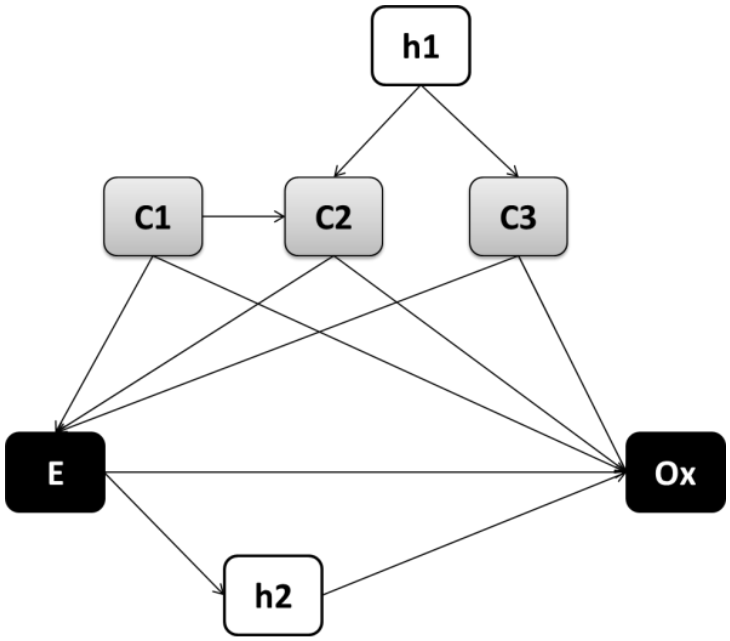
Response:

Providing causal interpretations to results of regression and other correlative studies in the absence of specific causal assumptions, constraints or background information, is tricky if not outright infeasible. Here how I read such results in the scientific and technical literature:

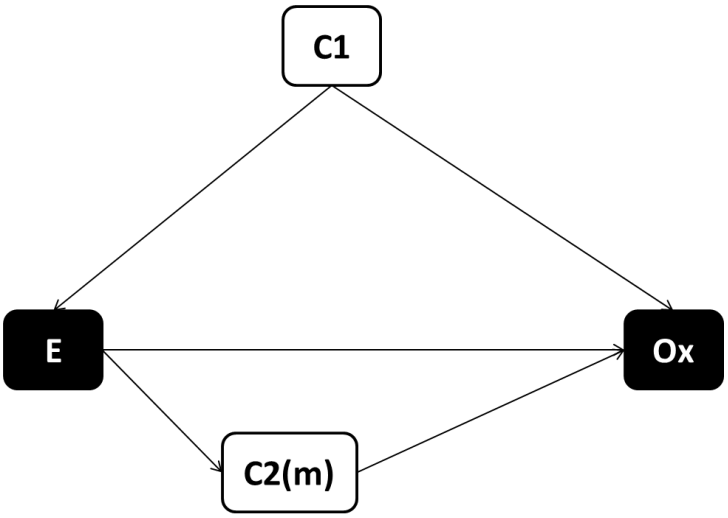
a: Yes, provided that we qualify by stating that all other covariate's values are held constant and that this beta value is strictly model-specific (i.e., lacks any invariant/inherent biomedical meaning in the sense that changing the model will change the beta).

b, c, d, e: Unclear - can be either yes or no depending on assumptions about the nature of the covariates inside and outside the model.

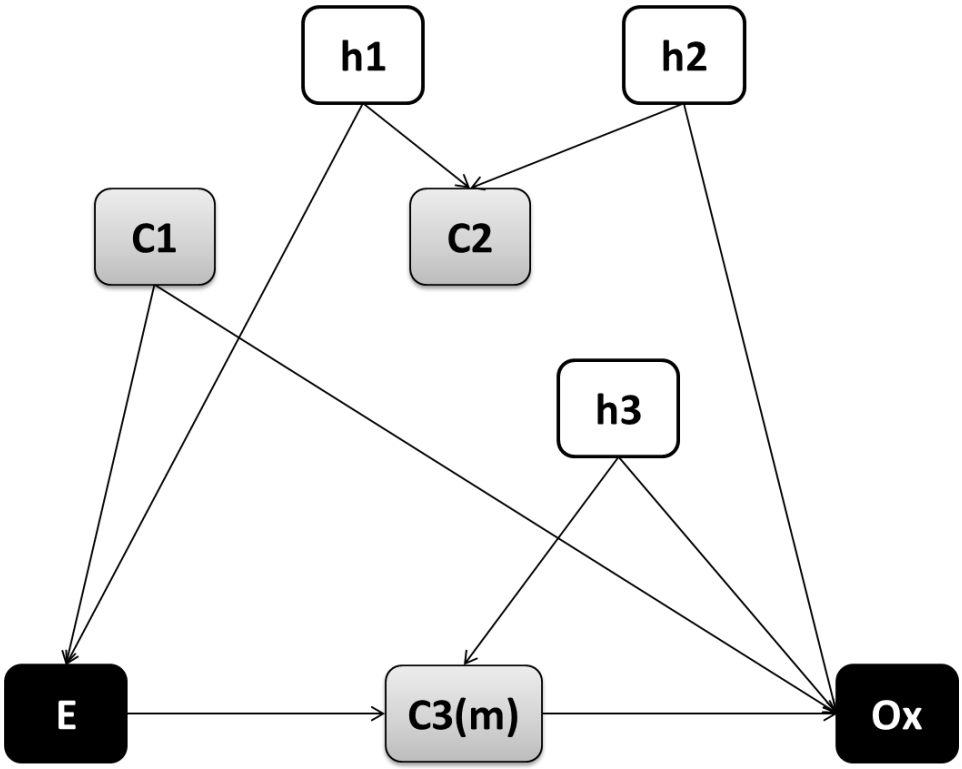
Understanding the last statement is key in clearing, in my opinion, much of the causal confusion in this and similar settings. Let's consider the following three examples (statements made hold under standard assumptions of causal modeling and for the vast majority of (i.e., Faithfull) distributions):



Example 1: this figure depicts an example true causal structure underlying some data. E= exposure, Ox=outcome, "C" variables are measured covariates, "h" variables are latent (unmeasured). If the regression model incorporates C1, C2, C3 as covariates, then the beta of E corresponds to the <u>total causal effect</u> of manipulating E (ie via both paths E→Ox and E→h2→ Ox) on Ox. This interpretation requires that no other unmeasured confounder exists that is not blocked by a measured covariate in the model. h2 is taken to change as a result of manipulating E. h1, C1, C2, C3 have their natural joint distribution.



Example 2: As before but I have marked C2 with "m" to point that it is mediating part of the causal influence of E on Ox. Here the beta of E corresponds to the direct causal effect of manipulating E (ie via path E→Ox only). This interpretation requires that no unmeasured confounder exists that is not blocked by a measured covariate in the model. C2 is taken to not change as a result of manipulating E. C1 has its natural joint distribution.



Example 3: Here the regression model incorporates C1, C2, C3(m) as covariates and the beta of E does not have a proper causal meaning of any kind. Specifically the indirect effect of E via path e→C3(m)→Ox is blocked by C3(m) in the model but the (confounding) path E→C3(m)←h3→Ox opens; the confounding path E←C1→Ox is blocked; using C2 as covariate opens a second biasing path via latent variables h1, h2 (E←h1→C2←h2→Ox). Hence <u>no unbiased estimate of causal effect of E on Ox is provided, either direct or indirect.</u> This analysis assumes that no other unmeasured confounder exists that will mitigate the biases (such an event would be extremely unlikely anyway). C3(m) is taken to not change as a result of manipulating E. h1, h2, h3, C1, have their natural joint distribution.

<u>These examples demonstrate a number of important facts of general significance:</u>

- For the regression analysis to produce a valid causal estimate (of any kind: total, direct, indirect, or other conceivable variants not usually discussed in the epidemiologic literature), a number of causal constraints (or assumptions) involving the measured and unmeasured variables must hold.

- These assumptions involve the following questions:
    - Are common causes of E and Ox sufficient for blocking all confounding, measured?

- Are paths involving common causes of E, Ox blocked or opened (or both) by measured variables?
- Are mediators of E to Px measured?
- Are common causes of E and mediators or mediators and Ox measured?

- If and only if we can guarantee the correct answers to these questions we can produce reliable causal effect estimates for E.

- On the basis of this information we can also decide which measured variables can and should be covariates. For example in the graph 2, we can elect to omit C2(m) from the model in order to estimate total causal effect of E on Ox, or to include it, in order to obtain the direct causal effect only.

- It does not matter whether we use Poisson regression, Cox regression, Logistic Regression, GLM etc in the sense that they all depend on the right choice of covariates (and are easily chosen on the basis of distribution and research design) *.

- As a matter of fact, 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.

- In addition, the causal discovery algorithms reveal which effects can be estimated without experiments and which cannot. For example, inference of the structure E←h→Ox and E→Ox informs us that we cannot estimate the causal effect of E and it should not be attempted because it cannot be done without experiments and/or additional assumptions.

- *Caution however: models that do not use conditioning (eg many machine learning and statistical models that use regularization, will systematically produce wrong causal estimates even with the right set of controlling covariates).

Question 4*: The PA and BenMAP-CE documentation repeatedly refers to relative risks, odds ratios, attributable risks, and regression coefficients as bases for quantifying causal effects of changes in air pollution on changes in population health responses. Some epidemiology methodologists and experts in causal analysis have distinguished between association and causation (and between "seeing" and "doing") and have warned that relative risks, odds ratios, attributable risks, regression coefficients, and related concepts (e.g., population attributable fractions, probabilities of causation, etiologic fractions) address associations but not causation (including causal effects of interventions), because measures of statistical association do not address how changing one variable would change another (e.g., Pearl J, 2009 Causal inference in statistics: An overview. Statistics Surveys 3: 96-146; Greenland and Robins*

2000, Epidemiology, justice, and the probability of causation; Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol. 2002 Apr;31(2):422-9.)

Do the beta coefficients in the PA overcome these methodological objections to using relative risks, regression coefficients, and related measures of association to predict (or simulate) effects of interventions? If so, how were they overcome? If not, does this imply that the simulations in the PA are not necessarily reliable or valid predictors of the real-world effects on public health of reducing PM2.5? Why or why not?

Response:

The previous analysis (question #2) applies to answer this question. In brief:

- 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.
- Stronger (less prone to error) approaches include, experiments and/or consistent causal discovery algorithms.
- Weaker (more prone to error) approaches include intuition, ad hoc synthesis of literature, application of "causal criteria" (Hill, Koch, Surgeon General etc), using non-causal predictive modeling to identify covariates, using clustering, etc.

Question 6*: More generally, do the Δy/Δx values calculated by BenMAP-CE have valid interpretations as causal impacts on y of interventions that change x? (If x represents daily ice cream consumption in a population and y represents daily cases of heat stroke, would the slope of an estimated regression line or C-R line relating them, β or Δy/Δx, necessarily provide valid predictions or simulations for how an intervention that changes ice cream consumption by 1 unit would change daily cases of heat stroke? Assuming the answer is no, how is the methodology of beta coefficients in Appendix C of the PA essentially different from this example?)*

Response:

The predictions would be causally valid, for a total causal effect for instance, if and only if a set of covariates was included in the model that would eliminate all confounding (by blocking all relevant paths) between ice cream consumption and heat stroke due to common causes of the two.

Essentially then the appraisal work done by the causal assessment methodology amounts, at least implicitly, to ensuring that this condition is likely enforced. It would be in my opinion much better if

stronger methods for enforcing such constraints that would ensure causal interpretation, were put in practice. Having said as much, I would not venture to dismiss the coefficients as uninformative or entirely unreliable, however (please see my response to Dr Frampton below for more details).

Question 16*: How can techniques of formal causal modeling and analysis best be applied to improve the clarity of definitions and communication and scientific soundness of simulations, inferences, causal interpretations, generalizations, and policy-relevant conclusions in the PA? Please comment on whether any aspects of the following (or other) causal model formalisms can substantially improve the clarity and scientific soundness of the analyses and simulations in the PA: causal graph and DAG methods, conditional independence tests, intervention and interrupted time series analyses, other quasi-experimental methods, Wiener-Granger causality and transfer entropy, causal dynamic Bayesian networks (DBNs), other information-theoretic and graph methods, Simon-Iwasaki causal ordering , non-parametric structural equations models, mediation analysis.*

Response:

The two most important steps toward progress, in my assessment, are:

A. Apply <u>valid</u> (theoretically consistent under well-defined and broad assumptions) and <u>scalable</u> causal structure discovery methods to infer the underlying qualitative causality. Then <u>use the results to drive the choice of confounding or mediator controlling covariates</u> in quantitative models of choice.

B. Use <u>Big Data designs</u> where as many relevant variables are induced *at the same time* in the analysis. It is very hard to capture the causal interactions if the variables are not measured together and included in the analysis and is also very hard to combine partial causal results from separate, small-scope studies.

<u>Questions from Dr. Frampton</u>

Question 3*: Dr. Cox, Chair of CASAC, has introduced concepts of causality determination that differ from the framework established by the EPA in prior ISAs, used in the current PM ISA, and reiterated in section 3.2 of the current PA. This was a topic of discussion during CASAC's recent review of the PM ISA. <u>Please opine on the adequacy of the causality analysis framework currently used by the EPA, and whether and how the concepts espoused by Dr. Cox should, or should not, be incorporated into the NAAQS causality framework.</u> Also <u>please comment on the implications, of any changes in the causality framework that you would recommend, for the analyses and conclusions in this current PA</u>. For background, see the following documents:*

1) Preamble to the Integrated Science Assessments, November 2015, pages 18 to 25 (https://yosemite.epa.gov/sab/sabproduct.nsf/78476291901FF5AE8525835F00634158/$File/ISA_PREAMBLE_FINAL2015.pdf);

2) Follow-up Questions for John Vandenberg from Dr. Cox, letter of 12/17/2018 (https://yosemite.epa.gov/sab/sabproduct.nsf/B28289529495F929852583660057A3EE/$File/Follow-up+questions+for+John+Vandenberg-rev.pdf);

3) Responses from John Vandenberg, letter of 2/20/2019 (https://yosemite.epa.gov/sab/sabproduct.nsf/B48131F413362439852583A7005FDFDA/$File/JVandenberg+response+to+TCox+ltr+of+121718.pdf);

4) CASAC letter to the Administrator, April 11, 2019 - Dr. Cox comments, pages A8 to A27 (https://yosemite.epa.gov/sab/sabproduct.nsf/264cb1227d55e02c85257402007446a4/6CBCBBC3025E13B4852583D90047B352/$File/EPA-CASAC-19-002+.pdf).

Response:

As a conceptual framework underlying my answer please review my responses to Dr. Cox since they are foundational to your question as well (I will not repeat for parsimony).

The causality analysis framework currently used by the EPA is, in my assessment, a reasonable framework given the limitations of causal discovery methodologies available until recently, the fact that it is very hard to conduct complex manipulations (policy) involving the environment, and that cause-effect horizons are lengthy.

Operationally its defining features are the following:

(a) Controlling for confounding using some form of regression.
(b) Choice of plausible confounders on the basis of domain theory and/or synthesis of prior studies.
(c) Application of causality criteria similar to the ones proposed by Hill, Koch, the Surgeon General and others.
(d) Creating a hierarchy of evidentiary support where results are placed.

The weaknesses of this framework (as also indicated in my responses to Dr Cox) are:

(a) Errors in the choice of confounders translate to errors in causal effect estimates. This is <u>the primary weaknesses.</u>
(b) Small-scale (narrow variables scope) studies produce partial results with large variance on effects as a consequence of Simpson's paradox.

(c) Because assumptions are not stated explicitly it is hard to state/understand/share/differentiate effects of mediators, total, indirect, direct, or combinatorial effects.
(d) The causality criteria, are in general highly heuristic and it is trivial to show that in an infinity of causal situations they do not hold. A very prominent book on formal causal discovery refers to such criteria as "indefensible" (from a technical perspective, to be clear) and conducts a thorough (and thoroughly convincing) analysis of this claim (Spirtes et al. Causation, Prediction and Search, 2000).
(e) Moreover, it is not known currently what is the heuristic value of applying such criteria in this domain (and others). For example, when these criteria are applied, what is the probability that qualitative correct inferences are made about direct and indirect causal edges? What is the probability that quantitative causal effect estimates are within epsilon from the true value (for some acceptable epsilon)? What are correct ways and incorrect ones to apply the criteria? How tolerant are they in misapplication situations? And so on.
(f) Because the evidentiary pyramid, a heuristic by itself, is built upon several component methods that are themselves heuristic (as explained), the EPA may be facing a "house of cards" situation where errors in several stages of the process accumulate uncontrollably.

Does these criticisms imply that:

(i) EPA should immediately abandon this framework?
(ii) Prior results using the framework are invalid and policies depending on these results must be reversed or abandoned?
(iii) Every method that has some uncertainties and unknowns is unscientific and useless?

In my assessment, the answers are emphatically "no" to all three questions. The main reasons are as follows:

- <u>First,</u> we know that imperfect causal inference methods often produce useful results. Characteristic example is the identification of smoking as causing cancer and CVD (one of the all-time triumphs of epidemiology and public health policy).
- <u>Second</u>, in data science several mitigation factors that are often in play allow imperfect methods to perform well (classic example: bias-variance error decomposition allows highly unlikely and simplistic models outperform in small samples more realistic but complex models).
- <u>Third,</u> recent empirical evidence suggests that correlative studies have good overall performance. Illustrative massive study to that effect:
  Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials.Cochrane Database of Systematic Reviews 2014, Issue 4. Art. No.: MR000034.DOI: 10.1002/14651858.MR000034.pub2.
- <u>Fourth</u>, from common sense we know that the majority of our practical, functional knowledge about the world comes neither from experiments, nor from formal designed studies and analysis.

Yet, it is highly useful and accepted by all (e.g., we all know that crossing the street when the traffic light is red is a bad idea; no experiments have been conducted and no cohorts were created and analyzed to that issue however).

Balance is needed: the above considerations do not mean that we should perpetuate an imperfect system of causal inference when new methods and tools have recently become available that give improved options for better understanding of environmental risk factors. As I mentioned above, recent methods for large scale causal inference are now available and when combined with big data designs can yield extraordinarily powerful results.

Such techniques have transformed for example cancer genomics and oncology and are beginning to do the same across several health sciences/disciplines. Indicatively in my group alone we have used such methods for discovery of causal knowledge, predictive modeling, treatment personalization and many other applications in many medical problems: predicting risk of death from community acquired pneumonia, predicting risk of neonatal sepsis at the ICU, determining which patients with ovarian cancer will benefit from bevacizumab, cancer diagnosis and outcome prediction (several types), factors determining regression of atherosclerosis, factors determining progression of osteoarthritis, genetic factors for rheumatoid arthritis causation and prediction, psoriasis diagnosis using microbiomic sequencing signatures, diagnosis of pre-clinical acute viral infections, differential diagnosis of stroke types and stroke-like syndromes, modeling of physician decision making and guideline compliance, predicting lab measurements in a hospital setting, predicting and understanding childhood PTSD, predicting suicidal thoughts and behaviors, modeling temporal brain connectivity, studying the mediating role of sRNA for longevity, as well as in several studies of mining bibliographies, text, citations, and the health WWW. Many others have contributed a large body of work with significant and highly novel results in numerous diseases and discovery settings.

Questions from Dr. Packham

Question 3: *Can you describe the research efforts necessary to provide the required information? To what extent has the needed research already been done, or started? For example, are there crucial experiments or research initiatives that could clarify the shape of the PM2.5-chronic inflammation causal dose-response relationships at relevant exposure concentrations? Are there specific data analyses (e.g., testing for confounding by weather variables over more days prior to mortality) that could clarify the causal interpretation of epidemiological associations relied on in the draft PA to simulate effects of interventions?*

Response:

A critical opportunity exists in the analysis of additional data forms in rich contexts (i.e., high-dimensional or "Big Data") that should be pursued using the latest, scalable and consistent causal discovery algorithms. I provide 3 illustrative examples of types of studies that have potential to make a positive difference:

a. Analysis of full medical record data that are now routinely mapped to common (and thus interoperable) data models and geo-located across top-tier institutions in the US; overlay and link geo-location to environmental data. This design can both test and generate hypotheses.
b. Analysis of total causal effects (and information content) of environment versus non-environmental factors for various outcomes on a large scale. For example studying pan-omic mediators and co-determinants, medications, occupational data, etc.
c. Systematically study the effectiveness of newer and/vs older causal methods by validating them against real-life effects of interventions against their predicted responses.

Dr. Dan Jaffe

In general: Many of these questions refer to the specifics of C-R models. Unfortunately, this is outside of my expertise. I will try to address as many of your questions as possible. Thank you for the opportunity to assist this round of the NAAQS review.

Questions from Dr. Tony Cox

Question 1

I do not see a definition for the beta coefficients in the document and so do suggest that these need further clarification.

Question 2

The BenMAP-CE model is outside of my expertise.

Question 3

Concentration-response (C-R) is a broad term that could refer to many different aspects of the analysis. I think it is fair to use here in a general sense, but specific applications need clarity so as the meaning is clear in each case.

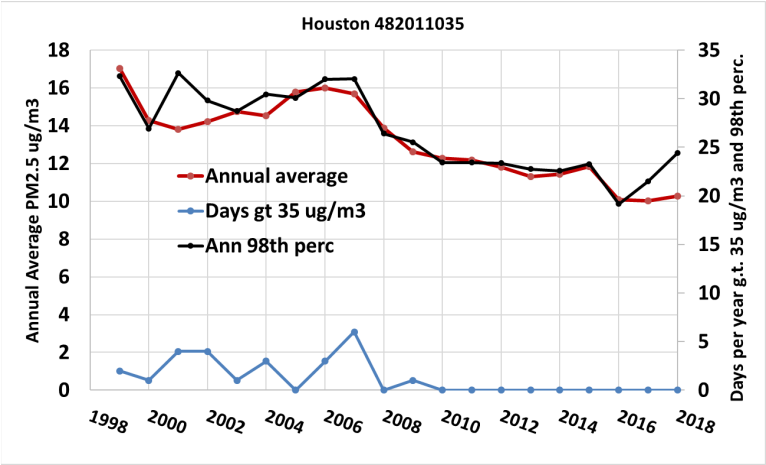
Other questions are outside my area of expertise.

Questions from Dr. Sabine Lange

Question 1

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 $PM_{2.5}$ standard. In general there is not a good relationship between the annual average $PM_{2.5}$ and the number of days over the daily standard (35 ug/m^3). 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/m^3 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.



Question 2

I agree that the short term studies are most relevant to the daily $PM_{2.5}$ 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 $PM_{2.5}$ 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.

Question 3

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 $PM_{2.5}$ 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.

The remaining questions are outside my area of expertise.

Questions from Dr. Mark Frampton

Question 1

The evidence for health impairment and premature mortality from $PM_{2.5}$ is extensive and fall into multiple categories. This includes epidemiological studies, mechanistic (biochemical) studies, clinical studies and animal studies. While the individual results from any one study could be questioned, the bulk of scientific papers provide strong evidence for health effects and premature mortality due to $PM_{2.5}$. The 2019 Draft Policy Assessment for $PM_{2.5}$ summarizes many of the scientific papers supporting this view.

Question 2

This question is very similar to that posed by Dr. Lange. Please see my answer above. All individuals are different. The question of heterogeneity across cities comes down to whether a key confounding variable has been missed. As a simple example, lets suppose that water quality in one region impacts health in a similar manner as $PM_{2.5}$. Clearly this could confound any results from the two cities. This is why the search for confounding variables is so important in epidemiology. To the best of my knowledge the major confounding variables have largely been addressed in air pollution epidemiology, but it is always possible that a new confounding variable could be identified.

Question 3

The preamble to the 2015 ISA seems like a valuable framework for judging both individual studies and the overall health impacts. While publication in peer-review journals is one important quality control step, it is not sufficient by itself to assess study quality. Section 4 of the ISA pre-amble seems especially important for judging individual studies and Table 1 suggests a useful framework for identifying causality. Given this framework, it would make sense to me that certain non-scientific analyses be excluded from the ISA. As one example, the outside consultants were provided a list of possible references to consider, and while I have not had time to examine them all, I would suggest that references such as this one, that we were provided, would not reasonably be included in the ISA or PA: https://doi.org/10.1016/j.gloepi.2019.100003.

The remaining questions are outside my area of expertise.

Questions from Dr. James Boylan

The discussion on primary sources seems accurate. The discussions on secondary aerosol is very cursory. This could be significantly improved.

On monitoring, yes this seems accurate. On trends there is too little discussion on the 98th percentile. We now that, at least for the western US, wildfires are increasing and having a significant influence on the 98th percentiles. See:

1. Abatzoglou, J. T. & Williams, A. P. Impact of anthropogenic climate change on wildfire across western US forests. *Proc. Natl. Acad. Sci.* 113, 11770–11775 (2016).
2. Dennison, P.E., Brewer, S.C., Arnold, J.D., Moritz, M.A., 2014. Large wildfire trends in the western United States, 1984–2011. Geophys. Res. Lett. 41, 2928–2933. https://doi.org/10.1002/2014GL059576
3. McClure C.D. and Jaffe D.A. US particulate matter air quality improves except in wildfire-prone areas. Proc.Natl.Acad.Sci., DOI: 10.1073/pnas.1804353115, 2018.
4. Westerling, A. L. Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. Phil. Trans. R. Soc. B 371: 1696 (2016).

On measurement biases: I am not aware of any studies showing consistently higher measured concentrations from FEMs compared to FRMs. (do you have a reference for this?)
I see this online from Minnesota Pollution Control Agency:
https://www3.epa.gov/ttn/amtic/files/2014conference/wedqamcmahon.pdf
there are many others. But none seem to suggest a persistent bias one way or the other.
Both FRM and FEM have potential weaknesses, particularly related to temperature and aerosol volatility. So where does this lead us? Maybe a better question is what is our uncertainty from both FRM and FEM measurements and are these uncertainties different? Based on the many comparisons that have been done, I suggest the uncertainties in $PM_{2.5}$ with either an FRM or FEM or of order 10%. These uncertainties are of similar magnitude as the uncertainties in spatial modeling.

Regarding background concentrations, I believe the discussion on background $PM_{2.5}$ is all reasonable, however I suggest more consideration be given to the implication of extreme events (international transport, wildfires, volcanoes, dust storms, etc) and their treatment in both health studies and exceptional events). While these infrequent events have little impact on the annual average, they can be especially important when considering the 98th percentile.

As for additional research:

I believe the spatial representation is a major uncertainty. It is very difficult to understand the exposure in densely populated urban areas that have few AQS monitoring sites. This would be especially true for neighborhoods in close proximity to freeways, other major roads and/or other sources. While the satellite fusion and hybrid modeling methods are all positive developments, I think even greater progress can be made by incorporation of low-cost sensor networks into fusion approaches. These are now being deployed around the country and world. There is a need for research to develop methods to utilize these data in some creative fusion approaches. Obviously sensor calibration and referencing to FRM and FEM

observations is critical, but I believe this problem can be overcome. Other areas for research that I believe are important are identification of biological agents in the air e.g. viruses, fungi and bacteria, both natural and human caused. These could represent a significant confounding variable in air quality studies and we know very little about these. Another large uncertainty is the contribution of secondary organic aerosols. These seem to be increasing in several parts of the country and our understanding of the sources of these is rather limited.

Questions from Dr. Steven Packham

To the extent that I can, I have addressed some of these topics in my responses above. Since most of these questions are directly related to the epidemiology and biological mechanisms these are outside my area of expertise. Regarding the role of natural vs anthropogenic activities, I agree with the document that, on average, natural sources are relatively small fraction of the current annual standard (e.g. 12 $\mu g/m^3$). However as discussed above, these can be very important on some days and thus more relevant to the 98th percentile.

Mr. John J. Jansen

Given the broad nature of the questions posed by the CASAC members, I read a good deal of the draft Policy Assessment (not chapter 4 and only skimmed chapter 5 and appendix D for personal interest). I have not had time to read the draft ISA. My more detailed responses are to Dr. Boylan's questions on Chapter 2 (my area of expertise). I also have more general observations touching on questions from Drs. Cox and Frampton based on my observation of the NAAQS and CASAC processes over the years.

Questions from Dr. Boylan

Chapter 2

As to the discussion of emission sources, the discussion is very summary in nature but reasonably complete. I believe EPA should provide some perspective on the magnitude of biogenic VOCs. While the magnitude of natural sources of SO2, NOx, and NH3 are small relative to anthropogenic sources, that is not the case for biogenic VOCs. I believe it would also be helpful for more information about the relative contribution of various source categories at the urban/regional vs. national scale (e.g., see p 2-5). Having similar observations/discussions of SO2, NOx, and NH3 could provide insight to those who will be implementing the standard. For example, I find it interesting that in the modeling to just attain the current and alternative standards, many areas controlled by the annual standard needed to resort to proportional scaling to achieve the goal (see page C-32), substantial SO2 plus NOx emissions reductions were not sufficient. It is not clear why EPA did not add the primary scenarios first before resorting to proportional scaling as they did for the areas in the Primary PM scenario. Assuming the models are correct, this could have important implications to the states in their attainment demonstrations. I recognize the purpose of the PA is on setting the NAAQS, not implementation. Nevertheless, highlighting these findings now can help with a step (i.e., providing implementation guidance) EPA is notoriously late at achieving.

The discussion on monitoring is reasonably complete for the compliance networks but is a bit weak on the "Additional PM Measurements and Metrics" (page 2-18). There is a wealth of data on PM, PM composition, and pollutant gases from research networks such as the SEARCH network which, although now shut down, can supply important insights for model development and application and measurement errors. There is no mention in this section of continuous NO3 and NH3, both of which existed in SEARCH and other networks.

The discussion of measurement correlations and trends is reasonably complete. The section on composition trends, focusing only on sulfate, leaves the misimpression that sulfate alone trended down. Nitrate and carbon also did so but to a smaller degree. A more complete summary here would help.

Further, the discussion of figure 2-11 on page 2-25 appears incorrect. The data for the Southeast is not an outlier. The Northwest and SoCal are. Clarification is needed.

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.

The hybrid modeling is becoming a more common tool for estimating exposures in health studies. And there have been several studies looking at their performance. And while the discussion in the PA seems reasonable, it needs to carry the evaluation through the health assessments. On page C-4, EPA places a priority on selecting studies that indeed use multiple hybrid modeling approaches and compare the health results. However, Table C-1 lists only 1 such study, Jerrett, et. al., 2016. However, I could find no discussion in the PA nor the draft ISA of this feature of the study. What were the results of the comparison? It seems that at a minimum, EPA would discuss these results to support their claim on page 2-48 that hybrid modeling results "are reliable for estimating PM2.5 exposure in many applications" and "also suggests that the fields are reliable for use in health studies."

As to model performance, as with all regional model performance approaches, average operational performance is the norm. Gone are the days when extensive diagnostic performance was included. Recognizing that such performance evaluation is not going to happen, at least some look at the range of performance across locations would be helpful. What is the range of overestimation? Under estimation? Is the range better or worse in populated areas than in unpopulated areas? On page C-23, Table C-6 model performance (not hybrid) shows over prediction in the east and under prediction in the west, in some cases by as much as 2 ug/m3. Knowing such information would help with interpreting the adequacy of the hybrid models for health purposes.

The discussion of background needs clarification. For example, statements on pages 2-49 (lines 28 to 32) and 2-52 (lines 14 to 16) leave the role of non-US anthropogenic emissions unclear given the examples shown. The discussion of such emissions on page 2-53 suggests they are small due to short residence times of PM and yet natural sources appear more prominent (see next section). Shouldn't the residence time argument apply to all PM, regardless of source? On page 2-54 EPA claims that the remaining sulfate at the remote sites is from domestic SO2 emissions. Why?

Chapter 3

See below for a more general response to the adequacy of EPA's evidence and risk based approaches to reviewing the primary PM2.5 standard. The criteria for selecting the study areas seem reasonable but entire central portion of the country is missing, why? EPA provides a brief rationale why it decided to focus only on all-cause mortality on page C-2. The explanation implies morbidity outcomes would be less reliable? Was time, resources, and/or insufficient "data" also a factor? EPA's criteria for choosing the concentration-response functions is provided in section C.1.1. I have always had concerns that studies for cancer do not adequately consider the long latency and lack knowledge of historical exposures and population migration. I do not agree with EPA's exclusion of "groups not differentiated by socio-economic . . . attributes." Studies that fail to account for socioeconomic status are missing a key factor affecting health. Seeking studies that use multiple approaches to estimating exposures and calculating health risks from each is laudable. Unfortunately, only one such study was included (Jerrett, et. al., 2016) and EPA fails to describe these results nor use them in the uncertainty discussion. On pages C-22 and C-23 EPA is asserting that an epidemiological study on a cause specific mortality provides biological plausibility for all-cause mortality (i.e., ischemic events and heart failure for cardiovascular mortality and COPD and asthma for respiratory mortality). I do not believe this is correct. Biological plausibility is provided through human exposure and animal toxicology studies (see page C-25 for a such a discussion). I recommend EPA refrain from blurring the definition of biological plausibility. As to additional research needs, the current debate over causality determinations is a clear area for EPA to explore. Also the approach reported in the studies by Dr. Anne Smith in Risk Analysis (see references provided by consultant Dr. Warner North) should have already been pursued by EPA since the last review.

Appendix C

See my comments on Chapter 2 above and the general observations below for some responses to the methods used in the risk assessment. I will say that the methods have become more complex over the years, supposedly to improve the credibility of the estimates. So many steps are added based on assertions of adding accuracy but little quantitative demonstration is made. Are we really getting more accurate estimates of exposure?

<u>Questions from Dr. Packham</u>

I have provided research recommendations elsewhere. As to the last question, effects from strategies to implement the NAAQS, EPA and CASAC have traditionally avoided this mandate. EPA by not providing the necessary information for CASAC to review and CASAC by not having the expertise, through its members or expert panels, on such issues to provide advice. Similar to conducting new causality determination approaches or new risk assessment approaches (e.g., suggested by Dr. Anne

Smith), EPA needs to take the issues seriously and undertake the work. I recognize that too often time and resources are limited but it is time to give these, and other, issues priority.

General Observations

The issue of causality determinations is currently a particularly contentious issue. EPA has updated and refined its approach to causality determinations over the years but it has always been qualitative in nature (i.e., 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.). Despite suggestions to alter and/or add to its approach (many to add quantification), EPA has continued to justify the status quo as consistent with past practices endorsed by CASAC.

I view the suggestions of Dr. Cox to be an attempt to add quantification with which I concur, but I am not a statistician. And I do not know whether applying these techniques will confirm causality or refute it. I will confess to believing that associations are not and cannot be causal. Something more is needed (e.g., human exposure and/or animal toxicological studies, more on that below). A key problem is that the work to apply these techniques has largely not been done. Therefore, such results are not available for this review. There is also, of course, the problem associated with applying the techniques to existing study data bases due to confidentiality issues. While the work suggested by Dr. Cox may be problematic in this review, EPA has failed to follow-up on past, certainly doable recommendations such as applying the risk assessment techniques recommended by Dr. Anne Smith during the last review or completing a quantitative uncertainty analysis. That leaves us with the status quo approaches and asking whether the evidence has changed significantly.

EPA has summarized the new studies, most conducted with concentration levels lower than before (certainly a change from the last review) and with new approaches to establishing exposures, and new approaches to conducting the risk assessment. EPA concludes that the new information reaffirms and strengthens previous conclusions of causality, that key uncertainties have been reduced, and a lowering of the annual standard is supported. Since I am not sure what strengthening means (more of the same?), I do not view it as justification to change the NAAQS. Looking at the reduction in key uncertainties, EPA finds no threshold (no change), the shape of the concentration response (C-R) curve at lower concentrations is linear to zero (no change), human exposure and animal toxicology studies (necessary for providing biological plausibility) are still conducted at concentrations above ambient (no change), and the uncertainties of the C-R has decreased at lower concentrations.

With respect to decreasing uncertainties in the C-R functions at lower concentrations, I ask whether this isn't expected. Since there is no threshold and the shape of the C-R is linear to zero, the uncertainty at the lower concentration levels must decrease as there is now more data at those levels. Does this really add to our confidence in the risk results? It is not apparent given the confidence intervals in the risk

results (see for example table C-5 on page C-85). EPA does not describe what the ranges are (e.g., uncertainty and/or variability). Nevertheless, they are quite large and call into question whether the reduced uncertainty has led to a different result than last time.

On the issue of biological plausibility, I reiterate that a) this is necessary to combine with associations to infer causality, and b) EPA should refrain from using an association to infer biological plausibility of another association. I also offer a different perspective on the results from human exposure and animal toxicological studies summarized in Table 3-2 on page 3-46. While these results still may be at levels higher than ambient (but closer than in the past), the trend in the results is intriguing. Of the studies with the lowest concentrations one shows no effect and one shows mixed results. Even the next few higher concentration studies show mixed results. And all of these studies are looking at biological markers that "demonstrate short-term exposures to PM2.5 *may* result in the types of cardiovascular endpoints that *could* lead to emergency department visits and hospital admissions in *some* people." (emphasis added, page 3-28). One might infer from this perspective that there are no adverse effects at ambient concentrations.

Finally, there are the risk estimates. As mentioned above, the confidence intervals (or whatever they are) in the tables such as 3-5 on page 3-85 are large. And the means of the alternative standards are within the range of the current standard. So are the changes really different than the current standard?

Dr. Frederick Lipfert

I organized my responses as follows. First, a background discussion based on observations that may differ from the PA and ISAs that are requisite to "policy relevant" discussions. Then I address specific questions and summarize my responses. Appendix A presents background issues in more detail. Appendix B contrasts short- and long-term studies. Appendix C lists my relevant publications that have not been included in an ISA or the PA.

Background on Air Pollution, $PM_{2.5}$, and Human Mortality

The Clean Air Act is not specific; pollutants of concern were defined in the 1970s and remain today although the definition of particulate matter has changed. Concentrations of CO and SO_2 have dropped to trivial levels but are still monitored. The most hazardous pollutants such as benzene, formaldehyde, or benzo(alpha)pyrene are not monitored in the atmosphere. $PM_{2.5}$ might be considered a "target of choice". "Ambient" was intended to represent outdoor air quality alone, perhaps because it was much worse than indoors at that time. This is no longer the case.

$PM_{2.5}$ is not really a pollutant in the sense of defined chemicals (CO, O_3, SO_2). It is a regulatory construct defined solely by particle size and based on the total mass of a mixture of many kinds of particles, some toxic, most not. Those of most concern (metals, ultrafines, carbon compounds) comprise a small fraction of the total mass. No sources emit $PM_{2.5}$ as measured in the ambient; source abatement strategies for specific constituents can thus be arbitrary without evidence of attained health benefits. We are exposed to $PM_{2.5}$ from both indoor and outdoor sources, regulated or not.

Excess mortality has been a focus of air pollution health effects since the advent of regulation; it is an accurate end point for which nationwide individual data have readily been available for decades. Monetary values may be assigned in various ways for cost-benefit analysis. Relative risk can be converted to longevity loss, which may be a more cogent endpoint than numbers of victims.

Epidemiology is concerned with establishing cause and effect, often beginning with statistical associations. However, some cautions are in order. Beyond statistical significance, we need to understand the entire system beginning with emissions. Replication of observational studies based on the same model does not confer validity to any of them. The mere presence of a pollutant in the atmosphere does not constitute a "cause", for which personal exposures and translation to a target organ are required. An "effect" should relate to a target organ and imply specific mechanisms; excess mortality or admissions to hospital may be *consequences* of such clinical effects.

In order to do its job, CASAC must evaluate the extant concentration-response (C-R) functions and the exposure data used to drive them, which should represent actual exposures if the responses are to represent actual health effects. This evaluation should consider three types of C-Rs:

Short-term (days) fluctuations in air quality that may trigger sudden death in vulnerable individuals. Causality in this mode has been demonstrated by heat wave mortality.

Long-term (decades) individual differences that affect incidence of new cases of chronic disease, for which the extended effects of cigarette smoking are a causal example.

Long-term differences among populations that may involve both of the above as well as attributes of their locations *per se*, such as climate, green spaces, traffic density, or characteristics of housing stock that affect indoor air quality.

It is possible for all three of these relationships to be involved in a given situation. These responses to the CASAC questions reflect these distinctions, specifically the timing of responses relative to exposures and the detailed nature of $PM_{2.5}$ exposures.

Responses to Questions from Dr. Boylan

Chapter 2 – PM Air Quality

- *Is the discussion on sources of emissions accurate and complete? If not, what additional information needs to be included?*

  No, indoor sources were not discussed.

- *Is the discussion on ambient monitoring accurate and complete? If not, what additional information needs to be included?*

  No, ultrafine PM (UFP) needs more attention. Given the importance of short-term effects, the need for more daily monitoring should be recognized. The need to understand indoor air quality should be recognized.

- *To what extent are biases associated with PM_{10}, $PM_{2.5}$, and ultrafine measurements discussed?*

  Insufficiently.

- *Is the discussion on background concentrations accurate and complete? If not, what additional information needs to be included?*
-
- I have no response here.

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

- *Is the evidence-based analysis presented in Chapter 3 scientifically sound?*

  No. Accountability was not established. Relationships between short- and long-term effects were not discussed.

- *Are the areas for additional research adequate and complete? If not, what additional areas need to be included?*

  The needs to consider latency and cumulative exposure were not discussed. The need to consider the sum of short-term effects over the lag period was not recognized. Difficulties in defining thresholds were not discussed**.**

Appendix C – Supplemental Information Related to the Human Health Risk Assessment

- *Is the air quality modeling approach to projecting $PM_{2.5}$ concentrations to correspond to just meeting the NAAQS (AQS, CMAQ, Downscaler, SMAT-CE, project monitors to just meet NAAQS, project spatial fields to correspond to just meeting the NAAQS) scientifically sound? If not, what are your concerns and how should they be addressed?*

Historical data are required to estimate cumulative exposures needed for valid long-term C-Rs.

- *Is the health risk modeling approach using BenMAP-CE appropriate for this application? If not, what are your concerns?*

I find the bulk of the PA dedicated to regulatory issues, ignoring the scientific issues that are central to supporting them. The Appendices are intended to discuss an illustrate those issues. I searched the PA to determine inclusion of specific terms central to my discussions and issues. The following were not mentioned in the PA text: indoor, frailty, latent, cumulative exposure, acute. The time-series studies by Murray and Lipfert (2012, 2013) were not cited. The only citation of the Veterans Cohort Study, for which 8 papers have been published (Lipfert et al., 2000, 2003, 2006a, 2006b, 2008a, 2008b, 2009, 2018, 2019 [accepted]) was for the only significant positive $PM_{2.5}$ risk among many negatives and was thus "cherry-picked".

Responses to Questions from Dr. Cox

1. *Are the beta coefficients in Table C-1 of the PA conceptually well defined? That is, are their intended conceptual meanings and causal interpretations clear and unambiguous?*

My answer is "no". Beta coefficients are outputs of models intended to describe the real world so that their increments may be interpreted as specifying effects of changes in inputs. Above I proposed 3 different C-R models that pertain to specific situations and differ from the PA. Their validity depends on the validity of inputs (exposures) and their timing. The PA does not consider indoor exposures and thus strictly applies only to the 10-15% of the time, if that. It does not consider timing of exposures, notably cumulative exposures. A more general discussion of causality and accountability is presented below. Suffice it to note that no victim of long-term exposure to ambient air pollution has ever been identified by autopsy, by contrast with 1952 London fog disaster (Bradley, 1957). The present C-Rs must be regarded as primarily a regulatory device.

2. *On the same topic of clear definitions, does the discussion of the BenMAP-CE beta coefficients in the PA and underlying documentation (described as typically representing "the percent change in a given adverse health impact per unit of pollution") unambiguously specify which of the following concepts the coefficients represent?*

My answer is "no". Those concepts are not physiologically defined, only operationally.

3. *Is the definition of "concentration-response (C-R) relationships" in the PA and its Appendices (cf p. C-38) adequately clear and unambiguous to support simulation of well-defined causal effects of interventions that change pollution levels?*

My answer is "no". Those interventions are not well defined. Ambient $PM_{2.5}$ is a mixture of various types of particles from specific types of sources, some from combustion, some from various types of processes, some from natural sources. $PM_{2.5}$ as measured in the ambient is not emitted from any source. The notion that a given percent reduction of emissions from any of those distinct types of sources would have the same benefit on the various health endpoints that have been considered is patently absurd.

4. *Do the beta coefficients in the PA overcome these methodological objections to using relative risks, regression coefficients, and related measures of association to predict (or simulate) effects of interventions? If so, how were they overcome? If not, does this imply that the simulations in the PA are not necessarily reliable or valid predictors of the real-world effects on public health of reducing PM2.5? Why or why not?*

5. My answer is "no". I get the impression that none of these basic flaws have been considered in conjunction with the establishment or modification of NAAQS.

6. *On the same methodological issue, which, if any, of these measures (relative risks, odds ratios, attributable risks, and regression coefficients) are currently generally accepted in contemporary causal analysis and epidemiology as valid measures of how changing one variable (e.g., exposure) will cause another (e.g., population health responses) to change?*

   My problems are with the underlying conceptual model, not the expressions of outputs. I would note however that no lives have ever been "saved" by air pollution control, but some may have been extended. Those longevity extensions under current conditions have ranged from a few days to a few months. The underlying health status of those putative victims has not been considered.

7. *More generally, do the $\Delta y/\Delta x$ values calculated by BenMAP-CE have valid interpretations as causal impacts on y of interventions that change x?*

As discussed elsewhere, my answer is "no". Such predictions have never been verified; they are statistical but not operational. As I understand it, BenMap would have us believe that moving from say, Steubenville, OH, to Madison, WI, would instantly clear our lungs, unclog our arteries, and stop our metastasizing lung cancer, which is clearly nonsense and statistical confidence intervals tell us nothing about this unreality. One might argue that only a few percent of those movers might actually benefit in this way, which would then raise the question of selection: random or according to pre-existing conditions? The more cogent questions, for which we have no answers, might be what role did Steubenville's dirty air play in creating those conditions and how long did it take? Common sense must prevail here.

8. *Have the beta coefficients in the PA been empirically validated?*

As discussed elsewhere, my answer is "no". Long-term effects have never been verified by intervention; short-term effects were validated in London 1952 (Bradley, 1957).

9. *Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately describe and discuss the extent (if any) to which their values reflect potential omitted confounders of the association between mortality risks and PM2.5 levels (e.g., lagged daily high and low temperatures and humidity in the weeks preceding mortality, if these contribute both to (possibly delayed) mortality and increased energy usage and PM2.5 pollution?)*

My answer is "no". Statistical confidence intervals portray the precision of a model and depend on sample size. Uncertainties about the validity of the model can be far more important and confounders

can depend on model structure and the data available. The sample size of a cohort can only be increased by extending follow-up time, which entails temporal gradients and depletion of its most vulnerable members. I would argue that the more a cohort study is extended the less we know about its members. Only the Veterans Cohort (described below) has considered nonlinear relationships with subjects' physiology (blood pressure) climate, poverty, neighborhood racial mix, exposure timing over 26 y, and age-specific risk estimates. It produced statistically significant *negative* risks of $PM_{2.5}$ and strong positive risks associated with vehicular traffic density. These findings have never been included in an ISA or PA.

10. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address the extent (if any) to which their values reflect residual confounding of the association between mortality risks and PM2.5 levels (e.g., by daily high and low temperatures and humidity in the weeks preceding mortality in models that only address seasonal, annual, or averaged temperatures)?

My answer is "no". The models used to set $PM_{2.5}$ NAAQS did not control for climate.

11. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address model uncertainty (e.g., the possibility that the linear no-threshold model specification is incorrect, e.g., because sufficiently low exposure concentrations do not cause pulmonary inflammation and adverse health effects that occur at higher concentrations)?

My answer is "no".

12. Does the PA's discussion of beta values, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address effects on the estimated values of the beta values of exposure uncertainties and estimation errors?

My answer is "no".

13. Does the PA adequately assess the suitability of the designs of the studies used to estimate beta values for purposes of valid causal inference and simulation?)

My answer is "no".

14. Does the PA's discussion of beta values adequately address attribution of risk in the presence of joint causes? Conversely, how well does the discussion in the PA make clear how much of the estimated beta value for PM2.5 is actually contributed by other variables (such as temperature extremes and poverty) that would not necessarily be changed by an intervention that reduces PM.5 levels?

My answer is "no".

15. How confident can policy analysts and decision makers be in the predictive validity of the simulated results?

There can be no confidence absent accountability through intervention analysis or autopsy; see Henneman et al. (2019). Below I summarized the current state of knowledge and what remains to be determined.

16. Are there other statistical or methodological issues that you would like to comment on that you believe might help the CASAC to assess the validity and soundness of the PA and its simulations for effects on health risks of changing PM2.5 levels, or that might help to improve the technical and scientific quality of the final PA?

I discuss these issues in detail below.

17. How can techniques of formal causal modeling and analysis best be applied to improve the clarity of definitions and communication and scientific soundness of simulations, inferences, causal interpretations, generalizations, and policy-relevant conclusions in the PA?

I conclude that this PA is essentially beyond redemption, as discussed below. More research on exposure timing and physiological modeling are required.

Responses to Questions from Dr. Frampton

1. Is there evidence that would support a reconsideration of the current and long-held views of this causal relationship as expressed in the PA and ISA?
3. Also please comment on the implications, of any changes in the causality framework that you would recommend, for the analyses and conclusions in this current PA.

It is my position that causality must be established through observations, in conjunction with statistical analyses. Causality was established in the short term when exposures and responses cycle in concert as in daily mortality or hospitalization. It was proven through autopsies in the 1952 London fog episode (Bradley, 1957). In long-term epidemiology, causality must meet five requirements, none of which has been established for $PM_{2.5}$:

Exposure. Indoor exposures and the contributions of daily peaks have not been considered in long-term studies. Initiation of chronic disease can only occur after a latency period and then depends on cumulative exposures. This has been established for smoking but not for air pollution epidemiology.

Toxicity. $PM_{2.5}$ is not really a pollutant in the sense of defined chemicals (CO, O_3, SO_2). It is a regulatory construct based on mixtures of many kinds of particles, some toxic, others not. Given the multiplicity of health endpoints that have been considered, a statistically significant combination with a PM constituent is likely.

Translocation. Inhalation of a pollutant is not sufficient to imply contact with an organ and initiation of chronic disease. $PM_{2.5}$ translocation from the lung to the bloodstream has not been demonstrated, by contrast with ultrafine particles, for example.

Susceptibility. Given the diversity of the population, the notion that a healthy person could be randomly selected to die coincidently with normal levels of ambient air quality is patently absurd. Time-series studies have shown that acute mortality is associated with both underlying frailty and daily air pollution peaks (Murray and Lipfert, 2012, 2013). Differential susceptibility is also likely the case for incidence of chronic disease, but making this case would require tracking individual cohort members during the latency period.

Accountability. The ultimate test of causality is whether public health has actually improved in response to reduced $PM_{2.5}$ (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; Henneman et al. (2019) showed no effect on total mortality (-0.011 [-0.92 – 0.71]) after reducing $PM_{2.5}$ from 10.0 $\mu g/m^3$ in 2005 $\mu g/m^3$ to 7.2 in 2012. However their analysis did not account for latency or cumulative exposures.

By contrast, the epidemiology of smoking has all of these elements. Exposure is self-reported on an individual basis, and the importance of cumulative exposure is shown by the use of pack-years as a predictor of disease after accounting for latency. Toxicity of smoke constituents has been determined by laboratory and animal tests, especially for benzo(alpha)pyrene (B[a)P) which is also an (unmonitored) ambient air pollutant. Translocation is not an issue for the gaseous pollutants and the most serious health effects (cancer, COPD) occur in the lung. Susceptibility comes into play when frail smokers are urged to quit. Accountability is shown by the improved health of quitters.

2. Please opine on the level of uncertainty that is represented by this heterogeneity, and the impact if any, on the conclusions in the PA.

Heterogeneity between cities must be controlled through the use of contextual variables in the regression models, in addition to personal information. This was first shown in the Veterans Cohort Study (Lipfert et al., 2000) in which the inclusion of variables dealing with climate, neighborhood poverty, and neighborhood racial distribution changed the mortality risks for $PM_{2.5}$ and SO_4^{2-} from positive to negative. However, those contextual variables have not been used in other cohort studies. Heterogeneity within cities includes house-to-house variability in indoor air quality and local features like green spaces and high-traffic areas. Figure 1 shows lack of correlation between indoor and outdoor across American cities (Lipfert, 2015) which has also been shown in EPA studies (Baxter et al.1994, 2007, 2010, 2013, 2014, 2017). While indoor $PM_{2.5}$ is about half of outdoor because of infiltration as shown in the Figure, personal exposure can be much higher because of indoor sources like environmental tobacco smoke. EPA has confirmed these observations but they have not been used in EPA-sponsored epidemiology.

4. The CASAC letter to the Administrator (April 11, 2019) states, "There is inadequate evidence for the 'likely to be causal' conclusion for long-term $PM_{2.5}$ exposure and cancer." exposure time frames for most of these studies are insufficient to draw conclusions about incident cancer. Do you agree with the CASAC's findings in this matter?

While response must always follow exposure, the delay depends on the latency period of the endpoint. Sudden death may occur within a week, as shown in time-series analysis; initiation of chronic disease may require decades. Figure 2 shows temporal trends for lung cancer, smoking, and $PM_{2.5}$. based on population-average data. Cancer lags smoking by about 18 years, consistent with cohort and individual data. We would expect a similar relationship for atmospheric particles, for which sufficient historical data are lacking. However, if latency were neglected, a (spurious) longitudinal relationship between lung cancer and $PM_{2.5}$ would be seen in Figure 2, as has been the case in long-term cross-sectional cohort studies. We would expect this concept to pertain to initiation of other diseases as well.

5. Please comment on the appropriateness and completeness of the approaches used in this PA to assess the risks of exposure, and the assessments of risk reduction of alternate standards, for $PM_{2.5}$ and PM_{10} (sections 3 and 4, respectively).

The concepts of causality and appropriate exposures are universal and thus would affect any differences resulting from alternative NAAQS levels. However, estimating any resulting monetary benefits must be revisited. The current practice is to assign a value of several million $ to each life "saved" regardless of the resulting change in life expectancy. Recent analyses have estimated longevity extensions from several days to several months that would have substantially lower value. The current monetization algorithm must be changed.

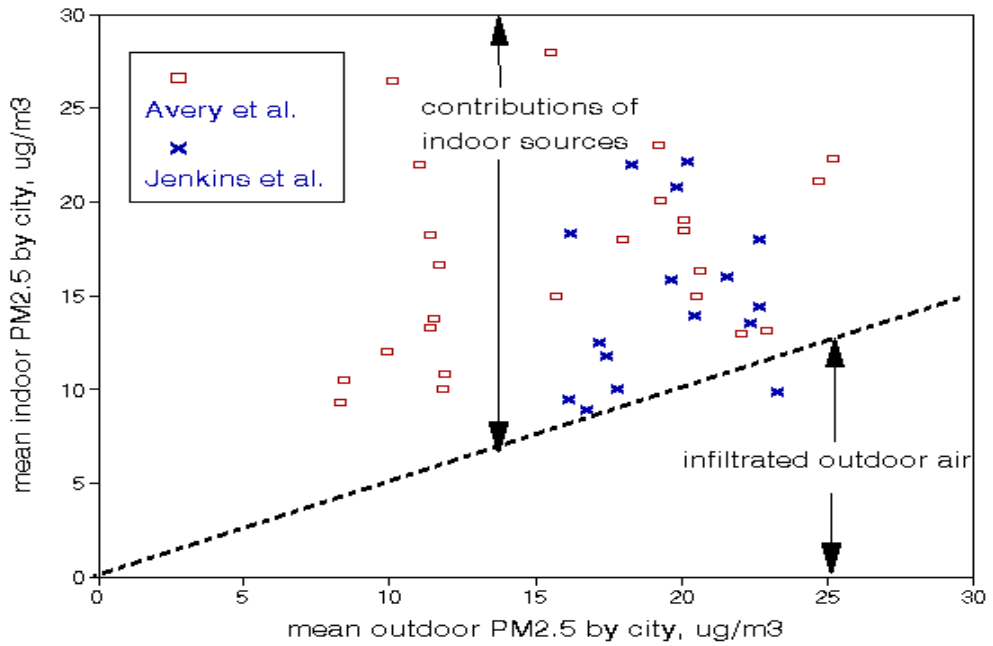


Figure 1. Relationships between indoor and outdoor $PM_{2.5}$ for selected U.S. cities.

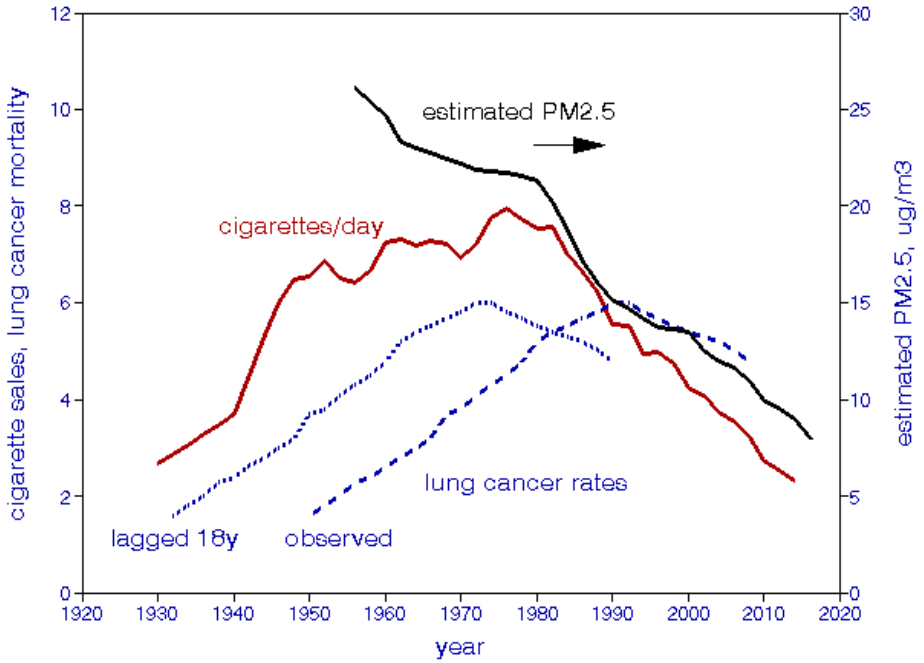


Figure 2. Comparison of trends in annual cigarette sales, estimated $PM_{2.5}$, and lung cancer mortality 18-y later (deaths/10,000 population)

6. Are there additional key studies that should be considered in sections 3 and 4 of the PA?

Here are citations and summaries of some of my relevant publications that have not appeared in the PA or an ISA.

<u>Measurement error and uncertainties</u>

These papers examine the effects of bias and noise in independent variables used in linear regression, especially from correlated air pollutants. They show that the pollutant with the least error will prevail in co-pollutant models. Random measurement error biases the slope of an exposure response function toward the null, which also biases the x-intercept toward the origin and obscures any threshold that may have existed in the absence of measurement error. “Measurement error” may be random (instrumental) or the result of mis-specification or mis-location of monitors.

F.W. Lipfert and R.E. Wyzga (1995), Uncertainties in Identifying "Responsible" Pollutants in Observational Epidemiology Studies, Inhalation Toxicology 7:671-89.

F.W. Lipfert and R.E. Wyzga, Air Pollution and Mortality: Issues and Uncertainties, J.AWMA 45:949-66 (1995).

F.W. Lipfert and R.E. Wyzga (1997), Air Pollution and Mortality: The Implications of Uncertainties in Regression Modeling and Exposure Measurement, J.AWMA 47 517-523.

F.W. Lipfert and R.E. Wyzga (1998), Effects of Exposure Error on Environmental Epidemiology, proc. International Symposium on the Health Effects of Particulate Matter, Prague. AWMA Publ. VIP-80, pp. 155-166.

F. W. Lipfert and R.E. Wyzga, Statistical Considerations in Determining the Health Significance of Constituents of Airborne Particulate Matter, J.AWMA. 49:PM-182-191 (1999).

Lipfert, F.W., The Use and Misuse of Surrogate Variables in Environmental Epidemiology, J. Environmental Medicine 1:267-278 (1999).

Time-series studies

Time-series studies of daily mortality predict acute responses on the first few days of exposure, usually persisting less than a week. Responses are limited to the elderly and are stronger among the most susceptible subjects. Similar risks have been associated with various air pollutants and are strongest for cardiac and respiratory deaths. Murray and Lipfert (2010) considered a new model that predicts the magnitude of an elderly subpopulation most at risk of imminent death on a daily basis, rather than the general population. In this model two conditions are required for an acute air pollution-related death: increased frailty (inability to maintain homeostasis) and increased pollution exposure. Extreme values of either parameter can result in death, thus precluding air pollution thresholds. Acute frailty is represented by membership in a small fluctuating sub-population (P_t) at high risk that is depleted by daily deaths (D_t) and replenished by new entries (N_t). D_t and N_t may be affected by current or previous environmental conditions. The ratio of P_t to D_t is a measure of survival time (i.e., daily life expectancy) in the frail state. The Kalman filter is used to solve these equations. P_t fluctuates with season and averages ~0.08% of the elderly. Life expectancies of subjects in the at-risk pool are about 5-7 days and robust to alternative models and environmental fluctuations. Daily mortality risk estimates are similar to those from conventional time-series analyses, but information on the accompanying life expectancies provided by this model is novel.

Murray CJ, Lipfert FW. Revisiting a Population-Dynamic Model of Air Pollution and Daily Mortality of the Elderly Population in Philadelphia. J Air Waste Manag Assoc. 2010 60:611-629.

Murray CJ, Lipfert FW. A new time-series methodology for estimating relationships between elderly frailty, remaining life expectancy, and ambient air quality. Inhalation Toxicology 2012 24:89-98.

Lipfert FW, Murray CJ. Air pollution and daily mortality: A new approach to an old problem. Atmos Environ 55; 467-74 (2012).

Murray CJ, Lipfert FW. Inferring frail life expectancies in Chicago from daily fluctuations in elderly mortality. Inhal Toxicol. 2013 Jul;25(8):461-79.

Other time-series papers.

These papers provided short-term risk estimates for numerous pollutants that have not been considered by others: a wide range of particulate measures in which TSP was the most important, and a complete set of PM chemical components from which carbon compounds were the most important.

Lipfert, F.W., Morris, S.C., and Wyzga, R.E. (2000), Daily Mortality in the Philadelphia Metropolitan Area and Size-Classified Particulate Matter, J.AWMA 50:1501-13.

Klemm, R.J., Lipfert, F.W., Wyzga, R.E., Gust, C. Daily mortality and air pollution in Atlanta: Two years of data from ARIES. Inhal Toxicology 16 (Suppl 1):131-41 (2004).

The Veterans Cohort Study

The Veterans Cohort Mortality Study began as a cooperative venture with Washington University at St. Louis (WUSTL), who developed the cohort and originated the study. It is based on a cohort of about 70,000 male veterans who were recruited in 1976 and followed for 26 y This is an important period of study because of the substantial improvements in ambient air quality that were achieved following passage of the 1970 Clean Air Act. The cohort was 37.7% black and 81% had smoked at one time; they were recruited at VA 32 clinics nationwide. it is the first cohort study to examine long-term all-cause mortality risks associated with air pollution for a variety of pollutants by race. The study considered the overall follow-up period (1976-2001) and four sequential mortality periods (1976-81, 1982-88,1989-96, and 1997-2001). The WUSTL proportional hazards model included non-linear terms for age, smoking, body-mass index (BMI), blood pressure, and interaction terms. The first paper showed that the relative importance of $PM_{2.5}$ and SO_4 is contingent on the inclusion of contextual predictor variables in the model including climate, poverty and neighborhood characteristics in the proportional hazards model. Those variables were retained in subsequent analyses. Later papers examined $PM_{2.5}$ constituents, hazardous pollutants, and county-level traffic density, which was the most important predictor along with NO_x and elemental carbon. The 2018 paper examined the effects of cohort aging with respect to temporal trends and found the former to be more important with stronger risks in middle age. $PM_{2.5}$ had consistently negative (beneficial) effects; no effects of accumulated exposures were evident. The most recent (2019) paper found that black veterans were more heavily exposed to traffic, with significantly negative effects of $PM_{2.5}$ but significantly positive risks for whites similar to those found in other cohorts. This explains the apparently anomalous (negative) $PM_{2.5}$ effects for the entire cohort. It also

supports the hypothesis that cross-sectional studies essentially consider differences among *places* per se (counties, zip-codes, etc.) rather than the personal exposures of residents. Traffic effects were found to be stable over the 26-y period; however, those actual risk agents remain unknown, comprising engine exhaust, noise, tire/road dust, undesirable residential locations, but not indoor air pollution.

Lipfert, F.W., Perry, H.M. Jr., Miller, J.P., Baty, J.D., Wyzga, R.E., Carmody, S.E. (2000) The Washington University-EPRI Veterans' Cohort Mortality Study: Preliminary Results, Inhalation Toxicology 12 (Suppl 4):41-73.

Lipfert, F.W., Perry, H.M. Jr., Miller, J.P., et al., 2003. Air Pollution, Blood Pressure, and Their Long-Term Associations with Mortality. Inhalation Toxicology 15, 493-512.

Lipfert, F.W., Wyzga, R.E., Baty, J.D., Miller, J.P., 2006a.Traffic Density as a Surrogate Measure of Environmental Exposures in Studies of Air Pollution Health Effects: Long-term Mortality in a Cohort of U.S. Veterans, Atmospheric Environment 40, 154-169.

Lipfert, F.W., Wyzga, R.E., Baty, J.D., Miller, J.P., 2006b. $PM_{2.5}$ Constituents and Related Air Quality Variables as Predictors of Survival in a Cohort of U.S. Military Veterans, Inhalation Toxicology 18:645-57.

F.W. Lipfert, R.E. Wyzga, Jack D. Baty, J. Philip Miller. Vehicular Traffic Effects on Survival within the Washington University - EPRI Veterans Cohort: New Estimates and Sensitivity Studies, Inhalation Toxicology 20:949-960 (2008).

F.W. Lipfert, R.E. Wyzga On Exposure and Response Relationships for Health Effects Associated with Exposure to Vehicular Traffic. J Expos Sci Environ Epidem 18: 588-599 (2008).

Lipfert FW, Wyzga RE, Baty JD, Miller JP. Air pollution and survival within the Washington University-EPRI Veterans Cohort: risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. J Air Waste Manag Assoc. 2009 59:473-89.

Lipfert FW, Wyzga RE. Revisiting the Veterans Cohort Mortality Study: New results and synthesis. J Air Waste Manag Assoc. 2018 Nov;68(11):1248-1268.

Lipfert FW, Wyzga RE. Environmental Predictors of Survival in a Cohort of U.S. Military Veterans: A Multi-level Spatio-temporal Analysis Stratified by Race. Envir Res (accepted 2019).

<u>Trend analysis</u>

Lipfert FW (1998), Trends in Airborne Particulate Matter in the United States. Applied Occupational and Environmental Hygiene 13:370-384

Lipfert FW, Morris SC, Temporal and spatial relationships between age-specific mortality and ambient air quality in the United States: Regression results for counties, 1960-97, Occup Env Med 59:156-74 (2002).

This paper used population data to show how mortality risks associated with air pollution decline with age and are stronger when mortality and exposure are measured coincidentally during a given follow-up period.

Review papers.

Lipfert, F.W., Commentary on the HEI Reanalyses of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, J. Toxicology and Environmental Health 66:1705-14 (2003).

Lipfert FW. An assessment of air pollution exposure information for health studies. *Atmosphere* 2015, *6*, 1736-1752.
Detailed analyses of indoor air quality data showing indoor infiltration rates and short- but not long-term relationships between indoor and outdoor concentration.
Lipfert FW. Long-Term Associations of morbidity with air pollution: A catalogue and synthesis. JAWMA 2018 68:12-26.
This catalog lists 417 long-term studies of cardiovascular, respiratory, cancer, diabetes, hospitalization, neurological, and pregnancy-birth endpoints with respect to $PM_{2.5}$, PM_{10}, TSP, carbon, ozone, sulfur, vehicular traffic, radon, and indoor air quality. Durations of exposure range from 60 d to 35 y. Most studies were cross-sectional over limited time spans with no consideration of lag or disease latency. 220 studies indicated significant adverse effects. Associations with cardiovascular indicators, lung function, respiratory symptoms, and low birth-weight were more likely to be significant than with disease incidence, heart attacks, diabetes, or neurological endpoints. Elemental carbon (EC), and NO_x were most likely to be significant for cardiovascular outcomes. For all diagnoses, the highest fractions of significant findings were for "other" pollutants including TSP, EC, and metals. $PM_{2.5}$ and PM_{10} had lower risk ratios. Overall, these studies support non-lethal physiological effects of various pollutants, more so for non-life-threatening endpoints and for non-criteria pollutants (TSP, EC, $PM_{2.5}$ metals).

Editorial comments

Lipfert FW. Clean air skepticism. Science. 1997 Oct 3;278(5335):19-20.

Lipfert FW. Air pollution and life expectancy. N Engl J Med. 2009 May 7;360(19):2033.

Lipfert FW. Air pollution and life expectancy. Epidemiology. 2014 Sep;25(5):776-7.

Lipfert FW. Letter to the Editor Re: Enstrom JE. Fine particulate and total mortality in Cancer Prevention Study cohort reanalysis. Dose-Response. 22;16(1):1559325817746304.

Lipfert FW. Re: Chemical Composition of Fine Particulate Matter and Life Expectancy in 95 US Counties. Epidemiology. 2017 Mar;28(2):e18.

Lipfert FW. Air Pollution and Mortality in the Medicare Population. JAMA. 2018 May 22;319(20):2133-2134.

Lipfert FW, Explain ill effects of airborne particles. Nature 2019 570:446.

Responses to Questions from Dr. Lange

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

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.

2) *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.*

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.

3) *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?

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).

(4) Results of Risk Analysis

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)*

Re: Table 1. "Causal Relationships" vary by PM composition, disease, age, race, season, and exposure time. Most of these factors have not been explored or discussed in the PA. "Air Monitoring" is inadequate because indoor exposures have not been considered and the databases for daily values and PM constituents are too sparse. There is virtually no ambient data on the ultrafine particles that can enter the bloodstream. I did not respond to questions on the present "C-R functions" because I do not find them credible, as discussed above.

(5) 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?

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. $PM_{2.5}$ 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 $PM_{2.5}$ constituents.

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

In general, the database for PM constituents is much more sparse than for $PM_{2.5}$ *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:

Olstrup H, Johansson C, Forsberg B. The Use of Carbonaceous Particle Exposure Metrics in Health Impact Calculations. Int J Environ Res Public Health. 2016 Feb 24;13(3). pii: E249.

Kim SY, Dutton SJ, Sheppard L et al. The short-term association of selected components of fine particulate matter and mortality in the Denver Aerosol Sources and Health (DASH) study. Environ Health. 2015 Jun 6;14:49.

Atkinson RW, Analitis A, Samoli E et al. Short-term exposure to traffic-related air pollution and daily mortality in London, UK. J Expo Sci Environ Epidemiol. 2016 Mar-Apr;26(2):125-32.

Ozkaynak H, Baxter LK, Dionisio KL, Burke J. Air pollution exposure prediction approaches used in air pollution epidemiology studies. J Expo Sci Environ Epidemiol. 2013 Nov-Dec;23(6):566-72.

Cakmak S, Dales RE, Rubio MA, Vidal CB. The risk of dying on days of higher air pollution among the socially disadvantaged elderly. Environ Res. 2011 Apr;111(3):388-93.

Mar TF, Norris GA, Koenig JQ, Larson TV. Associations between air pollution and mortality in Phoenix, 1995-1997. Environ Health Perspect. 2000 Apr;108(4):347-53.

Zhou J, Ito K, Lall R, Lippmann M, Thurston G. Time-series analysis of mortality effects of fine particulate matter components in Detroit and Seattle. Environ Health Perspect. 2011 Apr;119(4):461-6.

Ostro B, Roth L, Malig B, Marty M. The effects of fine particle components on respiratory hospital admissions in children. Environ Health Perspect. 2009 Mar;117(3):475-80.

Ostro BD, Feng WY, Broadwin R et al. The impact of components of fine particulate matter on cardiovascular mortality in susceptible subpopulations. Occup Environ Med. 2008 Nov;65(11):750-6.

Ostro B, Feng WY, Broadwin R, Green S, Lipsett M. The effects of components of fine particulate air pollution on mortality in California: results from CALFINE. Environ Health Perspect. 2007 Jan;115(1):13-9.

Sarnat SE, Suh HH, Coull BA et al. Ambient particulate air pollution and cardiac arrhythmia in a panel of older adults in Steubenville, Ohio. Occup Environ Med. 2006 Oct;63(10):700-6.

Burnett RT, Brook J, Dann T et al. Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. Inhal Toxicol. 2000;12 Suppl 4:15-39.

Klemm, R.J., Lipfert, F.W., Wyzga, R.E., Gust, C. Daily mortality and air pollution in Atlanta: Two years of data from ARIES. Inhal Toxicology 16 (Suppl 1):131-41 (2004).

Lipfert, F.W., Wyzga, R.E., Baty, J.D., Miller, J.P., 2006b. $PM_{2.5}$ Constituents and Related Air Quality Variables as Predictors of Survival in a Cohort of U.S. Military Veterans, Inhalation Toxicology 18:645-57.

Lipfert FW, Wyzga RE, Baty JD, Miller JP. Air pollution and survival within the Washington University-EPRI Veterans Cohort: risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. J Air Waste Manag Assoc. 2009 59:473-89.

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 $PM_{2.5}$ 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.

(7) *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?*

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*.

Responses to Questions from Dr. Packham

(1) *Are there areas (e.g., specific aspects of biological causation, pulmonary toxicology, or causality in epidemiology) in which additional knowledge is required to appraise the adequacy and basis of existing, new, or revised PM NAAQS to protect public health with an adequate margin of safety?*

Biological causation has not been addressed by health endpoint. A critical issue is translocation of PM from lung to bloodstream according to particle size and whether $PM_{2.5}$ as measured in the ambient is capable of doing so. Health effects centered in the respiratory system must be evaluated in terms of deposition, clearance, and importance of cumulative deposition. All of this must be characterized by PM constituent; if only EC could be shown to be harmful, then control strategies and accountability must focus on EC. This is especially important for ultrafines (UFP), for which exposures are essentially unknown.

The findings of strong effects of traffic density (the 3rd type of C-R mentioned in the Introduction) should be replicated in other cohorts. Such data are readily available.

(2) *Can you suggest additional specific scientific disciplines and areas of biomedical informatics and research (e.g., systems biology methods for clarifying biological causal pathways, mechanisms, modes of action, quantitative causal dose-response relationships) that should be included in future reviews of other criteria pollutants.*

Causality, speciation, and accountability go together. Short-term causality was shown in London 1952 by autopsy (Bradley, 1957) that indicated black carbon deposits. Accountability in short term effects is shown by synchronicity of exposure and response after controlling for weather, the only rational confounder. No such demonstrations have been reported for long-term effects. Animal tests in the 1970s were inconclusive and there have been no sufficiently long morbidity experiments that might demonstrate associations between $PM_{2.5}$ as measured in the ambient and development of atherosclerosis, for example. Human (or animal) experiments on concentrated particles (CAPS) in previous decades should be reviewed and possibly extended to PM components or UFPs.

(3) *Can you describe the research efforts necessary to provide the required information?*

Long-term morbidity experiments or evaluation of associations with long-term hospitalizations, nursing home stays, etc. Occupational studies on PM components, especially EC, OC, SO_4, NO_3. Existing cohort

studies should be reevaluated using lags and cumulative exposures. Within- and between-city data on indoor air quality are needed (See Figure 1, above).

(4) *To what extent has the needed research already been done, or started? For example, are there crucial experiments or research initiatives that could clarify the shape of the PM2.5-chronic inflammation causal dose-response relationships at relevant exposure concentrations?*

Not that I'm aware of.

(5) *Are there specific data analyses (e.g., testing for confounding by weather variables over more days prior to mortality) that could clarify the causal interpretation of epidemiological associations relied on in the draft PA to simulate effects of interventions?*

Yes. See sensitivity studies of temperature and air pollution lags in Murray and Lipfert (2012, 2013). Those studies should be replicated for other times and places. Short-term studies essentially comprise "interventions" (caused by atmospheric changes, week-days vs. weekends).

(6) *Can you suggest additional areas of scientific literature review on species and individual human organism's capacities of adaptation to inhaled environmental stressors that might help establish margins of safety when exposed to ambient levels of air pollution?*

Respiratory effects of cumulative exposures are well known for smoking and occupational exposures. Parallel information is needed for CVD, stroke, etc.

(7) Could you provide additional information on the relative contributions to air pollution concentrations and resulting health effects of natural and anthropogenic activity? For example, is either one alone, or are both natural and anthropogenic activities together, sufficient to cause the magnitudes of adverse health effects attributed to PM2.5 in the Draft PA?

This requires PM speciation and is important for indoor sources and wood smoke, including wildfires.

(8) *Could you provide additional information on any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards?* No.

Discussion of Issues

A. Causality

A causal relationship between exposure and response is required to justify regulation intended to benefit public health. Causality in air pollution epidemiology must rest on five requirements, none of which has been established for $PM_{2.5}$. The purpose of this exposition is not to demolish the existing regulatory structure but to advocate for reframing it by addressing each of the 5 points.

1. Exposure. Indoor exposures and daily peaks have not been considered in long-term studies. Initiation of chronic disease only occurs after a latency period and responds to cumulative exposures. This has been established for smoking but not for air pollution epidemiology.
2. Toxicity. $PM_{2.5}$ is not really a pollutant in the sense of defined chemicals (CO, O_3, SO_2). It is a regulatory construct based on mixture of many kinds of particles, some toxic, others not. These mixtures will have varied or time and among locations. Given the multiplicity of health endpoints that have been considered, some statistically significant combinations with a PM constituent are likely.
3. Translocation. Inhalation of a pollutant is not sufficient to imply contact with an organ and initiation of chronic disease. $PM_{2.5}$ translocation from the lung to the bloodstream has not been demonstrated, by contrast with ultrafine particles, for example.
4. Susceptibility. Given the diversity of the population, the notion that a healthy person could be randomly selected to die coincidently with normal levels of ambient air quality is patently absurd. Time-series studies have shown that acute mortality is associated with both underlying frailty and daily air pollution peaks. Such differential susceptibility is also likely the case for incidence of chronic disease, but cohort members would have to be tracked during the disease latency period to make this case.
5. Accountability. The ultimate test of causality is whether public health has actually improved in response to reduced $PM_{2.5}$ (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; Henneman et al. (2019) showed no effect on total mortality (-0.011 [-0.92 – 0.71]) after reducing $PM_{2.5}$ from 10.0 $\mu g/m^3$ in 2005 $\mu g/m^3$ to 7.2 in 2012. Note that their analysis did not account for latency or cumulative exposures.

By contrast, the epidemiology of smoking has all of these elements (Lipfert, 2019). Exposure is self-reported on an individual basis, and the importance of cumulative exposure is shown by the use of pack-years as a predictor of disease after accounting for latency. Toxicity of smoke constituents has been determined by laboratory and animal tests, especially for benzo(alpha)pyrene (B[a)P) which is also an (unmonitored) ambient air pollutant. Translocation is not an issue for the gaseous pollutants and the most serious health effects (cancer, COPD) occur in the lung. Susceptibility comes into play when frail smokers are urged to quit. Accountability is shown by the improved health of quitters.

B. Interactions between short- and long-term air pollution effects

Consider time-lines for daily deaths, frailty (morbidity), and air pollution (AP) exposure, for a population with no secular or seasonal trends over a period of decades. Each parameter progresses through daily peaks and valleys over time. When death and air quality peaks preceded by sufficient frailty coincide, a sudden AP death occurs. That person is replaced by a healthy person and the process continues along the time-lines. Frailty may also be affected by AP. The remaining deaths and air quality exposures accumulate, also with replacement. However, deaths would not be replaced in cohort analyses, leading to depletion of the most susceptible subjects. A non-peak AP death occurs when exposure accumulates to a level sufficient to initiate a specific disease. The total AP death count for the overall period includes peak and non-peak AP deaths.

This process is carried out in other locations and their total death counts are compared with their accumulated exposures to derive the ratio of total AP deaths to exposure for the group using cross-sectional analysis. Each AP death would have been preceded by an exposure of duration depending on frailty level and cause of death. The short-term peak deaths are of separate interest because they are limited to the most-frail subjects and do not constitute incidence of a new disease. Incidence of cancer or COPD may require 20-30 y exposures, cardiovascular diseases somewhat less. There may be other causes of death associated with intermediate exposure periods about which we have no information. These exposure differences suggest cause-specific analyses, each with its own exposure timing.

This scenario is straightforward, absent secular trends. However, air quality and medical care have been improving since about 1970, and relevant exposures for deaths in say, 2000, may reach back to the 1970s when air quality was much worse. The upshot is that we can't estimate appropriate long-term AP exposures for all-cause deaths, but they are surely higher than exposures at the time of death. Corresponding risk coefficients would thus be reduced. Table 1 compares the attributes and shortcomings of short- and long-term studies.

Table 1. Characteristics of time-series and cross-sectional mortality studies

	time-series	long-term populations	long-term cohorts
dependent variable	death count	mortality rate	dead (yes, no)
subjects	subpopulations	populations	individuals
likely causes	sudden death	---------- ischemic heart disease ----------	
confounders	weather climate	city characteristics contextual variables	personal characteristics holidays
exposures	daily	cumulative annual av'g	cumulative annual av'g
timing	lag days	after latent period	after latent period
indoor exposure effects	attenuation	confounding, bias	confounding, bias
secular trend	accounted for	exposure bias	cohort depletion by age
threshold	frailty dependent ----------obscured by measurement error----------		

Issues of exposure timing

Above I defined three types of C-R relationships that may be involved in a given situation:

(a) Short-term (days) fluctuations in air quality that may trigger sudden death in vulnerable individuals. Causality in this mode has been demonstrated by heat wave mortality.

(b) Long-term (decades) individual differences that affect incidence of new cases of chronic disease, for which the extended effects of cigarette smoking are a causal example.

(c) Long-term differences among populations that may involve both of the above as well as attributes of their locations *per se*, such as climate, green spaces, traffic density, or characteristics of housing stock that affect indoor air quality.

The sum of pollution-related deaths from (a) and (b) comprises the total long-term mortality effect, as derived from cross-sectional cohort analysis and reported by most of the $PM_{2.5}$ analyses used in the PA. However, there are no short-term effects involved with the contextual variables in type (c) and those effects are likely to have been stable over time so that cumulative exposures introduce no bias. The Veterans Cohort Study (Lipfert, et al., 2006 a,b) introduced the traffic density variable, the logarithm of which was a highly significant predictor of all-cause mortality over the 26-y follow-up period, better

than either population or housing densities. In a 2-pollutant model, both traffic density and ozone were significant, comprising a combination of all three C-R modalities. However, ozone effects were sensitive to exposure timing (Lipfert et al., 2000).

Including additional lags in a short-term study may attribute more deaths to the previous exposures, thus *in*creasing the risk coefficient. By contrast, considering prior cumulative exposures in a long-term study increases the *exposure* associated with the observed mortality, thus decreasing the risk coefficient. The effects of stable contextual variables are not sensitive to timing issues for long-term residents of a given location (e.g., county).

C. Summary

The current EPA paradigm linking mortality from chronic diseases with long-term air pollution exposure does not comport with observations. Here's what's been observed:

1. Exposure to high levels of black smoke and SO_2 for a few days has been linked with excess mortality by autopsies that show lung deposits of carbonaceous material. Those identified victims had suffered from preexisting cardiopulmonary diseases. Causality was supported by mortality levels returning to normal after air quality returned to normal.
2. Responses to daily fluctuations are included in annual variations.
3. Chronic effects should be characterized by cumulative exposures, similar to observations of smoking effects based on pack-years.
4. The timing of response following exposure depends on latency that varies by disease.
5. Adults typically spend 85-90% of their time indoors.
6. Long-term levels of outdoor air quality can be worse than indoors because of unregulated indoor sources that vary substantially within cities.
7. Short-term (daily) fluctuations in indoor air quality follow the outdoors with concentrations reduced by about 50%.
8. Some pollutants (O_3, SO_2) are adsorbed onto indoor surfaces and are thus not at issue.
9. Not all potentially important pollutants have been monitored adequately.
10. Frailty plays an important role in short-term health effects.
11. Ischemic heart disease (IHD) has been associated with long-term exposure to air pollution. Sudden death comprises about half of IHD and thus could be an outcome of short-term exposures consistent with relationships between the long- and short-term effects.

This is what has *not* been observed:

12. Deaths attributed to long-term exposure verified by autopsy.
13. Sufficient data on personal exposures to airborne pollutants adequate for epidemiology.
14. Translocation of inhaled particles from lung to blood stream.

15. Long-term morbidity effects that could support pollution-related incidence of chronic disease or onset of frailty.

The following may be inferred from statistical modeling:

16. Time-series analyses show similar short-term relationships with various pollutants including black smoke and other PM constituents, worldwide.
17. Short-term effects may persist for several days; total responses are given by the sum over the lag period.
18. A new time-series mortality model that includes frailty in conjunction with air quality shows that healthy individuals are not affected by air quality fluctuations. Since there is no upper limit on the severity of frailty (life expectancies can be as short as one day), there can be no short-term air quality threshold.
19. Remaining life expectancy is a better outcome metric than numbers of lives "saved" and has been estimated to range from days to months (but not years).
20. The validity of long-term dose-response modeling depends on controlling a wide variety of potential confounders including contextual variables that describe places of residence.
21. Long-term air pollution studies have inadequately controlled confounding including climate and personal income for cohort studies and smoking, obesity, and poverty for population-based studies.
22. Those confounders are not important for short-term effects.
23. Mortality differentials have been associated with places *per se* such as green spaces or traffic hot spots rather any accompanying air pollution exposure differentials. Such long-term associations may be affected by age but not by timing, confounded by indoor air pollution, nor limited by the available ambient air quality monitoring.

Here's what can be concluded from the above:

24. Short-term effects have been substantially underestimated since most have been based on 1-2 day average exposures rather than sums over lags. Short-term indoor exposures are attenuated, thus increasing their risk coefficients.
25. Long-term effects have been overestimated since cumulative exposures are much higher and latency pushes the applicable air quality levels back in time when they were much worse.
26. Any true long-term effects must be based on differences between estimates derived from annual vs. daily averages and would thus be problematic after adjusting for the above.
27. Long-term personal exposures are unknown because of variable contributions from indoor sources.
28. Responsible pollutants cannot be confirmed because of inadequate ambient monitoring.
29. Increased research on short-term effects will require substantial expansion of daily monitoring.
30. More attention should be given to ultrafine particles that can enter the blood stream.

31. "Safe" air quality levels cannot be defined because of the controlling effects of physical frailty. Society must judge how much (non-zero) risk can be tolerated.
32. Accountability for long-term air pollution abatement is problematic due to lingering effects of prior cumulative exposures.
33. Long-term health effects associated with places of residence rather air pollution exposures *per se* are not subject to these limitations and are thus more rational.
34. Future research should try to account for these limitations and focus on pollution-related morbidity that may lead to frailty and short-term effects.

References are listed after Appendix C

Appendix A. Background Issues

1.1 The potential lethality of ambient air pollution was established by the 1952 London fog disaster though autopsies that found black carbon deposits in the lungs of victims. Subsequent time-series analyses at various locations during normal atmospheric conditions found associations between mortality and daily air pollution peaks over several days, some of which also involved black carbon. Summing those deaths over lag periods provides larger estimates of excess deaths than previously reported. Time-series analyses using econometric methods found that these excess deaths were limited to previously frail subpopulations rather than the population at large. These procedures are consistent with analyses of heat wave mortality, when victims were also limited to the frail. Short-term effects are thus controlled by frailty as well as environmental conditions and are not mutually exclusive. A sufficiently frail subject may succumb to minor pollutant exposures and vice-versa, precluding observation of thresholds in either. Short-term effects have been well established; proof of existence through statistical testing is no longer required.

1.2 Particulate pollutants have been measured in various ways over time, beginning with filter staining, referred to as British smoke, black smoke, smokeshade, or coefficient of haze (COH) and equivalent to elemental carbon mass (EC). Early U.S. data were measured in terms of the total mass collected on filters: total suspended particulate matter. Atmospheric visibility has also been used. Sulfate content was also reported then but those early readings are suspect because gaseous SO_2 reacted passing through the alkaline filters, creating artifact sulfates. Beginning ca. 1980 better sampling methods were devised, with segregation by particle size, originally $PM_{3.5}$ and PM_{15} and later changed to $PM_{2.5}$ and PM_{10}. Daily TSP ranked highest in a comprehensive comparison of PM metrics in Philadelphia (Lipfert et al., 2000).

1.3 $PM_{2.5}$ infiltrates indoors but not uniformly (Sarnat et al., 2012). Penetration varies by season (and presumably climate), time of day, composition, and particle size. Submicron particles have the largest indoor/outdoor ratio, peaking at about 0.3 μm; EC penetrates indoors better than $PM_{2.5}$. Part of the focus on $PM_{2.5}$ came from the Acid Precipitaion Program that posited that acidity (H^+) of filter deposits was a human health hazard. Subsequent epidemiology ruled this out but concern about $PM_{2.5}$ remained, originally about respiratory effects that had been demonstrated in the Harvard Six Cities program. Based on subsequent epidemiology, emphasis shifter to heart disease for which mechanistic support remains incomplete.

1.4 Estimates of long-term effects of air pollution on specific cohorts or populations at large have been limited to cross-sectional comparisons across locations for which outdoor ambient air quality levels were estimated. These associations have been interpreted as evidence for causing new cases of chronic disease; however, neither latency periods nor cumulative exposures have been considered

nor have pollutants toxic to specific organs been identified. These cross-sectional analyses are thus inconsistent with well-established protocols for analyzing smoking effects (Lipfert, 2019). By contrast with effect of smoking, death by long-term air pollution exposure has never been confirmed by autopsy. In addition, there have been no sufficiently long-term morbidity or animal studies to provide support for the chronic effect hypothesis.

1.5 The timing of exposures relative to responses is important. Exposure must precede response, as has been shown in time-series studies. Long-term studies have been based on annual averages, but those averages *include* the daily peaks for which short-term responses were estimated separately. Any true long-term effects must thus be defined as *differences* between those based on annual averages and those based on daily peaks. Short-term effects should be summed over lags and long-term effects should account for latency and cumulative exposures. These fundamental requirements have not been recognized in the extant literature.

1.6 Only outdoor air pollution has been assumed to be relevant even though we spend ~85% of our time indoors, more for the severely frail. Indoor air pollution was not addressed in the Clean Air Act, is not regulated, and is thus not routinely monitored. In much of the United States, outdoor air may be dirtier than indoors for some pollutants, in which case further improvements in outdoor air quality would be moot. EPA has a separate entity for indoor air quality but those concerns have not been recognized in epidemiology and differ between short- and long-term exposures. About half of outdoor air pollution infiltrates indoors with the same frequency distributions; short-term effects may thus be experienced indoors but with larger risk coefficients. Long-term effects are dominated by unregulated indoor sources including environmental tobacco smoke, gas stoves, pets, candles, hobbies, etc., for which few data are available. Cross-sectionally, there are no long-term relationships between indoor and outdoor air quality, especially for particulate matter that has many types of indoor sources. The victims of air pollution occupy only a tiny fraction of a community's residences, practically precluding estimating their indoor exposures.

1.7 The implications of future climate change and of the likely efforts to ameliorate it should also be considered. There may be less vehicle exhaust but not tire and road dust. Rising temperatures will bring increased demand for residential air conditioning. The accompanying reduced infiltration of outside air will exacerbate effects of indoor air pollution sources.

1.8 The selection of criteria pollutants has involved circularity: The agency suspects health effects for a given pollutant based on prior information. It then deploys a monitoring network designed solely for that pollutant. The resulting data are used in epidemiology studies that find significant effects for that pollutant (alone), thus justifying the original decision. However, that decision would have been made in the absence of parallel counterfactual information for comparison, which would have required

monitoring other candidate pollutants in parallel. Epidemiology can only consider the pollutants that have been measured that are typically designed to support the *a priori* decision to regulate.

A good example involves constituents of $PM_{2.5}$, such as metals, carbon compounds (black smoke), or ultrafines (UFPs). The limited information available, including from foreign sources, suggests that each of these alternatives may be more toxic than $PM_{2.5}$ per se. But lacking parallel monitoring data, especially on a daily, these hypotheses cannot be tested. This lack of more specific information is critical in designing abatement strategies. There are no sources that directly emit the $PM_{2.5}$ used in epidemiology studies; they only emit specific constituents, making it difficult to design an efficient abatement strategy. This may help explain the lack of successful $PM_{2.5}$ accountability studies.

1.9 How can these concepts be reconciled with the extensive archive of existing studies that found significant risks? The subdivision of the total mortality effect into short-term and differential (true) long-term components does not invalidate the estimated total risks from existing epidemiology studies but drastically changes their interpretation. The total risk remains the same (those existing studies are not "wrong") but a portion of the risk must be ascribed to short-term daily effects, discussed below under "causality". Finding a threshold in the short-term portion is extremely unlikely since in a sufficiently large population during a sufficiently long lag period there will always be at least one person sufficiently frail so as to succumb to a minor excursion in ambient air quality. The difference between total cross-sectional risk and the acute portion of the risk cannot be assumed to have resulted in chronic disease incidence without considering cumulative exposures prior to the latency period. These determinations have not been made. Lacking defined thresholds, society must decide how much risk it will tolerate.

2.0 All epidemiology is concerned with both time and place. Results from time-series analyses are more credible when those from different cities agree. However, most long-term (i.e., spatial) studies have been limited to fixed time periods; longitudinal studies are rare but are required to establish accountability for the predicted benefits of improved air quality. Only one study has evaluated the effects of cohort aging separately from temporal trends (Lipfert and Wyzga, 2018). Cohort studies often give the impression that they are about people, but exposure to ambient air quality inevitably involves geography. In the absence of personal exposure data, the extant long-term cohort studies are about *places* rather than their residents. They cannot identify individual risks but may indicate that living in City A may be riskier than in City B for any of several possible reasons other than ambient air quality:

- climate, which can affect time spent outdoors and exposure to extreme weather conditions
- physical characteristics such as green space or traffic density
- differences in housing stock such as high-rises or dilapidated construction.

Literal interpretation of a cross-sectional study implies that moving from City A to City B would instantly confer health benefits. However, as shown by the delay involved in benefiting from smoking cessation, this could only be the case with acute effects.

2.1 The Clean Air Act requires that thresholds be identified with an "adequate margin of safety". Such thresholds have not been identified for particulate matter, for any or all of the following reasons:

- Underestimated exposures (indoor sources)
- PM constituents that change with $PM_{2.5}$ concentration level
- Mixtures involving different copollutants
- Daily effects on frail subjects that are embedded with long-term effects
- Measurement errors that bias the slope of the dose-response function downward.

This situation was inconceivable to the framers of the 1970 Clean Act or even its subsequent amendments, but it suggests that society may now have to decide how much risk may be tolerated; zero risk can never be achieved or maintained.

Appendix B. Air pollution epidemiology

Acute exposures

Following the disastrous London fog of 1952 that increased daily and air pollution mortality rates by an order of magnitude, accountability was obtained by autopsy. Under these conditions, the probability of having randomly selected a victim of the episode was around 90%, by contrast with current conditions (fortunately). The working hypothesis was that this extreme excess mortality had been caused by

a. black smoke.
b. SO_2.
c. dense fog.
d. adverse ambient temperature levels.
e. Combinations of the above.

Air quality including indoors was so bad that it was reported that cinemas had to close because the projected images could not reach the screen. This episode lasted for a few days and the excess death count was the sum over that period.

At that time, emissions from traffic (NO_x, CO, VOCs) would not have been involved because travel was curtailed by the dense fog. Ozone would not have formed due to the absence of sunlight. Particulate matter would have comprised elemental carbon (EC), organic compounds from incomplete coal combustion, and sulfates, and would have been ascertained by filter staining rather than by mass. Heavy metals such as Pb would not have been an issue. Autopsies found lungs that contained black carbon particles which then became the prime suspect although sulfates could not be ruled out. These sulfate compounds tended to be acidic and soluble and may have been overwhelmed the natural ammonia defenses. Individual victims could be identified with relative certainty and they were found to largely comprise persons with preexisting cardiorespiratory conditions.

The working hypothesis was then supported by analysis of subsequent but milder episodes in London, New York City and elsewhere. As air quality improved such episodes became rare and continuous time-series analyses were performed by OLS regression over periods long enough to attain statistical significance. They were limited to SO_2 and black smoke, the only pollutants monitored daily at that time. The most comprehensive such analysis was published by Schimmel (1980) and showed that black smoke was the likely culprit over 14 y in New York City. Heavy residual oil containing metals was the main fuel. Excess mortality was only of the order of a few percent, precluding the identification of specific victims.

Hundreds of similar daily mortality and hospitalization studies followed, limited mainly by the availability of daily air quality data. In the United States, particulate matter was then ascertained by the mass collected on filters without regard to particle size; sulfur was the only PM constituent identified. The only classification of victims was by age group, usually aged 65 and over. Allowance for coincident weather effects such as heat waves is also important, including lagged effects. The most recent nationwide analysis of daily mortality and $PM_{2.5}$ (Di et al., 2018) considered only the average of lags 0,1, thus underestimating the effects by about a factor of 3 that would have yielded a risk of 3.3% per 10 $\mu g/m^3$. Blomberg et al. (2019) investigated interactions between radon and $PM_{2.5}$ and estimated the $PM_{2.5}$ effect at about 2.8% per 10 $\mu g/m^3$ at the mean radon level but lag assumptions were not specified.

Around 2000, a new time-series methodology appeared that identified the subpopulations most at risk using Kalman filter methods as developed by econometricians. It postulates the existence of a fluctuating group of extremely frail elderly that is depleted by daily deaths and augmented by newly frail individuals. The life expectancies of such groups were of the order of weeks; the fractions of deaths associated with air pollution were similar to those from conventional time-series. Kundi et al. (2019) noted that prediction of post-hospital cardiovascular outcomes was significantly improved by considering a "frailty score" index. The important feature of the new model is that excess deaths are driven by the product of the population at risk and the air pollution level. An excess death can result from either extreme frailty or bad air quality or both. As a result, *there is no threshold for air pollution exposure alone*. Wen (2017) reported mortality hazard ratios around 5.0 for severe frailty while air pollution ratios are typically around 1.05.

This model has been applied to Philadelphia, Chicago, and Atlanta, with similar findings for several pollutants (Murray and Lipfert, 2012, 2013). For example, we typically found that these frail subpopulations had remaining life expectancies of a few weeks, a small fraction of which was associated with air pollution with about 2% excess mortality. For comparison, assuming that life expectancies for those aged 65 ranged from 1 day to 30 y with a log-normal distribution the bottom 2% of which would have a life expectancy of 24 days, in rough agreement with the new model. Another element of this methodology is the extent to which air pollution may been involved in worsening the onset of the severe frailty required for an acute. This possibility has been explored but not resolved. No other investigators have since adopted this new modeling protocol.

Another important aspect of time-series analysis is that of delayed responses. Some studies have considered lag periods up to 7 days, but most of them were aimed at identifying the lag days of greatest statistical significance, as if to establish the *existence* of such relationships. However, the validity of these findings was established by the 1952 London event; the magnitudes of subsequent events are now of primary interest based on excess deaths accumulated over the lag period as was the case in London. It

thus appears that most of the extant literature may have underestimated the severity of short-term effects on mortality, perhaps by a factor of 2-3.

Long-term exposures.

Long-term studies are based on annual average concentrations over various aggregation or follow-up periods. Such studies began in the 1970s based on differences in annual mortality rates among the populations of various cities or counties, for which personal data such as smoking habits or socioeconomic status were not available. Beginning with the Harvard Six Cities Study (Dockery et al., 1993), mortality rates were studied in defined cohorts having such personal data. However, no published long-term study has recognized that the deaths accumulated over a year in a given location include those precipitated by daily peak exposures, as discussed above. The true long-term effect is thus the difference between long-term and short-term analyses. The largest long-term mortality risks have been ascribed to ischemic heart disease (IHD), which includes "sudden" death and heart attacks, thus showing overlap between acute and chronic causes of death.

For example, daily mortality across those 6 cities increased by 1.5 (1.1-1.9) percent per 10 $\mu g/m^3$ for a 2-day average of $PM_{2.5}$ (Schwartz et al., 1996). The cumulative effect for the 2 days was thus 3.0 (2.2-3.8)% which might increase to 4.5 (3.3-5.7)% when additional days are considered (Staniswallis et al. 2009). For comparison, the long-term effect of $PM_{2.5}$ for these cities was reported as 12 (4-21)% per 10 $\mu g/m^3$. Dockery et al. (1993) commented that cumulative exposures over prior years should have been considered (especially for lung cancer) and that not all confounding variables had been controlled (notably climate and personal income). Both of these considerations would reduce the long-term effect estimates and render the difference between long- and short-term effects nonsignificant.

In general, short-term mortality effects have been underestimated by not summing over sufficient lag periods, and long-term effects have been overestimated by not considering prior cumulative effects and a complete suite of potential confounders. Most daily analyses have focused on statistical significance for the most lethal day instead of the total "yield". In a long-term study, the risk coefficient is reduced when the true exposure is increased. When the early studies based on annual averages found risk estimates that exceeded those from daily studies, they assumed that the effect pertained to the incidence of chronic disease. However, this hypothesis has not been tested or confirmed.

Latency is a further issue in long-term studies. Lung cancer incidence lags smoking by 15-30 y, perhaps less for heart disease. Thus, mortality from 1982-8 studied by Pope et al. (2002) should have been linked to pollution data as far back as 1952 when urban air was far more polluted (Lipfert, 1994). The total mortality risk coefficient should have been around 2% per 10 $\mu g/m^3$, again not significantly different than typical short-term values when summed over lag periods. Latency may also explain why it has been difficult to establish accountability for air quality improvement. Just as the benefits of

smoking cessation appear only after a 5-10 y abstinence, similar lags may delay the effects of cleaner air.

Long-term studies are generally conducted over decades to increase the number of deaths and for stability. Doing so invokes temporal trends in air quality and medical care, during which a cohort may be depleted of its most susceptible subjects. Lipfert and Wyzga (2018) found that cohort aging was more important than secular trends and that risks decreased with age. Lipfert et al. (2002) used population data to also show this aged effect and found stronger risks when mortality and exposure were measured coincidentally rather than lagged.

Classification of air pollution exposures is an important issue seldom considered in the extant literature, especially for cohorts; exposures have been based on measurements or estimates of outdoor air quality at fixed times and locations. They represent characterizations of those locations (places) but not the personal exposures of inhabitants; by contrast with smoking habits, education, body mass, etc; they do not travel with the subject during follow-up. Cumulative exposures have not been considered, by contrast with smoking. Some long-term analyses have considered characteristics directly and found significant associations, notably "green" spaces and vehicular traffic density (VMT). The latter was considered extensively in the Veterans Cohort Study (the only cohort analysis to do so), where it was the most important predictor of mortality risk (as ln[VMT]) along with NO_x and elemental carbon (EC). The geographic distribution of ln(VMT) has been stable over decades; temporal trends are thus not at issue. VMT is a true geographic descriptor that could entail engine exhaust and lubricant residues, tire and road dust, or socioeconomic effects associated with undesirable residential locations. It is integrated over counties so that local changes in residence are not an issue nor are commuting patterns. VMT per se is not amenable to regulatory abatement and is likely to increase over time. However, the engine exhaust pollution should diminish over time with increased use of electric vehicles. In addition to neglecting personal exposures, epidemiology studies have not considered the ultimate fates of inhaled particles, most of which are retained in the lung (Churg and Brauer, 1997). Only ultrafine particles (UFP) are likely to translate into the bloodstream and travel to other organs (Wichmann et al., 2000). UFPs are a tiny fraction of $PM_{2.5}$ and thus studies that attribute incidence cardiovascular, cancer, diabetes, or neurological effects to $PM_{2.5}$ are ipso facto problematic.

Appendix C. Publications by Lipfert and colleagues not considered in the ISA, by subject area.

Measurement error and uncertainties

These papers examine the effects of bias and noise in independent variables used in linear regression, especially correlated air pollutants. They show that the pollutant with the least error prevails in co-pollutant models. Random measurement error biases the slope of an exposure response function toward the null, which then biases the x-intercept toward the origin and obscures any threshold that may have otherwise existed. ”Measurement error” may be random (instrumental) or the result of mis-specification or mis-location of monitors.

F.W. Lipfert and R.E. Wyzga (1995), Uncertainties in Identifying "Responsible" Pollutants in Observational Epidemiology Studies, Inhalation Toxicology 7:671-89.

F.W. Lipfert and R.E. Wyzga, Air Pollution and Mortality: Issues and Uncertainties, J.AWMA 45:949-66 (1995).

F.W. Lipfert and R.E. Wyzga (1997), Air Pollution and Mortality: The Implications of Uncertainties in Regression Modeling and Exposure Measurement, J.AWMA 47 517-523.

F.W. Lipfert and R.E. Wyzga (1998), Effects of Exposure Error on Environmental Epidemiology, proc. International Symposium on the Health Effects of Particulate Matter, Prague. AWMA Publ. VIP-80, pp. 155-166.

F. W. Lipfert and R.E. Wyzga, Statistical Considerations in Determining the Health Significance of Constituents of Airborne Particulate Matter, J.AWMA. 49:PM-182-191 (1999).

F.W. Lipfert, The Use and Misuse of Surrogate Variables in Environmental Epidemiology, J. Environmental Medicine 1:267-278 (1999).

A new time-series model

Time-series studies of daily mortality predict acute responses on the first few days of exposure, usually persisting less than a week. Responses are limited to the elderly and are stronger among the most susceptible subjects. Similar risks have been associated with various air pollutants and are strongest for cardiac and respiratory deaths. Murray and Lipfert (2010) considered a new model that predicts the magnitude of an elderly subpopulation most at risk of imminent death on a daily basis, rather than the general population. In this model two conditions are required for an acute air pollution-related death: increased frailty (inability to maintain homeostasis) and increased pollution exposure. Extreme values of either parameter can result in death, thus precluding air pollution thresholds. Acute frailty is represented by membership in a small fluctuating sub-population (P_t) at high risk that is depleted by daily deaths

(D_t) and replenished by new entries (N_t). D_t and N_t may be affected by current or previous environmental conditions. The ratio of P_t to D_t is a measure of survival time (i.e., daily life expectancy [DLE]) in the frail state. The Kalman filter is used to solve these equations. P_t fluctuates with season and averages ~0.08% of the elderly. Life expectancies of subjects in the at-risk pool are about 5-7 days and robust to alternative models and environmental fluctuations. Daily mortality risk estimates are similar to those from conventional time-series analyses, but information on the accompanying life expectancies provided by this model is new.

Murray CJ, Lipfert FW. Revisiting a Population-Dynamic Model of Air Pollution and Daily Mortality of the Elderly Population in Philadelphia. J Air Waste Manag Assoc. 2010 60:611-629.

Murray CJ, Lipfert FW. A new time-series methodology for estimating relationships between elderly frailty, remaining life expectancy, and ambient air quality. Inhalation Toxicology 2012 24:89-98.

Lipfert FW, Murray CJ. Air pollution and daily mortality: A new approach to an old problem. Atmos Environ 55; 467-74 (2012).

Murray CJ, Lipfert FW. Inferring frail life expectancies in Chicago from daily fluctuations in elderly mortality. Inhal Toxicol. 2013 Jul;25(8):461-79.

Other time-series papers.

These papers provided short-term risk estimates for numerous pollutants that have not been considered by others: a wide range of particulate measures in which TSP was the most important in Philaelphia, and a complete set of PM chemical components was consideeed kn Atlanta of which carbon compounds were the most important.

Lipfert, F.W., Morris, S.C., and Wyzga, R.E. (2000), Daily Mortality in the Philadelphia Metropolitan Area and Size-Classified Particulate Matter, J.AWMA 50:1501-13.

Klemm, R.J., Lipfert, F.W., Wyzga, R.E., Gust, C. Daily mortality and air pollution in Atlanta: Two years of data from ARIES. Inhal Toxicology 16 (Suppl 1):131-41 (2004).

The Veterans Cohort Study

The Veterans Cohort Mortality Study began as a cooperative venture with Washington University at St. Louis (WUSTL), who developed the cohort and originated the study. It is based on a cohort of about 70,000 male veterans who were recruited in 1976 and followed for 26 y This is an important period of study because of the substantial improvements in ambient air quality that were achieved following passage of the 1970 Clean Air Act. The cohort was 37.7% black and 81% had smoked at

one time; they were recruited at VA 32 clinics nationwide. it is the first cohort study to examine long-term all-cause mortality risks associated with air pollution for a variety of pollutants by race. The study considered the overall follow-up period (1976-2001) and four sequential mortality periods (1976-81, 1982-88,1989-96, and 1997-2001). The WUSTL proportional hazards model included non-linear terms for age, smoking, body-mass index (BMI), blood pressure, and interaction terms. The first paper showed that the relative importance of $PM_{2.5}$ and SO_4 was contingent on the inclusion of contextual predictor variables in the model including climate, poverty and neighborhood characteristics in the proportional hazards model. Those variables were retained in subsequent analyses. Later papers examined $PM_{2.5}$ constituents, hazardous pollutants, and county-level traffic density, which was the most important predictor along with NO_x and elemental carbon. The 2018 paper examined the effects of cohort aging with respect to temporal trends and found the former to be more important with stronger risks in middle age. $PM_{2.5}$ had consistently negative (beneficial) effects; no effects of accumulated exposures were evident. The most recent (2019) paper found that black veterans were more heavily exposed to traffic, with significantly negative effects of $PM_{2.5}$ and significantly positive risks for whites similar to those found in other cohorts. This explains the apparently anomalous (ngative) $PM_{2.5}$ effects for the entire cohort. It also supports the hypothesis that cross-sectional studies essentially consider differences among *places* per se (counties, zip-codes, etc.) rather than the personal exposures of residents. Traffic effects were found to be stable over the 26-y period; however, those actual risk agents remain unknown, comprising engine exhaust, noise, tire/road dust, undesirable residential locations, but not indoor air pollution.

Lipfert, F.W., Perry, H.M. Jr., Miller, J.P., Baty, J.D., Wyzga, R.E., Carmody, S.E. (2000) The Washington University-EPRI Veterans' Cohort Mortality Study: Preliminary Results, Inhalation Toxicology 12 (Suppl 4):41-73.

Lipfert, F.W., Perry, H.M. Jr., Miller, J.P., et al., 2003. Air Pollution, Blood Pressure, and Their Long-Term Associations with Mortality. Inhalation Toxicology 15, 493-512.

Lipfert, F.W., Wyzga, R.E., Baty, J.D., Miller, J.P., 2006a.Traffic Density as a Surrogate Measure of Environmental Exposures in Studies of Air Pollution Health Effects: Long-term Mortality in a Cohort of U.S. Veterans, Atmospheric Environment 40, 154-169.

Lipfert, F.W., Wyzga, R.E., Baty, J.D., Miller, J.P., 2006b. $PM_{2.5}$ Constituents and Related Air Quality Variables as Predictors of Survival in a Cohort of U.S. Military Veterans, Inhalation Toxicology 18:645-57.

F.W. Lipfert, R.E. Wyzga, Jack D. Baty, J. Philip Miller. Vehicular Traffic Effects on Survival within the Washington University - EPRI Veterans Cohort: New Estimates and Sensitivity Studies, Inhalation Toxicology 20:949-960 (2008a).

F.W. Lipfert, R.E. Wyzga On Exposure and Response Relationships for Health Effects Associated with Exposure to Vehicular Traffic. J Expos Sci Environ Epidem 18: 588-599 (2008b).

Lipfert FW, Wyzga RE, Baty JD, Miller JP. Air pollution and survival within the Washington University-EPRI Veterans Cohort: risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. J Air Waste Manag Assoc. 2009 59:473-89.

Lipfert FW, Wyzga RE. Revisiting the Veterans Cohort Mortality Study: New results and synthesis. J Air Waste Manag Assoc. 2018 Nov;68(11):1248-1268.

Lipfert FW, Wyzga RE. Environmental Predictors of Survival in a Cohort of U.S. Military Veterans: A Multi-level Spatio-temporal Analysis Stratified by Race. Envir Res (accepted 2019)

Trend analyses

Lipfert FW (1998), Trends in Airborne Particulate Matter in the United States. Applied Occupational and Environmental Hygiene 13:370-384

Lipfert FW, Morris SC, Temporal and spatial relationships between age-specific mortality and ambient air quality in the United States: Regression results for counties, 1960-97, Occup Env Med 59:156-74 (2002).

This paper used population data to show how mortality risks associated with air pollution decline with age and are stronger when mortality and exposure are measured coincidentally during a given follow-up period.

Review papers

Lipfert, F.W., Commentary on the HEI Reanalyses of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, J. Toxicology and Environmental Health 66:1705-14 (2003).

Lipfert FW. An assessment of air pollution exposure information for health studies. *Atmosphere* 2015, *6*, 1736-1752.

Detailed analyses of indoor air quality data showing indoor infiltration rates and short- but not long-term relationships between indoor and outdoor concentration.

Lipfert FW. Long-Term Associations of morbidity with air pollution: A catalogue and synthesis. JAWMA 2018 68:12-26.

The catalog comprises 417 long-term studies of cardiovascular, respiratory, cancer, diabetes, hospitalization, neurological, and pregnancy-birth endpoints with respect to $PM_{2.5}$, PM_{10}, TSP, carbon, ozone, sulfur, vehicular traffic, radon, and indoor air quality. Durations of exposure range from 60 d to 35 y. Most studies were cross-sectional over limited time spans with no consideration of lag or disease latency. 220 studies indicated significant adverse effects. Associations with cardiovascular indicators, lung function, respiratory symptoms, and low birth-weight were more likely to be significant than with disease incidence, heart attacks, diabetes, or neurological endpoints. Elemental carbon (EC), and NO_x were most likely to be significant for cardiovascular outcomes. For all diagnoses, the highest fractions of significant findings were for "other" pollutants including TSP, EC, and metals. $PM_{2.5}$ and PM_{10} had lower risk ratios. Overall, these studies support non-lethal physiological effects of various pollutants, more so for non-life-threatening endpoints and for non-criteria pollutants (TSP, EC, $PM_{2.5}$ metals).

Editorial comments

Lipfert FW. Clean air skepticism. Science. 1997 Oct 3;278(5335):19-20.

Lipfert FW. Air pollution and life expectancy. N Engl J Med. 2009 May 7;360(19):2033.

Lipfert FW. Air pollution and life expectancy. Epidemiology. 2014 Sep;25(5):776-7.

Lipfert FW. Letter to the Editor Re: Enstrom JE. Fine particulate and total mortality in Cancer Prevention Study cohort reanalysis. Dose-Response. 22;16(1):1559325817746304.

Lipfert FW. Re: Chemical Composition of Fine Particulate Matter and Life Expectancy in 95 US Counties. Epidemiology. 2017 Mar;28(2):e18.

Lipfert FW. Air Pollution and Mortality in the Medicare Population. JAMA. 2018 May 22;319(20):2133-2134.

Lipfert FW, Explain ill effects of airborne particles. Nature 2019 570:446.

References

Baxter LK, Stallings C, Smith L, Burke J. Probabilistic estimation of residential air exchange rates for population-based human exposure modeling. J Expo Sci Environ Epidemiol. 2017 Mar;27(2):227-234.

Baxter LK, Sacks JD. Clustering cities with similar fine particulate matter exposure characteristics based on residential infiltration and in-vehicle commuting factors. Sci Total Environ. 2014 Feb 1;470-471:631-8.

Baxter LK, Burke J, Lunden M et al. Influence of human activity patterns, particle composition, and residential air exchange rates on modeled distributions of PM2.5 exposure compared with central-site monitoring data. J Expo Sci Environ Epidemiol. 2013 May-Jun;23(3):241-7.

Baxter LK, Wright RJ, Paciorek CJ et al.. Effects of exposure measurement error in the analysis of health effects from traffic-related air pollution. J Expo Sci Environ Epidemiol. 2010 Jan;20(1):101-11.

Baxter LK, Clougherty JE, Paciorek CJ et al. Predicting residential indoor concentrations of nitrogen dioxide, fine particulate matter, and elemental carbon using questionnaire and geographic information system based data. Atmos Environ (1994). 2007 Oct;41(31):6561-6571.

Baxter LK, Clougherty JE, Laden F, Levy JI. Predictors of concentrations of nitrogen dioxide, fine particulate matter, and particle constituents inside of lower socioeconomic status urban homes. J Expo Sci Environ Epidemiol. 2007 Aug;17(5):433-44.

Baxter LK, Crooks JL, Sacks JD. Influence of exposure differences on city-to-city heterogeneity in PM(2.5)-mortality associations in US cities. Environ Health. 2017 Jan 4;16(1):1.

Blomberg AJ, Coull BA, Jhun I et al. Effect modification of ambient particle mortality by radon: A time series analysis in 108 U.S. cities. J Air Waste Manag Assoc. 2019 Mar;69(3):266-276.

Bové H, Bongaerts E, Slenders E et al. Ambient black carbon particles reach the fetal side of human placenta. Nat Commun. 2019 Sep 17;10(1):3866.

Bradley WH, Logan WP, Martin AE. Mon Bull Minist Health Public Health Lab Serv. 1958 Jul;17:156-66.The London fog of December 2nd-5th, 1957.

Di Q, Dai L, Wang Y, Zanobetti A, Choirat C, Schwartz JD, Dominici F. Association of Short-term Exposure to Air Pollution With Mortality in Older Adults. JAMA. 2017 Dec 26;318(24):2446-2456.

Churg A, Brauer M. Human lung parenchyma retains PM2.5. Am J Respir Crit Care Med. 1997 Jun;155(6):2109-11.

Dockery DW, Cunningham J, Damokosh AI et al. Health effects of acid aerosols on North American children: respiratory symptoms. Environ Health Perspect. 1996 May;104(5):500-5.

Dockery DW, Pope CA 3rd, Xu X et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med. 1993 Dec 9;329(24):1753-9.

Henneman LRF, Choirat C, Zigler ACM. Accountability Assessment of Health Improvements in the United States Associated with Reduced Coal Emissions Between 2005 and 2012. Epidemiology. 2019 Jul;30(4):477-485.

Kundi H, Wadhera RK, Strom JB et al. Association of Frailty With 30-Day Outcomes for Acute Myocardial Infarction, Heart Failure, and Pneumonia Among Elderly Adults. JAMA Cardiol. 2019 Sep 25.

Lipfert FW. *Air Pollution and Community Health. A Critical Review and Data Sourcebook.* Wiley, New York. 1994.

Lipfert FW, Wyzga RE. Revisiting the Veterans Cohort Mortality Study: New results and synthesis. J Air Waste Manag Assoc. 2018 Nov;68(11):1248-1268.

Lipfert FW, Wyzga RE. Environmental predictors of survival in a cohort of u.s. military veterans: a multi-level spatio-temporal analysis stratified by race. Envir Res (accepted 2019)

Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA. 2002 Mar 6;287(9):1132-41.

Sarnat JA, Sarnat SE, Flanders WD et al. Spatiotemporally resolved air exchange rate as a modifier of acute air pollution-related morbidity in Atlanta. J Expo Sci Environ Epidemiol. 2013 Nov-Dec;23(6):606-15.

Sarnat SE, Raysoni AU, Li WW et al Air pollution and acute respiratory response in a panel of asthmatic children along the U.S.-Mexico border. Environ Health Perspect. 2012 Mar;120(3):437-44.

Schimmel H. Evidence for possible acute health effects of ambient air pollution from time series analysis: methodological questions and some new results based on New York City daily mortality, 1963-1976. Bull N Y Acad Med. 1978 Dec;54(11):1052-108.

Schwartz J, Dockery DW, Neas LM. Is daily mortality associated specifically with fine particles? J Waste Manag Assoc. 1996 Oct;46(10):927-939.

Staniswalis JG, Yang H, Li WW, Kelly KE. Using a continuous time lag to determine the associations between ambient PM2.5 hourly levels and daily mortality. J Air Waste Manag Assoc. 2009 Oct;59(10):1173-85.

Wen J, Shi G, Tian Y, Chen G et al. Source contributions to water-soluble organic carbon and water-insoluble organic carbon in PM(2.5) during Spring Festival, heating and non-heating seasons. Ecotoxicol Environ Saf. 2018 Nov 30;164:172-180.

Wichmann HE, Spix C, Tuch T, Wölke G, Peters A, Heinrich J, Kreyling WG, Heyder J. Daily mortality and fine and ultrafine particles in Erfurt, Germany part I: role of particle number and particle mass. Res Rep Health Eff Inst. 2000 Nov;(98):5-86; discussion 87-94.

Dr. D. Warner North

I have received questions from five CASAC members. I have had time to read much of the Policy Assessment, External Review Draft of September 2019, and also the CASAC letter report of April 11 and the Administrator's Reply of July 25.

I am providing a preliminary response stating my impressions based on my reading to date, with references to materials that I think CASAC members and other non-member consultants will find useful. I will welcome dialogue via e-mail among us. Part of my motivation for sending this response early is that I have a long-standing major commitment (a talk) on October 8, and I must spend much of my time between now and then preparing it rather than doing further reading and research as part of my recent engagement to help CASAC. But I believe we as a group will need to continue our dialogue via formal Questions and Responses, all made available to the public. The issues are deep. In my judgment, EPA's existing risk assessment process is in need of much correction from CASAC. That process has started with the review of the draft Integrated Science Assessment, and I can support most of what CASAC has already said in its April 25 letter. I also respect what the Administrator has said about meeting the deadlines. I welcome learning the views of others, and I reserve the right to change the opinions expressed below as I learn more.

The Questions from Dr. Steven Packham are among the most recent to arrive, so I will comment briefly on these first. I was going to start my response with a review of the charge to CASAC, and I am pleased to see his summary including the charge at the beginning of his Questions. I envision CASAC's main task as reviewing the scientific information and analysis supporting the revision of the numerical NAAQS standards for PM. But CASAC should also provide guidance on needed research and analysis methods (ii), and address the effects of various strategies for attainment and maintenance of NAAQS so as to protect public health (iv).

The historical context is important. I cannot claim to know it well by reading all five documents, but I think my background gives me the opportunity to offer evaluation and advice. Let me list three main themes before I get into the wording of questions.

I. EPA's determinations of causality seem to me seriously flawed by confounding. The confounding factors appear to be socioeconomic status (SES) and weather variables such as extremes of heat and cold. Using lay language, heat waves in summer result in increased mortality, and cold waves in winter result in influenza and other conditions leading to increased mortality and morbidity. Low-income and especially elderly and disabled people may be especially susceptible to these extreme temperature effects, so these factors are not readily separable.

II. Particulate matter is not one substance, but a mixture. It should be expected that since sources differ, the concentration- response (C-R) relationship may therefore differ from place to place. Further, there is important uncertainty in the shape of the C-R relationships, which may differ from place to place because of differences in sources, weather, atmospheric chemistry, and the exposed populations. Using the Cox proportional hazard model with national averaged data is essentially equivalent to assuming a single, linear-through-zero relationship, and it suppresses uncertainty and variability. The uncertainty should be made explicit. Epidemiological and other data addressing place-to-place variability should be sought and used. Some studies with such spatial heterogeneity have been recommended to EPA by CASAC as additions to the ISA. More may be available or become available. Sensitivity analysis and weighting on C-R relationships at different concentration levels should be a useful supplement for EPA to include in its risk assessments. What EPA has done in Chapter 3 and Appendix C of the draft PA seems to me extremely simplistic, and potentially misleading on the extent of lives and disease that would be avoided by more stringent standards.

III. Fires, especially wildfires, contribute a large portion of fine particulate matter ($PM_{2.5}$) in the west and especially in California, the state in which I live. EPA should work with other federal agencies and state and local agencies to assess the public health consequences of fires and foster the development of strategies to reduce wildfires as a source of particulate matter air pollution.

Rather than organizing by charge questions and by the questions posed by the five members of CASAC, it may be most effective if I first develop these three main themes with references. Then I will respond to the CASAC questions.

Theme I. Causality and confounding. This is a major concern in the questions of Dr. Mark Frampton. Tony Cox has just sent out references to links to four of his papers. I have reviewed Tony Cox's book, *Causal Analytics*, and I was a reviewer for an earlier draft of his paper in *Critical Reviews in Toxicology*. I am in general agreement with Tony Cox on the importance of confounding, the need to consider the Bradford Hill Criteria as out-of-date, and the importance of obtaining data that will allow discriminating among multiple potentially causal factors.

My review of Tony Cox's Book and of two books by Judea Pearl is recommended as a starting point to those unfamiliar with this area of scholarship. Professor Pearl at UCLA is considered one of the pioneers of artificial intelligence methods in computer science, which are now very extensively used in information technology commerce, sometimes under the heading of data science. His *Book of Why* is an easy read and a good source of instruction on the limits of regression-based statistical techniques still in wide use. And I urge those on CASAC not familiar with this material to read the papers Tony Cox has sent out to the group, and as well as Pearl's *The Book of Why*, for the reasons I gave to readers of my book review.

On page 3-5, lines 8-12, the draft PA quotes from EPA's earlier 2011 document as follows:

> An analysis of air quality and population demographic information indicated that the highest $PM_{2.5}$ concentrations in a given area tended to be measured at monitors in locations where the surrounding populations were more likely to live below the poverty line and to include larger percentages of racial and ethnic minorities (U.S. EPA, 2011, p. 2-60).

I have not reviewed the analysis referred to in this passage, but what the passage tells me that there is an association, perhaps a very strong one, between socioeconomic status (SES) and concentration levels of $PM_{2.5}$. Might the "causal relationship" go the other way? Poorer people tend to live in more polluted areas, but it is access to medical care, lifestyle issues such as diet and drugs, and perhaps poorer housing conditions including lack of air conditioning and inadequate heating/insulation, which may cause the increased mortality. I did not find ANY discussion of this important question in the draft PA. A Google search identified multiple studies on SES and mortality, without consideration of PM air pollution as an important causal factor. These studies (Crimmins and Saito, 2001; Adler, 1994; Anderson, 1997; Bosworth, 2018) indicate that SES has a strong association with life expectancy. There is a large literature from studies on this issue. Why is this literature not addressed in the draft PA – and in other, earlier EPA documents?

The draft PA lists $PM_{2.5}$ has having a "causal" relation to both short-term and long-term mortality in Table 3.1 , page 3-18. Page 3-16, lines 11-13 defines the top of the five EPA categories as:

> 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.

Maybe I have missed something from the history. I have not read the 2009-11 documents, but what I have read persuades me that confounding via factors associated with SES should be addressed thoroughly in order to satisfy the condition that "confounding and other biases can be ruled out." I have not yet found that EPA has satisfied that condition. That seems to me a very serious flaw in the draft PA.

Let's consider cardiovascular effects disease rather than mortality as the end point. This is listed as causal for $PM_{2.5}$ in Table 3-1, Might there be confounding by socioeconomic status here as well? Again, a Google search leads to a large literature, none of which I have found referenced in the draft PA. A recent example is Schultz, 2018.

I cannot claim to be an expert on this literature on SES and public health. But I can claim some expertise on the analysis techniques, and until I learn more, 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 $PM_{2.5}$ in the present NAAQS. At very high concentrations, I am convinced: People die from smoke inhalation.

Theme II. The concentration response (C-R) relationship. There is a need to address uncertainty and variability and especially at the low concentration levels of the present annual average and 24 hour average $PM_{2.5}$ NAAQS and below. The place for this need to be met is in Chapter 3 and Appendix C.

I am moderately familiar with EPA's BenMAP-EC computer program, having managed the review of papers dealing with it for the journal *Risk Analysis*. The March 2015 issue includes a Smith and Gans paper, a response to it from members of EPA's staff, and a rebuttal from Dr. Smith. In my judgment these three papers present a good introductory discussion of the issues on the limitations of BenMAP-EC for dealing with epistemic uncertainty, such as the shape of the C-R relationship. In September 2016 *Risk Analysis* published a special issue on air pollution risk, which highlighted the need for better characterization of uncertainties, especially on the C-R relationship. I have included these papers in the references at the back of this Response. I managed the peer review of these papers and wrote the introduction to this issue (North, 2016). Contributors included two former chairs of CASAC, Christopher Frey and Roger McClellan, as well as the current chair, Tony Cox. To my surprise and disappointment, I have found NONE of the papers mentioned in this paragraph included in the references for Chapter 3 and Appendix C in the draft PA.

Yes, there is not clear evidence of a threshold for mortality or for cardiovascular (CV) effects. But NEITHER is there clear evidence for linearity at very low concentrations. Most of the data from deaths and CV effects comes at the high dose levels, and a straight line is estimated from these data. (See for example, the acknowledgment of this pattern in Footnote 40, page 3-67.) The procedure EPA is using is an extrapolation from the concentrations at which there are extensive data to levels at which data are sparse. And by combining data from different areas of the country where $PM_{2.5}$ comes from different sources, data on possible heterogeneity (different slopes in different locations) is being lost.
At the CASAC meeting in 2018, a presentation was made with a follow-up written submission from the Electric Power Research Institute about epidemiological methods to disagggregate spatial variation, with four recent papers describing and applying these techniques. CASAC in its April 11 letter recommended that three of these papers be added in revising the ISA draft. The fourth paper, Eum et al. (2018) was an attachment to EPRI's written submission. I believe methods and applications for spatial disaggregation should be included in the ISA and the PA. The Chapter 3 risk assessments should at least discuss these methods. It would be even better to use the data and results from published applications.

Dr. Anne Smith has written a Perspective paper which has just been accepted for publication in *Risk Analysis* after multiple rounds of peer review and revision. I have communicated with Dr. Smith and received a manuscript copy of the final text and figures. The paper is now being typeset and will appear within a few weeks on the Early View website that provides web access to papers accepted for *Risk Analysis* but not yet published in a print issue. In this paper Dr. Smith proposes a weighting of the C-R relationship below the range with extensive data for a judgment-based estimate of premature deaths avoided by changes in the standard. This weighting is a matter of judgment in the absence of data. It should be recognized that the Administrator must rely on judgment as the basis for his decision because

of the sparseness of the data and the uncertainty in the C-R relationship at low concentrations. I urge that Dr. Smith's proposal be considered by CASAC. It seems far better to me than the extensive tables of numbers in Chapter 3 and Appendix C, which, to many readers, might provide an appearance of scientific information and rigor. To me these tables are no better than repeated back-of-the-envelope calculations made with the highly questionable assumption of a linear-through-zero C-R relationship. CASAC has suggested a quantal and graded C-R assumption. A sigmoid shape would seem much more sensible on biological grounds than a linear-to-zero shape. Why assume that many deaths and cardiovascular effects will occur from exposure to concentrations below the present NAAAQS, and below potential new NAAQS at the lower values used in Chapter 3 and Appendix C of the PA draft?

The extreme levels of detail in modeling air quality at the level of a 12 km grid make little sense to me. The mass of detail will confuse readers rather than help them to understand what has been done in the quantitative assessment of risk. I would much prefer a set of regional scenarios, showing how specific sources of PM might change under different standards and strategies for attaining these new standards. That brings me to theme III.

Theme III. Wildfires as a source of PM, especially $PM_{2.5}$.

Chapter 2 presents a summary of the evolving information on sources of particulate matter in the various size classes. It seems to me very important for CASAC members to consider Figures 2-2, 2-3, and 2-4. Fires provide more than half of the national emissions of organic carbon. In urban areas a significant portion of ambient $PM_{2.5}$ and PM_{10} come from mobile sources and stationary fuel combustion, but the organic carbon and elemental carbon portions of PM deserve serious attention as a public health hazard. In the past the attention in the air pollution community was on sulfur oxides and nitrogen oxides, particularly the secondary particulates undergoing long range transport. Figures 2-2 and 2-3 and accompanying text tell that now fires, dust, and agriculture account for the majority of the national emissions – about 2/3 for $PM_{2.5}$ and more than 80% for PM_{10}.

When a large wildfire occurs and the wind blows the smoke into populated areas, the levels of $PM_{2.5}$ go far above the current standards to levels listed by EPA as "unhealthy" and "very unhealthy." I and many others in the San Francisco Bay Area experienced such levels after the Camp Fire in November 2018. I found a link to my local public television station, whose news coverage included a summary of 24-hour readings at a single monitoring station (Goldberg, KQED, 2018). Shown with this story (you can follow the link in the Reference section) is a set of 24-hour $PM_{2.5}$ readings ranging from 158 to 249, over multiple adjacent days. Compare these big numbers to the 24-hour NAAQS standard of 35. Of the 30 highest readings, 18 were due to two recent wildfires, and 9 others were due to residential wood smoke. Only three of the 30 were from other sources. Similar high concentrations occur, quite frequently, from wildfires in Southern California.

Footnote 21 page 2-21 for Figure 2-7 is one example where wildfire contributions to $PM_{2.5}$ have been removed from a presentation of $PM_{2.5}$ data over time. Has EPA done more of this in its risk assessment? I believe the answer may be yes, and that concerns me. See 2-41, line 23. The text on 2-47 reports that in 2015 the areas with annual mean concentration of $PM_{2.5}$ greater than 11 $\mu g/m^3$ in 2016 are limited to California and southwest Arizona. These are potential annual standard violations with a lowered standard. The text on 2-44, line 10-11 notes that some of the largest difference [among projections from modeling] occurred over southwest Arizona. What source would be causing these high levels? Smoke from fires? Windblown dust, perhaps associated with agriculture? Why are there large differences from different modeling assumptions? When I look at the plots in Chapter 3 for where the annual average and 24-hour NAAQS standard may be exceeded, I see areas in California where smoke is frequently present from large wildfires. In Chapter 3 and Appendix C I find suggestions that rather than focus on the infrequent wildfire events that result in these persistent high levels, far above existing standards and clearly a threat to public health, these episodes are either averaged in, or possibly deleted from the data being used. (See for example, the text on C-16 and 17). I have not taken the time to understand the details, but I would like to understand better the data and the projections by modeling of concentrations used in the risk assessment. Why should California and southwest Arizona be leading candidates for non-compliance under stricter standards?

President Donald Trump tweeted about the cause of California's massive wildfires: "poor forest management" (Pierre-Louis, 2018). As one who has consulted to the US Forest Service on risk management for wildland fires, I am in agreement with the President - after a clarification that many fires occur near the urban-wildland interface on land that would not be readily described as forest, but rather bushy hillsides or even grassland. Congressional law and practices of the U.S. Forest Service and many fire agencies in California and elsewhere in the country have emphasized putting fires out when they occur, rather than managing the fuel – the material that is available to burn. The result of fire exclusion is fuel build-up. Then when a fire gets going under conditions of high heat, high wind and low humidity, the fire is so intense that it is often unstoppable: It burns until the weather changes or it runs out of fuel. Strategies to reduce the risk of large fires are readily available, for example use of grazing animals (Daniels, 2016), mechanical brush and selective tree clearance, and use of prescribed fires under mild weather conditions (Fernandes, 2003). But these strategies come at a cost and someone must provide the funds. EPA and other federal agencies could act to improve this situation, and so benefit public health by reducing the frequency and severity of high PM concentrations from wildfires. CASAC might want to comment on strategies to achieve public health goals by managing wildfires and perhaps other sources of PM, such as from dust and agriculture, shown in the Figures and text of Chapter 2. Evaluating the public health benefits of these strategies to reduce PM from wildfires would be a good addition to Chapter 3. Perhaps other strategies for other sources might be good additions as well. The PA might summarize what states are doing in state implementation plans to reduce emissions from sources such as diesel engines by retiring older vehicles with high emissions. Suggestions might be listed for specific strategies that might be added in the local areas where standards could be exceeded.

Now let me turn to the questions from the CASAC Members.

Questions from Dr. Steven Packham:

I will take your bullets in order.

1. I think biological causation and causality in epidemiology are areas of weakness in the PA draft, and that these weaknesses go back to earlier documents and earlier rounds of CASAC review. I am not qualified to speak on pulmonary toxicology. I hope others can do this.
2. I have provided references that I hope will be helpful. I think the confounding by weather and SES is a most urgent need for further attention. I am surprised and disappointed that relevant recent peer-reviewed publications have not been included in this "last and final" draft PA. While CASAC comments on the ISA are acknowledged in footnotes, references recommended by CASAC have not been included in this draft PA. EPA should have had time to do this.
3. I would like to see more analysis using spatial and source disaggregation. Weather data and SES characteristics of the exposed populations should be collected and used in the analysis of the epidemiological data.
4. Humans have been breathing wood smoke since we learned to use fire millions of years ago. The data on physiological impacts from wood smoke inhalation ought to be extensive.
5. I would also like to see dust separated into natural and anthropogenic sources. I wonder about the high levels in the Salt Lake City area in the figures in Chapter 3. Is this windblown salt from salt flats and from the Great Salt Lake? Case studies in Chapter 3 would be helpful. EPA should be encouraged to separate the sources by type of PM, which is done in Chapter 2 but not in Chapter 3.
6. I have discussed wildfires as an example. If subsidizing landowners to run goats, sheep, or cattle on land that will otherwise grow flammable bushes and trees will reduce the risk of massive fires, EPA ought to be encouraging such strategy. I have been involved on private land in getting a plan for mechanical clearing approved and implemented, as part of timber and fire management under California law and regulation. I'm proud of it. I wish many others would learn and do something similar.

Questions from Dr. James Boylan

Chapter 2. My expertise on PM Air Quality is rather marginal. I learned a good deal from this chapter, and from what I could tell, the material seems scientifically sound. Your questions probe at a depth beyond my knowledge.

Chapter 3 (and 4). I found Chapter 3 and the supporting Appendix C very disappointing. The methods are old and in my judgment, out of date. See my comments on themes I and II above. I will not attest

that the material is scientifically sound by current standards. I believe the material is out of date and needs to be redone. The research needs discussion of Chapter 4 seemed sparse and unimaginative. I've provided a few ideas. My favorite would be a serious effort to sort out the contributions of PM, SES and weather variables. I think leading experts in statistics and analyses of big data sets could do it. It would not be for this round on PM NAAQS but perhaps for one five years hence.

Appendix C. In my available time I have not tried to learn about Downscaler and other modifications on the data to "just meet" current and alternative standards. I judge these to be bells and whistles rather than the essence of the needed analysis, which should address confounding, causality, and the uncertainty and variability of the concentration-response relationship for different sources (different chemical species) of PM and in different locations, with different weather and populations with different SES characteristics. I regard BenMAP-CE as far too simple on the C-R portion and far too complex in its adjustments on air quality data. I would like to see case studies on specific areas where the $PM_{2.5}$ standards might be exceeded, such as the areas east of San Francisco and Los Angeles, southwest Arizona, the Salt Lake City area, and selected eastern urban areas. The estimates of mortality with current and alternative standards, based on old studies with the Cox proportional hazard model – that is, linear- to- zero concentration response - are not anything I could endorse or defend.

Questions from Dr. Mark Frampton.

1. For reasons I have discussed under Theme I, I am not persuaded that the 2009 ISA or the 2018 conclusions of a "causal relationship" as EPA has defined this term, are valid. Now that I have found the EPA has found an association between PM levels and SES, I believe that the appropriate action is not to make low SES a sensitive group, but consider that SES has an observed association with PM ambient concentration levels and that this is a potential confounding factor. I don't claim any expertise on the SES and mortality relationship. But it looks to me like there may be an association that might be causal for mortality from other factors, and independent of air pollutant concentration at low levels. Tony Cox's example of association and not causation is whether ice cream sales predict heat stroke. Judea Pearl's example is whether children's shoe sizes predict their reading ability. These are examples of strong association where common sense would say there is no causal relation, that the causal relationship is elsewhere, and what we have is confounding from another factor (temperature; children's age) that is causally related to the end point of interest and the predicting variable.
2. I believe the impact of heterogeneity should be studied. This should give insights on the location and character of the risks that need attention from our national leaders. I think reducing $PM_{2.5}$ from future wildfires is a good candidate activity, based on readily available present information. I believe the level of uncertainty on mortality and other health effects from PM addressed in the PA from exposures at, below, or slightly above the current NAAQS is high. These effects might be essentially negligible as a public health issue – at this point we do not know. But there is no

doubt that people died of smoke inhalation in California's recent wildfires, and many more may have suffered health impairment. At the exposures that occurred in California in November 2018, causality seems quite clearly established: People with compromised lung function were asked to wear masks, stay indoors, or leave the area. Members of my family followed these instructions.

3. As you can ascertain from reading my book review, written well before my engagement as a consultant to CASAC, I am a fan of Dr. Cox's concepts of causality determination, although I may disagree with him on details and on how to explain these concepts in simple language. I highly recommend to you reading "The Book of Why" by Judea Pearl. Dr. Cox learned of these concepts from Dr. Pearl's earlier writings, which are much harder to read. For "The Book of Why" Professor Pearl was aided by a good science writer. With my limited available time I am going to skip the 4 documents you listed for comments by the non-member consultants. I did read Dr. Vandenberg's response to Dr. Cox and found it unpersuasive. Dr. Vandenberg seems intent on defending past EPA practices, without going into the issues that were discussed in the papers in the September 2016 Special Issue of *Risk Analysis*. I will be pleased to read the other three documents and respond on them in a future communication, when there is a sharper focus on specifics, and I have more available time.
4. I see the same confounding issues with SES and weather as applying to cancer as well as mortality and cardiovascular effects. Cancer rates vary in different societies with different diets and other exposure to biologically active substances. I am aware that there is evidence that cooking with some types of coal in enclosed household spaces in China led to an elevated incidence of lung cancer (Seow et al, 2014). Was this lung cancer response from PM or volatile organics released from the coal? Some coals appeared worse than others. In a situation like this with very high exposure levels, I would be strongly inclined to endorse a causal relationship. I would speculate that if people were cooking indoors with wood fuel and with the same lack of ventilation as in these rural Chinese homes, there might be elevated levels of lung cancer.
5. Because of Theme I and Theme II, I do not view the risk assessments of Chapter 3 in this draft PA as either complete or appropriate. These seemed to me a barely warmed-over version of what was done in 2009-11. I believe these risk assessments are superficial and out-of-date. New epidemiological studies are available, and spatial disaggregation might be done and used by EPA. Some references have already been recommended by CASAC to EPA for including in the revised ISA. These should be discussed in the PA as well.
6. I expect there are lots of key studies that ought to be considered in Chapters 3 and 4 of the PA. Other may be better at identifying them than I am, with limited time. I would like to learn more about the data and analysis underlying the quote from EPA 2011 that is at the top of page three of this communication. What is the strength of the observed association between $PM_{2.5}$ exposures and the SES level of the populations receiving these exposures?

Questions from Dr. Sabine Lange:

1. I do not believe I understand EPA's motivation for the "pseudo-design value" adjustments. Your Houston data illustrates that on a day-to-day basis, the data points are highly variable and a seasonal pattern is not evident. I would like to see more attention to patterns of exposure, such as an occasional week-long exposure to levels three times the top value on your figure, which many of us in my part of Northern California experienced in November 2018. Is there a cumulative effect from repeated daily exposures to high levels, on the biological basis that some fraction of $PM_{2.5}$ material inhaled is retained in the lungs? Does toxicity depend on solubility, or on the composition: elemental carbon, organic carbon compounds, oxides of sulfur and nitrogen, etc. I was surprised reading the extent of wood smoke and dust in $PM_{2.5}$ in the west, compared to other constituents. My overall impression is that much more could be learned about the constituents of PM, because there is better data on what is in $PM_{2.5}$ in different locations. There is not uniformity in the national exposure. Yet an averaged-over-the-nation concentration response relationship is used, with the only variation being which epidemiological study is used for the slope coefficient, e.g., the beta value.
2. 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.
3. 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. The greater Salt Lake City area shows up green in the Chapter 3 - Figure C-17, page C-18. Green represents exceedance of the 24 hour standard. What is going on here? Is this from wind-blown salt/grit off of the Bonneville Salt Flats or the Great Salt Lake? I doubt it is from fires and dust, as in California and Oregon. The draft PA text does not help me at all to understand what is going on. I found a local news story (Jacobs, 2016) that indicates lowered levels from evaporation from the Great Salt Lake are creating an extensive dry area with a salty "crust," and resuspension of materials from this crust may be adding to the level of dust, including toxic materials from industrial sources deposited in earlier times. Such local stories suggest that trying to regulate based on national standards may miss locally important air quality problems. EPA ought to be more sensitive to this, and evaluate regulatory strategies that fit the local problems.

4. 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.
5. 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!
6. You ask great questions! 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.

Questions from Dr. Tony Cox

1. I don't believe the C-R coefficients are conceptually very useful, because of what I have described under themes I and II. These coefficients may be well-defined in terms of the Cox proportional hazard model from the listed epidemiological studies. But given the adjustments on the concentration side, the potential extent of confounding by weather and SES, and the uncertainty in response at low concentrations, I could not defend the calculations of these betas as a proper basis for setting NAAQS.
2. I would chose a, association, over b,c, and d. I have trouble imaging what intervention means for the aggregated PM concentration across the entire nation. Maybe e is the best answer, something different than any of the above. I am puzzled by why EPA chose spatial aggregation rather than trying to learn from disaggregated C-R data where available.
3. My theme I is evidence of potential confounding from SES and possibly weather variables as well. I can't attest to more than slopes of regression lines (association) from the epidemiological data. The basis is association, not a causal relationship that I can endorse.
4. As a follow-on to the previous question, I do not believe that a causal relationship has been established for any of the endpoints in Chapter 3, Table 3-1.
5. I'm not sure I can judge what is "generally accepted." I can judge what I think is acceptable and believe should be generally accepted. I believe that the strong association implies causality aspect of the Bradford Hill Criteria is out-of-date and should no longer be used. But the Bradford Hill Criteria may still be "generally accepted" by many epidemiologists and public health experts. Scientific revolutions happen slowly, according to Thomas Kuhn's book.
6. The problem with $\Delta y/\Delta x$ is that there may be multiple factors $x_1 \ldots x_n$ that affect y. Why should the other factors be excluded when there is evidence that these other factors may be important influences on y? A one-dimensional regression analysis of health effects on one air pollution concentration doesn't seem like a definitive basis for prediction. Your example is a good one.

Temperature should be included in the analysis. Changes in ice cream consumption may have had a strong association with heat stroke – certainly in my youth, before the widespread availability of air conditioning. I remember as a kid waiting eagerly for the arrival of the Good Humor Man on hot summer days.

7. If there is any empirical validation of the beta coefficients, which would support approximately accurate predictions or simulations on how air pollution interventions would change population health responses, I have missed it in my reading of the draft PA. I think such estimates could be developed with spatially disaggregated data. If SES and weather variables were included, this analysis might be very instructive on the potential result of such interventions. I recall you have such an example in your *Causal Analytics* book. I would strongly advocate to EPA trying this type of analysis on regional problems, based on larger data sets than may be currently available.
8. I did not find more than a few dismissive words about "potential omitted confounders" in Chapter 3 and Appendix C. The absence of discussion on SES and weather was disappointing. An at Dr. Frampton's urging I read the response letter to you from Dr. Vandenberg. I thought he was defending old EPA practices with little indication that EPA is open to new ideas. And I am appalled that there are no references in the draft IPA to the many relevant papers published in the past few years in *Risk Analysis*.
9. If there was much sensitivity analysis beside which location and which (old) epidemiological study, I missed it.
10. Model uncertainty on the C-R relationship was not adequately discussed or included.
11. Yes, the discussion of exposure estimation uncertainties and estimation errors was also inadequate.
12. Themes I, II and III motivate me to withhold any endorsement that the draft PA provides an adequate basis for assessing causal impacts on public health of changing $PM_{2.5}$ levels. And I do not understand, nor do I believe the public will understand, the complex and convoluted way that EPA describes changes in concentration with alternative standards. What sources of $PM_{2.5}$, as described in Chapter 2, will be reduced, and by what amounts?
13. Your question addresses my Theme I concerns. I did not find these addressed in the draft PA.
14. My response is a strong NO.
15. I think that CASAC should be advising EPA on how to do a better analysis. It may not be possible to get such analysis done by the time of finalization of the PA. But EPA should at least include discussions of new epidemiological methods and recent studies with spatial disaggregation, such as those discussed in the Comments submitted by EPRI at the last CASAC meeting.
16. I am pessimistic that that the techniques of formal causal modeling can be incorporated in the final version of the PA. Maybe Chapter 4 can be expanded into a research plan that describes these methods and lays out steps toward using them in future PM reviews. Developing the detail you ask for is far more than I can describe in the very limited time I have to write these comments. (It took me a good deal more time to read and review your *Causal Analytics* book, which touches on many of these details.) The hardest aspect may be in collecting the needed

disaggregated data – on sources, concentrations, health endpoints, SES factors, and weather factors. I would recommend some regional or local case studies to EPA before trying to do a national analysis that includes such disaggregated data.

I'll add a summary statement of my own. I think the Administrator may be given relatively little useful new information on the impact of changing the Primary PM standards from what has been assembled in Chapters 3 and 4 of this draft PA. (I have not read the supporting material on Secondary PM standards.) If CASAC should agree with me, then that judgment needs to be explained to the Administrator. Perhaps a new and much simpler analysis, such as in Anne Smith's Perspective piece, might be used in a greatly revised final PA, with uncertainties and sensitivity analysis highlighted. I am not persuaded that a case has yet been made for stricter national standards, but I believe that more aggressive strategies to control sources may be needed in certain locations, such as wildfire-prone areas in California and areas affected by wind-blown PM from land once covered by the Great Salt Lake. There may be other important candidates for additional study of source control strategies.

References:

Bosworth, B., "Socioeconomic status and mortality," *Annual Review of Public Health*, 39:237-251 (2018).

Adler, N.E. et al., "Socioeconomic Status and Health," *American Psychologist,* **49**:15-24 (1994).

Anderson, R.T. et al., "Mortality Effects of Socioeconomic Status," *Epidemiology* **8**:42-47 (1997).

Anenberg, N.C. et al., "Survey of Ambient Air Pollution Health Risk Assessment Tools," *Risk Analysis* 36: 1718-1736 (2016).

Chapman, Robert, Comments of the Electric Power Research Institute on Environmental Protection Agency 83 FR 53471 [EPA-HQ-ORD-2014-0859;FRL-9985-ORD, Integrated Science Assessment for Particulate Matter (External Review Draft), December 10, 2018. Available at: https://yosemite.epa.gov/sab/sabproduct.nsf/A89F40762DF41E4385258360006404B4/$File/EPRI+Comments-EPA-HQ-ORD-2014-0859_web.pdf

Cox, L.C. (Tony), Jr. "Invited Commentary, Rethinking the Meaning of Concentration-Response Functions and the Estimated Burden of Adverse Health Effects Attributed to Exposure Concentrations," *Risk Analysis* **36**:1770-1779 (2016).

Crimmins, E.M., and Saito, Y., "Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences, "*Social Science and Medicine*, **52**:1629-1641 (2001).

Daniels, J. "Goats, sheep reducing California's fire risk one bite at a time," CNBC, July 19, 2016. Available at: https://www.cnbc.com/2016/07/19/goats-sheep-reducing-californias-fire-risk-one-bite-at-a-time.html.

Evans, J.S., "Invited Commentary: Characterizing Uncertainty in Estimates of Mortality Risk for Exposure to Ambient Fine Particulate Matter," *Risk Analysis* **36**:1748-1750 (2016).
Fann, N.L., Lamson, A.D., Luben, T.J., and Hubbell, B.J., "Commentary: Response to 'Enhancing the Characterization of Epistemic Uncertainties in PM2.5 Risk Analysis,'"*Risk Analysis* **35**:379-380 (2015).

Fann, N., Gilmore, E.A., and Walker, K., "Characterizing the Long-Term PM2.5 Concentration Response Function: Comparing the Strengths and Weaknesses of Research Synthesis Approaches," *Risk Analysis* **36**:1693- 1707 (2016).

Fernandes, P.M., and Botelho, "A review of prescribed fire effectiveness in fire hazard reduction," *International Journal of Wildland Fire*, 12:117-128 (2003). Available at: https://www.fs.fed.us/rm/pubs/rmrs_gtr292/2003_fernandes.pdf.

Frey, H.C., "Invited Commentary: Dose-Response Models are Conditional on Determination of Causality," *Risk Analysis* **36**:1751-1754 (2016).

Goldberg, T., "Came Fire Caused Nearly 2 Straight Weeks of Bay Area's Worst Air Quality on Record," KQED News, December 19, 2018. Available at: https://www.kqed.org/news/11712211/the-camp-fire-caused-nearly-two-straight-weeks-of-the-bay-areas-worst-air-quality-on-record.

Jacobs, J.P., "Great Salt Lake faces ruin,"*E&E News, Greenwire*, July 11, 2016. Available at: https://www.eenews.net/stories/1060040048.

Levy, J.I., "Invited Commentary: Fine Particulate Matter, Risk Assessment, and Risk Management," *Risk Analysis* **36**:1745-1747 (2016).

McClellan, R.O., "Invited Commentary: Providing Context for Ambient Particulate Matter and Estimates of Attributable Mortality," *Risk Analysis* **36**:1755-1765 (2016).

Moogavkar, S.H., "Invited Commentary: Fine Particulate Matter Pollution and Mortality," *Risk Analysis* **36**:1766-1769 (2016).

North, D.W., "Introduction to the Special Issue on Air Pollution Risk," *Risk Analysis* **36**:1688-1692 (2016).

North, D.W., "Mega-Review: Causality Books. Louis Anthony Cox, Jr., *Causal Analytics for Applied Risk Analysis*, Judea Pearl and Dana MacKenzie, *The Book of Why: The New Science of Cause and Effect*, Judea Pearl, *Causality: Models, Reasoning, and Inference, Risk Analysis* **37**(7):1647-1654 (2019).

Pierre- Louis, K., "Trump's Misleading Claims About California's Fire 'Management,'" *The New York Times*, November 12, 2018. Available at: https://www.nytimes.com/2018/11/12/us/politics/fact-check-trump-california-fire-tweet.html.

Schultz, W.M., et al., "Socioeconomic Status and Cardiovascular Outcomes," *Circulation*, **137**: 2166-2178 (2018).

Seow, W.J., et al., "Household air pollution and lung cancer in China: a review of studies in Xaunwei," *Chinese Journal of Cancer*, **33**:471-475 (2014).
Shin, H.H., et al., "Meta-Analysis Methods to Estimate the Shape and Uncertainty in the Association Between Long-Term Exposure to Ambient Fine Particulate Matter and Cause-Specific Mortality Over the Global Concentration Range," *Risk Analysis* **36**:1813-1825 (2016).

Smith, A.E. and Gans, W., "Enhancing the Characterization of Epistemic Uncertainties in $PM_{2.5}$ Risk Analyses, *Risk Analysis*, **35**:361-378 (2015).

Smith, A.E., "Response to Commentary by Fann et al. on 'Enhancing the Characterization of Epistemic Uncertainties in $PM_{2.5}$ Risk Analyses,'" *Risk Analysis* **35**:381-384 (2015).

Smith, A.E., "Inconsistencies in Risk Analyses for Ambient Air Pollutant Regulations," *Risk Analysis* **36**:1737-1744 (2016).

Smith, A.E., "Response: Author Synthesis and Response," *Risk Analysis* **36**:1780-1792688-1692 (2016).

Smith, A.E., "Using Uncertainty Analysis to Improve Consistency in Regulatory Assessments of Criteria Pollutant Standards," Perspective, accepted for publication in *Risk Analysis* (2019). (To become available as soon as typesetting and proofing are done on Early View, https://onlinelibrary.wiley.com/toc/15396924/0/0).

Dr. David Parrish

Questions from Dr. Steven Packham

The first 5 questions deal with health effects of PM. These are far outside my area of expertise, so I cannot respond.

The final question is more general: *"Could you provide additional information on any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards?"*

I am not knowledgeable regarding general adverse effects on public health, welfare, social, economic, or energy effects due to attainment and maintenance of national ambient air quality standards. If there were some specific adverse energy effect that was of particular focus, I could possibly contribute.

Questions from Dr. James Boylan

Chapter 2 – PM Air Quality

Is the discussion on sources of emissions accurate and complete? If not, what additional information needs to be included?

I would characterize Section 2.1.1 as an accurate and complete summary of PM emission sources. It is certainly not a comprehensive discussion, as that would require many volumes of the size of the PM PA.

One significant shortcoming is that the uncertainty of the emissions estimates should be clearly discussed and defined to the extent possible. The words "approximately" and "estimate" are used throughout (e.g., The National Emissions Inventory (NEI) is a comprehensive and detailed *estimate* of air emissions of criteria pollutants, criteria precursors, and hazardous air pollutants …."), but such terms are not clearly defined. The following sentence is the first quantitative emission estimate in the Section: "Based on the 2014 NEI, approximately 5.4 million tons/year of PM2.5 were estimated to be directly emitted to the atmosphere from a number of source sectors in the U.S." In the absence of defined confidence limits for quantitative numbers, specifying 5.4 million tons/year with one decimal place generally indicates that the number is 5.4 ± 0.1 million tons, which corresponds to an accuracy of about 2%. This is not an accurate description of the uncertainty embedded in the NEI, which I believe is much larger. I think that a few paragraphs should be devoted to a discussion of emission inventory uncertainty; this discussion might be based on material in Miller et al. (2006) or a similar, more recent emission inventory assessment.

Is the discussion on ambient monitoring accurate and complete? If not, what additional information needs to be included?

As above, Section 2.2 gives an accurate and complete summary of PM ambient monitoring, but cannot be considered comprehensive given the tremendous amount of available material. The discussion does appropriately focus on the routine monitoring networks, but in addition to the instrumentation utilized in these networks, there also exists a wide variety of research instrumentation designed for characterizing many properties of ambient PM. These instruments are generally deployed in short-term intensive field studies, but their description cannot be realistically discussed in this PM PA.

Again, the one significant shortcoming that I can identify is the lack of a general overview of the uncertainty (accuracy and precision) of the ambient measurements produced by the monitoring networks. It would be useful to include a few paragraphs devoted to this issue.

Is the discussion on ambient measurement correlations and trends accurate and complete? If not, what additional correlations and trends need to be included?

The trends in emissions discussed in Section 2.3.1 are based on emission inventories, and hence subject to the same uncertainties embedded in the NEI that are mentioned above. However, the relative trends (e.g., those given in Table 2-1) are expected to be more accurate than the absolute trends illustrated in Figure 2-7).

Figure 2.8 seems to indicate that the accumulated PM2.5 monitoring data is a tremendous resource for informing us about the sources of PM2.5 in the US. For example, the 2015-2017 annual average PM2.5 is remarkably uniform (8-10 µg/m3) across the state of Texas, including the Houston and Dallas urban areas, the Haynesville oil and gas basin on the Louisiana border, and the relatively rural southern Gulf Coast. Perhaps a more in-depth analysis of such spatial patterns could inform the most effective approach to reducing PM2.5 concentrations.

Figure 2.9 would be more informative if the meaning of the upper and lower limits of the shaded region were defined in the figure caption. (Similarly for Figure 2.17.)

To what extent are biases associated with PM_{10}, $PM_{2.5}$, and ultrafine measurements discussed? How would differing $PM_{2.5}$ biases associated with FRM vs. FEM continuous measurements (e.g., FEMs typically show higher $PM_{2.5}$ concentrations compared to FRMs) impact the evidence-based and risk-based $PM_{2.5}$ assessments in Chapter 3?

As noted above, discussion of the accuracy (which would include biases) and precision of the measurements are missing from this section. If it is true that FEMs typically show higher PM2.5 concentrations compared to FRMs, then that is an important issue to include in the discussion of

measurement accuracy. That discussion of biases would provide the basis from which the health effects community could address the impact on the evidence-based and risk-based PM2.5 assessments in Chapter 3.

Is the discussion on hybrid modeling approaches accurate and complete? If not, what additional information needs to be included?

I am ignorant regarding hybrid modeling approaches. Judging from the literature references, this is a rapidly developing research field, so it is difficult to judge the accuracy and completeness of the discussion. A discussion that is accurate and complete at the time of writing, may be out of date quickly. From my reading of Section 2.3.3 and my understanding of ambient PM2.5 concentrations, the summary of the hybrid modeling approaches given in Section 2.3.3.1.4 is accurate and complete, at least at the time of its writing.

Is the discussion on performance methods for evaluating hybrid modeling methods accurate and complete? If not, what additional information needs to be included?

Figure 2-28 gives the annual average PM2.5 predictions throughout the U.S. from the four hybrid modeling examples (for 2011?). It is possible to match these predictions with the annual average measured ambient PM2.5 on a site-by-site basis for all U.S. monitoring sites. Then a direct model-measurement comparison for each of the models could be shown as correlation plots and those comparisons discussed from a quantitative statistical basis. I believe that this additional information needs to be included. However, it must be recognized that this is not an independent test of the model performance, as the measurements themselves were incorporated into the modeling that gave the results, and the comparisons will not provide an objective basis for selecting the "best" model, because the models incorporate the measurements in different ways, and are measurement dependent to different degrees. Nevertheless, such a comparison is worth including.

Is the discussion on background concentrations accurate and complete? If not, what additional information needs to be included?

Section 2.4 is an accurate and reasonably complete summary of PM emission sources. One shortcoming of this section is a failure to clearly distinguish between the background concentration as measured as an annual average (relevant for the annual NAAQS) and as the 98th percentile of 24-hour averages (relevant for the 24- hour NAAQS). It is likely that very different sources contribute the background contribution to these two NAAQS relevant averages. Most of the discussion is focused on the background contribution to the annual average, with little discussion on the shorter term average. For the annual average, the measurement based discussion in Section 2.4.3 provides the best guidance, as the CTMs only poorly simulate SOA formation. The lower limit of the zero-out modeling estimate quoted from the last review of the PM NAAQS (a range from 0.5 - 3 μg/m3) is likely too low for this

reason. Similarly, Section 2.4.2.2 is based on such CTM calculations, and do not include SOA in these estimates.

I do not believe that the statement on page 2-49 accurately applies to 24-hour averages: "As described further below, contributions to background PM in the U.S. result mainly from sources within North America. Contributions from intercontinental events have also been documented (e.g., transport from dust storms occurring in deserts in North Africa and Asia), but these events are less common and represent a relatively small fraction of background PM in most places." For 24-hour periods, dust transport (and wildfire smoke transport as well) can contribute a large fraction of the background PM.

I suggest that Section 2.4 be revised, either by or in consultation with a PM expert, to provide a more up-to-date summary of background PM contributions, with separate discussion of the two statistics (annual average and 98th percentile of 24-hour averages). In addition, a short discussion of the likely influence of the changing global climate on background PM should be included. This influence is highly uncertain, but the background PM will depend on changing land-use (including possible desertification), occurrence and intensity of wildfires, agricultural practices, emissions of biogenic hydrocarbons, and others, both within and outside the US.

Chapter 3 – Review of the Primary $PM_{2.5}$ Standards and Appendix C - Supplemental Information Related to the Human Health Risk Assessment

Chapter 3 and Appendix C deal with health effects of PM. These are far outside my area of expertise, so I cannot respond.

References

Miller, C.A., G. Hidy, J. Hales, C.E. Kolb, A.S. Werner, B. Haneke, D. Parrish, H.C. Frey, L. Rojas-Bracho, M. Deslauriers, B. Pennell, J.D. Mobley (2006) Air Emission Inventories in North America: A Critical Assessment, *J. Air & Waste Manage. Assoc., 56*, 1115–1129.

Dr. Lorenz Rhomberg

Questions from Dr. Cox

Dr. Cox's questions do not seem to be designed to be answered individually. Together, they present a set of concerns about causal interpretation of associations and the consequent impact on using association measures as indices for the amount of risk reduction that might accompany specific changes in PM measures that could result from altered standards and subsequent controls. Accordingly, I have answered the set of questions in one answer.

There are several aspects to be addressed. First is the purely technical question of whether Dr. Cox's expressed concerns have merit – are the difficulties in interpretation he raises indeed present? Second, if so, what as a technical matter can be done to alter the analysis so as to avoid or at least minimize the impact of these concerns as they might affect judgments about the impact of changes in PM exposures. Third, in view of the limits to what can be done in this regard, how can achievable analyses of observational data be used to address the fundamental objective of the whole document's analysis: to inform judgments by the Administrator as to what PM standards – and in particular, what changes to those standards – would achieve the objective of protection of public health with an adequate margin of safety. As noted on p.3-15 of the Policy Assessment, "'adequate margin of safety' is meant to address uncertainties associated with inconclusive scientific and technical information."

To start with the first question, Dr. Cox is correct that the measures of association developed in modeling observational studies are not directly measures of causative actions by the factors measured in individual variables, including variables describing PM exposures. They are contingent on the whole array of conditions, influences (known and unknown), as they vary person-by-person and place-by-place and time-by-time in their antecedent states and combinations as potential contributors to the effect or modifiers of the impact of still other factors. (All these are in addition to measurement errors in the variables used.)

The data going into a regression model represent a snapshot of the overwhelmingly complex state of the world as it was upon data collection, with the model used being an attempt to identify the major patterns of association as they were then seen. Of course, the regression may include covariates that aim to "adjust for" the contribution of some factors, but even these adjustments will be contingent on the particular states of the array of the covariate factors at the times of observation. They also depend on the chosen methods for adjustment. They necessarily make hypothetical decisions about the functional form of covariate contribution, how measured averages over space, time, and individuals succeed in representing average impacts, whether or how the adjustment factors covary, and whether they influence one another's' impacts in ways that can be represented by simple model equations. In short, the fitted model represents, in effect, an assertion that the chosen functional forms for interactions among

variables are exactly correct, that the way the variables are measured (and how they average over the complexity of observed instances) completely capture the effect (irrespective of unmeasured patterns of underlying variability), and that all underlying variability, unmeasured influences, and contingencies among varying states can be lumped together in a simple unbiased "error" term that needs to represent not only measurement error but also the influence of all the unmeasured contingencies and interactions among particulars.

All this is an unavoidable aspect of observational studies. All one can ask for is that the asserted functional structure of the model is reasonable in view of our broader understanding of what kinds of causative processes might operate (or be of interest), that the simplifications of the actual causal nexus are not obviously misleading or ignoring known or plausible important influences, and (to speak directly to Dr. Cox's questions) that we do not over-interpret the functional forms as actual measures of causal impacts without considering the pitfalls and limits to such interpretation.

In particular, as Dr. Cox notes, one cannot simply change the value of one variable and suppose that the model's result represents what would be expected in a real setting if that variable had the different value. To do so presupposes that the functional form of the model has captured all influences correctly (and not simply in an average way), and that all the other variables would indeed be unchanged. This is different from changing the value and allowing for the corresponding changes in other variables that might result from the same intervention (or from patterns of interactions among factors). And these are different from asserting that the parameter characterizing the functional dependence of the outcome on a variable is a measure of that variable's causal effect, much less a measure of how a different level of that variable would exert an altered causal effect.

In short, Dr. Cox's concerns have technical merit in terms of demonstrating how translating measures of association can be misleading if interpreted as generally applicable measures of the degree of causal effect imparted by the factor, and how generalizing from the world-state measured in the study to other states entails buying in to the actual truth of the simplifying assertions about functional relationships, lack of interactions, and sufficiency of averaging assumptions that are entailed in any regression approach to observational data.

An important part of the interpretation of causality arguments is to consider the plausibility (in view of our wider biological knowledge and other kinds of data) of the operation of some of the potential complications. For instance, one can note (as I have done) that using group averages for factors that vary among individuals makes assertions that the average effect of the variations is given by the average of the variations themselves. One can sometimes judge which potential complications are serious by using outside information on such things as age dependence, dependence of effects on time patterns of exposure, and so on to judge whether an exposure measure that averages these things out could be potentially misleading.

Another important consideration is confounding factors. As one not steeped in the air pollution epidemiology literature, I was surprised to see that many PM studies do not account for the levels of other air pollutants. Since several important pollutants are markedly correlated in their levels across time and space, and since, when each of these is studied alone, associations with a similar set of outcomes is asserted, it would seem necessary that a study not controlling for copollutants would overestimate the effect attributable to the measured one. If the pollutants act independently, then one is "explaining" the same interlocality variation in endpoint levels as attributable to each one if studied alone, overcounting the total effect. If they interact (additively or synergistically), then the ability to sort out contributions is further complicated and attribution to any one alone is again a significant overestimate.

The second of my questions noted above is what can be done in terms of technical analytical approaches to ameliorate this situation. Dr. Cox names several candidate approaches in his questions (specifically, in his Question 16). I am somewhat familiar with a few of these, but I do not feel able to comment in detail, especially about how well the approaches would resolve the issues for the currently applicable data. I think one would need to try them out to see.

I have an overarching observation, however. As I have set out above, the problem arises because any realized dataset is the product of myriad factors, measured and unmeasured, known or suspected or unrecognized, that have all contributed in a complex nexus of interactions, causes, prior states, and influences. This array will never be completely specifiable, and in any case, any application of the findings to another time or place will entail a different set of such influences, also not fully specifiable. In short, one cannot generalize an empirical finding as such; one can only hypothesize what abstraction of the particular case might be generalizable and then (1) use whatever means at hand to evaluate how compelling that hypothetical generalization might be, and (2) apply it to make predictions that nonetheless need to be couched in examination of the soundness and sufficiency of the asserted generalizable assertions.

This applies to any deeper analysis of causal patterns. The proffered methods in Dr. Cox's Question 16 might aid in finding bases for stronger hypothesizing of causal influences, but any such effort will ultimately be less than fully sufficient, since any feasible explanatory model must necessarily include unverified assertions about functional relationships among variables (that may be more complicated than is feasible to estimate from real data), about the sufficiency of measurements that average over time or members of populations, *etc*. That is, the challenge of near-infinite complexity simplified for tractability and for focus on main effects is still an issue. Therefore, what can be hoped for is improvement in the ability to hypothesize causal relationships with greater dependability, but with no absolute solution to the challenges.

The third major question I set out at the start of this comment is how this state of affairs should affect how we interpret the association relationships observed as we grapple with finding a basis for projecting what risk changes might arise from alterations in exposures.

It should be stated at this point that, just because it is difficult to rigorously identify and pin down a measured causal effect, this does not mean that there is no such causal effect. It would not be prudent (or consistent with the CAA mandates) to reject a causal effect simply because it cannot be unambiguously measured and characterized. There is no basis to default to a conclusion of no causation unless a rigorous demonstration of its particulars is at hand. We have overall patterns of outcomes in many studies that certainly suggest causative roles for PM in risks of health impacts, and these patterns are repeated with some level of consistency across studies.

It is important to remember that, if one denies the causal influence of a variable (owing to the difficulty in unambiguously demonstrating it or measuring it), then one is faced with proposing some alternative hypnotical causes for the patterns that were found in the data. That is, if PM does not increase mortality as a function of the levels prevailing in each city, but there are mortality differences among the cities, then some other factor must operate to produce the observed differences. The fuller picture of attribution of causality is not just the ability unambiguously to demonstrate and measure it, but also the comparative plausibility of alternative explanations of the patterns that are actually observed. Something must be causing the patterns, and one asks if it is more plausible to accept the causal role of the variable being examined or to assert some one among a set of conceivable alternatives, with evidence for and against the alternatives needing to be weighed.

The analogy is less to a significance test (if the evidence is not strong enough, one defaults to the null) and more like a Bayesian or Likelihood approach (what causal model is most plausible, measured by how likely the results would be under each competing hypothetical attribution of the reasons behind the observed patterns. This is an instance of the larger weight-of-evidence methodology issue, and I refer the CASAC to some publications of mine that set out the above line of argument (e.g., Rhomberg LR. 2014. "Hypothesis-based weight-of-evidence: An approach to assessing causation and its application to regulatory toxicology." Risk Analysis 35(6):1114-1124).
I would propose that it is useful to separate the following questions:

1. What is the basis for asserting that an exposure has some causative effect on an outcome of interest? (This is the qualitative question of whether any causal influence exists.)
2. What is the basis for asserting that such a causal effect has actually operated so as to affect the magnitude of the outcome of interest in a particular setting that has been observed? (This is a retroactive and case-specific assertion of the qualitative operation of the cause.)
3. What is the basis for measuring the degree of effect in such an observed setting? (That is, what change in the outcome measurement is being attributed to the causal factor as manifested in the already observed dataset, *i.e.*, that particular instance of causation?)
4. What is the basis for projecting the degree of change in an outcome of interest as a result of a particular change in the causal variable? (This is a prediction about the causal influence in a new, as yet unobserved dataset.)

a. Presuming that all other variables are similar to their state in an existing dataset? (This is to predict the setting-specific consequence of change in a causal factor, even though one may not have a real manifestation to test the prediction.)

b. Under a (possibly) different set of states of the other variables? (This is the prediction of an influence in a future setting where other factors may have changed, needing the ability to sort out the effects of the causative variable in question from other potential influences on the outcome as well as the ability of other states to affect the magnitude of causal influence of the target variable.)

These get progressively more challenging to answer fully and unambiguously as one goes down the list. The list represents progressively more generalizability of the observed patterns (or rather of the explanation proffered for those patterns) to other settings, where the level of the causative agent is different and then where the values of other factors that may enhance, inhibit, or interact are also different from the setting observed. But it is the final question that one wishes to answer in order to judge the impact of alteration of an exposure standard.

In my view, the important final lesson is to avoid conflating the above set of questions. We can develop evidence for some causal role, largely on the grounds of repletion of apparent effects across studies with different conditions, or unambiguous effects at high dose, plus biological understanding of the plausibility of toxicity-generating processes. To judge the present impact of present levels is a step farther, and to project the changes in impacts from possible changes in levels is yet a different and more challenging question. It is not that possible answers cannot be identified, but that they should be interpreted with awareness of how the limits to interpretation of the causation process and its generalizability can lead to uncertainties that need to be factored in to decisions.

<u>Questions from Drs. Packham, Frampton, and Boylan</u>

I have taken my role to be to provide focused comments on particular technical issues, rather than to act as a reviewer of the document as a whole. Many of the questions from these CASAC members are of a more general reviewing nature, and I have not been able to address them. Some questions from Dr. Frampton (Questions 1, 3 and 4) bear on issues similar to those raised by Dr. Cox, and my response above to those is meant to cover the same ground.

<u>Questions from Dr. Lange</u>

I fear that lack of sufficient time adequately to delve into the calculations referred to in Dr. Lange's questions preclude me from addressing them with sufficient rigor at this time.

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. I addressed this general issue in a paper on acute inhalation toxicity (Rhomberg, LR. 2009. "Uptake kinetics, species differences, and the determination of equivalent combinations of air concentration and exposure duration for assessment of acute inhalation toxicity." Hum. Ecol. Risk Assess. 15(6):1099-1145), but the issues explored therein also apply to the comparison of short-term and long-term exposure as causes of effects.

Dr. Sonja Sax

Questions from Dr. Tony Cox

https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD010919.pub2/abstract

Are the C-R models in Table C-1 appropriate, logically valid, and empirically well-validated, for answering the causal question of how changes in PM2.5 levels would change health risks?

To answer this first question generally, the methodology used in the Draft PA[2] (i.e., the BenMAP model) to estimate the human health risks (i.e., mortality) associated with the current and alternative PM NAAQS is not meant to establish causality, but to provide quantitative estimates of risks and provide information regarding the adequacy of the current NAAQS. The causality is determined by application of the EPA NAAQS framework in the ISA (discussed below).

As such, the risk assessment (RA) generally considers health endpoint for which EPA has already determined that the endpoint is "causal"or "likely causal" based on integration of the evidence in the ISA. It is noteworthy, however, that in the PA, EPA evaluated only mortality (all-cause, ischemic heart disease or IHD, and lung cancer) from long-term exposures and all-cause mortality from short-term exposures, as it noted that these health endpoints reflect "clear public health importance, the large number of epidemiologic studies available for consideration, and the broad availability of baseline incidence data (Draft PA, C-2)."[3] In fact, mortality does drive the health impacts and particularly the valuation estimates in the regulatory impact analyses. However, I would argue that it is still of interest to assess the health impacts for other endpoints that EPA deems are causally associated as this provides a wider perspective, and provides context. For example, one would expect to see more emergency room visits for cardiovascular endpoints, followed by hospital admissions, and lastly mortality attributable to PM exposures for this endpoint.

In the risk assessment, EPA seeks to quantify the health impacts from current conditions or modeled estimates for just meeting the PM NAAQS vs. zero levels of $PM_{2.5}$ (which is not realistic) compared to modeled conditions for achieving alternative NAAQS vs. a zero PM concentration. This approach yields the elevated "avoided" mortalities reported in the PA, which is misleading because they are not realistic

[2] Like the ISA, these analyses were previously presented as a separate risk assessment document that included more health outcomes and more sensitivity analyses. In addition, the document underwent its own review process and was therefore subject to more scrutiny than in the current evaluation.
[3] It is important to note also that the ISA has not been finalized and it is unclear whether there will be changes to the causal classifications that would also impact the risk assessment (e.g., for lung cancer mortality)

(i.e., there is always some background PM exposure and a zero PM level is unattainable) and as EPA acknowledges the uncertainty below background PM levels is very large since there is little data. From the point of view of assessing the risk differences, it is unclear why EPA would not simply compare the current or just meeting the NAAQS with the alternative NAAQS scenarios as this difference is in essence what the risk assessment is addressing.

The estimates of health risks (i.e., mortality) or risk reductions (i.e. "avoided deaths") are, solely based on the associations observed in selected epidemiological studies (i.e., relative risks, or hazard ratios). Therefore, the risk assessment, which is used to answer the question of how changes in $PM_{2.5}$ levels change health risks, already assumes that there is a causal link (that is that the concentration-response function or CRF can be used to quantify the change in health outcome per some change in air pollution concentration). For most studies, this is based on the relative risk (RR), which EPA converts to a beta health function for use in BenMAP. The methodology is described in the BenMAP manual. But in short for a log-linear relationship:

$$ß = \ln (RR)/ \Delta PM$$

The ΔPM is typically the difference between the baseline scenario and the control scenario. As noted above, EPA actually calculated the ΔPM as the difference between meeting the NAAQS and zero PM ambient concentrations and this is compared to the ΔPM of alternative NAAQS and zero PM ambient concentrations.

While EPA details the criteria for the selection of epidemiological studies to include in the risk assessment (see page C-3 of the Draft PA), it is noteworthy that there is no consideration of study quality in the selection of studies or any consideration for any potential biases that may be associated with these studies. That is, EPA appears to consider these to be high quality studies for which bias, including any confounding has been ruled out. Again, this analysis assumes a causal association has been established.

Another noteworthy aspect of the analysis is that EPA relies on studies that report National estimates, yet conducts an analysis for specific geographic locations. It would be more appropriate for EPA to apply specific CRFs for each location, as there is substantial heterogeneity in the CRFs both spatially (regional and city-specific estimates) as well as temporally (i.e., different seasons). Most studies would be able to provide regional or city specific CRFs for use in BenMAP. This would reduce the uncertainty of using a National estimate to estimate effects in a specific region or city.

An additional observation is that given a positive RR and the assumption of linearity, BenMAP will always estimate a reduction of risk with reduced PM concentrations, so even without doing the BenMAP analysis you would expect reductions in risk with reduced air pollutant concentrations (i.e., lower alternative NAAQS), but it is unclear how informative this really is. That is, since you have a

relationship that is assumed to be linear and it does not matter where on the concentration-response curve you are, you will see a risk reduction with reduced air concentrations down to zero. Therefore, for this analysis to be useful, it is important to consider the uncertainty. That is, if you already assume a linear response, and a causal association, then the question becomes: are the observed reductions in risk statistically significantly different between the current and the alternative NAAQS standards?

While this deviates somewhat from the questions posed by Dr. Cox with regards causality, it is important to note that EPA does presents uncertainty associated with the risk estimates. This uncertainty, however, is based only on the statistical uncertainty of the CRF (the standard error), but does not include uncertainty associated with the modeled air quality estimates (i.e., estimates for attaining the current standard as well as alternative standards), uncertainties associated with the baseline health statistics or population estimates. All of these variables factor into the BenMAP estimates. Therefore, the uncertainty estimates presented by EPA only reflect uncertainty in one aspect of the model variables (i.e., the statistical imprecision of the model), although it is likely that this variable is the most uncertain. Still, if we just consider the CRF uncertainty (i.e., the 95% confidence intervals) presented by EPA in comparing the estimated mortality from just meeting the current standard to meeting alternative standards (both based on comparing to a zero PM concentration) we find that the mean mortality estimates for just meeting the current standard are within the range of uncertainty for meeting alternative standards – that is that the reduction in risk is not likely to be statistically significantly different because of overlapping confidence intervals. This indicates that even though risk reductions are observed with alternative standards, these risk reductions are within the uncertainty bounds of the analyses. EPA needs to make this point in the PA, instead of focusing on only the point estimates. In addition, it is noteworthy that although EPA discusses qualitatively many additional sources of uncertainty (see pages C-77 to C-80) and many are rated by EPA as having a medium-high degree of uncertainty with the impact on both directions, the EPA does not discuss this uncertainty in the context of interpreting the risk results.

With regards to causality, as noted above, EPA determines causality as part of the evaluation presented in the ISA. As noted by several reviewers and in the ISA comments, the EPA causal framework could be significantly improved with the application of modern day systematic review practices. 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. Given the very small relative risks from largely ecologic epidemiological studies with serious limitations, it is critical to assess all lines of evidence to determine causality.

In the Draft PA, there is some limited discussion of lines of evidence that EPA relies on other the epidemiological evidence. I would argue that the epidemiological evidence alone should not be the basis for assessing a causal association, and this clearly has implication for the risk assessment (i.e., if the

epidemiological evidence cannot establish causality then it maybe should not be used to quantify health risks). On the one hand, EPA specifically states that evidence from controlled exposure studies and animal studies provide supportive evidence for the associations observed in epidemiological studies, but on the other hand EPA readily dismisses this evidence for providing information regarding the adequacy of the current NAAQS noting that because controlled exposure studies and animal studies were conducted at levels above the current NAAQS they were not informative. I disagree that this evidence is not informative, instead it is important to consider that if effects are not consistent or coherent at levels above the NAAQS in experimental studies, this may indicate that the observations in the epidemiological studies may be due to other variables that are unaccounted for.

Specific questions:

1. Are the beta coefficients in Table C-1 of the PA conceptually well defined?

In reviewing the PA, EPA does not explicitly define the beta coefficient that is used in BenMAP, or how it is derived, instead referring to prior evaluations and the BenMAP manual. EPA should include a definition of the beta coefficient and how it is derived for clarity, especially because both prior assessments and the BenMAP manual do not have specific information related to the studies that are considered in the PA. That is, prior assessments and information in the BenMAP manual include different epidemiology studies than those used in the PA evaluation. Similarly, EPA should define the concentration-response function and how it is used to derive the beta coefficient. In short, in the PA the beta coefficient is not well defined.

2. On the same topic of clear definitions, does the discussion of the BenMAP-CE beta coefficients in the PA and underlying documentation (described as typically representing "the percent change in a given adverse health impact per unit of pollution") unambiguously specify which of the following concepts the coefficients represent?

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. One must refer to the BenMAP manual to understand how the beta coefficients are derived, and as noted previously the manual (or prior assessments) do not include derivation of the coefficients for the specific studies included in the PA evaluation. Based on review of the BenMAP model, it is my understanding that the beta estimates are more in line with the first choice, that is "Beta estimates the percent change in the conditional expected *observed* value of the health impact *associated with* a unit change in the *observed* value of the pollution variable." Or as noted this would be most similar to a regression interpretation.

3. Similarly, is the definition of "concentration-response (C-R) relationships" in the PA and its Appendices (cf p. C-38) adequately clear and unambiguous to support simulation of well-defined causal effects of interventions that change pollution levels?

As with the beta coefficients, 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 underly 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.

4. Do the beta coefficients in the PA overcome these methodological objections to using relative risks, regression coefficients, and related measures of association to predict (or simulate) effects of interventions?

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.

8. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately describe and discuss the extent (if any) to which their values reflect potential omitted confounders of the association between mortality risks and PM2.5 levels (e.g., lagged daily high and low temperatures and humidity in the weeks preceding mortality, if these contribute both to (possibly delayed) mortality and increased energy usage and PM2.5 pollution.

The discussion of uncertainty in the PA is limited. The PA only discusses potential confounding by co-pollutants qualitatively, noting the confounding could impact results in both directions and is of low-medium magnitude and further that most studies evaluated in the ISA found “relatively unchanged” effect estimates in co-pollutant models. 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.

There are several examples in the literature that suggest confounding by co-pollutants or other variables and to my knowledge this issue is far from being resolved as suggested by the discussions in the PA. For example, Burnett *et al*. (1997) evaluated confounding by gaseous co-pollutants in a study of hospital admissions (HA) and cardiovascular and respiratory disease in Toronto, Ontario from 1992 to 1994. The authors observed that in multi-pollutant models, PM contributions were reduced and not statistically significant. The authors concluded that "PM mass and chemistry could not be identified as an independent risk factor for the exacerbations of cardiorespiratory diseases in this study beyond those attributable to climate and gaseous air pollution." Similarly, in an independent mortality study of short-term NO_2 exposure and non-accidental mortality, $PM_{2.5}$ was associated with mortality in single-pollutant models, but not when it was adjusted for NO_2 (Burnett *et al*., 2004). Lastly, Klemm *et al*. (2004)

demonstrated that all-cause mortality estimates associated with short-term $PM_{2.5}$ exposures in counties in Georgia were highly sensitive to temporal smoothing and confounding by co-pollutants. The authors concluded that "none of these findings supports a causal interpretation of the association between $PM_{2.5}$ and all-disease elderly mortality." The authors suggested that $PM_{2.5}$ was acting as a surrogate for other environmental agents.

Burnett, RT; Cakmak, S; Brook, JR; Krewski, D. 1997. "The role of particulate size and chemistry in the association between summertime ambient air pollution and hospitalization for cardiorespiratory diseases." *Environ. Health Perspect.* 105(6):614-620

Burnett, RT; Stieb, D; Brook, JR; Cakmak, S; Dales, R; Raizenne, M; Vincent, R; Dann, T. 2004. "Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities." *Arch. Environ. Health* 59(5):228-236.

Klemm, RJ; Lipfert, FW; Wyzga, C; Gust, C. 2004. "Daily mortality and air pollution in Atlanta: Two years of data from ARIES." *Inhal. Toxicol.* 16(Suppl. 1):131-141

It is more complicated to control for confounding in long-term studies than in short-term (time-series) studies because inferences from long-term studies are based on differences in pollution levels between cities, as opposed to day-to-day differences in pollution levels in a single city. Thus, factors such as socioeconomic (such as poverty levels) or general lifestyle factors (such as smoking or obesity rates) that vary from city to city could be a potential confounders. Controlling for these factors is difficult and residual confounding is likely almost always present because of lack of data, even if surrogates or some form of the variable can be included in the statistical models. For example, analyses of the ACS data (e.g., Krewski *et al*. 2009), featured analyses that addressed potential biases and confounding associated with various sociodemographic ecological factors and spatial model specifications, which included controlling for individual-level covariates. The information for the individual, however, were taken from the original 1982 enrollment questionnaire, and no follow-up was conducted. In addition, neighborhood-level socioeconomic covariates (*e.g.*, education, poverty levels, and unemployment) were based on census data that also likely varies over time. EPA continues to heavily rely on the ACS cohort analyses, although issues with confounding have not been resolved.

Krewski, D; Jerrett, M; Burnett, RT; Ma, R; Hughes, E; Shi, Y; Turner, MC; Pope, AC III; Thurston, G; Calle, EE; Thun, MJ. 2009. "Extended Follow-up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." HEI Research Report 140. Health Effects Institute, Cambridge, MA

9. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address the extent (if any) to which their values reflect residual confounding of the association between mortality risks and PM2.5 levels (e.g., by daily high

and low temperatures and humidity in the weeks preceding mortality in models that only address seasonal, annual, or averaged temperatures)?

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.

10. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address model uncertainty (e.g., the possibility that the linear no-threshold model specification is incorrect, e.g., because sufficiently low exposure concentrations do not cause pulmonary inflammation and adverse health effects that occur at higher concentrations)?

Model uncertainty remains one of the largest sources of uncertainty in observational air pollution studies. For example, adjustments for temporal trends (with respect to other population behaviors or episodic diseases such as viral infections), control for climate and weather effects, and confounding from co-pollutants have been addressed in a comprehensive re-analysis of seminal time-series studies (*e.g.,* HSC study and NMMAPS), which include extensive sensitivity analyses. The results from these analyses showed that methods used for controlling temporal trends and weather can yield large variations in effect estimates (HEI, 2003). Both HEI and Moolgavkar (2005) stressed that there is no objective statistical test to determine the best method to control for these factors. Analyses of $PM_{2.5}$ effects have been shown to be particularly sensitive to selection of degrees of freedom used in smoothing functions to control for temporal confounding in time-series studies (Dominici *et al*., 2007; Ostro *et al*., 2008).

Dominici, F; Peng, RD; Zeger, SL; White, RH; Samet, JM. 2007. "Particulate air pollution and mortality in the United States: Did the risks change from 1987 to 2000?" *Am. J. Epidemiol.* 166(8):880-888.

Health Effects Institute (HEI). 2003. "Revised Analyses of Time-Series Studies of Air Pollution and Health: Revised Analyses of the National Morbidity, Mortality and Air Pollution Study, Part II. Revised Analyses of Selected Time-Series Studies." Health Effects Institute, Cambridge, MA, May.

Moolgavkar, SH. 2005. "A review and critique of the EPA's rationale for a fine particle standard." *Regul. Toxicol. Pharmacol.* 42:123-144.

Ostro, B; Broadwin, R; Green, S; Wen-Ying, F; Lipsett, M. 2008. "Fine particulate air pollution and mortality in nine California Counties: Results from CALFINE." *Environ. Health Perspect*. 114(1):29-33.

11. Does the PA's discussion of beta values, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address effects on the estimated values of the beta values of exposure uncertainties and estimation errors (e.g., the possibility that individual exposure concentrations among people with adverse health responses tend to be higher than those among people who did not respond, even when they have the same estimated exposure values)?

Exposure measurement error is one of the largest sources of bias in long- and short-term ecological epidemiological studies. Most studies rely on central-site monitors as surrogates for personal PM exposure or newer hybrid models that rely on monitored and modeled concentration, which do not necessarily accurately capture population mobility, or capture uneven distribution of PM attributable to local sources (monitoring sites may represent a nearby source and not human exposures a small distance away), pollution patterns that can be affected by terrain features and weather, or daily variations in PM concentrations or composition that may differ from variations experienced by individuals. Sarnat *et al.* (2009) found that personal exposures to sulfate (a major component of $PM_{2.5}$ in certain parts of the country) average over time; varied by individual, city, and season; and that this variability can lead to C-R functions that do not represent the true relationship between exposure and outcome.

> Sarnat, JA; Brown, KW; Bartell, SM; Sarnat, SE; Wheeler, AJ; Suh, HH; Koutrakis, P. 2009. "The relationship between averaged sulfate exposures and concentrations: results from exposure assessment panel studies in four U.S. cities." *Environ. Sci. Technol.* 43(13):5028-5034.

12. Does the PA adequately assess the suitability of the designs of the studies used to estimate beta values for purposes of valid causal inference and simulation? Does the PA's discussion of uncertainty and sensitivity analyses adequately address the internal validity and external validity (generalizability) of the estimated beta values used to simulate the causal impacts on public health risks of changing PM2.5 levels?

As noted previously, in the 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.

13. Does the PA's discussion of beta values adequately address attribution of risk in the presence of joint causes? For example, if a unit change in PM2.5 levels has different expected effects on mortality risk for a person below the poverty line and during extremely hot or cold weather than it would for an initially similar (exchangeable) person with higher income and no exposure to extreme temperatures, then how much of the statistical "effect" of PM2.5 on mortality risk, as reflected in the beta values in Table C-1, should be attributed to income and weather variables? Conversely, how well does the discussion in the PA make clear how much of the estimated beta value for PM2.5 is actually contributed

by other variables (such as temperature extremes and poverty) that would not necessarily be changed by an intervention that reduces PM.5 levels?

Although sensitivity analyses have been conducted to evaluate whether differences in socioeconomic status or extreme temperature explain the heterogeneity in the mortality risk estimates across regions and seasons, the issue of observed heterogeneity across regions and seasons has not been resolved. Also, using National estimates from multi-city studies tends to mask this heterogeneity and is likely inappropriate for use when evaluating effects in specific cities. For example, in an analysis conducted by Krewski *et al.* (2009), the authors evaluated mortality in two cities, Los Angeles and New York City using estimates of PM2.5 exposure based on land-use regression (LUR) models and/or kriging methods. Despite the use of similar methods to estimate $PM_{2.5}$ exposures in the two cities, the city-specific mortality risks differed. The LA mortality risks were only statistically significant for all-cause and ischemic heart disease mortality for estimates were adjusted for 44 individual-level covariates, but not for COPD or lung cancer. None of the mortality risk estimates were statistically significant in NYC. These results indicate that mortality risks are not consistently positive or significant across cities and the CR functions (or betas) would differ and even be null if city-specific estimates were used in the BenMAP analysis rather than National estimates. In fact, based on the intra-urban study results, the authors concluded that "comparing the mortality risk estimates obtained from the Nationwide Analysis with those from the Intra-Urban Analyses indicates that the Nationwide risk estimates cannot be directly applied to all urban areas within the United States and that mortality risk estimates can vary appreciably among large urban areas with different characteristics" (Krewski *et al.*, 2009)

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. This is an aspect of uncertainty in the risk assessment that needs to be further discussed in the PA with regards to interpreting the results.

Furthermore, going back to the ISA, EPA needs to include a discussion of alternative hypotheses that could explain the observed associations in observational epidemiological studies, particularly given the very small effect estimates. For example, in all-cause mortality, a large portion of the mortality is attributed to cardiovascular mortality. Cardiovascular mortality has a large number of risk factors and it is well-known that activity and stress can trigger heart attacks. It is plausible that stress can confound relationships between day to day air pollution and mortality (i.e., related to driving and commuting) particularly in urban areas.

Krewski, D; Jerrett, M; Burnett, RT; Ma, R; Hughes, E; Shi, Y; Turner, MC; Pope, AC III; Thurston, G; Calle, EE; Thun, MJ. 2009. "Extended Follow-up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." HEI Research Report 140. Health Effects Institute, Cambridge, MA

14. Overall, does the PA and its underlying documents (e.g., the BenMAP-CE documentation) make a convincing technical case that its simulated health impacts of reductions in PM2.5 are trustworthy and usefully accurate? How confident can policy analysts and decision makers be in the predictive validity of the simulated results?

The PA would benefit from a more detailed discussion of the uncertainties in the risk estimates that are presented, and in particular it 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. Also presenting risk estimates down to a zero level of air pollution is misleading and unrealistic, given that there is always some level of PM2.5 that can't realistically be controlled. If the risk assessment is included in the PA, the focus should be on the comparison between current or just meeting the NAAQS with alternative NAAQS and should include a statistical analysis to assess whether *differences* in risk estimates are statistically different between different scenarios given full consideration of the uncertainties (not limited to the statistical uncertainty in the beta coefficient).

15. Are there other statistical or methodological issues that you would like to comment on that you believe might help the CASAC to assess the validity and soundness of the PA and its simulations for effects on health risks of changing PM2.5 levels, or that might help to improve the technical and scientific quality of the final PA?

As noted previously, the uncertainties in the BenMAP analyses are limited and should be expanded to include other sources of uncertainty, including but not limited to the modeling uncertainty in the ambient air concentrations for the scenarios of just meeting the NAAQS or alternative NAAQS (see for example, Table 1 from Dr. Lange's questions).

16. How can techniques of formal causal modeling and analysis best be applied to improve the clarity of definitions and communication and scientific soundness of simulations, inferences, causal interpretations, generalizations, and policy-relevant conclusions in the PA? Please comment on whether any aspects of the following (or other) causal model formalisms can substantially improve the clarity and scientific soundness of the analyses and simulations in the PA: causal graph and DAG methods, conditional independence tests, intervention and interrupted time series analyses, other quasi-experimental methods, Wiener-Granger causality and transfer entropy, causal dynamic Bayesian networks (DBNs), other information-theoretic and graph methods, Simon-Iwasaki causal ordering , non-parametric structural equations models, mediation analysis.

There are many aspects of the EPA causal framework that could be improved to provide greater transparency and scientific soundness to the NAAQS process. For example, the improvements would start as part of the development of systematic review protocol that follows modern systematic review methodology, including a solid approach for assessing study quality and risk of bias and ranking a weighing studies based on study quality and risk of bias findings. In addition, EPA's integration of the

evidence could be improved such that no one line of scientific evidence (e.g., epidemiological evidence) is used exclusively to draw conclusions, especially when other lines of evidence are not consistent or coherent with the selected line of evidence.

Also, with regards to specific lines of evidence, in this case epidemiological evidence, EPA needs to provide more clarity regarding the inclusion and exclusion of studies. As was noted by CASAC in it review of the draft PM ISA, EPA excluded many studies that were informative regarding causal associations between PM and various health effects, including many so called "accountability" studies. Similarly, EPA tends to exclude occupational studies, even though these studies can also provide important information regarding health effects at high levels of PM exposure and whether this evidence is consistent and coherent with observational population-based studies.

I cannot comment specifically on formal causal modeling and analyses techniques as these are outside my expertise.

Questions from Dr. Sabine Lange

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

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.

2. 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.

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.

3. 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?

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.

4. 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?

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.

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

I am not sure if I understand the question. 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.

6. 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?

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.

Stieb, DM; Judek, S; Burnett, RT. 2002. "Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season." *J. Air Waste Manage. Assoc.* 52:470-484.

<u>Questions from Dr. Steven Packham</u>

Are there areas (e.g., specific aspects of biological causation, pulmonary toxicology, or causality in epidemiology) in which additional knowledge is required to appraise the adequacy and basis of existing, new, or revised PM NAAQS to protect public health with an adequate margin of safety?

The NAAQS process in general could benefit from a more systematic process of evaluating individual studies assessing study quality and weighing and ranking studies. This would allow for a more balanced consideration of the evidence and a more transparent and objective analysis. Similarly, better integration of the evidence is needed in order to assess consistency and coherence in the data from different lines of evidence (epidemiological and experimental). The evidence is currently presented in the PA, heavily weighed in favor of the epidemiological evidence. These studies are not sufficient for assessing causality because of large inherent limitations including exposure measurement errors, uncontrolled confounding, among other issues. Information derived from controlled exposure studies and animal studies is

apparently dismissed in the PA. For example, EPA notes that animal toxicology studies assess effects to PM2.5 concentrations ”well-above the concentrations likely to be allowed by the current PM2.5 standards (pg. 3-49 in the PA)” and therefore based on this and challenges with from extrapolating from effects in animals to humans, EPA concludes that “animal toxicology studies are of limited utility in informing conclusions on the public health protection provided by the current or alternative primary PM2.5 standards,” similarly, the for controlled human exposure studies EPA concluded:

“the PM2.5 exposure concentrations evaluated in most of these studies are well-above the ambient concentrations typically measured in locations meeting the current primary standards. Therefore, 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 standard.”

I don’t agree with the EPA conclusions. Instead, I think there is utility in evaluating all lines of evidence to determine if these lines of evidence provide consistent and coherent evidence of an effect at concentrations lower than the current NAAQS. Both animal studies and controlled exposure studies have been used in numerous evaluations (by EPA and others) for assessing levels of no or low adversity.

Can you suggest additional specific scientific disciplines and areas of biomedical informatics and research (e.g., systems biology methods for clarifying biological causal pathways, mechanisms, modes of action, quantitative causal dose-response relationships) that should be included in future reviews of other criteria pollutants?

If EPA deems that the animal and controlled exposure studies are inadequate for informing the adequacy of the level of the NAAQS, it stands to reason that more studies of this kind at relevant ambient exposure levels are needed. It would also be important to conduct animal studies that evaluate not only markers of early effects (such as inflammatory reactions), but the complete chain of events that might result in overt disease or even mortality at relevant ambient exposures (i.e., equivalent levels below the current NAAQS).

Can you describe the research efforts necessary to provide the required information? To what extent has the needed research already been done, or started? For example, are there crucial experiments or research initiatives that could clarify the shape of the PM2.5-chronic inflammation causal dose-response relationships at relevant exposure concentrations? Are there specific data analyses (e.g., testing for confounding by weather variables over more days prior to mortality) that could clarify the causal interpretation of epidemiological associations relied on in the draft PA to simulate effects of interventions?

There is a large extensive body of literature that I believe already can be used if integrated properly and evaluated objectively to assess consistency and coherence across scientific lines of evidence. Solely relying on the epidemiological findings is not sufficient to establish causality or to determine a level of exposure with an adequate margin of safety (especially because as EPA notes there is no discernible threshold level that can be identified in epidemiology studies). The epidemiology studies are hindered because both the exposure and outcome are very common (we are all exposed to PM and we all die) and this makes the analysis challenging in the face of many variables that can explain the observed associations. The issue is further complicated by the fact that PM composition varies and there is no single or group of constituents that has been found to be the 'causative' component.

EPA identifies areas of future research and I agree that these areas of research could potentially aid in assessing future evaluations. These include (as noted above) toxicological studies that evaluate the full chain of events that would results in disease, long-term animal studies at low PM levels, and longer controlled human exposure studies at low PM levels. These studies could include PM of varying composition to help elucidate whether there are differential effects depending on composition.

Can you suggest additional areas of scientific literature review on species and individual human organism's capacities of adaptation to inhaled environmental stressors that might help establish margins of safety when exposed to ambient levels of air pollution?

One scientific line of evidence that I can think of that is missing in the evaluation of PM effects is the occupational literature. This literature has been included in past assessments to a limited degree (for example workers exposed to diesel particulate matter). I think inclusion of these studies can provide an important perspective and can also help identify levels of exposure that have been shown to be hazardous in these population groups (compared to recent observational population studies).

Could you provide additional information on the relative contributions to air pollution concentrations and resulting health effects of natural and anthropogenic activity? For example, is either one alone, or are both natural and anthropogenic activities together, sufficient to cause the magnitudes of adverse health effects attributed to PM2.5 in the Draft PA?

This is an interesting question, and I am not sure what the answer is. EPA does not distinguish between concentrations of PM from natural and anthropogenic sources and in fact assessed risks down to a zero (non-existent) PM concentration. EPA does discuss "background" level of PM in the PA, but these background levels are not considered in the interpretation of the scientific evidence or in the risk assessment. In addition, PM from different sources will vary in composition and to date there is no specific PM component of group of components that have been identified as the causative agents that could account for the observed associations between PM and mortality.

Could you provide additional information on any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards?

This is outside my area of expertise, however, along with the NAAQS process, EPA also conducts a Regulatory Impact Assessment (also using BenMAP) that not only quantifies the health impacts, but also calculated the value associated with the health impacts (or avoided health impacts from for example lowering the NAAQS). These health impacts are driven by mortality and the value of a "statistical life" and far exceed estimated costs of complying with the NAAQS regulations.

Questions from Dr. Mark Frampton

1. Some CASAC members are of the opinion that the scientific evidence base is insufficient to support a causal relationship between PM exposure and mortality. Is there evidence that would support a reconsideration of the current and long-held views of this causal relationship as expressed in the PA and ISA?

I agree that the observational epidemiological evidence alone is insufficient to establish a clear causal relationship between PM2.5 and health effects at levels below the current NAAQS. The study designs cannot overcome the challenges of exposure measurement errors and potential confounding by a myriad of variables – particularly when evaluating a very common exposure and a very common outcome (i.e., mortality). All that the epidemiological literature can tell us is that there is a relatively small statistical association between estimated ambient PM2.5 concentrations and various health outcomes. These associations vary in precision and magnitude on both temporal and spatial scales (i.e., by city or region and by season), and yet this heterogeneity does not appear to be related with PM composition, or even with relative PM concentrations (i.e., large health effects are often observed in areas with lower PM concentrations compared to higher PM concentrations). The observational epidemiology literature also is not consistent with the occupational epidemiological literature. For example, in occupations associated with high PM exposures (carbon black, coal, talc, etc.) workers may develop respiratory illnesses, but high rates or cardiovascular diseases or all-cause mortality are not observed in these populations, despite being exposed to levels several fold higher than ambient concentrations. Similarly, controlled exposure studies and animal studies indicate inconsistent results at high PM exposures.

2. In the long-term epidemiological studies of PM mortality, heterogeneity between and within cities has been cited as a source of uncertainty in drawing conclusions about causality. Please opine on the level of uncertainty that is represented by this heterogeneity, and the impact if any, on the conclusions in the PA.

As noted above and copied here:

Although sensitivity analyses have been conducted to evaluate whether differences in socioeconomic status or extreme temperature explain the heterogeneity in the mortality risk estimates across regions and seasons, the issue of observed heterogeneity across regions and seasons has not been resolved. Also, using National estimates from multi-city studies tends to mask this heterogeneity and is likely inappropriate for use when evaluating effects in specific cities. For example, in an analysis conducted by Krewski *et al.* (2009), the authors evaluated mortality in two cities, Los Angeles and New York City using estimates of PM2.5 exposure based on land-use regression (LUR) models and/or kriging methods. Despite the use of similar methods to estimate $PM_{2.5}$ exposures in the two cities, the city-specific mortality risks differed. The LA mortality risks were only statistically significant for all-cause and ischemic heart disease mortality for estimates were adjusted for 44 individual-level covariates, but not for COPD or lung cancer. None of the mortality risk estimates were statistically significant in NYC. These results indicate that mortality risks are not consistently positive or significant across cities and the CR functions (or betas) would differ and even be null if city-specific estimates were used in the BenMAP analysis rather than National estimates. In fact, based on the intra-urban study results, the authors concluded that "comparing the mortality risk estimates obtained from the Nationwide Analysis with those from the Intra-Urban Analyses indicates that the Nationwide risk estimates cannot be directly applied to all urban areas within the United States and that mortality risk estimates can vary appreciably among large urban areas with different characteristics" (Krewski *et al.*, 2009)

3. Please opine on the adequacy of the causality analysis framework currently used by the EPA, and whether and how the concepts espoused by Dr. Cox should, or should not, be incorporated into the NAAQS causality framework. Also please comment on the implications, of any changes in the causality framework that you would recommend, for the analyses and conclusions in this current PA.

As noted above and copied here (with some additions):

The NAAQS process in general could benefit from a more systematic process of evaluating individual studies for assessing study quality and weighing and ranking studies and this is currently lacking the NAAQS framework. This would allow for a more balanced consideration of the evidence and a more transparent and objective analysis (i.e. instead of relying so heavily on one line of evidence – such as the epidemiology evidence). Similarly, better integration of the evidence is needed in order to assess consistency and coherence in the data from different lines of evidence (epidemiological and experimental) and this is also lacking as applied under the current framework. The evidence is currently presented in the PA, heavily weighed in favor of the epidemiological evidence. These studies are not sufficient for assessing causality because of large inherent limitations including exposure measurement errors, uncontrolled confounding, among other issues. Information derived from controlled exposure studies and animal studies is apparently dismissed in the PA, even though this type of evidence is often used to derive health-based limits for numerous chemicals.

I cannot specifically comment on the causal methodologies that Dr. Cox has recommended as this is outside my area of expertise. However, a more balanced assessment of the scientific evidence would likely result in changes to the overall conclusions in the PA, or at least provide more context that would allow for a more critical interpretation of the data.

4. "There is inadequate evidence for the 'likely to be causal' conclusion for long-term PM2.5 exposure and cancer." This is based on epidemiological studies that do not appear to adequately differentiate incident cancer and cancer-related mortality because the exposure time frames for most of these studies are insufficient to draw conclusions about incident cancer. Do you agree with the CASAC's findings in this matter? Please discuss the evidence (or lack thereof) that supports your opinion.

In the previous EPA PM evaluation, EPA concluded that evidence was suggestive but not enough to conclude that it is "likely to be causal" and this is largely because of the lack of consistent finding in the epidemiological literature and the lack of support of *in vivo* mutagenicity or carcinogenicity in experimental animals. I agree that the literature still does not support a "likely to be causal" conclusion and that many studies do not provide an adequate lag time to establish temporality of effect (that is sufficient time that has elapsed between exposure and effect). As noted by Dr. Lange in her CASAC comments this includes the seminal studies of the H6C and CCHS cohorts.

5. Please comment on the appropriateness and completeness of the approaches used in this PA to assess the risks of exposure, and the assessments of risk reduction of alternate standards, for PM2.5 and PM10 (sections 3 and 4, respectively).

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). EPA also estimates effects down to an unrealistic zero level of PM, which is misleading. Another important issue is that EPA applies National-level effect estimates to specific cities, even though studies have observed substantial heterogeneity in effect estimates across cities. Lastly, EPA only presents a limited uncertainty analysis that incorporates only the statistical uncertainty in the effect estimate derived from the epidemiological study. Other important sources of uncertainty are not quantified.

Questions from Dr. James Boylan

Chapter 2 – PM Air Quality

Is the discussion on sources of emissions accurate and complete? If not, what additional information needs to be included?

In general, the discussion is accurate, although somewhat limited and basic. Additional information that should be included in this section is information on how the PM emissions vary by region and also by season. In addition, there should be some discussion about the uncertainties in estimating the sources and emissions of PM.

Is the discussion on ambient monitoring accurate and complete? If not, what additional information needs to be included?

In general, this section is also accurate, but also somewhat limited. It would be useful to have information regarding the sampling schedule and how this has changed (or not) over time. For example, PM measurements taken every 3 or 6 days. It would also be useful to have information on how the ambient monitors are used for determining compliance with the NAAQS. I did not see this information in this section.

Is the discussion on ambient measurement correlations and trends accurate and complete? If not, what additional correlations and trends need to be included?

I found this section to be useful and adequate.

To what extent are biases associated with PM10, PM2.5, and ultrafine measurements discussed? How would differing PM2.5 biases associated with FRM vs. FEM continuous measurements (e.g., FEMs typically show higher PM2.5 concentrations compared to FRMs) impact the evidence-based and risk-based PM2.5 assessments in Chapter 3?

This is an area of uncertainty that is not included in either the epidemiological studies or in the risk assessment, although I would imagine that other sources of uncertainty would likely be much higher than the biases related to measurement errors. For example, exposure measurement errors are likely to be higher in epidemiological studies, and modeling uncertainty and potential biases are likely to be higher in the risk assessment.

Is the discussion on hybrid modeling approaches accurate and complete? If not, what additional information needs to be included?

Overall, I think this section was adequate. However, one area that I think is critical for evaluating exposure measurement error in the epidemiological studies is how ambient concentrations (whether measured or modeled) relate to personal exposures in individuals that spend time in different microenvironments. This is additional information that could be included in this section.

Is the discussion on performance methods for evaluating hybrid modeling methods accurate and complete? If not, what additional information needs to be included?

I am not familiar enough with these models to opine on the adequacy of the performance evaluations.

Is the discussion on background concentrations accurate and complete? If not, what additional information needs to be included?

I was surprised by the low background levels of PM reported in this section. While I have not reviewed the literature on this issue, as with the question above, one important questions is how ambient air pollution levels relate to personal exposures. One can imagine that any individual person may have very different background levels of exposure depending on their personal activities, workplace exposures, and proximity to different natural sources of PM (e.g., wildfires).

Chapter 3 – Review of the Primary PM2.5 Standards

Is the evidence-based analysis presented in Chapter 3 scientifically sound?

As noted above and copied here:

The NAAQS process in general could benefit from a more systematic process of evaluating individual studies for assessing study quality and weighing and ranking studies and this is currently lacking the NAAQS framework. This would allow for a more balanced consideration of the evidence and a more transparent and objective analysis (i.e. instead of relying so heavily on one line of evidence – such as the epidemiology evidence). Similarly, better integration of the evidence is needed in order to assess consistency and coherence in the data from different lines of evidence (epidemiological and experimental) and this is also lacking as applied under the current framework. The evidence is currently presented in the PA, heavily weighed in favor of the epidemiological evidence. These studies are not sufficient for assessing causality because of large inherent limitations including exposure measurement errors, uncontrolled confounding, among other issues. Information derived from controlled exposure studies and animal studies is apparently dismissed in the PA, even though this type of evidence is often used to derive health-based limits for numerous chemicals.

Is the risk-based analysis presented in Chapter 3 and Appendix C scientifically sound?

As noted above and copied here:

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). EPA also estimates effects down to an unrealistic zero level of PM, which is misleading. Another important issue is that EPA applies National-level effect estimates to specific cities, even though studies have observed substantial heterogeneity in effect estimates across cities. Lastly, EPA only presents a limited uncertainty analysis that incorporates only the statistical uncertainty in the effect estimate derived from the epidemiological study. Other important sources of uncertainty are not quantified.

Is the discussion on following topics adequate and complete? If not, what additional information needs to be included?

o Study area selection,

I did not have any specific issues with the selection of study areas.

o Health outcomes (e.g., decision to focus on mortality and ignore cardiovascular and respiratory effects),

Previous evaluations have included morbidity outcomes and these are useful to provide context (i.e., higher ER visits and hospital admissions would be expected compared to mortality)

o Concentration-response functions,

While EPA provides some justification for the selection of studies, the justification does not include an evaluation of study quality or risk of bias. EPA should include more information on why it considered these studies to be of higher quality than other available studies.

o PM2.5 air quality scenarios evaluated,

It is unclear why EPA evaluates separately the scenarios for just meeting the NAAQS (vs. zero) and meeting alternative standards (vs. zero). EPA should consider evaluating the baseline vs. the control scenarios to estimate the difference directly.

o Model-based approach to adjusting air quality,

I am not familiar enough with the methodology to opine.

o Linear interpolation/extrapolation to additional annual standard levels, and

I am not familiar enough with the methodology to opine.

o Characterization of variability and uncertainty in the risk estimates.

The quantitative uncertainty characterization is extremely limited and should be expanded to include other sources of uncertainty (see for example Dr. Lange's questions).

Are the areas for additional research adequate and complete? If not, what additional areas need to be included?

Appendix C – Supplemental Information Related to the Human Health Risk Assessment

Is the air quality modeling approach to projecting PM2.5 concentrations to correspond to just meeting the NAAQS (AQS, CMAQ, Downscaler, SMAT-CE, project monitors to just meet NAAQS, project spatial fields to correspond to just meeting the NAAQS) scientifically sound? If not, what are your concerns and how should they be addressed?

I am not sufficiently familiar with the air quality modeling to opine on whether it is scientifically sound, however, the modeling uncertainty should be quantified and this uncertainty should be incorporated in the risk assessment.

Is the CMAQ model configuration and input files used in the air quality modeling appropriate for this application? If not, what updates are recommended?

I am not sufficiently familiar with the air quality modeling to opine on whether the input files are appropriate.

Are the CMAQ model performance metrics that were evaluated appropriate and adequate for this application? Are there any concerns with the model performance in any of the study areas used in the human health risk assessment? If so, how should these concerns be addressed in the health risk assessment?

I am not sufficiently familiar with the air quality modeling to opine on whether the model performance metrics are appropriate, but as noted above the modeling uncertainty should be incorporated in the calculation of the risk estimates.

Is the health risk modeling approach using BenMAP-CE appropriate for this application? If not, what are your concerns?

As noted above and pasted here:

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). EPA also estimates effects down to an unrealistic zero level of PM, which is misleading. Another important issue is that EPA applies National-level effect estimates to specific cities, even though studies have observed substantial heterogeneity in effect estimates across cities. Lastly, EPA only presents a limited uncertainty analysis that incorporates only the statistical uncertainty in the effect estimate derived from the epidemiological study. Other important sources of uncertainty are not quantified.

Do the risk summary tables showing the impact of alternative PM2.5 standards and graphical plots showing the distribution of risk across ambient PM2.5 levels clearly and accurately summarize the results of the health risk analysis? If not, what additional information should be included?

Even though EPA only presents a limited uncertainty analysis, in the PA this uncertainty is presented as 95% confidence bounds around point estimates. When comparing results between attainment of current NAAQS and alternative NAAQS, however, EPA focuses on the point estimates and does not discuss the uncertainty bounds. EPA should provide statistical analyses to evaluate whether the calculated effect estimates differ significantly between meeting the current NAAQS and alternative NAAQS given the overlapping confidence bounds. The implications of this analysis on the risk conclusions should be included in the PA.

Dr. Duncan Thomas

Before answering the specific questions posed by chairman Dr. Tony Cox and other members of CASAC, I'd like to offer some general observations.

First, I have reviewed substantial portions of the draft PA, focusing particularly on my area of expertise, Chapter 3, as well as some parts of the External Review Draft ISA dated October 2018 and the CASAC Draft Report (03/07/19) to Assist Meeting Deliberations. In general, I find both the draft PA and the draft ISA to be well written, authoritative, and comprehensive reviews of the literature and thoughtful discussion of the policy implications, including limitations thereof. Two main themes seem to me to emerge from the questions that Drs. Cox and Lange have posed, and as reflected in the CASAC critique of the ISA.

The first is an emphasis on the difference between association and causation, as discussed in the epidemiologic literature for well over a half century, and the limitations of epidemiologic evidence for drawing causal conclusions. In the last couple decades, this has spurred the development of novel methods of "causal inference", i.e., statistical methods for analysis of observational data to address the expected changes in outcomes that might result from hypothetical changes in exposures (often called the Average Causal Effect). Despite the growing interest in such approaches, they have only recently been applied in the field of air pollution epidemiology (e.g., (Moore et al. 2008; Moore et al. 2010; Zigler et al. 2012; Wang et al. 2016; Zigler et al. 2018)). These are important contributions with direct policy applicability, but the vast majority of epidemiologic studies report standard measures of association (e.g., relative risks, etc.). This is not to disparage the bulk of epidemiologic studies, only to emphasize that it would be inappropriate to dismiss them as not addressing causation, given their concordance and the general conformity with the criteria used by epidemiologists for decades to qualitatively evaluate causation.

The consultants received a list of additional citations by Dr. Cox on 9/24/19 about his views on causality. These were provided as background, but the consultants were not asked specifically to comment on them. I have not read these four papers, although I have read others by him on this topic, have peer-reviewed one of them in detail, and am generally familiar with his views. If requested, I can review them and elaborate further, but the main points are covered in the responses to the questions below.

The second recurring theme is an emphasis on uncertainty. All scientific evidence — observational or experimental — is always fraught with some measure of uncertainly, generally addressed qualitatively in the Discussion section of scientific papers and quantitatively in terms of such statistical measures as standard errors of estimates, measurement error adjustments, etc. Air pollution epidemiology is no exception. Critical reviews of the literature and quantitative meta-analyses pay particular attention to

both quantitative measures of uncertainty as well as judgments of study quality, potential bias, etc. In this regard, my impression from what I've read of the draft ISA and PA documents is that the authors have done an excellent job of addressing these uncertainties and their policy implications.

I will elaborate on these general points in responding to the specific questions from CASAC members below. I am limiting my answers to a subset of the questions that are closest to my area of expertise in biostatistics methods, rather than the substance matter of air pollution epidemiology (with which I have some familiarity from my work in the field for about 25 years, but would not consider myself having the comprehensive knowledge of a substance matter expert).

<u>Questions from Dr. Anthony Cox</u> [edited for brevity]

Are the C-R models in Table C-1 appropriate, logically valid, and empirically well-validated, for answering the causal question of how changes in PM2.5 levels would change health risks? Both specific and general versions of this question are of interest:

- *Are the specific regression models in Table C-1 known to produce valid predictions or simulations of how changing PM2.5 exposure would change population health effects or risks?*
- *In general, are such regression models appropriate for predicting how an intervention or manipulation that changes the value of a predictor (right-hand side variable) would cause the value (or frequency distribution) of the dependent variable to change? If certain conditions must hold for this to be an appropriate use of regression models, what are they, and has the PA shown that they are satisfied for the C-R regression models in Table C-1?*

RESPONSE: These general questions are adequately covered in my responses to the specific questions below.

1. Are the beta coefficients in Table C-1 of the PA conceptually well defined? That is, are their intended conceptual meanings and causal interpretations clear and unambiguous? If so, can the definition of beta be expressed using standard epidemiological terms (e.g., controlled direct effects, natural direct effects, total effects, indirect effects, and so forth?) [references omitted]

RESPONSE*:* The "standard epidemiological terms" mentioned in the question relate to mediation analysis, which is not relevant here. If the goal were to assess whether the effects of air pollution on mortality were mediated through some intermediate variable (say, inflammation), then the natural indirect effect would describe the effect of the pathway pollution → inflammation → mortality; the natural direct effect that of pollution → mortality through mechanisms other than inflammation; and the total effect the sum of the two. Likewise if the goal were to assess the effect of socioeconomic effect on mortality through air pollution, the indirect effect would describe the pathway socioeconomic status → pollution → mortality; the direct effect that of socioeconomic status on mortality through mechanisms

other than pollution; and the total effect the sum of the two. In either case, the terms refer to the expected causal effect of some intervention. For example, Zigler et al. (2012) partition the average causal effect of a hypothetical intervention to regulate air pollution into "associative" and "dissociative" components: regulation that has an effect in areas where it causes improvement in air quality, vs regulation that improves mortality without reducing pollution, e.g., through some other mechanism such as changes in unemployment.

Other than the handful of references cited above, most epidemiologic analyses report measures of association, such as relative risks, or derived quantities such as the change in relative risk per unit change of exposure. These are well defined quantities, although they lack a direct causal interpretation (more on this below).

*2. On the same topic of clear definitions, does the discussion of the BenMAP-CE beta coefficients in the PA and underlying documentation (*described *as typically representing "the percent change in a given adverse health impact per unit of pollution") unambiguously specify which of the following concepts the coefficients represent?*

a) Beta estimates the percent change in the conditional expected observed value of the health impact associated with a unit change in the observed value of the pollution variable. …

b) Beta estimates the percent change in the mean value of the health impact variable caused by a manipulation or intervention that changes the value of the pollution variable by 1 unit …, while holding the values of all other variables fixed at the values they had before the intervention. …

c) Beta estimates the percent change in the mean value of the health impact caused by a manipulation or intervention that changes the value of the pollution variable by 1 unit … , while allowing the values all other variables to change in response to the changes in pollution. …

d) Beta estimates the percent change in the mean value of the health impact caused by a manipulation or intervention that changes the value of the pollution variable by 1 unit … , while holding the values of all other variables … fixed at the values they would be expected to have after the intervention.

e) Beta means something different from any of the above.

Do all of the beta values in Table C-1 refer to the same one of these concepts, or might they refer to different ones (or is the answer not clear)?

RESPONSE: Choice a) would be close, except that it lacks the qualification (except as implied by the word "conditional") provided in choices b) through d) that all other variables are held constant, so I suppose the correct answer would be e) something different. Choices b) through d) are not appropriate because in most analysis of observational epidemiology there is no intervention being studied. Instead, beta coefficients (and the risk estimates from which they are derived) represents a form of weighted average of the observed association between exposure and outcome within strata defined by levels of all

other variables. That said, the usual interpretation is that it represents the expected reduction in an individual's risk of the outcome if exposure had been reduced by one unit, holding everything else in the regression model fixed. The expected value of a regulation that would affect the distribution of exposures in a population on health outcomes can be derived from the beta coefficients using the techniques described in the PA. Since, to my knowledge, none of the studies described in Table C-1 are intervention studies, the beta coefficients given would all have the same "regression interpretation".

3. *Similarly, is the definition of "concentration-response (C-R) relationships" in the PA and its Appendices (cf p. C-38) adequately clear and unambiguous to support simulation of well-defined causal effects of interventions that change pollution levels? …*

RESPONSE: Yes, clear and appropriate for use in the simulation of the effects of hypothetical interventions.

4. … Do the beta coefficients in the PA overcome these methodological objections to using relative risks, regression coefficients, and related measures of association to predict (or simulate) effects of interventions? If so, how were they overcome? If not, does this imply that the simulations in the PA are not necessarily reliable or valid predictors of the real-world effects on public health of reducing PM2.5? Why or why not?

RESPONSE: I don't believe there are serious methodological objections to the use of such measures of association for the purpose of risk assessment, so nothing to overcome. The simulations in the PA appear to me to be valid.

5. *… which, if any, of these measures (relative risks, odds ratios, attributable risks, and regression coefficients) are currently generally accepted in contemporary causal analysis and epidemiology as valid measures of how changing one variable (e.g., exposure) will cause another (e.g., population health responses) to change? …*

RESPONSE: All these epidemiologic measures of risk are widely accepted as valid measures of the association between differences in exposure and outcomes across individuals (in the case of cohort studies) or of the association between differences in exposure and disease rates in a population across time (in the case of time series studies). Their causal interpretation as the expected effect of a hypothetical intervention to change exposure (among individuals or across a population) requires additional assumptions, but these may be reasonable depending on the context.

6. More generally, do the $\Delta y/\Delta x$ values calculated by BenMAP-CE have valid interpretations as causal impacts on y of interventions that change x? …

RESPONSE: Same response as to question 5.

7. Have the beta coefficients in the PA been empirically validated as providing approximately correct or usefully accurate predictions or simulations of how interventions that change air pollution would change population health responses? … To what extent are such findings [from a recent non-EPA review] from actual interventions consistent with the assumptions, confidence intervals, and simulations in the PA?

RESPONSE: Same response as to question 5. I am aware of various quasi-experimental ("accountability") studies of the health effects of changes in air pollution but am not in a position to be able to compare their estimates quantitatively with those from purely observational studies. This seems to be the closest one could hope to come to "empirical validation" of risk reduction predictions.

8. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately describe and discuss the extent (if any) to which their values reflect potential omitted confounders of the association between mortality risks and PM2.5 levels …

RESPONSE: I found the discussion of these points in the PA to be thorough and well-reasoned. Uncontrolled confounding can never be eliminated in any observational study; the best one can do is control for known measured confounders. The same could be said of causal inference analyses of observational data. Only randomized experiments can eliminate confounding (both known and unknown) and then only in expectation, and this is not feasible for air pollution studies for obvious reasons.

9. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address the extent (if any) to which their values reflect residual confounding of the association between mortality risks and PM2.5 levels…

RESPONSE: Same as for question 8.

10. Does the discussion of beta values in the PA, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address model uncertainty (e.g., the possibility that the linear no-threshold model specification is incorrect…)?

RESPONSE: Yes, but worth elaborating. The observed range of risks is generally very small, albeit frequently highly significant because of the large sample sizes. Over this range, differences in the shape of the C-R relationship are virtually indistinguishable. Any C-R function can be approximated by a Taylor series, of which the first term describes the slope of the relationship at zero exposure (Crump et al. 1976); if nonzero, then there is no "threshold" below which there is absolutely no effect (even if very small). Statistical methods are available to test for nonlinearity (as described in chapters 6 and 15 of my

environmental epidemiology textbook (Thomas 2009)), including the existence of a threshold and inter-individual variability, but when the range of risks is small, such methods have very low power.

11. Does the PA's discussion of beta values, including the discussions of uncertainty, confidence intervals, and sensitivity analyses, adequately address effects on the estimated values of the beta values of exposure uncertainties and estimation errors…?

RESPONSE: Yes, same as for question 8.

12. Does the PA adequately assess the suitability of the designs of the studies used to estimate beta values for purposes of valid causal inference and simulation? Does the PA's discussion of uncertainty and sensitivity analyses adequately address the internal validity and external validity (generalizability) of the estimated beta values used to simulate the causal impacts on public health risks of changing PM2.5 levels?

RESPONSE: Yes, the cohort and time series study designs used to assess chronic and acute effects of air pollution are standard in the field and widely accepted as providing valid estimates of association parameters. There relevance for estimating the average causal effect of hypothetical interventions has been addressed above. The response to Dr. Cox's queries by Dr. Vandenberg dated 2/20/19 cogently explains why such hypothetical questions are beyond EPA's mandate.

13. Does the PA's discussion of beta values adequately address attribution of risk in the presence of joint causes? … Conversely, how well does the discussion in the PA make clear how much of the estimated beta value for PM2.5 is actually contributed by other variables … that would not necessarily be changed by an intervention that reduces PM.5 levels?

RESPONSE: Yes. See my response to Dr. Lange's question 5 below for further details.

14. Overall, does the PA and its underlying documents (e.g., the BenMAP-CE documentation) make a convincing technical case that its simulated health impacts of reductions in PM2.5 are trustworthy and usefully accurate? How confident can policy analysts and decision makers be in the predictive validity of the simulated results?

RESPONSE: Yes.

15. Are there other statistical or methodological issues that you would like to comment on that you believe might help the CASAC to assess the validity and soundness of the PA and its simulations for effects on health risks of changing PM2.5 levels, or that might help to improve the technical and scientific quality of the final PA?
RESPONSE: No nothing further.

16. How can techniques of formal causal modeling and analysis best be applied to improve the clarity of definitions and communication and scientific soundness of simulations, inferences, causal interpretations, generalizations, and policy-relevant conclusions in the PA? …

RESPONSE: Formal causal inference methods do indeed have a potentially important role in the future, as exemplified in some of the references cited above, but at present very few studies have exploited them. Without access to the raw data, it would not be reasonable to expect others conducting meta-analyses (e.g., EPA staff) to reanalyze published studies using these techniques (which is not to say that studies for which raw data are not accessible — e.g., for appropriate human subjects reasons — should be excluded).

<u>Questions from Dr. Sabine Lange</u> [edited for brevity]

1. Is it appropriate to compare daily PM2.5 concentrations to the annual average? …. 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…

RESPONSE: I'm not sure I understand this question. 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.

2. Is it informative to derive annual average pseudo-design values for study areas in short-term studies …, 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.

RESPONSE: Both are possible, as discussed in my response to the previous question.

3. 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 … over long periods of time. … 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?

RESPONSE: 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.

4. 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 …?

RESPONSE: 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.

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

RESPONSE: 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.

6. What happens when multiple potential explanatory factors are included in a single variable in an already-complex multiple regression system? 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?

RESPONSE: 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).

Questions from Dr. Mark Frampton (edited for brevity)

1. Is there evidence that would support a reconsideration of the current and long-held views of this causal relationship as expressed in the PA and ISA?

RESPONSE: In my view, none of the new epidemiologic studies would alter the fundamental approach EPA has been using to assess the causality of air pollution health effects, although estimates of the magnitude of effects may be strengthened by new data, particularly on the question of effects below the current standards. (More on this in response to question 3).

2. In the long-term epidemiological studies of PM mortality, heterogeneity between and within cities has been cited as a source of uncertainty in drawing conclusions about causality. Please opine on the level of uncertainty that is represented by this heterogeneity, and the impact if any, on the conclusions in the PA.

RESPONSE: Heterogeneity between and within cities is to be expected and can be useful to explain, e.g., whether variation between cities is due to differences in pollutant mix, for example, or variation within cities is due to age, socioeconomic or health status, etc. Such factors are relevant to the EPA's mandate to protect the most vulnerable segments of the population, and the current PA adequately discusses these issues.

3. ... Please opine on the adequacy of the causality analysis framework currently used by the EPA, and whether and how the concepts espoused by Dr. Cox should, or should not, be incorporated into the NAAQS causality framework. Also please comment on the implications, of any changes in the causality framework that you would recommend, for the analyses and conclusions in this current PA.

RESPONSE: In my view, while modern causal analysis methods being developed in the statistical literature have potential utility in the future, they are not presently amenable to the kinds of reanalysis of primary data by EPA staff that Dr. Cox is calling for (see my response to Dr. Cox's question 16). The few examples of recent applications of such methods by investigators with access to the raw data that I am familiar with have been cited above and merit consideration and discussion in the PA.

4. The CASAC letter to the Administrator (April 11, 2019) states, "There is inadequate evidence for the 'likely to be causal' conclusion for long-term $PM_{2.5}$ exposure and cancer." This is based on epidemiological studies that do not appear to adequately differentiate incident cancer and cancer-related mortality because the exposure time frames for most of these studies are insufficient to draw conclusions about incident cancer. Do you agree with the CASAC's findings in this matter? Please discuss the evidence (or lack thereof) that supports your opinion.

RESPONSE: The challenge of establishing causal long-term effects of air pollution of cancer incidence or mortality are not fundamentally different. Cohort studies that are based on prospective follow-up of outcomes using only air pollution exposure information available at entry to the cohort would obviously require many years of observation, whether incidence or mortality is being studied. However, many cohorts are able to take advantage of the decades of air quality monitoring data that precedes the initiation of the cohort; those studies that have only place of residence at or since enrollment would obviously be subject to some degree of measurement error for those subjects who had moved previously, but some cohort studies have collected detailed residential histories that would allow comparisons within cities based on individual exposure histories. That said, the only fundamental difference between incidence and mortality is that data on the latter are more widely available on much larger populations, so there are many more such studies.

5. Please comment on the appropriateness and completeness of the approaches used in this PA to assess the risks of exposure, and the assessments of risk reduction of alternate standards, for $PM_{2.5}$ and PM_{10} (sections 3 and 4, respectively).

RESPONSE: I think my other comments here adequately address the first part of the question. I found the assessment of the implications of alternate standards to be a particularly interesting and compelling part of the report.

6. Are there additional key studies that should be considered in sections 3 and 4 of the PA?

RESPONSE: Not to my knowledge, but I do not consider myself a substance matter expert in epidemiology.

Questions from Dr. Steven Packham (edited for brevity)

1. Are there areas (e.g., specific aspects of biological causation, pulmonary toxicology, or causality in epidemiology) in which additional knowledge is required to appraise the adequacy and basis of existing, new, or revised PM NAAQS to protect public health with an adequate margin of safety?

RESPONSE: I can only speak to the field of epidemiology, which has been the workhorse for providing the evidence in support of NAAQSs. Obviously, such research is on-going and will continue to contribute to future risk assessments. The current ISA and PA adequately cover the current state of epidemiologic knowledge. Future efforts to characterize effects below the current standard, as well as the application of causal inference methods to analyze observational epidemiology data and "accountability" studies to evaluate the effects of actual changes in pollution in response to new regulations will be important in the future. Closer integration of experimental toxicology and observational epidemiology would also be helpful, as mentioned below.

2. Can you suggest additional specific scientific disciplines and areas of biomedical informatics and research (e.g., systems biology methods for clarifying biological causal pathways, mechanisms, modes of action, quantitative causal dose-response relationships) that should be included in future reviews of other criteria pollutants?

RESPONSE: Beyond the application of novel statistical methods of causal inference to epidemiologic data, methodological advances for studying mechanisms of action (systems biology, etc.) are out of my area of expertise. That said, I am aware of the excitement over -omics methods (metabolomics, etc., not just genomics) and am currently leading an NCI program project grant on "statistical methods for integrative genomics in cancer" which aims to develop such methods and am collaborating with my epidemiologic colleagues on applications to metabolomic data on air pollution.

3. Can you describe the research efforts necessary to provide the required information? To what extent has the needed research already been done, or started? For example, are there crucial experiments or research initiatives that could clarify the shape of the PM2.5-chronic inflammation causal dose-response relationships at relevant exposure concentrations? Are there specific data analyses (e.g., testing for confounding by weather variables over more days prior to mortality) that could clarify the causal interpretation of epidemiological associations relied on in the draft PA to simulate effects of interventions?

RESPONSE: See my responses to questions 1, 2, and 4.

4. Can you suggest additional areas of scientific literature review on species and individual human organism's capacities of adaptation to inhaled environmental stressors that might help establish margins of safety when exposed to ambient levels of air pollution?

RESPONSE: Generally beyond my expertise, other than to comment that there is growing interest in methods of formally integrating experimental animal and other mechanistic studies into the analysis of epidemiologic data (e.g., through hierarchical Bayes models).

5. Could you provide additional information on the relative contributions to air pollution concentrations and resulting health effects of natural and anthropogenic activity? For example, is either one alone, or are both natural and anthropogenic activities together, sufficient to cause the magnitudes of adverse health effects attributed to PM2.5 in the Draft PA?

RESPONSE: Beyond my expertise.

6. Could you provide additional information on any adverse public health, welfare, social, economic, or energy effects which may result from various strategies for attainment and maintenance of such national ambient air quality standards?

RESPONSE: Beyond my expertise.

Questions from Dr. James Boylan (subset, Chapter 2 questions are outside of my expertise)

Chapter 3:

1. Is the evidence-based analysis presented in Chapter 3 scientifically sound?

RESPONSE: Yes, see comments above.

2. Is the risk-based analysis presented in Chapter 3 and Appendix C scientifically sound?

RESPONSE: Yes, see comments above.

3. Is the discussion on following topics adequate and complete? If not, what additional information needs to be included?

- *Study area selection,*
- *Health outcomes (e.g., decision to focus on mortality and ignore cardiovascular and respiratory effects),*
- *Concentration-response functions,*
- *PM2.5 air quality scenarios evaluated,*
- *Model-based approach to adjusting air quality,*
- *Linear interpolation/extrapolation to additional annual standard levels, and*
- *Characterization of variability and uncertainty in the risk estimates.*

RESPONSE: No further comments on any of these points beyond my responses elsewhere in this document. In general, I found the PA's treatment of these issues adequate and complete.

4. Are the areas for additional research adequate and complete? If not, what additional areas need to be included?

RESPONSE: See my response to Dr. Packham's questions 1-4.

Appendix C

1. Is the air quality modeling approach to projecting PM2.5 concentrations to correspond to just meeting the NAAQS (AQS, CMAQ, Downscaler, SMAT-CE, project monitors to just meet NAAQS, project spatial fields to correspond to just meeting the NAAQS) scientifically sound? If not, what are your concerns and how should they be addressed?

RESPONSE: Beyond my expertise

2. Is the CMAQ model configuration and input files used in the air quality modeling appropriate for this application? If not, what updates are recommended?

RESPONSE: Beyond my expertise.

3. Are the CMAQ model performance metrics that were evaluated appropriate and adequate for this application? Are there any concerns with the model performance in any of the study areas used in the human health risk assessment? If so, how should these concerns be addressed in the health risk assessment?

RESPONSE: Beyond my expertise.

4. Is the health risk modeling approach using BenMAP-CE appropriate for this application? If not, what are your concerns?

RESPONSE: While I am not personally familiar with the BenMAP-CE program, the description of the methodology in Appendix C seems clear and appropriate for this task.

5. Do the risk summary tables showing the impact of alternative PM2.5 standards and graphical plots showing the distribution of risk across ambient PM2.5 levels clearly and accurately summarize the results of the health risk analysis? If not, what additional information should be included?

RESPONSE: I found the presentation of the risks from alternative PM2.5 standards to be clear.

REFERENCES

Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, Godleski JJ, Coull BA (2014). Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. Biostatistics.

Crump KS, Hoel DG, Langley CH, Peto R (1976). Fundamental carcinogenic processes and their implications for low dose risk assessment. Cancer Res, 36:2973-9.

Kunzli N, Medina S, Kaiser R, Quenel P, Horak F, Jr., Studnicka M (2001). Assessment of deaths attributable to air pollution: Should we use risk estimates based on time series or on cohort studies? Am J Epidemiol, 153:1050-5.

Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcorn S, Tager I (2008). Ambient ozone concentrations cause increased hospitalizations for asthma in children: An 18-year study in southern california. Environ Health Persp, 116:1063-70.

Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcorn S, Tager I (2010). Ambient ozone concentrations and cardiac mortality in southern california 1983-2000: Application of a new marginal structural model approach. Am J Epidemiol, 171:1233-43.

Thomas D (2005). Why do estimates of the acute and chronic effects of air pollution on mortality differ? J Toxicol Environ Health A, 68:1167-74.

Thomas DC (2009). Statistical methods in environmental epidemiology. Oxford,Oxford University Press.

Wang Y, Kloog I, Coull BA, Kosheleva A, Zanobetti A, Schwartz JD (2016). Estimating causal effects of long-term pm2.5 exposure on mortality in new jersey. Environ Health Persp, 124:1182-8.

Zeger SL, Thomas D, Dominici F, Samet JM, Schwartz J, Dockery D, Cohen A (2000). Exposure measurement error in time-series studies of air pollution: Concepts and consequences. Environ Health Persp, 108:419-26.

Zigler CM, Choirat C, Dominici F (2018). Impact of national ambient air quality standards nonattainment designations on particulate pollution and health. Epidemiology, 29:165-74.

Zigler CM, Dominici F, Wang Y (2012). Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. Biostatistics, 13:289-302.