UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

April 11, 2019

EPA-CASAC-19-002

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 *Integrated Science Assessment for Particulate Matter (External Review Draft – October 2018)*

Dear Administrator Wheeler:

The Chartered Clean Air Scientific Advisory Committee (CASAC) met on December 12-13, 2018, and March 28, 2019, to peer review the EPA's *Integrated Science Assessment for Particulate Matter (External Review Draft – October 2018)*, hereafter referred to as the Draft ISA. The CASAC's consensus responses to the agency's charge questions and individual review comments from members of the CASAC are enclosed. Major comments and recommendations are highlighted below and detailed in the consensus responses to charge questions.

Overall, the CASAC finds that the Draft ISA does not provide a sufficiently comprehensive, systematic assessment of the available science relevant to understanding the health impacts of exposure to particulate matter (PM). The CASAC recommends that the following fundamental limitations be remedied in a second draft of the ISA for CASAC review.

- *Lack of comprehensive, systematic review* - some of the relevant and important scientific literature is not reviewed and study quality is not systematically considered. The revised ISA should provide a clearer and more complete description of the process and criteria for study quality assessment, including an explanation of how systematic assessments of individual study quality were used in preparing the ISA and the causality determinations.
- *Inadequate evidence for altered causal determinations* - the CASAC finds that the Draft ISA does not present adequate evidence to conclude that there is likely to be a causal relationship between long-term $PM_{2.5}$ exposure and nervous system effects; between long-term ultrafine particulate (UFP) exposure and nervous system effects; or between long-term $PM_{2.5}$ exposure and cancer.
- *Clearer discussion of causality and causal biological mechanisms and pathways* - specifically including pulmonary inflammation.

Some members of the CASAC strongly recommend that all key conclusions in the final ISA be supported by explicit, and in principle, verifiable tests (e.g., statistical tests or experimental results).

The need for substantial revisions to the Draft ISA to provide clearer definitions, and technical details and methods in order to enable meaningful independent scientific review leads to the following two process recommendations:

1. The CASAC recommends development of a Second Draft ISA for CASAC review.
2. The CASAC recommends that the EPA reappoint the previous CASAC PM panel (or appoint a panel with similar expertise) as well as adding expertise in biological mechanisms of causation, causal inference, multi-stressor interactions, and potentially others such as: epidemiology, human clinical studies; comparative toxicology, dosimetry, and extrapolation of findings in animals to humans; characterization of sampling errors and biases from continuous ambient PM measurements and satellite remote sensing aerosol optical depth (AOD) analysis; errors and biases in dispersion modeling and photochemical grid modeling; errors-in-variables methods and effects of exposure (and covariate) estimation errors on epidemiologic study results; epidemiology of low-dose causal concentration-response functions; and effects of PM on visibility impairment, climate, and materials. The panel should be appointed in time to review the Second Draft ISA.

Turning to the parts of the Draft ISA, the CASAC finds that the Executive Summary provides a concise and accessible summary of many of the key findings and conclusions of the Draft ISA for a broad range of audiences. It does not accurately represent the totality of available high-quality scientific evidence. The CASAC recommends that the EPA indicate when exposures referred to are estimates. Some members of the CASAC recommend that the EPA consider distinguishing between the following in statements of key findings and conclusions: effects of PM and effects of confounders (such as poverty and temperature); observed changes and model-predicted changes in public health risks following changes in exposures; assumptions and data on shapes of concentration-response (C-R) functions; results from the total body of scientific evidence and results from selected subsets of evidence; and association and causation.

The CASAC finds that Chapter 1, similar to the Executive Summary, provides an effective summary of material from subsequent chapters. In presenting key conclusions from the other chapters, Chapter 1 should characterize uncertainty about them and their sensitivity to assumptions. Chapter 1 should also discuss inconsistencies and discordant data, as detailed in the consensus responses.

The CASAC finds that Chapter 2 adequately characterizes the sources, chemistry, measurements and modeling of UFP, $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$ (coarse fraction) and usefully describes the extent of available information on the spatial and temporal trends of ambient PM concentrations at various scales. However, clarification is required on some of the PM monitoring and modeling information. Chapter 3, for the most part, clearly and accurately describes methods for exposure measurement and modeling although corrections are needed for some of the modeling information. Errors in exposure estimates arising from different methods, and their effects on risk estimates and on estimates of C-R functions, should be characterized and discussed more fully. Recommendations for several additions and clarifications for both chapters are detailed in the consensus responses to charge questions.

Chapter 4 provides a useful, thorough review of the deposition, clearance, retention, and translocation of inhaled PM, but the CASAC recommends additional discussion of dosimetry exposure concentrations and of how dosimetry study results can be translated to humans exposed to ambient PM concentrations.

The CASAC finds that Chapters 5-12 do not provide adequate discussions of biological plausibility, omit several relevant studies, and mischaracterize others. The CASAC recommends that study inclusion and exclusion criteria for literature referenced in Chapters 5-12 should be more explicitly stated and systematically applied. Chance, bias, and confounding should be more explicitly and completely addressed in presenting and evaluating study results. The CASAC recommends several refinements, improvements, and extensions in the presentation of biological information in Chapters 5-12. The CASAC's recommendations include improving the organization and presentation of the document; revising the biological plausibility sections to clarify and correct several pathways; including concentration information when discussing study results; identifying no- and low-adverse effect levels from the human controlled exposure studies; addressing discrepant results between studies; and further integrating study results in Chapter 12. The CASAC did not reach consensus on whether the EPA had adequately considered and caveated the presented C-R functions. Details about the CASAC's consensus recommendations for Chapters 5-12 are in the consensus responses to charge questions.

For the causality determinations described in Chapters 5-11, the CASAC finds 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 nervous system effects; between long-term UFP exposure and nervous system effects; or between long-term $PM_{2.5}$ exposure and cancer. The CASAC members had varying opinions on whether there is robust and convincing evidence to support the EPA's conclusion that there is a causal relationship between $PM_{2.5}$ exposure and mortality.

The CASAC finds that Chapter 13 provides evidence supporting a causal relationship between PM and visibility impairment, climate effects, and effects on materials, but recommends that the Draft ISA include more analyses for different size fractions and add discussion of the direct effects of PM or other pollutants (e.g., photochemical oxidants) on visual acuity.

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

Sincerely,

/S/

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, Rochester, NY

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

Dr. Timothy E. Lewis, U.S. Army Corps of Engineers (Retired), Vicksburg, MS

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

Consensus Responses to Charge Questions on the EPA's
Integrated Science Assessment for Particulate Matter (External Review Draft – October 2018)

Overall Comments and Recommendations on the Integrated Science Assessment (ISA)

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. For example, the chartered CASAC has found it difficult to achieve consensus in some areas (summarized below), and to do so likely requires further scientific expertise from, and discussion with, epidemiologists and additional experts in human clinical studies and toxicology. Some of the proposed changes in causality determinations in the Draft ISA, for example changing the causality designation of long-term exposure to ultrafine particles (UFP) on nervous system outcomes from "inadequate" to "likely," are driven primarily by animal toxicology studies. Therefore, additional expertise is needed in comparative toxicology, dosimetry, and extrapolation of findings in animals to humans.

Over the past 30 years, the CASAC's advice to the EPA on NAAQS reviews has been assisted by expert review panels that supplement and expand the scientific expertise brought to bear. Such a review panel was appointed by the EPA for the current PM review. However, the panel was disbanded by the EPA prior to the release of the Draft ISA.

The CASAC recommends that the EPA reappoint the previous CASAC PM panel or appoint a panel with similar expertise, as well as adding expertise in biological mechanisms of causation, causal inference, multi-stressor interactions, and potentially others such as: epidemiology, human clinical studies; comparative toxicology, dosimetry, and extrapolation of findings in animals to humans; characterization of sampling errors and biases from continuous ambient PM measurements and satellite remote sensing aerosol optical depth (AOD) analysis; errors and biases in dispersion modeling and photochemical grid modeling; errors-in-variables methods and effects of exposure (and covariate) estimation errors on epidemiologic study results; epidemiology of low-dose causal concentration-response functions; and effects of PM on visibility impairment, climate, and materials. The panel should be appointed in time to review the Second Draft ISA.

<u>Causality Determination of Mortality from $PM_{2.5}$ Exposure</u>

The CASAC did not reach consensus on the causality determination of mortality from $PM_{2.5}$ exposure.

Some members of the CASAC think that the EPA must better justify their determination that short-term or long-term exposure to $PM_{2.5}$ causes mortality. The EPA should address the following considerations:

- <u>Biological action of PM.</u> How do low concentrations of $PM_{2.5}$ cause mortality? The EPA should discuss not just general, possible mechanisms, but specifically how ambient concentrations of $PM_{2.5}$ can move into and through the biological systems in the body to activate a cascade of effects that ultimately lead to a person's death.

- Heterogeneity. The EPA should also address the substantial unexplained geographic heterogeneity in effect estimates between $PM_{2.5}$ exposure and mortality (e.g. Eftim et al., 2008, Baxter et al., 2017, and many others). In the previous PM NAAQS review, the EPA noted that uncertainty remained in the form of unexplained within- and between-city heterogeneity in responses to PM. The EPA also asked several policy-relevant questions related to geographical heterogeneity in the Integrated Review Plan for this current PM NAAQS review. Given the emphasis that the EPA has placed on this topic, they should include more discussion of geographic and other types of heterogeneity in this ISA. The implications of unexplained heterogeneity need to be discussed for those endpoints where many potential explanations have been tested, but none has been able to explain the observed heterogeneity (e.g. short-term $PM_{2.5}$ exposure and total mortality). At what point does heterogeneity move from being an uncertainty, to impacting the causality conclusion or other policy-relevant issues such as the use of a single effect estimate for the whole nation?

- Concentration Concordance. When discussing the continuum of effects from $PM_{2.5}$ exposure, the EPA should include a discussion of how this continuum is impacted by the concentrations at which different effects have been observed. For example, when the EPA states that mortality evidence provides coherence for a continuum of effects, this should be considered within the context of whether more serious effects occur at higher, lower, or similar concentrations as more mild effects. This comparison of concentrations of effect should be extended to comparisons between epidemiology, animal, and human controlled exposure studies.

- Concentration-response (C-R) functions and thresholds. Various epidemiology studies have concluded that the C-R function describing the association between $PM_{2.5}$ and mortality is linear with no threshold. Statistical analysis shows that epidemiology studies cannot determine the true C-R function shape (discussed elsewhere in this document), and the use of linear-no-threshold C-R functions is also inconsistent with human and animal experimental data demonstrating a threshold of $PM_{2.5}$ exposure concentrations below which no health effects are seen. The likely modes of action of $PM_{2.5}$ effects on the body also support a threshold (discussed elsewhere in this document). Therefore, epidemiology studies that draw the conclusion that there is an effect of $PM_{2.5}$ on mortality at concentrations down to zero are not consistent with animal and human data or with our knowledge of adverse effect pathways.

  The EPA's mortality causality determination appears to be based almost exclusively on epidemiology studies, which cannot be used in isolation to determine causation. Further integration amongst epidemiology studies showing logical patterns in magnitude and types of health effects, as well as demonstrations of substantial health effects in animals exposed to high concentrations could provide some of the necessary justification for this causality conclusion (discussed below).

- Comparing results between and within studies. The EPA could improve the integration of evidence in this ISA by hypothesis-testing its conclusions by comparing $PM_{2.5}$ effect estimates within and between studies. For example, if one expects that some subset of mortality is more affected by $PM_{2.5}$ (e.g. cardiovascular mortality), then that mortality should have a larger and more significant association with $PM_{2.5}$ than total mortality. Similarly, if all these effects are occurring at the same concentrations, then one would expect more mild effects (e.g. symptom exacerbation) to be more common and more likely to show an association than the more serious

effects (e.g. hospital admission or mortality). One would also expect that long-term effects would occur at lower concentrations and would show stronger effects than short-term, because of cumulative exposure (if $PM_{2.5}$ has an impact via cumulative exposure); and that health risks associated with $PM_{2.5}$ would be higher in places with higher $PM_{2.5}$ concentrations. Investigating these types of patterns could be done with the study information that the EPA has already collected for this ISA and would greatly strengthen the conclusions that are drawn.

- Mortality in animals. If the EPA has identified any short-term or long-term exposure studies in animals where PM exposure increased mortality, that would be a useful addition to the discussions in Chapter 11. If none has been identified, that would also be useful information, if put into the appropriate context of aging and differential susceptibility of rodents.

Other members of the CASAC are of the opinion that, although uncertainties remain, the evidence supporting the causal relationship between $PM_{2.5}$ exposure and mortality is robust, diverse, and convincing. The epidemiological observations have been reproduced around the world in communities with widely varying exposures. The findings of many of the largest studies have been repeatedly reanalyzed, with confirmation of the original findings. The EPA's causality determination, rather than considering the epidemiological evidence "in isolation," includes a wide range of evidence from a variety of sources, including human clinical exposure and animal toxicology studies that have provided rational biological plausibility and potential mechanisms. This causality determination was first clearly promulgated in the 2009 ISA, with full CASAC support. It is widely accepted by the scientific community and many public health organizations, including the World Health Organization. There is no credible or convincing new evidence since 2009 to question or refute this determination. Indeed, there is new evidence from epidemiological studies supporting the relationship between $PM_{2.5}$ and mortality, and new toxicology studies informing the mechanisms involved and supporting their plausibility. The evidence supporting a causal relationship between $PM_{2.5}$ and mortality is even more robust now than it was in 2009.

Uncertainties clearly remain: for example, the specific PM characteristics responsible for health effects, dose-response relationships at low ambient concentrations (the threshold issue), explanations for the observed heterogeneity in effect sizes across geographical locations, and whether (or to what degree) particle translocation away from the lung mediates health effects. These uncertainties have been for the most part thoroughly discussed in the draft ISA, as well as in previous PM ISAs. The fact that there is uncertainty with regard to specific issues does not negate the overwhelming evidence that $PM_{2.5}$ exposure increases mortality.

Executive Summary

The Executive Summary is intended to provide a concise synopsis of the key findings and conclusions of the PM ISA for a broad range of audiences. Please comment on the ***clarity*** *with which the Executive Summary communicates the key information from the PM ISA. Please provide* ***recommendations on information that should be added*** *or information that should be left for discussion in the subsequent chapters of the PM ISA.* (Emphases added.)

The CASAC finds that the Executive Summary provides a concise and accessible summary of many of the key findings and conclusions of the Draft ISA for a broad range of audiences. However, it does not

accurately represent the totality of available high-quality scientific evidence. The CASAC recommends that the EPA indicate when exposures referred to are estimates. Some members of the CASAC recommend that the EPA consider distinguishing between the following in statements of key findings and conclusions: effects of PM and effects of confounders (such as poverty and temperature); observed changes and model-predicted changes in public health risks following changes in exposures; assumptions and data on shapes of C-R functions; results from the total body of scientific evidence and results from selected subsets of evidence; and association and causation.

Integrated Synthesis (Chapter 1)

Chapter 1 presents an integrated summary and the overall conclusions from the subsequent detailed chapters of the PM ISA and characterizes available scientific information on policy-relevant issues. Please comment on the ***usefulness and effectiveness of the summary*** *presentation. Please provide* ***recommendations on approaches that may improve the communication*** *of key findings to varied audiences* ***and the synthesis*** *of available information across subject areas.* ***What information should be added*** *or is more appropriate to leave for discussion in the subsequent detailed chapters?* (Emphases added.)

The CASAC finds that Chapter 1, similar to the Executive Summary, provides an effective summary of material from subsequent chapters. In presenting key conclusions from the other chapters, Chapter 1 should characterize uncertainty about them and their sensitivity to assumptions. Chapter 1 should also discuss inconsistencies and discordant data, as detailed in the consensus response to the Chapter 5 charge questions and to the individual comments.

Sources, Atmospheric Chemistry, and Ambient Concentrations (Chapter 2)

To what extent is the information presented in Chapter 2 regarding sources, chemistry, and measurement and modeling of ambient concentrations of PM clearly and accurately conveyed and appropriately characterized?

Overall, Chapter 2 does a sufficient job of conveying and characterizing the sources, chemistry, and measurements and modeling of UFP, $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$ (coarse fraction).

Section 2.3 discusses primary sources of PM. Figures 2-2, 2-3, and 2-6 show the importance of various types of dust to total $PM_{2.5}$ and PM_{10} primary emissions based on the U.S. EPA 2014 National Emissions Inventory. However, when these emissions are used as inputs to chemical transport models (CTMs), the modeled concentrations are significantly higher than the observed concentrations at the speciation monitors. The reason for the overprediction is that there is no adjustment for near-source removal due to small sub-grid scale turbulence and impaction on building and vegetative surfaces (Pouliot et al., 2012). It is estimated that local source removal typically accounts for 75% of total removal of fine particulate matter nationally (Pace, 2005). This removal factor is defined as a "capture fraction" and varies by location. The amount that is not removed is defined as the "transportable fraction." A discussion of capture fraction and transportable fraction should be included in this chapter to help place the importance of dust emissions into proper perspective.

Although the predominant sources of coarse PM primary emissions are thoroughly explained, there should also be discussion on the potential for the formation and/or sources of formation of secondary coarse PM. This chapter does a good job of identifying the sources of UFP; however, there should also be discussion on the transport or the potential for transport of UFP and on the possible existence (or lack thereof) of biogenic, natural background concentrations of UFP.

With respect to monitoring of $PM_{2.5}$ and PM_{10}, Section 2.4.1 does a sufficient job of discussing the difference between Federal Reference Method (FRM) and Federal Equivalency Method (FEM) monitors and describes the three most widely used FEMs. FRMs typically measure 24-hour integrated samples every third or sixth day. Short time resolution automated FEMs can measure hourly samples every day. In the past, FEMs typically measured higher $PM_{2.5}$ concentrations than FRMs; therefore, some states were reluctant to switch to FEMs. However, the new Teledyne optical spectrometer FEMs are much more reliable and consistent and many states are now converting their FRMs to FEMs. In July 2017, the Georgia Environmental Protection Division (EPD) ran two regulatory FEMs. As of March 2019, the Georgia EPD runs nine regulatory FEMs and will be running twelve regulatory FEMs by June 2019. A similar trend is occurring across many parts of the country which will produce significantly more $PM_{2.5}$ data at hourly resolution. In addition, the measurement technology for $PM_{10-2.5}$ is considerably improved compared to previous methods (i.e., subtraction methods). As a result of the improved accuracy and reliability in measurements of $PM_{2.5}$, PM_{10}, and $PM_{10-2.5}$ using FEM monitors, exposure and health effects assessments should consider the availability of these continuous and robust datasets.

With respect to the utilization of satellite remote sensing for obtaining ambient concentrations of PM, caution should be given to any results obtained from such techniques because the computational algorithms use a range of assumptions to obtain estimates of $PM_{2.5}$ concentrations. These inferred measurements involve potential errors that are not encountered with the FRM or other ground-based $PM_{2.5}$ measurements. In addition, data cannot be collected when clouds and snow are present or from excessive amounts of smoke being mistaken for clouds. Conclusively, the many factors that impact the relationship between AOD and $PM_{2.5}$ concentrations sometimes lead to widely varying and relatively low correlations when linear relationships are developed.

Discussions on the limitations and/or uncertainties of utilizing CTM to estimate ambient concentrations of PM should be added to Chapter 2, although Section 2.4.7 does a good job of documenting the relative scientific advances in CTMs.

Please comment on the extent to which available information on the spatial and temporal trends of ambient PM concentrations at various scales has been adequately and accurately described.

Overall, Chapter 2 does a sufficient job of adequately and accurately describing the extent to which information is available on the spatial and temporal trends of ambient PM concentrations at various scales.

There are a few noted discrepancies in the figures. Figure 2-14 shows the 98^{th} percentile 24-hour $PM_{2.5}$ concentrations for 2013-2015. The red monitor (indicating concentrations of 35-40 $\mu g/m^3$) in southern Georgia appears to be Albany (13-095-0007). However, according to certified U.S. EPA Air Quality System (AQS) data, the 24-hour 2013-2015 design value for Albany is 23 $\mu g/m^3$ (should be a blue dot, not a red dot). Figure 2-15 shows the 98^{th} percentile PM_{10} concentrations for 2013-2015. There are no measurements shown in Georgia, although Georgia has three PM_{10} monitors (13-089-0002, 13-121-

0039, and 13-245-0091) with certified data in AQS from 2013-2015. The 98th percentile PM_{10} concentrations for all three PM_{10} monitors in Georgia are well below 75 $\mu g/m^3$ (designated by blue dots).

Regarding spatial trends, there does not appear to be an adequate discussion about the regional (state-to-state) transport of PM. In addition, Figures 2-13 through 2-16 show 2013-2015 PM design values. These figures should be updated with 2015-2017 design values.

The CASAC should be given ready access to one or more experts in ambient PM measurements and satellite remote sensing AOD analysis to assist in review of the next iteration of the ISA. This would allow for a better understanding of sampling errors and biases associated with integrated and continuous ambient PM measurements and satellite data. This is particularly important since this information will be used to characterize ambient concentrations in the Risk and Exposure Assessment (REA) document.

Exposure to Ambient Particulate Matter (Chapter 3)

Chapter 3 describes scientific information on exposure to ambient PM and implications for epidemiologic studies. To what extent is the discussion on methodological considerations for exposure measurement and modeling clearly and accurately conveyed and appropriately characterized?

Overall, this chapter does a good job describing methods for exposure measurements and modeling. New developments in PM exposure assessment methods have reduced bias and uncertainty in health effect estimates by improving the spatial resolution and accuracy of exposure predictions. High correlations of $PM_{2.5}$ with some gaseous copollutants necessitate evaluating the impacts of confounding on health effect estimates. There is typically more uncertainty for health effects estimates caused by exposure to $PM_{10–2.5}$ and UFP than for health effects associated with $PM_{2.5}$.

Section 3.3.1.2 and Table 3-1 discuss personal monitoring and measurement error characteristics. Additional discussion of the personal-exposure measurement literature is warranted, and the ISA should include information from two key systematic reviews published in 2010 (Avery et al., 2010a, 2010b). These reviews describe the variability in personal-ambient relationships. They state that "The wide range in estimated correlations between personal and ambient $PM_{2.5}$, as well as the associations with participant, study and environment characteristics, suggest that the potential for exposure misclassification can be substantial." This should be further discussed in the ISA and used to better inform interpretations of studies that assume a simple, or even perfect, relationship between ambient and personal $PM_{2.5}$ concentrations. There are a wide variety of personal samplers and some perform better than others. Therefore, a detailed evaluation of the sampler performance compared to FEM monitors or FRM monitors should be performed before using personal sampling data as the definitive estimate of exposure. In some cases, the data may be better suited for examining gradients in $PM_{2.5}$ exposures rather than directly using the measured $PM_{2.5}$ concentrations.

Table 3-2 compares models for estimating "exposure concentrations or exposure." The EPA should clarify the difference between "exposure concentration" and "exposure." If they are used interchangeably, only one term should be used rather than "exposure concentrations or exposure." Under the column for "Dispersion" there is an "X" for Chemistry. However, it is stated on page 3-28 "Dispersion models…typically have limited ability to model chemistry (if any)." A footnote should be

added to the table to indicate that many dispersion models do *not* account for chemistry (alternatively, the "X" could be removed from the "Chemistry" row in the "Dispersion" column). Table 3-3 presents statistical measures used for air quality model performance evaluations. While the four performance measures listed are commonly used, the table should also include normalized mean bias and normalized mean error. These are also commonly used and will provide information about percent differences in addition to absolute differences (mean bias and mean estimate).

Section 3.3.2.4 addresses "Mechanistic Models." The summary paragraph at the beginning of the section discusses CTMs, but does not mention dispersion models, which are discussed later in this section (3.3.2.4.2). All the relevant models should be included in the summary section. In Section 3.3.2.4.1, information should be added to describe how Eulerian CTMs work (e.g., grid structures, finite difference, solving advection diffusion equations, etc.).

Page 3-27 states that "Differential bias may also be observed across regions in space. Many such biases can be corrected for using adjustment factors based on comparisons of simulation results with observational data." However, it should be noted that adjusting modeling results to match observations can lead to the right answer for the wrong reasons. This is important if emission sensitivities or source apportionment is being used to look at alternative levels of the standard, since the model may not correctly predict the response to emission controls.

The bottom of page 3-27 discusses the Lagrangian trajectory model (which lacks chemistry) by Stanier et al. (2014). Typically, Lagrangian models are not classified as CTMs, but as dispersion models. This discussion should be moved to Section 3.3.2.4.2. Section 3.3.2.4.2 emphasizes dispersion models for near-road modeling of mobile sources. However, dispersion models are much more widely used for permitting industrial point sources. The ISA should add further discussion of dispersion modeling for point source emissions. Section 3.3.2.4.2 should be updated to include Lagrangian dispersion models. A discussion on the difference between a Lagrangian puff model and a steady-state Gaussian plume model should also be added. Under the section on Lagrangian puff models, a discussion on CALPUFF (limited chemistry), SCIPUFF (no chemistry), and SCICHEM (full chemistry) should be added.

Table 3-4 and other parts of Section 3.3.2.4.2 discuss model performance of dispersion models. Most published dispersion model performance evaluations are associated with using the model for compliance assessments. In these cases, the model's ability to capture the high end of the concentration distribution is evaluated with Q-Q plots, where the highest data point from the model is compared to the highest data point from the observations even if they occur at different locations, time of day, and/or season of the year. When the model is being used to support health studies, spatial and temporal accuracy are much more important than they are in compliance assessments. Therefore, dispersion modeling results need to be evaluated against observations paired in time and space, especially if they are being used as inputs to exposure models such as SHEDS, APEX, or EMI. The ISA should add a discussion of the performance of dispersion models in modeling exposure distributions for health protection, rather than compliance assessments.

Pages 3-32 to 3-34 discuss fusion of CTM predictions with surface observation data. This section does not discuss EPA's recommended approach to States for estimating ozone and $PM_{2.5}$ concentrations at unmonitored locations (U.S. EPA, 2014 and U.S. EPA, 2018a). These guidance documents discuss EPA's Modeled Attainment Test Software (MATS) and the Software for Modeled Attainment Test-Community Edition (SMAT-CE) tool. MATS and SMAT-CE will spatially interpolate data using the

Voronoi Neighbor Averaging (VNA) technique and adjust the spatial fields based on model output gradients. These tools can be applied to annual $PM_{2.5}$ design values, daily $PM_{2.5}$ design values, or 24-hour $PM_{2.5}$ values.

The last paragraph on page 3-34 states "Hybrid approaches can involve merging CTMs with dispersion and/or LUR models, merging CTMs with observational data, or some combination therein." However, there are no references showing how CTMs can be merged with dispersion models. Below are two examples that could be referenced:

- EPA's 2014 National Air Toxics Assessment (NATA) Technical Support Document (U.S. EPA, 2018b) which merged CMAQ and AERMOD modeling results to determine cancer risks for HAPs and diesel PM.
- Wesson et al., 2010: This study merged CMAQ and AERMOD modeling results to determine exposure to HAPs, ozone, and $PM_{2.5}$. Data from the air quality modeling was used as input into the environmental Benefits Mapping and Analysis Program (BenMAP) and the Human Exposure Model-3 (HEM-3) to assess how the control strategies could affect human health.

It would be helpful for the CASAC to have ready access to an expert in dispersion modeling and photochemical grid modeling used in health effects analyses. This would allow for a better understanding of errors and biases associated with models that are used to characterize ambient concentrations in the REA document.

Please comment on the extent to which the discussion regarding exposure assessment and the influence of exposure error on effect estimates in epidemiologic studies of the health effects of PM has been adequately and accurately described.

In general, the chapter provides useful discussions of exposure assessments and the influence of errors on effect estimates in epidemiologic studies. However, the final ISA should either modify, or provide citations or explanations to support, the statement that exposure error tends to produce underestimations of health effects in epidemiologic studies of PM exposure. In general, this is not true. Estimation errors typically lead to overestimates of low-dose risks and underestimates of high-dose risks if the true manipulative causal C-R function has a threshold or threshold-like nonlinearity. Many studies have shown that bias or error in the exposure or outcome assignment can cause the estimated C-R function to flatten and appear linear even if the true C-R function has a well-defined threshold or other non-linear shape (Brauer et al., 2002; Cox, 2018; Lipfert and Wyzga, 1996; Rhomberg et al., 2011; Watt et al., 1995; Yoshimura, 1990). On page 3-76, the EPA states that "If this occurs, the health effect related to PM exposure would be underestimated or potentially not detected. Positive correlation between PM and the copollutant and between the exposure measurement errors of PM and the copollutant can add more negative bias to the PM health effect estimate. Spatial variability of concentration differs among the particle size spectrum, and this may cause more exposure measurement error in $PM_{10–2.5}$ or UFP compared with $PM_{2.5}$ (Section 3.4.2.2). Hence, if $PM_{2.5}$ is measured with less error than copollutants, it is likely that the effect will be attributed to $PM_{2.5}$." This means that in copollutant models, whichever pollutant is measured with the least error is most likely to be ascribed the positive effect. This phenomenon has been demonstrated by several groups (Carrothers and Evans, 2000; Fewell et al., 2007; Lipfert and Wyzga, 1996) and makes interpreting copollutant models quite challenging. It requires considerations of joint exposure measurement errors for each component. Studies that have investigated

the effects of better exposure estimates on health effect estimates (e.g., Ebelt et al., 2005; Hart et al., 2015; McGuinn et al., 2017; Trenga et al., 2006) have demonstrated that there is no or little difference in health effects estimates or width of confidence intervals with different (presumably better) exposure estimates.

It would be helpful for the CASAC to have ready access to an expert in errors-in-variables methods and effects of exposure (and covariate) estimation errors in epidemiology to allow for a better understanding of the impact of exposure errors on epidemiologic study results.

Dosimetry of Particulate Matter (Chapter 4)

Chapter 4 characterizes scientific evidence on the dosimetry of PM. To what extent does the discussion clearly and accurately convey the dosimetry of inhaled PM and the processes of deposition, clearance, retention, and translocation?

Organization

This chapter provides a very important and thorough review of the deposition, clearance, retention, and translocation of inhaled PM. However, the text would benefit from careful copy editing. In addition, the EPA could streamline the chapter by removing some extraneous information, such as discussions of the history of scientific views on post-natal alveolar development, and ventilation distribution in dogs and horses.

Additional Information

Some additional information would improve the translation of these dosimetric study results to human-relevant exposures. The EPA should include the concentrations at which the dosimetry exposures were conducted, along with a discussion of the impact of the concentration on the measured dosimetric observations. Concentration information is particularly important for the translocation studies that tend to use very high exposure concentrations. The CASAC also recommends that the EPA add a discussion about how the dosimetry study results can be translated to humans exposed to ambient PM concentrations. For example, in the section on interspecies clearance and retention, the EPA could add the PM dose or dose-rate at which particle overload occurs in rats to provide a reference dose at which extrapolation to humans would become inaccurate. In the section about translocation of soluble versus insoluble components, the EPA could add information about how big a contribution is made by soluble particles to total particles. This information is important because these particles could have a more direct or obvious linkage to systemic effects than insoluble particle translocation (which occurs at a very low frequency). The EPA should also provide greater consideration of the impact of exercise on dose-rate, and how this may affect study interpretation in subsequent health effect chapters. In addition, particle deposition density per cm^2 of surface area in various anatomical regions is an important factor, particularly in evaluating studies of pulmonary defense mechanisms. A table of particle deposition densities should be added.

Another important addition to this chapter would be a reference to the study by Kendall et al. (2002) in which $PM_{2.5}$ samples were immersed in normal lung lining liquid (surfactant). The small (~35 nm) particles aggregated into larger (>5 um) structures when immersed in lung lining fluid, compared with

samples in air or saline, specifically due to interaction with the protein-rich surfactant solution. The probability - and physical possibility - of a particle breaching alveolar epithelial cell membranes is inversely related to the size of the particle. The possibility that small particles aggregate into larger particles upon contact with lung lining liquid could impact the number of particles available for translocation into the circulation. This may impact hypotheses about extra-pulmonary effects of small particles, although how much this aggregation occurs at ambient concentrations in humans is unknown.

Accuracy

The EPA largely conveys the information about dosimetry of inhaled PM accurately. However, in the section on transplacental barrier transport, it is not accurate to portray the available data on fetal translocation of particles as providing "biological plausibility of effects during pregnancy," for two reasons: 1) The only information available for fetal translocation is from oral or intravenous animal studies that are not relevant to inhalation exposures (as evidenced by data showing that the extra-pulmonary distribution of particles from inhalation is different compared to intravenous or oral administration); and 2) The administered doses generate systemic particle concentrations that are orders of magnitude higher than would be attained via inhalation. Although these studies provide some information about the plausibility of translocation to the fetus at high concentrations, there is no evidence provided in this chapter to support fetal translocation of particles at ambient concentrations.

Some parts of the dosimetry chapter require a more accurate discussion of the uncertainty of the available data. In particular, the translocation of UFPs into the human brain is quite uncertain: the human autopsy studies by Calderón-Garcidueñas et al. (2010, 2013) do not provide definitive evidence that ambient UFPs translocate to the brain (they lack proper controls and don't determine the source of the UFPs found in brain tissue). Even if there is translocation of UFPs to the brain, it is likely to be a very tiny fraction of particles, as estimated by Garcia et al. (2015), with only 0.001% of 20 nm particles being deposited on the human olfactory mucosa (and presumably far fewer particles actually translocating from the mucosa to the olfactory bulb). Therefore, although these studies show that translocation to the brain may occur at high doses of UFPs, the EPA should note the uncertainty about this translocation, and how much it occurs at ambient concentrations in humans.

The EPA should reflect these changes in the summary statements at the end of this chapter. In addition, the EPA should review the summary statements to ensure that they accurately reflect the information presented in this chapter. For example, "The fraction of nanoparticles translocating from the peripheral lung into circulation is generally low (less than a fraction of a percent) for larger nanoparticles (18−80 nm) but can approach several percent for extremely small particles (1.4−2.8 nm)." This statement should be caveated with the information that although several percent of extremely small particles may translocate into the peripheral circulation in rodent studies with exposure by lung instillation, there is no evidence that this much translocation occurs with exposure to even very small particles (4-5 nm) via inhalation in humans. Similarly, when discussing results showing fetal translocation of particles, the EPA should state that the evidence is based on high-dose oral or intravenous particle administration of UFPs with an unknown relationship to human inhalation of ambient concentrations of PM. Although studies have shown that translocation can occur under some exposure scenarios at high concentrations, information is lacking about translocation of particles outside of the respiratory tract at relevant ambient concentrations in humans.

Health Effects Associated with Short-term and Long-term Exposure to PM (Chapters 5-11)

Please comment on the identification, evaluation and characterization of the available scientific evidence from epidemiologic, controlled human exposure, toxicological and associated human exposure and atmospheric sciences studies and the application of information from these studies to inform causality determinations and uncertainty characterizations for human health outcomes.

Chapters 5 – 11 present assessments of the health effects associated with short-term and long-term exposure to PM. The discussion is organized by PM size fraction, exposure duration, broad health effects (e.g., asthma, ischemic heart disease, etc.), and scientific discipline. Please comment on the characterization of the evidence within these chapters.

Please comment on the portrayal and discussion of the biological plausibility evidence presented at the outset of Chapters 5 – 11 and the extent to which: (1) the organization adequately captures the current state of the science with respect to potential pathways by which PM could impart health effects, and (2) as currently constructed, inform causality determinations.

Organization and presentation:

These chapters would be improved by reducing redundancy. For example, the biological plausibility sections repeat many discussions verbatim. A judicious edit of all the sections would help to streamline the document.

The document would benefit from a careful editorial review. There are missing words, incomplete sentences, and grammatical errors that in some places confuse the meaning.

The quality of some of the figures is poor. For example, much of the text in Fig. 5-4, page 5-25 is blurry and difficult to read.

The section references within the text need to be checked and edited, to ensure that the correct section is being referenced. In addition, links to "Section 0" need to be fixed.

Section 6.2.6 is "Cardiac Electrophysiology and Arrhythmia", and section 6.2.11 is "Heart Rate (HR) and Heart Rate Variability (HRV)." These sections should be combined, retaining the electrophysiology and arrhythmia heading. Cardiac electrophysiology encompasses cardiac conduction abnormalities, repolarization, HRV, and arrhythmia. All are measured using electrocardiography (ECG). Having these widely separated sections is confusing. An alternative would be to rename each of these sections and present them sequentially. Similar comments apply to Sections 6.1.4 and 6.1.10.

Chapter 7, Metabolic Effects - A better distinction needs to be made between the potential metabolic effects of PM, and metabolic abnormalities as markers of susceptibility to cardiovascular (CV) effects of PM. These two issues are inappropriately thrown together in this chapter. Metabolic effects include increased insulin resistance, blood glucose, hemoglobin A1c, and incident diabetes. Alternatively, having diabetes, obesity, or metabolic syndrome could render increased susceptibility to the CV effects of PM exposure. These are separate, important questions. However, the latter should not be described as or included with "metabolic effects," but considered with other susceptibility factors, in Chapter 12.

Additional suggestions for editorial changes and corrections can be found in comments of individual CASAC members.

Biological plausibility:

Possible pulmonary vascular effects of PM, and cardiopulmonary interactions - In general, the background sections of Chapters 5 and 6 ignore the importance of inter-relationships between respiratory and cardiac function. The mechanistic figures showing potential pathways for PM pulmonary and CV effects should be modified to reflect these considerations. Acute PM-related effects on left ventricular (LV) ischemia or function, or effects on pulmonary artery pressure, could present as respiratory effects, with dyspnea. This is especially true for COPD; many COPD patients have co-existing cardiac disease and/or pulmonary arterial hypertension, and acute exacerbations often have a major cardiac contribution.

Pulmonary vascular and right ventricular (RV) effects are likely pathways, in addition to inflammation and translocation, for both acute and long-term PM effects. Pulmonary hypertension and RV failure are briefly discussed in Section 6.2.5, under long-term effects, and Ohlwein et al. (2016) shows evidence for $PM_{2.5}$ effects on cardiac diastolic function, which is discussed on page 6-167, line 10. But these effects are intertwined, because RV dysfunction can worsen LV diastolic dysfunction by encroachment on the LV, with impaired filling of the LV and consequent respiratory effects. This is a pathway leading to clinical findings of acute heart failure, but with preservation of LV systolic function. This is a very common occurrence in COPD patients, and a major contributor to exacerbations.

There is epidemiological, clinical, and toxicological evidence to support a pathway of pulmonary vascular effects for PM (Aaron et al., 2016; Grunig et al., 2014; Leary et al., 2014; Liu et al., 2018; Park et al., 2014; Rich et al., 2008; Wauters et al., 2015). Aaron et al. (2016) showed that long-term $PM_{2.5}$ exposures were associated with increased RV mass and RV mass/LV end-diastolic volume; Grunig et al. (2014) provided a perspective on this issue; Leary et al. (2014) identified NO_2 as a marker of traffic-related air pollution linked with increased RV mass (see also the accompanying editorial); Rich et al. (2008) reported a panel study of patients with heart failure showing an acute increase in PA pressure in association with $PM_{2.5}$ exposure). Others include Liu et al. (2018), Park et al. (2014), and Wauters et al. (2015). Only one of these studies (Aaron et al., 2016) was cited in the ISA, and that was in the context of heart failure in general.

Page 5-6, line 5 - "Activation of sensory nerves in the respiratory tract can trigger local reflex responses resulting in lung irritation." The term "lung irritation" lacks specificity and may have different meanings for different people. The more accurate term is "airway irritant response" which refers to this sensory-mediated process, not just its result. The CASAC suggests replacing "lung irritation" in this sentence with "lung function decrements and airway inflammation." Elsewhere, "lung irritation" could be replaced with "airway irritant response."

Figure 6-1 - The potentially important role for endogenous nitric oxide (NO) and endothelins in PM effects on vascular function are not adequately covered in the figures or the paragraphs on biological plausibility. There is evidence that PM may act through both, with reduction in NO bioavailability (for example, reduced NO-mediated arterial vasodilatation in response to diesel exhaust particles, Mills et al., 2011) and increased production of endothelins (Kumarathasan et al., 2015; Lund et al., 2009). There

is also the possibility that translocated particles or their components may directly injure the vascular endothelium.

Figure 7-2 should be revised. It is unclear what is meant by "peripheral inflammation" in Figure 7-2 and the accompanying text, and the CASAC recommends not using this term. In section 7.1.3, "peripheral inflammation" seems to be referring to increased inflammation in adipose tissue in various organs, which could have important implications for obesity and metabolic responses. This should be stated more clearly. In addition, this evidence would support a pathway that differs from the current pathways in the biological plausibility figures, suggesting that PM exposure may lead to focal or organ-specific inflammation/oxidative stress that could also possibly be mediated by translocated PM or their components.

The introductory and biological plausibility sections of Chapter 7 do not adequately address the distinctions and differences between type 1 and type 2 diabetes. This is addressed regarding insulin therapy in Figure 7-1, but most of the rest of the text simply refers to "diabetes." It should be made clear that almost all diabetes that is incident in older adults is type 2, and that the glucose intolerance in metabolic syndrome is related to the development of type 2 diabetes, rather than type 1.

The first paragraph on page 7-18 describes a potential pathway for metabolic effects involving the hypothalamus of the brain, and this important pathway is represented by a blue box (an intermediate event) in Figure 7-3. However, there is no initial event that links the exposure to this intermediate effect. These should be linked with a green box (initial event) of UFP translocation to the brain via the nasal mucosa and olfactory nerves.

Figure 9-1 - The CASAC recommends that erectile dysfunction be removed from the biological plausibility figure because only a single paper is cited in this ISA about an association between PM and erectile dysfunction, and it was not statistically significant (OR=1.26, ICI 0.81-1.96).

Page 9-4 - "Inhalation of $PM_{2.5}$ can result in translocation of particles or soluble factors from the lungs (see Chapter 5) which then can increase respiratory tract inflammation…" The sequence is likely wrong here. Particles in contact with airway epithelium initiate airway inflammation, in part via chemokine production by the epithelial cells. That takes a few hours to develop, while transport of particles likely starts before airway inflammation is fully developed. If translocation occurs, it would occur rapidly as the particles enter the pulmonary capillary bed and are quickly transported to the left heart and then the systemic circulation. This sentence seems to make the assumption that translocation causes pulmonary inflammation, which does not accurately represent the pathophysiology.

Figure 10-2 does not accurately reflect the likely pathways for lung cancer. The current emphasis in the figure is on transport of particles and systemic or brain effects. However, the most relevant pathway is direct effects of PM or its components on the airway epithelium. Although airway inflammation may be involved, direct mutagenic, genotoxic, and epigenetic effects on the airway epithelium are likely more important. Systemic inflammation and particle translocation away from the lung are not relevant for lung cancer.

Page 10-3 - The discussion of the 10 key characteristics of carcinogens should note that these characteristics may be necessary but are not sufficient for cancer formation (i.e., all carcinogens may have these characteristics, but substances with these characteristics are not all carcinogens).

Identification of the available scientific evidence:

Chapter 5, Short-Term Exposure to $PM_{2.5}$, Section 5.1.2.3, Lung Function Changes in Asthmatics - The first sentence states: "Studies evaluating the effects of short-term $PM_{2.5}$ exposure on lung function consisted solely of epidemiologic studies." However, Gong et al. (2004b) and Urch et al. (2010), both human controlled exposure studies, investigated effects of fine concentrated ambient particle exposures on lung function in asthmatics.

Chapter 6, section 6.1.4.3, Controlled Human Exposure Studies for Arrhythmia and Conduction Abnormalities - Langrish et al. (2014) should be included in the discussion of potential effects of PM on cardiac arrhythmias.

Study inclusion information - Although the EPA provides a broad overview of the study search strategy and the general inclusion choices in the ISA Preamble and the Preface, they should provide more specific inclusion information for studies in those chapters and sections where specific characteristics are being used to exclude studies. For example, on pg. 5-8 (similarly on pp. 5-52, 5-66, 10-37, and 10-59 among others), the EPA states that "Other recent studies of asthma hospital admissions and emergency department (ED) visits are not the focus of this evaluation because they did not address uncertainties and limitations in the evidence previously identified, and therefore, do not directly inform the discussion of policy-relevant considerations detailed in Section 5.1.10." Because it is not clear how those studies listed in Table 5-10 do address previously identified uncertainties and limitations, the EPA should provide an explicit list of those criteria used in these sections to include or exclude studies, and/or should provide a list of those studies that were not the focus of this evaluation (perhaps as a sub-category in the HERO database).

Causality studies - Several studies that conduct causal-type analyses should be included in this ISA: Cox et al. (2017) (for short-term effects of $PM_{2.5}$ on cardiovascular disease), and Greven et al. (2011), Cox and Popken (2015), and Pun et al. (2017) (for long-term effects of $PM_{2.5}$ on mortality). Greven et al. (2011) and Pun et al. (2017) use a method called "a difference-in-differences analysis" that has been used by others to conduct causal-type analyses.

Evaluation and characterization and of the available scientific evidence:

Dosimetric extrapolation for particle translocation - When addressing the possibility of particle translocation to the brain, the EPA should include an explicit discussion of the relevant differences between animals and humans and how this affects the interpretation of animal study results showing particle translocation. For example, if the putative pathway of PM effects on the brain is translocation from the nasal epithelium to the nasal bulb, then how does the difference in percent of nasal deposition and nasal epithelium between rodents and humans impact the interpretation of the animal study results? This discussion should be included when drawing conclusions that rely on extrapolating animal particle translocation results to humans.

Study quality - A clear and detailed explanation should be provided of how study quality criteria were used and applied to the reviewed studies. Study selection and quality assessment are described in general in the ISA Preamble, and more specifically for the Draft ISA in Appendix 1. That appendix indicates that studies are not necessarily excluded from consideration based on quality assessment. What is missing from the current Draft ISA, as well as the preamble and Appendix 1, is a description of how

quality assessments are used in the review process. Although the methods for assessing quality appear appropriate, there is currently a gap between study quality assessment and its application in the ISA preparation and subsequent risk assessment process. Are the quality reviews performed by the ISA section author(s) themselves or independently? If independently, are there written quality assessments for each study that are available to the author(s)? The ISA occasionally provides quality-related commentary in the text and/or tables, but this seems to be left up to the individual author of that section, and overall there seems to be little application of study quality considerations in the document. Significant weakness or strengths could be added to the tables listing the studies in each section. Chance, bias, and confounding are all potential reasons for a study to observe an association between two variables (Zaccai, 2004) and therefore should be more explicitly considered when presenting and discussing study results. In addition, more factors than just copollutants should be considered as important confounders in the referenced epidemiology studies.

Effects of chance - Results that are not statistically significant should be indicated as such in the ISA discussion. If there is a reason why statistical significance may not have been achieved (e.g., low sample size), this should be included in the discussion of the study results. An example of the importance of considering statistical significance of results is given on p. 5-118 (Section 5.1.10.2) where the EPA discusses the results of epidemiology studies that used lag -1 as a negative control (i.e., the relationship between asthma ED visits and PM2.5 concentrations *the day after* the ED visit). Strickland et al. (2010) found associations at lag 0-2 RR = 1.05 (1.02, 1.08), and at lag -1 RR = 1.03 (1.00, 1.05); Sarnat et al. (2015) found associations at lag 0-2 RR = 1.04 (1.01, 1.06) and at lag -1 RR = 1.02 (0.99, 1.04). In the absence of statistical significance, the lag -1 results would look like they were providing evidence for a positive association, even though they break the rule that cause must come before effect. In addition to demonstrating the importance of considering statistical significance, the minimal differences between the RR at lag 0-2 compared to lag -1 calls into question the judgement that there is a real association between PM2.5 and asthma ED visits in these studies.

Exposure concentrations - When results of a study are discussed in the text, it would be helpful for the reader if exposure concentrations were also included in that discussion.

Human equivalent concentrations - Extrapolating animal exposure concentrations to human equivalent concentrations for key studies using particulate dosimetric adjustment models (e.g., MPPD, RDDR models) is standard practice in risk assessment when deriving a toxicity factor for a particulate chemical. If the EPA is drawing a conclusion based on an animal toxicology study, then they should conduct dosimetric adjustments to convert the animal exposure concentrations to human equivalent concentrations, and then determine the likely effects and the relationship to ambient concentrations from those calculated concentrations.

Identifying no- and lowest-observed adverse effect levels - The controlled human exposure studies provide a wealth of information about potential $PM_{2.5}$ effects generated in an experimental setting. Further integration and discussion of this evidence may demonstrate that there are exposure concentrations of effect and no effect (i.e., low- and no-observed (adverse) effect levels – LOEL/LOAELs and NOELs/NOAELs), which would be very informative in determining thresholds of effect and may identify likely mechanistic pathways. Identifying these levels is a standard practice in toxicity factor derivation.

- For example, Ghio et al. (2000) observed an increase in neutrophils in bronchoalveolar lavage (BAL) fluid with exposure to (on average) 120 μg/m^3 fine Chapel Hill concentrated ambient particles (CAPs), but Huang et al. (2012) did not observe an increase in neutrophils in BAL fluid with an exposure to (on average) 90 μg/m^3 fine Chapel Hill CAPs. Neither observed any change in soluble inflammatory cytokines in BAL fluid or blood. Both studies exposed ~25-year-old healthy adults for 2 hours with 1 hour of exercise and took measurements at 18 hours after exposure. This suggests a NOEL at 90 μg/m^3, and a LOEL at 120 μg/m^3 for increased neutrophils in BAL fluid. The question of the adversity of the effect would still need to be discussed.
- Animal studies can also provide evidence of NOELs and LOELs. For example, Harkema et al. (2009) exposed rats to Detroit fine CAPS with and without Ovalbumin (OVA)-sensitization. The authors did not observe independent effects of 600 ug/m^3 CAPs (8 hrs per day for 3 days) on pulmonary endpoints but found that fine CAPs enhanced OVA-induced bronchopneumonia. This did not happen with the animals exposed to 356 ug/m^3 CAPs, demonstrating a potential threshold of effects. When modified with a dosimetric adjustment to a human equivalent concentration, and with appropriate uncertainty factors, this information may be relevant to standard setting.

Dose-responses in experimental studies - Dose-responsiveness of effects of PM exposure in experimental studies can be used to identify relevant biological plausibility pathways and exposure-specific responses. For example, in the Ghio et al. (2000) controlled human exposure study, the increase in neutrophils in bronchial lavage fluid was not dose-responsive (the highest infiltration was in the 3rd quartile dose), but the increase in neutrophils in the BAL fluid was dose-responsive (neutrophil infiltration increased with every dose). In the same study, fibrinogen concentrations increased in the bronchoalveolar lavage fluid as well, but it was not dose-responsive.

Adversity - The ISA should address the adversity and clinical significance of the presented health effects. For example, there should be consideration of the adversity of a 1 mm Hg change in blood pressure, or a 0.8 μg/dL change in glucose levels.

Chapter 5 - The ISA should address possible reasons for the discrepancy in findings from epidemiological and human clinical studies of PM exposure. Despite strong evidence for increased respiratory morbidity and mortality in epidemiology studies, clinical studies that often use PM concentrations much higher than ambient generally show little or no effects on lung function (Bräuner et al., 2007; Ghio et al., 2000; Gong et al., 2003; Gong et al., 2004a; Huang et al., 2012; Sivagangabalan et al., 2011; or Urch et al., 2010), and somewhat variable findings in terms of airway inflammation (e.g. Holgate et al., 2003; Huang et al., 2012). There may be a number of reasons for this, including the fact that clinical studies involve generally healthy people, or involve those with relatively mild respiratory disease, with brief durations of exposure. The ISA would be strengthened by addressing this, especially considering that the biological plausibility sections repeatedly indicate that airway inflammatory effects may be driving systemic and cardiovascular effects. One possible explanation is that the respiratory effects of ambient PM are enhanced by co-pollutants that are not present in clinical exposure studies, although several controlled human exposure studies that exposed people to fine CAPs + NO_2 or ozone have not shown impacts on airway inflammation, systemic inflammation, or lung function (Gong et al., 2005; Huang et al., 2012; Sivagangabalan et al., 2011; Urch et al., 2010).

Figure 6-2, page 6-13 - The depiction of the associations with myocardial infarction (MI) in the Gardner et al. (2014) study is incorrect. The ISA figure appears to show only data for the 72- or 96-hour lag, neither of which was significant. As shown in the figure below from the publication, ST-elevation MI (STEMI) was significantly increased with a 1-hr lag.

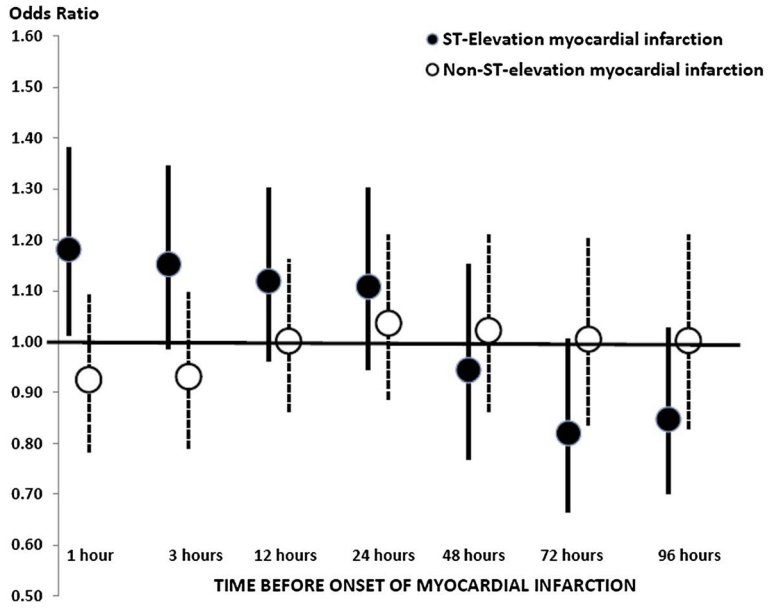


Figure 1 from Gardner et al. (2014)

Page 6-14, line 18 - "There were generally consistent results across recent studies looking specifically at MI, and registry studies, which are likely to reduce outcome misclassification, report evidence of positive associations with MI subtypes." This sentence seems at odds with the first paragraph on this page, which indicates inconsistencies, especially in the European studies. The interpretation should be further clarified, with justification for disregarding the negative European studies.

Page 6-176, line 3 - "A study of newborns in Massachusetts found elevated [systolic blood pressure] SBP with higher $PM_{2.5}$ averages over the 30-, but not 60- or 90-day periods before birth (van Rossem et al., 2015) while trimester specific associations between $PM_{2.5}$ and increased SBP increased but confidence intervals were wide…" This is a run-on sentence and needs clarification. The description should make clear that the exposure estimates were during the 90 days before birth, but the BP measurements were 30 hours after birth.

Section 6.2.8, Peripheral Vascular Disease (PVD), Venous Thromboembolism (TE), Pulmonary Embolism - The diagnosis "peripheral vascular disease" generally refers to disease in the peripheral arterial system, rather than venous disease. The discussion in this section is limited to venous thromboembolism, and does not address arterial PVD, so this term should be removed from the title. In any case, PVD should not be lumped together with venous TE disease; they have different etiologies, pathophysiology, and treatments.

Page 6-196, line 31 - In the description of the Wilker et al. (2014) study, the ISA states, "Only hyperemic flow velocity was additionally associated with $PM_{2.5}$ [-1.80 % change (95% CI: -3.45, -0.15)]

These effects are relatively large given that normal ranges are between 5-10% (Järhult et al., 2009)." This description of the findings of the Wilker et al. (2014) study is incorrect. The normal range for FMD% is 5-10. Hyperemic flow velocity is expressed in the units of cm/s, not %. Also, it is not clear where the "-1.80% change" comes from. The Wilker et al. (2014) abstract states: "An inter-quartile range difference in $PM_{2.5}$ (1.99 ug/m^3) was associated with -0.16% (95% confidence interval [CI] -0.27%, -0.05%) lower flow-mediated dilation% and -0.72 (95% CI -1.38, -0.06) cm/s lower hyperemic flow velocity%."

Page 6-283, line 26 - "Weichenthal et al. (2014a) reported positive associations between 2-hour averages of NCs with SBP measurements taken 3 hours post-exposure, but associations with SBP were null." This sentence is contradictory and needs clarification. Associations of UFP with SBP were not significant in this study.

Section 7.1.3, Other Indicators of Metabolic Function - This section should be re-thought and re-organized. The subheading topics of systemic inflammation and blood pressure have already been reviewed as outcomes, and it is redundant to revisit them here. It is enough for the background to point out the interplay of inflammation in metabolic effects and in obesity, as well as hypertension as a clinical component of the metabolic syndrome, and then just reference the previous sections.

Section 7.2.10, Metabolic Disease Mortality - The title of this section, and some of the text, are a bit misleading. People don't often die of "metabolic disease" (although there are certainly deaths from diabetic ketoacidosis). Their metabolic conditions increase risk for mortality from a variety of causes, from cardiovascular deaths to pneumonia and other infections. The Pope et al. (2014) paper described in this section looks at cardiovascular mortality, and examines whether metabolic disease such as diabetes, contribute to the PM risk for CV mortality. This issue fits best in Chapter 12.

Chapter 8, pages 8-5 and 8-67 - The ISA cites a human clinical study, Liu et al. (2017), as evidence that $PM_{2.5}$ causes perturbation of the blood-brain barrier (BBB). However, the study did not observe a significant change in the BBB biomarkers S100B, NSE, or UCHL1 with exposure to fine, coarse, or UFP CAPs. The p-values in some of these comparisons were less than 0.1, but not less than 0.05. There was a significant relationship between a component of coarse PM and S100B, but this is difficult to interpret in the absence of a total PM effect. There were no effects of concentrated UFP on any marker. Since UFP is the size fraction most capable of transport to the brain, this finding is counter-intuitive. This study should not be interpreted as showing an effect of PM on the BBB, especially considering the many comparisons made in this study.

Chapter 8, UFP short-term nervous system effects, page 8-82 - The Liu et al. (2017) human clinical study is mistakenly characterized as showing an effect of UFP on the hypothalamic-pituitary-adrenal (HPA) axis. Again, the p-value was <0.1, not <0.05. From the abstract of the Liu study: "Ultra fine CAP was not significantly associated with changes in any blood and urinary neural biomarkers examined."

Chapter 9, Table 9-6, page 9-29 - For the Kloog et al. (2012) study, the last column indicates the effect estimate is 1.03, with 95% CIs of 0.54, 0.63. These values are obviously incorrect, because the CI does not include the estimate. According to the abstract and Table 3 of the study, the odds ratio of premature birth was 1.06 (1.01 to 1.13).

Page 9-6, line 23 - The Tallon et al. (2017) study is described as showing "…positive associations between exposure to annual $PM_{2.5}$ concentrations and erectile dysfunction in men aged 57−85 years (OR: 1.26; 95% CI: 0.81, 1.96)." Although the OR is positive, the 95% CI includes 1, so the findings are not statistically significant. Highlighting this in the ISA as a positive study, without further qualification, is misleading.

The words fecundity and fecundability are used interchangeably in this section. The CASAC suggests changing the latter to the former wherever it occurs.

Chapter 10, Lung Cancer - This chapter reviews new studies addressing lung cancer incidence and mortality in relation to long-term PM exposure. 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 $PM_{2.5}$ 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.

The cancer section notes that many studies find the greatest effects of $PM_{2.5}$ in non-smokers. The possible biological reasons for this pattern should be discussed, as well as how it fits in with other information. For example, is it consistent with evidence that animal studies with $PM_{2.5}$ exposure have not shown increased carcinogenesis, except with animals that were pre-initiated with urethane (Pereira et al., 2011)?

Page 10-9 - "…an in vivo study by Sato et al. (2003) reported increased DNA adducts in lung, liver, and nasal mucosal tissues after inhalation exposure to urban roadside air. Because this study evaluated effects of exposure to a mixture of PM and gases, it does not inform the current ISA, which identifies the hazard for effects after exposures to only the PM component of complex mixtures…" Virtually all epidemiological studies involve exposures to mixtures of PM and gases, and yet they can and do inform the Draft ISA. The issue with the study in this case is not the exposure to a mixture, but that PM concentrations in the roadside air were not quantified. One could therefore argue that this study should not be included in the ISA, since it does not meet the screening criteria stated in the Preface, page P-14, indicating the focus is on studies that "…(1) include a composite measure of PM or (2) characterize PM and apply some approach to assess the direct effect of PM when the exposure of interest is a source-based mixture…."

A similar issue requires correction on page 10-35, line 3. "Because these in vivo studies evaluated effects of exposure to mixtures of PM and gases, they do not directly inform the current ISA, which identifies the hazard for effects after exposures to only the PM component of complex mixtures." Again, the problem is not the exposure to mixtures, but the absence of quantification or characterization of the PM exposure. The CASAC suggests removing these studies from consideration in the ISA, because it is true that they do not inform this ISA.

Page 10-49, line 27 - This statement is incorrect: "Specifically, an assessment of adenocarcinoma, the only subtype that develops in nonsmokers…" Adenocarcinoma is not the only type of lung cancer that occurs in nonsmokers. "Only" should be changed to "predominant." The same applies to page 10-53, line 11, and page 10-74, line 29.

Application of scientific information to inform causal conclusions:

Nervous System Effects

Chapter 8, Nervous System Effects – $PM_{2.5}$ - The EPA does not provide adequate evidence for the conclusion that there is likely to be a causal association between long-term $PM_{2.5}$ 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.

Chapter 8, Nervous System Effects - UFP - 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.

Cancer

There is inadequate evidence for the "likely to be causal" conclusion for long-term $PM_{2.5}$ exposure and cancer. This determination relies largely on epidemiology studies that, as noted above, do not provide exposure time frames that are appropriate for cancer causation. There are no animal studies showing direct effects of $PM_{2.5}$ on cancer formation, with the only positive animal results coming from a group that pre-initiated the animals with urethane.

Conflicting Evidence

The ISA would be strengthened by more justification of decisions in the face of conflicting evidence.

For example:

- In controlled human exposure studies that investigated blood pressure (BP), Bellavia et al. (2013) found increased SBP with exposure to 242 $\mu g/m^3$ Toronto CAPs; Brook et al.

(2009) showed increased DBP (not SBP) with exposure to 148 $\mu g/m^3$ Toronto CAPs; and Sivagangabalan et al. (2011) showed increased DBP (not SBP) with exposure to 154 $\mu g/m^3$ Toronto CAPs. No effects on BP were seen with fine CAPs exposure in Bräuner et al. (2008), Brook et al. (2002), Gong et al. (2003, 2004, or 2005), Hemmingsen et al. (2015), Huang et al. (2012), or Mills et al. (2008). These studies exposed individuals who were healthy, elderly, overweight, with COPD, asthma, or CHD, to $PM_{2.5}$ CAPs concentrations up to 207 $\mu g/m^3$.

- Section 6.1.2.1, ED visits and hospital admissions. This section concludes by saying that recent studies "continue to provide evidence for positive associations between short-term $PM_{2.5}$ exposure and IHD ED visits and HA." However, the preceding text and Figure 6-2 (please note the separate comment about the error in this figure) show considerable heterogeneity in the findings of the studies conducted since the 2009 ISA. There was one study with a positive but not statistically significant result (Bell et al., 2015), one with a positive statistically significant result (Kloog et al., 2014), one with associations only in NYC but not the rest of the state (Hsu et al., 2017), one with associations in 2 of 7 states (Talbot 2014), one with a negative association (Milojevic et al., 2014), and two single city studies with opposite results (Kim et al., 2012, Sarnat et al., 2015). This section should strengthen the rationale for the conclusions.
- Section 6.1.12, Coagulation. The statement is made in the second paragraph that "When considered as a whole, these recent studies do provide additional evidence that short-term exposure to $PM_{2.5}$ can promote clot formation." However, the section goes on to describe considerable inconsistency in the findings from epidemiological, human clinical, and toxicology studies. The conclusion drawn in this section should be reconsidered and provided in a summary paragraph at the end of the section.
- Section 6.1.1, Short-Term Exposure to $PM_{2.5}$, Biological Plausibility. The results of Langrish et al. (2014) do not support the conclusion that $PM_{2.5}$ has effects on heart conduction abnormalities. The EPA should clarify how they choose biological plausibility endpoints in the presence of conflicting evidence.

Concentration-Response Functions

CASAC members were unable to reach consensus on the ISA assessment of concentration-response functions.

Some members of CASAC think that the EPA should do further work on C-R functions. In the Draft ISA, the EPA concludes that the evidence from epidemiology studies largely supports a linear, no-threshold association between $PM_{2.5}$ and various health effects. However, a number of statistical studies have shown that the error (e.g., measurement error) in these types of epidemiology studies lead study authors to the erroneous conclusion that C-R functions are linear with no threshold when that is not, in fact, the case (Rhomberg et al., 2011; Brauer et al., 2002; Cox, 2018; Lipfert and Wyzga, 1996; Watt et al., 1995; Yoshimura, 1990). Therefore, the EPA should not be using these epidemiology studies to draw conclusions about the true shape of the relationship between $PM_{2.5}$ and health effects, unless it can strongly argue (and provide evidence) that the referenced epidemiology studies can produce an unbiased estimate of the true shape of the C-R function. In addition, this conclusion is not consistent with the evidence of a threshold of effects demonstrated in human controlled exposure and animal toxicology studies (discussed above).

The EPA should consider deriving C-R relationships from animal and human controlled exposure studies, where we can be more certain that the effects are caused by the exposure, and there is less error to bias the shape of the relationship. Interpretation and extrapolation from either epidemiology or experimental C-R relationships is impacted by whether the relationship is quantal or graded, and so the EPA should include this information in their discussion of these responses.

If the EPA does use C-R functions derived from epidemiology studies with binary outcomes (assuming, importantly, that there is a causal relationship between the concentration and the response), they should consider these functions as quantal relationships. Quantal relationships generally have Gaussian distributions and describe a continuum of a population response to an exposure where the response is binary (e.g., percent of population who experienced an asthma attack, or who died). Quantal relationships asymptotically approach a response of 0 percent as the concentration decreases and 100 percent as the concentration increases. The smallest effective concentration of any chemical or substance that causes a pre-determined amount of an all-or-none response may be referred to as a *threshold concentration* even though it cannot be determined experimentally. Therefore, by their asymptotic Gaussian nature quantal dose responses cannot identify a concentration where 0% of people are responders, and so it is standard practice in toxicology-risk assessment to set an effect level (such as a 10%, 1%, or 0.1% response level) and designate the concentration that causes that effect level to be the threshold dose or concentration. Using this method, the EPA could dictate the threshold response, and therefore concentration, from a type of relationship that otherwise by its nature does not allow the identification of a concentration that causes a response in 0% of the population.

Other members of CASAC think that the ISA contains a reasonably balanced assessment of the new data on C-R relationships, with appropriate caveats about the uncertainties, especially at low concentrations. However, these members also think that these important considerations need further input from experts in epidemiology. In the case of PM, understanding of C-R relationships at low exposure concentrations must come from epidemiology. Toxicological and human clinical studies have a limited role, especially with regard to mortality. For example, mortality and morbidity are not outcomes of human clinical studies, by design. Additionally, clinical studies are generally conducted at concentrations higher than ambient concentrations, in order to provide a contrast with prior ambient exposures of the subjects. Further, clinical studies involve relatively small numbers of subjects and generally do not include individuals with severe disease or frailty that may make them more susceptible to effects from relatively low PM concentrations. For these reasons, clinical studies unfortunately provide little help in informing thresholds of effect for PM.

Populations and Lifestages Potentially at Increased Risk of a Particulate Matter-Related Health Effect (Chapter 12)

Chapter 12 evaluates scientific information and presents conclusions on factors that may contribute to specific populations or lifestages being at increased risk of a PM-related health effect. Please comment on the extent to which the available scientific evidence from epidemiologic, controlled human exposure, and toxicological studies been integrated to inform conclusions on populations and/or lifestages potentially at increased risk of a PM-related health effect. Is there information available on other key factors that is not included in the draft PM ISA that inform differential risk that should be added?

This chapter delineates the approach to considering the evidence for at-risk groups and populations. It is an improvement on the approach taken in the 2009 ISA, and, aside from the comments below, clearly presents the rationale and evidence base for the conclusions. The 4-level grading of the conclusions is logical and reasonable, and parallels the approach taken for causality determinations.

Inclusion

The ISA should include lung cancer as an endpoint in the smoking section (12.6.1), with a brief summary of the findings presented in Chapter 10 and any additional information.

Evidence integration

Better integration of the information presented in this chapter is needed, beyond listing of study results. The results within a section should be integrated, for example in the genetic factors section. When considering genetic factors such as glutathione polymorphisms, the EPA should look at the effect of the polymorphism (i.e., does it increase or decrease the effectiveness of the glutathione system), and not just whether there is an association with any glutathione polymorphism. Similarly, the EPA should integrate their conclusions between chapters. For example, one would expect that older adults (compared to younger adults) would be more susceptible to the toxic effects of PM (because, as a group, older people tend to be more frail and have more diseases). However, the section on older adults (for which there have been many studies) does not find that older age is consistently a risk factor.

Causal Determinations

The ISA should better justify the different causal conclusions for various risk factors. For example, diabetes, obesity, and elevated cholesterol all seem to have a similar amount of data about risk, and generally show inconsistent study results, but it is not clear why diabetes and elevated cholesterol were categorized as "insufficient," and obesity was designated as "suggestive."

Welfare Effects (Chapter 13)

Please comment on the identification, evaluation and characterization of the available scientific evidence from studies of PM on non-ecological welfare effects of visibility impairment, climate, and materials and the application of information from these studies, as presented in Chapter 13, to inform causality determinations and uncertainty characterizations for these welfare outcomes.

The information presented in Chapter 13 supports a causal relationship between PM and visibility impairment, climate effects, and effects on materials.

The evaluation of welfare effects often lumps PM together as a whole without considering different size fractions. It is recommended that the EPA perform more analyses for different size fractions to determine whether various effects on visibility, climate, and materials are observed. Specific quality criteria targets for inclusion or exclusion of welfare effects studies should be articulated up front in each section. The studies presented in this chapter are mostly descriptive with little reference to quality. There is little discussion of how study findings that consist of different PM concentrations, different mixtures, different experimental design questions, and different ambient conditions apply directly to non-

ecological welfare effects in the United States. A "Research Needs" section should be added to the final ISA. In addition, line numbers should be added for pages 13-1 through 13-56.

For visibility effects, a thorough discussion of the instrumentation used for measuring visibility is provided. It would be very useful if the instruments were shown in a table with the figures of merit associated with each, and how well each instrument provides the most policy relevant measurements. The distinction between anthropogenic PM impairment versus natural PM impairment needs to be more clearly separated and explained. How this distinction can or will be used for setting a secondary standard needs to be included in the document.

The document does a good job more firmly establishing a causal relationship between PM and visibility. However, it is challenging to tease out the complex nature of PM across the country and how the variation in PM composition affects visibility differently. Setting a secondary standard given such variability will be very difficult. A discussion on the direct effect of PM or other pollutants (e.g., photochemical oxidant) on visual acuity should be included. Instruments would not be responsive to these eye irritants. Also, comparing perceived visibility impairment of urban versus more "bucolic" settings may have inherent biases. Some viewers of these scenes may not find urban viewscapes to be very appealing no matter how clear the image may be. Moreover, regional differences in perceived visibility may be due to societal differences. Westerners may have greater expectations of clear mountain vistas than Easterners.

On page 3-6, the haze index is presented as $dv = 10 * ln\ (b_{ext}/0.01\ km^{-1})$. However, the units for b_{ext} are typically Mm^{-1}. Therefore, the correct equation should be $dv = 10 * ln\ (b_{ext}/10\ Mm^{-1})$. The original IMPROVE algorithm (Equation 13-6) and the modified original IMPROVE algorithm (Equation 13-7) are presented on page 13-10. These equations tend to underestimate the highest light scattering values and overestimate the lowest values at IMPROVE monitors throughout the U.S. To resolve these biases, a revised IMPROVE equation was developed (Pitchford et al., 2007) that divides PM components into small and large particle sizes with separate mass scattering efficiencies and hygroscopic growth functions for each size. The revised IMPROVE equation was described in detail in the 2009 PM ISA, and it both reduced bias at the lowest and highest scattering values and improved the accuracy of the reconstructed b_{ext}. However, the revised IMPROVE equation is not presented in this document:

b_{ext} = 2.2 x fs(RH) x [Small Sulfate] + 4.8 x fL(RH) x [Large Sulfate]
+ 2.4 x fs(RH) x [Small Nitrate] + 5.1 x fL(RH) x [Large Nitrate]
+ 2.8 x [Small Organic Mass] + 6.1 x [Large Organic Mass]
+ 10 x [Elemental Carbon]
+ 1 x [Fine Soil]
+ 1.7 x fss(RH) x [Sea Salt]
+ 0.6 x [Coarse Mass]
+ Rayleigh Scattering (site specific)
+ 0.33 x [NO2 (ppb)]

This equation should be added to the document and discussed in detail. Use of this equation is recommended by EPA in their modeling guidance (U.S. EPA, 2018a).

The color maps, bar charts, and other graphical data presentations are very helpful. The uncertainty associated with the size fraction and visibility impairment needs to be stated clearly. Figures 13-1

through 13-14 should use the same colors to represent the same species in the stacked bar charts when comparing across years (2005-2008 vs. 2011-2014). The figures that show 2011-2014 data use light blue for POM while the figures that show 2005-2008 data use green for POM. In addition, Figures 13-2, 13-4, 13-6, 13-8, 13-10, 13-12, 13-13, and 13-14 are low resolution and very difficult to read.

For climate effects, uncertainty in the effects of complex aerosol composition on climate needs to be better resolved. If there is new evidence that increased atmospheric turbidity is increasing cloud-to-ground lightning strikes and hence increased forest fires, that information should be added to the document.

For effects on materials, it was difficult to determine from the literature review presented in the ISA at what level damage to materials was unacceptable and how that relates back to PM concentration, size, and mixture. It is laudable that data from other countries were included in the assessment. The document should discuss if there is sufficient meta data available to fully characterize the data quality attributes associated with these data.

It would be helpful for the CASAC to have ready access to an expert that studies the effects of PM on visibility impairment, climate, and materials. This would allow for additional insight into the non-ecological welfare effects and better inform our recommendations on the appropriate level for the secondary PM standard.

References

Aaron, C. P., Chervona, Y., Kawut, S. M., Roux, A. V., Shen, M., Bluemke, D. A., Van Hee, V. C., Kaufman, J. D., and Barr, R. G. (2016). Particulate matter exposure and cardiopulmonary differences in the Multi-Ethnic study of atherosclerosis. *Environmental Health Perspectives*, 124(8):1166-1173.

Avery, C. L., Mills, K. T., Williams, R., McGraw, K. A., Poole, C., Smith, R. L., and Whitsel, E. A. (2010a). Estimating error in using ambient PM2.5 concentrations as proxies for personal exposures: a review. *Epidemiology*, 21(2):215-223.

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 PM2.5 concentrations as proxies for personal exposures: a meta-analysis. *Environmental Health Perspectives*, *118*(5), 673-8.

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. *Environmental Health*, 16(1).

Bell, M. L., Son, J.-Y., Peng, R. D., Wang, Y., & Dominici, F. (2015). Brief Report: Ambient PM2.5 and risk of hospital admissions: do risks differ for men and women? *Epidemiology*, 26(4), 575–579.

Bellavia, A., Urch, B., Speck, M., Brook, R. D., Scott, J. A., Albetti, B., Behbod, B., North, M., Valeri, L., Bertazzi, P. A., Silverman, F., Gold, D., and Baccarelli (2013). DNA hypomethylation, ambient particulate matter, and increased blood pressure: Findings from controlled human exposure experiments. *Journal of the American Heart Association*, 2(3).

Brauer, M., Brumm, J., Vedal, S., and Petkau, A. J. (2002). Exposure misclassification and threshold concentrations in time series analyses of air pollution health effects. *Risk Analysis*, 22(6):1183-1193.

Bräuner, E. V., Forchhammer, L., Møller, P., Barregard, L., Gunnarsen, L., Afshari, A., Wåhlin, P., Glasius, M., Dragsted, L. O., Basu, S., Raaschou-Nielsen, O., and Loft, S. (2008). Indoor particles affect vascular function in the aged. *American Journal of Respiratory and Critical Care Medicine*, 177(4):419-425.

Bräuner, E. V., Forchhammer, L., Møller, P., Simonsen, J., Glasius, M., Wåhlin, P., Raaschou-Nielsen, O., and Loft, S. (2007). Exposure to ultrafine particles from ambient air and oxidative Stress–Induced DNA damage. *Environmental Health Perspectives*, 115(8):1177-1182.

Brook, R. D., Brook, J. R., Urch, B., Vincent, R., Rajagopalan, S., and Silverman, F. (2002). Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation*, 105(13):1534-1536.

Brook, R. D., Urch, B., Dvonch, J. T., Bard, R. L., Speck, M., Keeler, G., Morishita, M., Marsik, F. J., Kamal, A. S., Kaciroti, N., Harkema, J., Corey, P., Silverman, F., Gold, D. R., Wellenius, G., Mittleman, M. A., Rajagopalan, S., and Brook, J. R. (2009). Insights into the mechanisms and mediators of the effects of air pollution exposure on blood pressure and vascular function in healthy humans. *Hypertension*, 54(3):659-667.

Calderón-Garcidueñas, L., Franco-Lira, M., Henríquez-Roldán, C., Osnaya, N., González-Maciel, A., Reynoso-Robles, R., Villarreal-Calderon, R., Herritt, L., Brooks, D., Keefe, S., Palacios-Moreno, J., Villarreal-Calderon, R., Torres-Jardón, R., Medina-Cortina, H., Delgado-Chávez, R., Aiello-Mora, M., Maronpot, R. R., and Doty, R. L. (2010). Urban air pollution: influences on olfactory function and pathology in exposed children and young adults. *Experimental and Toxicologic Pathology*, 62(1):91-102.

Calderón-Garcidueñas, L., Serrano-Sierra, A., Torres-Jardón, R., Zhu, H., Yuan, Y., Smith, D., Delgado-Chávez, R., Cross, J. V., Medina-Cortina, H., Kavanaugh, M., and Guilarte, T. R. (2013). The impact of environmental metals in young urbanites' brains. *Experimental and Toxicologic Pathology*, 65(5):503-511.

Campbell, A., Oldham, M., Becaria, A., Bondy, S. C., Meacher, D., Sioutas, C., Misra, C., Mendez, L. B., Kleinman, M. (2005). Particulate matter in polluted air may increase biomarkers of inflammation in mouse brain. *Neurotoxicology,* 26:133-140.

Carrothers, T. J. and Evans, J. S. (2000). Assessing the impact of differential measurement error on estimates of fine particle mortality. *Journal of the Air & Waste Management Association (1995)*, 50(1):65-74.

Cox, L., Liu, X., Shi, L., Zu, K., and Goodman, J. (2017). Applying nonparametric methods to analyses of Short-Term fine particulate matter exposure and hospital admissions for cardiovascular diseases among older adults. *International Journal of Environmental Research and Public Health*, 14(9):1051+.

Cox, L. A. (2018). Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. *Environmental Research*, 164:636-646.

Cox, L. A. and Popken, D. A. (2015). Has reducing fine particulate matter and ozone caused reduced mortality rates in the united states? *Annals of Epidemiology*, 25(3):162-173.

Ebelt, S. T., Wilson, W. E., and Brauer, M. (2005). Exposure to ambient and nonambient components of particulate matter: a comparison of health effects. *Epidemiology*, 16(3):396-405.

Eftim, S. E., Samet, J. M., Janes, H., McDermott, A., & Dominici, F. (2008). Fine Particulate Matter and Mortality. *Epidemiology*, 19(2), 209–216.

Fewell, Z., Davey Smith, G., and Sterne, J. A. (2007). The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *American Journal of Epidemiology*, 166(6):646-655.

Garcia, G. J. M., Schroeter, J. D., and Kimbell, J. S. (2015). Olfactory deposition of inhaled nanoparticles in humans. *Inhalation Toxicology*, 27(8):394-403.

Gardner, B., Ling, F., Hopke, P. K., Frampton, M. W., Utell, M. J., Zareba, W., Cameron, S. J., Chalupa, D., Kane, C., Kulandhaisamy, S., Topf, M. C., Rich, D. Q. (2014). Ambient fine particulate air pollution triggers ST-elevation myocardial infarction, but not non-ST elevation myocardial infarction: a case-crossover study. *Particle and Fibre Toxicology*, 11(1).

Ghio, A. J., Kim, C., and Devlin, R. B. (2000). Concentrated ambient air particles induce mild pulmonary inflammation in healthy human volunteers. *American Journal of Respiratory and Critical Care Medicine*, 162(3):981-988.

Gong, H., Linn, W. S., Clark, K. W., Anderson, K. R., Geller, M. D., and Sioutas, C. (2005). Respiratory responses to exposures with fine particulates and nitrogen dioxide in the elderly with and without COPD. *Inhalation Toxicology*, 17(3):123-132.

Gong, H., Linn, W. S., Sioutas, C., Terrell, S. L., Clark, K. W., Anderson, K. R., and Terrell, L. L. (2003). Controlled exposures of healthy and asthmatic volunteers to concentrated ambient fine particles in Los Angeles. *Inhalation Toxicology*, 15(4):305-325.

Gong, H., Linn, W. S., Terrell, S. L., Anderson, K. R., Clark, K. W., Sioutas, C., Cascio, W. E., Alexis, N., and Devlin, R. B. (2004a). Exposures of elderly volunteers with and without chronic obstructive pulmonary disease (COPD) to concentrated ambient fine particulate pollution. *Inhalation Toxicology*, 16(11-12):731-744.

Gong, H., Linn, W. S., Terrell, S. L., Clark, K. W., Geller, M. D., Anderson, K. R., Cascio, W. E., and Sioutas, C. (2004b). Altered Heart-Rate variability in asthmatic and healthy volunteers exposed to concentrated ambient coarse particles. *Inhalation Toxicology*, 16(6-7):335-343.

Greven, S., Dominici, F., and 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.

Grunig, G., Marsh, L. M., Esmaeil, N., Jackson, K., Gordon, T., Reibman, J., Kwapiszewska, G., and Park, S.-H. (2014). Perspective: Ambient air pollution: Inflammatory response and effects on the lung's vasculature. *Pulmonary Circulation*, 4(1):25-35.

Harkema, J.R., Wagner, J.G., Kaminski, N.E., Morishita, M., Keeler, G.J., McDonald, J.D., Barrett, E.G. (2009). Effects of concentrated ambient particles and diesel engine exhaust on allergic airway disease in Brown Norway rats. *Research Report (Health Effects Institute)*, (145):5–55.

Hart, J. E., Liao, X., Hong, B., Puett, R. C., Yanosky, J. D., Suh, H., Kioumourtzoglou, M. A., Spiegelman, D., and 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.

Hemmingsen, J. G., Rissler, J., Lykkesfeldt, J., Sallsten, G., Kristiansen, J., Møller, P., and 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):6+.

Holgate, S. T., Sandström, T., Frew, A. J., Stenfors, N., Nördenhall, C., Salvi, S., Blomberg, A., Helleday, R., and Söderberg, M. (2003). Health effects of acute exposure to air pollution. part i: Healthy and asthmatic subjects exposed to diesel exhaust. *Research Report (Health Effects Institute)*, (112).

Hsu, W.-H., Hwang, S.-A., Kinney, P. L., & Lin, S. (2017). Seasonal and temperature modifications of the association between fine particulate air pollution and cardiovascular hospitalization in New York state. *Science of The Total Environment*, 578, 626–632.

Huang, Y.-C. T., Rappold, A. G., Graff, D. W., Ghio, A. J., and Devlin, R. B. (2012). Synergistic effects of exposure to concentrated ambient fine pollution particles and nitrogen dioxide in humans. *Inhalation Toxicology*, 24(12):790-797.

Järhult, S. J., Sundström, J., & Lind, L. (2009). Brachial artery hyperemic blood flow velocities are related to carotid atherosclerosis. *Clinical Physiology and Functional Imaging*, 29(5), 360–365.

Kendall, M., Tetley, T. D., Wigzell, E., Hutton, B., Nieuwenhuijsen, M., & Luckham, P. (2002). Lung lining liquid modifies PM2.5 in favor of particle aggregation: a protective mechanism. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 282(1), L109–L114.

Kim, S.-Y., Peel, J. L., Hannigan, M. P., Dutton, S. J., Sheppard, L., Clark, M. L., & Vedal, S. (2012). The Temporal Lag Structure of Short-term Associations of Fine Particulate Matter Chemical Constituents and Cardiovascular and Respiratory Hospitalizations. *Environmental Health Perspectives*, 120(8), 1094–1099.

Kloog, I., Coull, B. A., Zanobetti, A., Koutrakis, P., & Schwartz, J. D. (2012). Acute and Chronic Effects of Particles on Hospital Admissions in New-England. *PLoS ONE*, 7(4), e34664.

Kloog, I., Nordio, F., Zanobetti, A., Coull, B. A., Koutrakis, P., & Schwartz, J. D. (2014). Short Term Effects of Particle Exposure on Hospital Admissions in the Mid-Atlantic States: A Population Estimate. *PLoS ONE*, 9(2), e88578.

Kumarathasan, P., Blais, E., Saravanamuthu, A., Bielecki, A., Mukherjee, B., Bjarnason, S., Guénette, J., Goegan, P., and Vincent, R. (2015). Nitrative stress, oxidative stress and plasma endothelin levels after inhalation of particulate matter and ozone. *Particle and Fibre Toxicology*, 12(1):28+.

Langrish, J. P., Watts, S. J., Hunter, A. J., Shah, A. S. V., Bosson, J. A., Unosson, J., Barath, S., Lundbäck, M., Cassee, F. R., Donaldson, K., Sandström, T., Blomberg, A., Newby, D. E., and Mills, N. L. (2014). Controlled exposures to air pollutants and risk of cardiac arrhythmia. *Environmental Health Perspectives*, 122(7):747-753.

Leary, P. J., Kaufman, J. D., Barr, R. G., Bluemke, D. A., Curl, C. L., Hough, C. L., Lima, J. A., Szpiro, A. A., Van Hee, V. C., and Kawut, S. M. (2014). Traffic-related air pollution and the right ventricle. the multi-ethnic study of atherosclerosis. *American Journal of Respiratory and Critical Care Medicine*, 189(9):1093-1100.

Lipfert, F.W., Wyzga, R.E. (1996). The effects of exposure error on environmental epidemiology. In RF Phalen; RC Mannix; MC Tonini (Eds.), The second colloquium on particulate air pollution & human mortality & morbidity (pp. 4-295 - 4-302). Sacramento, CA: California Air Resources Board.

Liu, J., Ye, X., Ji, D., Zhou, X., Qiu, C., Liu, W., and Yu, L. (2018). Diesel exhaust inhalation exposure induces pulmonary arterial hypertension in mice. *Environmental Pollution*, 237:747-755.

Liu, L., Urch, B., Szyszkowicz, M., Speck, M., Leingartner, K., Shutt, R., Pelletier, G., Gold, D. R., Scott, J. A., Brook, J. R., Thorne, P. S., and Silverman, F. S. (2017). Influence of exposure to coarse, fine and ultrafine urban particulate matter and their biological constituents on neural biomarkers in a randomized controlled crossover study. *Environment International*, 101:89-95.

Lund, A. K., Lucero, J., Lucas, S., Madden, M. C., McDonald, J. D., Seagrave, J.-C., Knuckles, T. L., and Campen, M. J. (2009). Vehicular emissions induce vascular MMP-9 expression and activity associated with Endothelin-1–Mediated pathways. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 29(4):511-517.

McGuinn, L. A., Ward-Caviness, C., Neas, L. M., Schneider, A., Di, Q., Chudnovsky, A., Schwartz, J., Koutrakis, P., Russell, A. G., Garcia, V., Kraus, W. E., Hauser, E. R., Cascio, W., Diaz-Sanchez, D., and Devlin, R. B. (2017). Fine particulate matter and cardiovascular disease: Comparison of assessment methods for long-term exposure. *Environmental Research*, 159:16-23.

Mills, N. L., Miller, M. R., Lucking, A. J., Beveridge, J., Flint, L., Boere, A. J. F., Fokkens, P. H., Boon, N. A., Sandstrom, T., Blomberg, A., Duffin, R., Donaldson, K., Hadoke, P. W. F., Cassee, F. R., and Newby, D. E. (2011). Combustion-derived nanoparticulate induces the adverse vascular effects of diesel exhaust inhalation. *European Heart Journal*, 32(21):2660-2671.

Mills, N. L., Robinson, S. D., Fokkens, P. H. B., Leseman, D. L. A. C., Miller, M. R., Anderson, D., Freney, E. J., Heal, M. R., Donovan, R. J., Blomberg, A., Sandström, T., MacNee, W., Boon, N. A., Donaldson, K., Newby, D. E., and Cassee, F. R. (2008). Exposure to concentrated ambient particles does not affect vascular function in patients with coronary heart disease. *Environmental Health Perspectives*, 116(6):709-715.

Milojevic, A., Wilkinson, P., Armstrong, B., Bhaskaran, K., Smeeth, L., & Hajat, S. (2014). Short-term effects of air pollution on a range of cardiovascular events in England and Wales: case-crossover analysis of the MINAP database, hospital admissions and mortality. *Heart*, 100(14), 1093–1098.

Ohlwein, S., Klümper, C., Vossoughi, M., Sugiri, D., Stolz, S., Vierkötter, A., Schikowski, T., Kara, K., Germing, A., Quass, U., Krämer, U., Hoffmann, B. (2016). Air pollution and diastolic function in elderly women – Results from the SALIA study cohort. *International Journal of Hygiene and Environmental Health*, 219(4–5), 356–363.

Pace, T.G. (2005). Methodology to Estimate the Transportable Fraction (TF) of Fugitive Dust Emissions for Regional and Urban Scale Air Quality Analyses. U.S. EPA, Research Triangle Park, NC, August 2005, https://www.nrc.gov/docs/ML1321/ML13213A386.pdf

Park, S.-H., Chen, W.-C., Esmaeil, N., Lucas, B., Marsh, L. M., Reibman, J., and Grunig, G. (2014). Interleukin 13– and interleukin 17A–induced pulmonary hypertension phenotype due to inhalation of antigen and fine particles from air pollution. *Pulmonary Circulation*, 4(4):654-668.

Pereira, F. A., Lemos, M., Mauad, T., Assunção, J. V., and Saldiva, P. H. (2011). Urban, traffic- related particles and lung tumors in urethane treated mice. *Clinics*, 66(6):1051-1054.

Pitchford, M., Maim, W., Schichtel, B., Kumar, N., Lowenthal, D., and Hand, J. (2007). Revised algorithm for estimating light extinction from IMPROVE particle speciation data. *Journal of the Air & Waste Management Association*, 57(11):1326-1336.

Pope, C. A., III, Turner, M. C., Burnett, R. T., Jerrett, M., Gapstur, S. M., Diver, W. R., Krewski, D., Brook, R. D. (2014). Relationships Between Fine Particulate Air Pollution, Cardiometabolic Disorders, and Cardiovascular Mortality. *Circulation Research*, 116(1), 108–115.

Pouliot, G., Simon, H., Bhave, P., Tong, D., Mobley, D., Pace, T., and Pierce, T. (2012). *Assessing the Anthropogenic Fugitive Dust Emission Inventory and Temporal Allocation Using an Updated Speciation of Particulate Matter*, chapter 97, pages 585-589. Springer Netherlands, Dordrecht.

Pun, V. C., Kazemiparkouhi, F., Manjourides, J., and 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.

Reardon, P. K., Seidlitz, J., Vandekar, S., Liu, S., Patel, R., Park, M. T., Alexander-Bloch, A., Clasen, L. S., Blumenthal, J. D., Lalonde, F. M., Giedd, J. N., Gur, R. C., Gur, R. E., Lerch, J. P., Chakravarty, M. M., Satterthwaite, T. D., Shinohara, R. T., and Raznahan, A. (2018). Normative brain size variation and brain shape diversity in humans. *Science*, 360(6394):1222-1227.

Rhomberg, L. R., Chandalia, J. K., Long, C. M., and Goodman, J. E. (2011). Measurement error in environmental epidemiology and the shape of exposure-response curves. *Critical Reviews in Toxicology*, 41(8):651-671.

Rich, D. Q., Freudenberger, R. S., Ohman-Strickland, P., Cho, Y., and Kipen, H. M. (2008). Right heart pressure increases after acute increases in ambient particulate concentration. *Environmental Health Perspectives*, 116(9):1167-1171.

Sarnat, S. E., Winquist, A., Schauer, J. J., Turner, J. R., & Sarnat, J. A. (2015). Fine Particulate Matter Components and Emergency Department Visits for Cardiovascular and Respiratory Diseases in the St. Louis, Missouri–Illinois, Metropolitan Area. *Environmental Health Perspectives*, 123(5), 437–444.

Sato, H., Suzuki, K. T., Sone, H., Yamano, Y., Kagawa, J., & Aoki, Y. (2003). DNA-adduct formation in lungs, nasal mucosa, and livers of rats exposed to urban roadside air in Kawasaki City, Japan. *Environmental Research*, 93(1), 36–44

Sivagangabalan, G., Spears, D., Masse, S., Urch, B., Brook, R. D., Silverman, F., Gold, D. R., Lukic, K. Z., Speck, M., Kusha, M., Farid, T., Poku, K., Shi, E., Floras, J., and Nanthakumar, K. (2011). The effect of air pollution on spatial dispersion of myocardial repolarization in healthy human volunteers. *Journal of the American College of Cardiology*, 57(2):198-206.

Stanier, C. O., Lee, S.-R. R., and HEI Health Review Committee (2014). Development and application of an aerosol screening model for size-resolved urban aerosols. *Research Report (Health Effects Institute)*, (179):3-79.

Strickland, M. J., Darrow, L. A., Klein, M., Flanders, W. D., Sarnat, J. A., Waller, L. A., Sarnat, S.E., Mulholland, J.A., Tolbert, P. E. (2010). Short-term Associations between Ambient Air Pollutants and Pediatric Asthma Emergency Department Visits. *American Journal of Respiratory and Critical Care Medicine*, 182(3), 307–316.

Talbott, E. O., Rager, J. R., Benson, S., Ann Brink, L., Bilonick, R. A., & Wu, C. (2014). A case-crossover analysis of the impact of PM2.5 on cardiovascular disease hospitalizations for selected CDC tracking states. *Environmental Research*, 134, 455–465.

Tallon, L. A., Manjourides, J., Pun, V. C., Mittleman, M. A., Kioumourtzoglou, M.-A., Coull, B., & Suh, H. (2017). Erectile dysfunction and exposure to ambient air pollution in a nationally representative cohort of older men. *Environmental Health*, 16(1).

Tonne, C., Elbaz, A., Beevers, S., & Singh-Manoux, A. (2014). Traffic-related Air Pollution in Relation to Cognitive Function in Older Adults. *Epidemiology*, 25(5), 674–681.

Trenga, C. A., Sullivan, J. H., Schildcrout, J. S., Shepherd, K. P., Shapiro, G. G., Liu, L.-J. S. J., Kaufman, J. D., and Koenig, J. Q. (2006). Effect of particulate air pollution on lung function in adult and pediatric subjects in a seattle panel study. *Chest*, 129(6):1614-1622.

Tyler, C. R., Zychowski, K. E., Sanchez, B. N., Rivero, V., Lucas, S., Herbert, G., Liu, J., Irshad, H., McDonald, J. D., Bleske, B. E., Campen, M. J. (2016). Surface area-dependence of gas-particle interactions influences pulmonary and neuroinflammatory outcomes. Particle and Fibre Toxicology, 13(1).

Urch, B., Speck, M., Corey, P., Wasserstein, D., Manno, M., Lukic, K. Z., Brook, J. R., Liu, L., Coull, B., Schwartz, J., Gold, D. R., and Silverman, F. (2010). Concentrated ambient fine particles and not ozone induce a systemic interleukin-6 response in humans. *Inhalation Toxicology*, 22(3):210-218.

U.S. EPA (U.S. Environmental Protection Agency). (2014). Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5, and Regional Haze, Research Triangle Park, NC.

U.S. EPA (U.S. Environmental Protection Agency). (2018a). Modeling Guidance for Demonstrating Air Quality Goals for Ozone, PM2.5, and Regional Haze, Research Triangle Park, NC.

U.S. EPA (U.S. Environmental Protection Agency). (2018b). Technical Support Document EPA's 2014 National Air Toxics Assessment. Research Triangle Park, NC.

van Rossem, L., Rifas-Shiman, S. L., Melly, S. J., Kloog, I., Luttmann-Gibson, H., Zanobetti, A., Coull, B.A., Schwartz, J.D., Mittleman, M.A., Oken, E., Gillman, M.W., Koutrakis, P., Gold, D. R. (2015). Prenatal Air Pollution Exposure and Newborn Blood Pressure. *Environmental Health Perspectives*, 123(4), 353–359.

Watt, M., Godden, D., Cherrie, J., and 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.

Wauters, A., Vicenzi, M., De Becker, B., Riga, J.-P., Esmaeilzadeh, F., Faoro, V., Vachiéry, J.-L., van de Borne, P., and Argacha, J.-F. (2015). At high cardiac output, diesel exhaust exposure increases pulmonary vascular resistance and decreases distensibility of pulmonary resistive vessels. *American Journal of Physiology-Heart and Circulatory Physiology*, 309(12):H2137-H2144.

Weichenthal, S., Hatzopoulou, M., & Goldberg, M. S. (2014). Exposure to traffic-related air pollution during physical activity and acute changes in blood pressure, autonomic and micro-vascular function in women: a cross-over study. *Particle and Fibre Toxicology*, 11(1).

Wesson, K., Fann, N., Morris, M., Fox, T., Hubbell, B. (2010). A multi–pollutant, risk–based approach to air quality management: Case study for Detroit. *Atmospheric Pollution Research*, 1(4):296–304.
Weuve, J. (2012). Exposure to Particulate Air Pollution and Cognitive Decline in Older Women. *Archives of Internal Medicine*, 172(3), 219.

Wilker, E. H., Ljungman, P. L., Rice, M. B., Kloog, I., Schwartz, J., Gold, D. R., Koutrakis, P., Vita, J.A., Mitchell, G.F., Vasan, R.S., Benjamin, E.J., Hamburg, N.M., Mittleman, M. A. (2014). Relation of Long-Term Exposure to Air Pollution to Brachial Artery Flow-Mediated Dilation and Reactive Hyperemia. *The American Journal of Cardiology*, 113(12), 2057–2063.

Yoshimura I. (1990). The effect of measurement error on the dose-response curve. *Environmental Health Perspectives*, 87, 173-8.

Appendix A

Individual Comments by CASAC Members on the EPA's
Integrated Science Assessment for Particulate Matter (External Review Draft - October 2018)

Dr. James Boylan

General Comment

I recommend that EPA reconvene the PM Review Panel. I believe that a PM Review Panel would provide the 7-member chartered CASAC with additional insight and expertise to allow for a more thorough and in-depth review of the relevant science and policy documents. My experience on the most recent SO_2 Review Panel has shown me the importance and value of having multiple independent experts (who are at the leading edge of research in their respective fields) thoroughly reviewing each chapter.

The proposed review schedule is very aggressive and allows for one draft of the ISA and one draft of the PA. Also, EPA is planning to incorporate the REA analysis into the PA. EPA should recognize the possibility that second drafts of these documents might be necessary after CASAC and the public review the first drafts. In addition, the REA should not be included as part of the PA. Instead, the REA should be a stand-alone document that is reviewed by CASAC and the public prior to the release of the first draft of the PA. This will allow scientific review of risk and exposure metrics prior to developing policy recommendations. This review should not be strictly tied to the proposed schedule since getting high quality documents is more important than meeting the statutory deadline.

Executive Summary

The Executive Summary is intended to provide a concise synopsis of the key findings and conclusions of the PM ISA for a broad range of audiences. Please comment on the clarity with which the Executive Summary communicates the key information from the PM ISA. Please provide recommendations on information that should be added or information that should be left for discussion in the subsequent chapters of the PM ISA.

The Executive Summary did a good job of communicating the key information from the PM ISA. I have no recommendations for information that should be added or deleted.

Chapter 1 (Integrated Synthesis)

Chapter 1 presents an integrated summary and the overall conclusions from the subsequent detailed chapters of the PM ISA and characterizes available scientific information on policy-relevant issues. Please comment on the usefulness and effectiveness of the summary presentation. Please provide recommendations on approaches that may improve the communication of key findings to varied audiences and the synthesis of available information across subject areas. What information should be added or is more appropriate to leave for discussion in the subsequent detailed chapters?
Chapter 1 provides a comprehensive overview of each chapter in the PM ISA and the policy-relevant issues. The summary tables in Section 1.7 (Tables 1-5, 1-6, and 1-7) are very useful for presenting the causality determinations in the current PM ISA and the previous PM ISA. I have no recommendations for information that should be added or deleted.

Chapter 2 (Sources, Atmospheric Chemistry, and Ambient Concentrations)

To what extent is the information presented in Chapter 2 regarding sources, chemistry, and measurement and modeling of ambient concentrations of PM clearly and accurately conveyed and appropriately characterized? Please comment on the extent to which available information on the spatial and temporal trends of ambient PM concentrations at various scales has been adequately and accurately described.

In general, Chapter 2 does a good job of presenting sources, chemistry, and measurements and modeling of ambient PM concentrations. The spatial and temporal trends of ambient PM concentrations have been accurately described.

Section 2.3 discusses primary sources of PM. Figures 2-2, 2-3, and 2-6 show the importance of various types of dust to total $PM_{2.5}$ and PM_{10} primary emissions based on the U.S. EPA 2014 National Emissions Inventory. However, when these emissions are used as inputs to CTMs, the modeled concentrations are significantly higher than the observed concentrations at the speciation monitors. The reason for the overprediction is that there is no adjustment for near-source removal due to small sub-grid scale turbulence and impaction on building and vegetative surfaces (Pouliot G., *et al*., Assessing the Anthropogenic Fugitive Dust Emission Inventory and Temporal Allocation Using an Updated Speciation of Particulate Matter, January 2012, DOI: 10.1007/978-94-007-1359-8_97). It is estimated that local source removal typically accounts for 75% of total removal of fine particulate matter nationally (T.G. Pace, "Methodology to Estimate the Transportable Fraction (TF) of Fugitive Dust Emissions for Regional and Urban Scale Air Quality Analyses", U.S. EPA, Research Triangle Park, NC, August 2005, https://www.nrc.gov/docs/ML1321/ML13213A386.pdf). This removal factor is defined as a "capture fraction" and varies by location. The amount that is not removed is defined as the "transportable fraction." A discussion of capture fraction and transportable fraction should be included in this chapter to help place the importance of dust emissions into proper perspective.

Section 2.4.1 discusses the difference between FRM and FEM monitors and describes the three most widely used FEMs. FRMs typically measure 24-hour integrated samples every third day. Short time resolution automated FEMs can measure hourly samples every day. In the past, FEMs typically measured higher $PM_{2.5}$ concentrations than FRMs; therefore, some states were reluctant to switch to FEMs. However, the new Teledyne optical spectrometer FEMs are much more reliable and more consistent and many states are now converting their FRMs to FEMs. In July of 2017, Georgia EPD ran two regulatory FEMs. Currently, Georgia EPD runs nine regulatory FEMs and will be running twelve regulatory FEMs by June of 2019. A similar trend is occurring across many parts of the country which will produce significantly more $PM_{2.5}$ data at hourly resolution.

Section 2.4.7 does a good job of documenting the scientific advances in CTMs.

Figures 2-13 through 2-16 for PM concentrations show 2013-2015 design values. Need to update figures with 2015-2017 design values.

Figure 2-14 shows the 98^{th} percentile 24-hour $PM_{2.5}$ concentrations for 2013-2015. The red monitor in southern Georgia appears to be Albany (13-095-0007). However, according to certified AQS data, the 24-hour 2013-2015 design value for Albany is 23 □g/m^3 (should be a blue dot, not red dot).

Figure 2-15 shows the 98th percentile PM_{10} concentrations for 2013-2015. There are no measurements shown in Georgia although Georgia has three PM_{10} monitors (13-089-0002, 13-121-0039, and 13-245-0091) with certified data in AQS from 2013-2015. The 98th percentile PM_{10} concentrations for all three PM_{10} monitors in Georgia are well below 75 □g/m^3 (blue dots).

It would be helpful to have a member on the PM Review Panel that is an expert in ambient PM measurements and satellite remote sensing AOD analysis. This would allow for a better understanding of sampling errors and biases associated with integrated and continuous ambient PM measurements and satellite data. This is important since this information will be used to characterize ambient concentrations in the REA document.

Chapter 3 (Exposure to Ambient Particulate Matter)

Chapter 3 describes scientific information on exposure to ambient PM and implications for epidemiologic studies. To what extent is the discussion on methodological considerations for exposure measurement and modeling clearly and accurately conveyed and appropriately characterized? Please comment on the extent to which the discussion regarding exposure assessment and the influence of exposure error on effect estimates in epidemiologic studies of the health effects of PM has been adequately and accurately described.

In general, Chapter 3 does a good job of describing the latest scientific information on exposure to ambient PM and implications for epidemiologic studies, methodological considerations for exposure measurement and modeling, and the influence of exposure error on effect estimates in epidemiologic studies.

Section 3.3.1.2 and Table 3-1 discuss personal monitoring and error characteristics. Some personal samplers perform better than others. Before using personal sampling data to estimate exposure, a detailed evaluation of the sampler performance compared to FEMs should be performed. In some cases, the data may be better suited for looking at gradients in $PM_{2.5}$ exposure rather than directly using the measured $PM_{2.5}$ concentrations.

Table 3-2 contains a comparison of models used for estimating exposure concentrations or exposure. Under the column for "Dispersion", there is an "X" for Chemistry. However, it is stated on page 3-28 "Dispersion models…typically have limited ability to model chemistry (if any)." A footnote should be added to the table to indicate that many dispersion models do not account for chemistry.

Table 3-3 contains statistic measures used for air quality model performance evaluations. While the four performance measures listed are commonly used, the table should also include normalized mean bias (NMB) and normalized mean error (NME) since these are also commonly used and will show percent differences in addition to absolute differences (MB and ME).

Section 3.3.2.4 discusses "Mechanistic Models". The first paragraph in this section discusses CTMs, but does not mention dispersion models which are also discussed in this section (3.3.2.4.2).

In Section 3.3.2.4.1, additional information should be added to describe how Eulerian CTMs work (e.g., grid structures, finite difference, solving ADE).

On page 3-27, it is stated "Differential bias may also be observed across regions in space. Many such biases can be corrected for using adjustment factors based on comparisons of simulation results with observational data." However, it should be noted that arbitrarily adjusting modeling results to match observations can lead to the model getting the right answer for the wrong reasons. This is important if emission sensitivities or source apportionment is being used to look at alternative levels of the standard since the model will not respond appropriately to emission controls.

The bottom of page 3-27 discusses the Lagrangian trajectory model (which does not have any chemistry) by Stanier et al. (2014). Typically, Lagrangian models are not classified as CTMs, but rather they are classified as dispersion models. This discussion should be moved into Section 3.3.2.4.2.

Section 3.3.2.4.2 seems to mostly focus on using dispersion models for near-road modeling of mobile sources. However, dispersion models are much more widely used for modeling industrial point sources. Additional focus should be added for this aspect.

Section 3.3.2.4.2 should be updated to include Lagrangian dispersion models. The difference between a Lagrangian puff model and a steady-state plume model should be added. Under the section on Lagrangian puff models, a discussion on CALPUFF (limited chemistry), SCIPUFF (no chemistry), and SCICHEM (full chemistry) should be added.

Table 3-4 and other parts of Section 3.3.2.4.2 discuss model performance of dispersion models. Most published dispersion model performance evaluations are associated with using the model for compliance assessments. In these cases, the model's ability to capture the high end of the concentration distribution is evaluated with Q-Q plots where the highest data point from the model is compared to the highest data point from the observations even if they occur at different locations, time of day, and/or season of the year. When the model is being used to support health studies, spatial and temporal accuracy is much more important compared with compliance assessments. Therefore, dispersion modeling results need to be evaluated against observations paired in time and space, especially if they are being used as inputs to an exposure model such as SHEDS, APEX, or EMI.

Pages 3-32 to 3-34 discuss fusion of CTM predictions with surface observation data. This section does not discuss EPA's recommended approach to States for estimating ozone and $PM_{2.5}$ concentrations at unmonitored locations contained in their "Draft Modeling Guidance for Demonstrating Attainment of Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze" (December 3, 2014) located at https://www3.epa.gov/scram001/guidance/guide/Draft_O3-PM-RH_Modeling_Guidance-2014.pdf and "Modeling Guidance for Demonstrating Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze" (November 29, 2018) located at https://www3.epa.gov/ttn/scram/guidance/guide/O3-PM-RH-Modeling_Guidance-2018.pdf. In these guidance documents, EPA discusses their "Modeled Attainment Test Software" (MATS, Abt Associates, Inc., 2014. Modeled Attainment Test Software: User's Manual. https://www3.epa.gov/ttn/scram/guidance/guide/MATS_2-6-1_manual.pdf) and the Software for Modeled Attainment Test-Community Edition (SMAT-CE) tool (https://www.epa.gov/scram/photochemical-modeling-tools). MATS and SMAT-CE will spatially interpolate data using the Voronoi Neighbor Averaging (VNA) technique and adjust the spatial fields based on model output gradients. These tools can be applied to annual $PM_{2.5}$ design values, daily $PM_{2.5}$ design values, or 24-hour $PM_{2.5}$ values.

The last paragraph on page 3-34 states "Hybrid approaches can involve merging CTMs with dispersion and/or LUR models, merging CTMs with observational data, or some combination therein." However, there are no references showing how CTMs can be merged with dispersion models. Below are two examples that could be referenced:

- EPA's 2014 National Air Toxics Assessment (NATA) Technical Support Document (August, 2018) which merged CMAQ and AERMOD modeling results to determine cancer risks for HAPs and diesel PM (https://www.epa.gov/national-air-toxics-assessment/2014-nata-technical-support-document).
- K. Wesson, *et al*. (2010), A multi-pollutant, risk-based approach to air quality management: Case study for Detroit, Atmospheric Pollution Research 1, 296-304. This study merged CMAQ and AERMOD modeling results to determine exposure to HAPs, ozone, and $PM_{2.5}$. Data from the air quality modeling was used as input into the environmental Benefits Mapping and Analysis Program (BenMAP) and the Human Exposure Model-3 (HEM-3) to assess how the control strategies affect human health.

I am in agreement with the following EPA conclusions in Chapter 3:

- New developments in PM exposure assessment methods have reduced bias and uncertainty in health effect estimates.
- High correlations of $PM_{2.5}$ with some gaseous copollutants necessitate evaluation of the impact of confounding on health effect estimates.
- There is typically more uncertainty for health effect estimates for exposure to $PM_{10–2.5}$ and UFP.

However, the following EPA conclusion is not completely accurate. EPA should add references and/or an explanation justifying this conclusion. At a minimum, appropriate caveats need to be added, such as a linear, no-threshold (LNT) C-R assumption:

- Exposure error tends to produce underestimations of health effects in epidemiologic studies of PM exposure, although bias in either direction can occur.

It would be helpful to have a member on the PM Review Panel that is an expert in dispersion modeling and photochemical grid modeling used in health effects analyses. This would allow for a better understanding of errors and biases associated with models that are used to characterize ambient concentrations in the REA document. In addition, it would be helpful to have a member on the PM Review Panel that is an expert in epidemiology to allow for a better understanding of the impact of exposure errors on epidemiologic study results.

Chapter 13 (Welfare Effects)

Please comment on the identification, evaluation and characterization of the available scientific evidence from studies of PM on non-ecological welfare effects of visibility impairment, climate, and materials and the application of information from these studies, as presented in Chapter 13, to inform causality determinations and uncertainty characterizations for these welfare outcomes.

The information presented in Chapter 3 supports a causal relationship between PM and visibility impairment, climate effects, and effects on materials.

On page 3-6, the haze index is presented as $dv = 10 * ln\ (b_{ext}/0.01\ km^{-1})$. However, the units for b_{ext} are typically Mm^{-1}. Therefore, the correct equation should be $dv = 10 * ln\ (b_{ext}/10\ Mm^{-1})$.

The original IMPROVE algorithm (Equation 13-6) and the modified original IMPROVE algorithm (Equation 13-7) are presented on page 13-10. These equations tend to underestimate the highest light scattering values and overestimate the lowest values at IMPROVE monitors throughout the U.S. To resolve these biases, a revised IMPROVE equation was developed (Pitchford et al., 2007) that divides PM components into small and large particle sizes with separate mass scattering efficiencies and hygroscopic growth functions for each size. The revised IMPROVE equation was described in detail in the 2009 PM ISA, and it both reduced bias at the lowest and highest scattering values and improved the accuracy of the reconstructed b_{ext}. However, the revised IMPROVE equation is not presented in this document:

b_{ext} = 2.2 x fs(RH) x [Small Sulfate] + 4.8 x fL(RH) x [Large Sulfate]
+ 2.4 x fs(RH) x [Small Nitrate] + 5.1 x fL(RH) x [Large Nitrate]
+ 2.8 x [Small Organic Mass] + 6.1 x [Large Organic Mass]
+ 10 x [Elemental Carbon]
+ 1 x [Fine Soil]
+ 1.7 x fss(RH) x [Sea Salt]
+ 0.6 x [Coarse Mass]
+ Rayleigh Scattering (site specific)
+ 0.33 x [NO2 (ppb)]

This equation should be added to the document and discussed in detail. Use of this equation is recommended by EPA in their "Modeling Guidance for Demonstrating Air Quality Goals for Ozone, $PM_{2.5}$, and Regional Haze" (November 29, 2018), pages 146-148.

Figures 13-1 through 13-14 should use the same colors to represent the same species in the stacked bar charts when comparing across years (2005-2008 vs. 2011-2014). The figures that show 2011-2014 data use light blue for POM while the figures that show 2005-2008 data use green for POM. In addition, Figures 13-2, 13-4, 13-6, 13-8, 13-10, 13-12, 13-13, and 13-14 are very low resolution and very difficult to read.

Dr. Tony Cox

Comments on Preface, Executive Summary, and Chapter 1 for PM Draft ISA

Overall Comments and Recommended Additions to the ISA

Scientific statements make testable, and potentially falsifiable, predictions. Policy statements make recommendations about what to do next and how to proceed in doing it. The current Draft ISA states major conclusions in terms of causal determination categories that make no clear testable or falsifiable predictions. No tests for the predictive validity of the causal determinations are proposed and no results of validation tests are reported. In this sense, the causal determinations in the Draft ISA are not scientific statements. Yet, they are interpreted as having clear implications for further evaluations leading up to regulatory policies and actions. This process risks conflating science and policy judgments.

To bolster the role of objective science in the ISA and to enable it to better inform the subsequent REA and PA, the following changes should be made:

- Define key terms.
    - The term "causal" as used in the Draft ISA (as in the phrase "Likely to be a causal relationship") is ambiguous. For risk analysis and risk management policy purposes, the most relevant definition is that an exposure is a cause of a harm (e.g., an adverse health effect or loss of welfare) if and only if reducing the exposure would reduce risk of the harm. The ISA should therefore specifically address harms and risks that can be reduced or prevented by reducing exposures.
    - Throughout the ISA, use modern epidemiological concepts, terms, definitions, and methods to describe causal impacts of PM on human health more precisely than in the Draft ISA.
    - Revise definitions of causal determination categories for clarity, correctness, and consistency.
- Select evidence to evaluate using explicitly stated criteria and independently reproducible methods or rules for applying them to individual studies.
    - Include relevant high-quality studies of observed changes in health effects caused by changes in PM, specifically including accountability studies, natural experiments, intervention studies, and other quasi-experiments.
    - Provide explicit, objective, independently verifiable criteria for how individual studies and evidence are to be selected, evaluated, combined or synthesized, resolved when they conflict, and summarized in the ISA.
- Evaluate and interpret evidence using explicitly stated, independently verifiable criteria and methods to reach the ISA's conclusions.
    - Derive all conclusions via explicit, independently verifiable derivations using stated criteria and methods from stated premises (facts, data, and assumptions)
    - Validation: State the testable predictions implied by the conclusions and assumptions. Discuss the extent to which these testable implications have been tested and verified.
    - Interpretation: Discuss the extent to which alternative explanations and interpretations of the same facts and data are supported or refuted by available data.

- State and use criteria for assessing the *validity of evidence* accepted for use in informing causal determinations.

These recommendations are discussed in more detail next.

Address harm that can be prevented by reducing exposures.

The Draft ISA seeks to characterize harms to human health and welfare that are caused by PM exposures, but the meaning of "caused by" is left unspecified. During the public meeting on the PM Draft ISA in December, 2018, the following interpretations of "causal relationship" were discussed:

- *Explanatory causation.* CASAC's deliberation touched on an explanatory view of causation, articulated in part by Dr. Frampton: that it implies that "one thing leads to another," and that exposure has an (explanatory) causal relationship to harm if exposure leads to, or could lead to (and hence help to explain the occurrence of) harm in some cases. In this view, PM exposure might appropriately be said to have a "causal relationship" to harm even if it is neither necessary nor sufficient for the harm to occur and even if it does not change the probability or risk of harm. (For example, if PM exposure sometimes triggers an adverse health response in exposed individuals, but only under conditions where the same response would have been triggered by something else, e.g., copollutants, in the absence of the PM exposure, then PM might have an explanatory causal relationship to the harm without changing the frequency or severity of harm.)
- *Predictive causation.* Likewise, CASAC's discussion mentioned predictive (e.g., Granger) causation: that observed changes in exposure help to predict and explain subsequent changes in effects (whether or not reducing exposure would reduce the effects).
- *Similarity-based causation.* In public comments, Professor Corwin Zigler suggested that studies should be treated as causal based on whether their designs and analyses are similar to, or seek to "approximate," those for idealized randomized studies. This point of view is further expressed in Dominici and Zigler (2017), as follows: "We argue that evidence of causality should be gauged by a critical evaluation of design decisions such as 1) what actions or exposure levels are being compared, 2) whether an adequate comparison group was constructed, and 3) how closely these design decisions approximate an idealized randomized study. We argue that air pollution **studies that are more scientifically rigorous in terms of the decisions made to approximate a randomized experiment are more likely to provide evidence of causality** and should be prioritized among the body of evidence for regulatory review accordingly." (Dominici F, Zigler C. Best Practices for Gauging Evidence of Causality in Air Pollution Epidemiology. Am J Epidemiol. 2017 Dec 15;186(12):1303-1309. doi: 10.1093/aje/kwx307.) However, attempts to approximate randomized studies without actually doing randomization may be misguided: "**Seeking to approximate idealized randomized designs in observational studies is unsound**: observations are not actions, associations are not effects, and approximation is not randomization. Instead, modern causal analysis offers well-developed concepts, theory, and algorithms for estimating identifiable causal effects from data and applying them in new settings even without randomization." (Cox (2018) *American Journal of Epidemiology*, 187(6):1338–1339, https://doi.org/10.1093/aje/kwy034); see also Pfister N, Bühlmann P, Peters J (2018) Invariant causal prediction for sequential data. https://arxiv.org/pdf/1706.08058.pdf.) The opinion that "studies that are more scientifically rigorous in terms of the decisions made to approximate a randomized experiment are more likely to provide evidence of causality" lacks theoretical or empirical support. Even an idealized randomized study

that is used to address associational questions does not thereby provide evidence of causality (Pearl J, (2009) Causal inference in statistics: An overview. Statistics Surveys 3: 96-146). In practice, attempts to approximate idealized random experiments without actually doing randomization are subject to errors and biases that can make them incapable of providing reliable information about causality. For example, Wang et al. (2016), cited by the Draft ISA in Figure 11-18 ("Associations between long-term exposure to PM2.5 and total (nonaccidental) mortality in recent North American cohorts") state that "We applied a variant of the **difference-in-differences approach, which serves to approximate random assignment of exposure across the population and hence estimate a causal effect**. Specifically, we estimated the association between long-term exposure to PM2.5 and mortality while controlling for geographical differences using dummy variables for each census tract in New Jersey, a state-wide time trend using dummy variables for each year from 2004 to 2009, and **mean summer and winter temperatures for each tract in each year**." One obvious problem with this approach is that using mean summer and winter temperatures leaves unaddressed substantial potential (and likely) residual confounding by temperature (e.g., by daily temperature extremes over the two weeks preceding death), thereby invalidating any straightforward causal interpretation of observed PM2.5-mortality associations. In addition, the difference-in-differences approach is notoriously unreliable, e.g., producing up to 45% false positives (misidentifying an "effect" as significant at the 5 percent level for up to 45 percent of placebo interventions with no true causal impact, in Monte Carlo experiments) (Bertrand M, Duflo E, Mullainathan S (2004). How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics 119(1): 249-275. https://doi.org/10.1162/003355304772839588).

- *Labeling-based causation.* Professor Zigler's public comments also suggested that "causality" is largely a matter of labeling intended to connote methodological rigor. This point of view is expressed in Dominici and Zigler (2017), as follows: "The contentious political climate surrounding air pollution regulations has brought some researchers and policy-makers to argue that evidence of causality is necessary before implementing more stringent regulations**. Recently, investigators in an increasing number of air pollution studies have purported to have used 'causal analysis,' generating the impression that studies not explicitly labeled as such are merely 'associational' and therefore less rigorous."** (Dominici F, Zigler C. Best Practices for Gauging Evidence of Causality in Air Pollution Epidemiology. Am J Epidemiol. 2017 Dec 15;186(12):1303-1309. doi: 10.1093/aje/kwx307.) However, the very real, important, and fundamental distinction between associational and causal methods is not at all merely a matter of labeling: "Concern about 'a false message that studies using causal inference methods should always be considered more credible than studies using more traditional statistical approaches' is puzzling: **causal and associational studies address different questions, not the same questions with different degrees of credibility.** Associational methods such as regression infer conditional distributions of responses given *observed* values of explanatory variables. Causal methods analyze how response distributions change if explanatory variables are *changed* or *set* to new levels. This distinction is fundamental" (Cox (2018) *American Journal of Epidemiology*, 187(6):1338–1339, https://doi.org/10.1093/aje/kwy034). This fundamental distinction is well known in other areas of epidemiology and statistics (e.g., Pearl J, (2009) Causal inference in statistics: An overview. Statistics Surveys 3: 96-146).
- *Associational causation.* EPA's presentation referred to a weight-of-evidence (WoE) approach and to WoE considerations similar to the Bradford Hill considerations for determining whether to call associations causal. These considerations do not define what the label "causal" means, however, nor assign to it any specific practical implications for protecting or improving health by reducing exposures (e.g., that reducing exposures would improve or protect health).

- *Manipulative causation.* CASAC's discussion also mentioned the mainstream view from outside the air pollution area, that a study is "causal" if and only if it is designed and analyzed to address causal questions, such as the manipulative causal questions of whether and how changes in exposure would change harms, holding specified other variables fixed at specified levels (e.g., Pearl J, (2009) Causal inference in statistics: An overview. Statistics Surveys 3: 96-146; 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).

While these and perhaps other possible interpretations and meaning of "causal relationship" are all consistent with the ambiguous language in the Draft ISA, it is striking that most of them have no necessary implications for protecting public health. Explanatory, predictive, similarity-based, labeling-based, and associational causation do not imply that reducing exposure would reduce the harm said to be caused by exposure. Hence they leave risk managers uninformed about what actions are needed to reduce harm, and thereby protect public health and welfare.

The final ISA should clearly define the intended meaning(s) of its key terms, including "causal" and "causal relationship." To best inform policy deliberations and decisions, it should specifically characterize the types of human health harm *preventable* or reducible by reducing PM exposures (manipulative causation), since this is ultimately what policy makers need to be informed about in deciding whether further reductions in PM are needed to protect human health with an adequate margin of safety. Questions about preventable harm that the ISA should address to support scientifically well-informed decision-making include the following:

a. *Are the sizes of human health effects (e.g., risk reductions) that could be caused by further reductions in exposure de minimis, or are they large enough to warrant further characterization?* The ISA should address which preventable harmful effects of PM exposure are large enough, as well as certain enough, to warrant further consideration in this review cycle. Although the Draft ISA appropriately notes that "In the current review, quantitative assessments for health-related exposures and risks, if warranted, would be presented in the Health Risk and Exposure Assessment (HREA)" and that "In the current review, quantitative assessments for welfare effects, if warranted, would be presented in the Welfare Risk and Exposure Assessment (WREA)," the ISA should indicate *which adverse effects might be reduced enough by reducing PM exposures to materially affect human welfare or protection of human health*. "Materially" here means that the effect could be large enough to change a decision about whether a reduction in PM is needed to protect human health or welfare with an adequate margin of safety. Where the answers are unknown or depend on further information that is not yet available but that is expected soon, or that could be obtained by further research in time to inform this review cycle, those conditions should be noted. To take two unrealistically extreme cases as examples to clarify this point, a finding that reducing current ambient PM exposure by 1% would eliminate all lung cancer in the US would clearly make this a very desirable change and one well worth evaluating further. Conversely, a finding that eliminating all remaining PM exposure would increase expected cancer-free life-years by less than one trillionth of a life-second in the entire population over the next century would show that this change does not materially affect protection of human health. The ISA should provide at least enough information to clearly distinguish among such extreme cases, and among others that are less extreme. It should indicate whether current knowledge is sufficient to conclude with high confidence that reducing current ambient PM levels would materially affect human health and welfare, for each adverse effect

considered; or, conversely, whether current knowledge is sufficient to conclude with high confidence that reductions in current ambient PM levels would not materially affect each human health and welfare adverse effect considered. Rough order-of-magnitude estimates and uncertainty ranges for the amount of harm (e.g., fractions or numbers of mortalities or morbidities per year in the US population or in identified subpopulations) preventable by reducing PM exposure are adequate for this purpose. There is no need to anticipate or duplicate the more refined quantitative assessments that belong in the REA or HREA.

It is not clear whether or to what extent the current causal determination framework is intended to reflect any information about effect sizes, as opposed to whether a non-zero effect exists. It is therefore important to explicitly address what is known about rough (e.g., order-of-magnitude) effect sizes in the ISA. If needed, the current causal determination framework should be extended to address the approximate sizes of effects that could be caused by reducing exposure – and specifically whether they could materially affect human welfare and health – as well as indicating the degree of certainty that some effect exists.

b. *What else* materially affects health responses to changes in ambient PM levels? The Draft ISA discusses sociodemographic, meteorological, genetic, co-morbidity, and co-pollutant factors in several places, notably in Chapter 12 and parts of Chapter 5-11. How sensitive to the levels of these other factors is the approximate reduction in each type of harm that would be caused by reducing PM? Again, for the ISA, rough order-of-magnitude information is sufficient. The goal is to identify factors or combinations of factors that affect whether a given reduction in PM is necessary – and also sufficient, with high confidence – to achieve a given reduction in risks of adverse effects. As an extreme example, suppose that

*Risk = Susceptibility*Exposure,*

where "*Susceptibility*" might depend on many other factors. If improvements in these other factors (e.g., reductions of copollutants and comorbidities) were to reduce *Susceptibility* to 0, then no reduction in *Exposure* would be necessary to protect public health. Conversely, if *Susceptibility* were to increase, then perhaps levels of *Exposure* that are sufficient to protect public health now would no longer be sufficient. In the real word, are there factors or combinations of factors that modify the causal relationship between reductions in PM and resulting reductions in risks by enough to be important for informing decisions about what changes in PM, if any, are needed to protect human health an welfare? What are the approximate sizes of direct, indirect, and total effects of changes in PM exposure on changes in health risks for relevant values of other causally relevant factors? How heterogeneous are resulting risk reduction across individuals and sub-populations?

c. *What changes are expected* over time in the other factors on which the preventable harm from PM materially depends (e.g., sociodemographic, meteorological, and copollutant factors)? The Draft ISA provides useful information on several trends already. Which ones, if any, are expected to change by enough so that the changes materially affect the changes in PM required to protect human health and welfare? How, if at all, do the answers to these questions about future changes in other relevant variables depend on future changes in PM NAAQS? What are the direct, indirect, and total effects of these predicted changes in other factors on the approximate sizes of health and welfare changes from reducing PM exposures?

d. *How sure are we* at present about the answers to the preceding questions? How are they derived, from what data, using what models and assumptions? How well validated are the models,

assumptions, data, and calculations? What uncertainty bounds, intervals, distributions, or other uncertainty characterizations should be attached to identification of adverse effects that would be prevented by different reductions in PM, taking into account model uncertainty as well as data uncertainty?

EPA's presentation at the public meeting in December mentioned (bottom of p. 8) that "CASAC reviewed the Agency's causal framework **~13 times** by **~90** CASAC charter and ad hoc panel members in the process of reviewing ISAs from 2008-2015; its use was supported in all ISAs." This history and appeal to authority certainly inspire sympathy for EPA's attempt to implement a conceptual framework that previous CASACs have supported *en masse* over the past decade. However, such support does not address the remaining gaps in the conceptual framework, including its lack of any clear definition for key terms such as "causal relationship;" lack of testable and potentially falsifiable implications of these terms; lack of scientific validation for the correctness of the causal determination labels assigned to evidence (including reports on tests of internal and external validity of causal conclusions); and lack of operational procedures for determining objectively when each causal determination category label should be used or what it means. The high level of comfort expressed by former CASAC members with this state of affairs suggests that different expertise may be needed to fix these problems, and perhaps even to recognize why they matter in enabling a genuine science-based approach to regulation.
The current review cycle provides an opportunity to restore a practice of scientific information review, critical assessment, and synthesis that is primarily driven by independently verifiable facts and data and by clearly stated, tested, and validated assumptions and methods, rather than by comfort with consensus opinions about the assignment of ambiguous causal terms to selected conclusions from the literature.

Use modern epidemiological concepts, terms, and methods to describe causal impacts of PM on human health more precisely throughout the ISA.

Standard modern epidemiological causal terms and concepts should be used in place of, or in addition to, vague and ambiguous terms such as "causal relationship," "likely to be causal," and "concentration-response relationship." Crucially vague, ambiguous, undefined, or imprecisely defined terms in the Draft ISA include the following:

- "causal" (in phrases and classifications such as "likely to be causal" and "causal relationship")
- "causal relationship"
- "result in"
- "effect"
- "independent effect"
- "adverse effect"
- "concentration-response relationship"
- "the relationship" between exposure and response
- "evidence" and "scientific evidence" (as contrasted with unverified assumptions, models, opinions, or judgments)
- The five causal determination categories used throughout the ISA.

Some specific questions about the meanings of the five causal determination categories are listed below. As currently described, e.g., in Table P-2 of the ISA, these categories appear to be overlapping,

ambiguous, and contradictory in some places. Their use risks obscuring, rather than accurately communicating, policy-relevant scientific information. This should be fixed in the final ISA by clearly defining key terms used related to causality, effects, and risks.

To illustrate the type of policy-relevant confusion that use of such vague terms generates, consider the following two simplified hypothetical cases:

- Case A: Exposure to a pollutant has zero direct effect on health (i.e., changing just the pollutant level while holding other explanatory variables fixed has no effect on health); but it has a large total effect mediated by a copollutant (i.e., changing the pollutant level causes the level of the copollutant to change, which directly affects health).
- Case B: Exposure to a pollutant has a large direct effect on health but no positive total effect (e.g., because people stay indoors on smoggy days or otherwise avoid exposures that threaten health; or because increases in the pollutant create offsetting reductions in another pollutant).

Should case A or case B or both or neither be categorized as "Causal"? At present, the answer is unclear. Moreover, whatever the answer is, policy makers cannot ascertain from use of the five causal categories whether or in which directions alternative policy choices that affect the pollutant level and copollutant level differently will affect public health. Thus, *the vague term "Causal" fails to convey the essential information needed to inform decision-making.*

If the above-listed terms continue to be used, quantitative definitions should be specified for the boundaries of each descriptive category. For example, consider the simple hypothetical causal model described by the two structural (causal) equations

$$Risk = w*Poverty + (1 - w)*Exposure$$

$$Exposure = 1*Poverty$$

where w is a number between 0 and 1. These two structural equations imply the empirically observed (reduced-form) C-R relationship

$$Risk = Exposure,$$

The definition of the "Causal" category in Table P-2 of the ISA should be clear enough to allow users to independently determine the largest value of w that warrants a determination that the C-R relationship $Risk = Exposure$ is "Causal." (If there is no such value, then that should be made clear.) Similarly, if the probability that the relationship is causal is known to be p and there are no other relevant considerations to complicate selection of a label, then the criteria for causal determination categories should be clear enough so that users can independently determine the range of probability values p for which each causal determination category label, such as "Likely to be a causal relationship," applies. More generally, the ISA should add a section or glossary that clearly defines all key terms used to communicate policy-relevant information. The definitions should be clear enough so that different users of these terms can independently determine the same answer for which terms describe any given, precisely described situation.

Preferably, however, the final ISA should replace vague and undefined terms with standard, well-defined epidemiological concepts and terminology to communicate more precisely what is currently known about the extent to which, and the conditions under which, reducing PM exposures materially reduces associated risks of adverse health and welfare effects. Terms and concepts that the ISA should adopt, where appropriate, to communicate more precisely about risk, effects, and causality include the following:

- "Controlled direct effect"
- "Natural direct effect"
- "Interventional direct effect"
- "Natural indirect effect"
- "Indirect effect mediated by a specific variable"
- "Total effect"

These terms can be illustrated using the following simple structural equation model (SEM):

$$Risk = 0.4*Poverty + 0.6*Exposure + 2*Exposure*Poverty$$

$$Exposure = 1*Poverty,$$

corresponding to the following causal graph diagram:

Standard terms and concepts include the following.

- The *controlled direct effect* of a specified change in *Exposure* on *Risk* is the change in *Risk* if *Exposure* is changed (from a specified initial value to a specified final value) while *holding other explanatory variables fixed* at specified levels. For example, holding *Poverty* fixed at 0, the controlled direct effect of exposure on risk is a decrease of 0.6 units of *Risk* per unit decrease in *Exposure*. (More generally, the controlled direct effect of *Exposure* on *Risk* can be calculated by integrating $\partial Risk/\partial Exposure$ from the initial to the final value of *Exposure*, holding other variables fixed at specified levels.)
- The controlled direct effect of *Exposure* on *Risk*, holding *Poverty* fixed at 1, is a decrease of 2.6 units of *Risk* per unit decrease in *Exposure*.
- In general, the controlled direct effect of *Exposure* on *Risk*, holding *Poverty* fixed is given in this example by 0.6 + 2**Poverty*. Thus, the controlled direct effect of *Exposure* on *Risk* depends on the level of *Poverty*.
- The controlled direct effect of *Poverty* on *Risk* while holding *Exposure* fixed is 0.4 + 2**Exposure* units of risk reduced per unit reduction in *Poverty*.
- The *natural direct effect* (NDE) of *Exposure* on *Risk* in a population is found by calculating the controlled direct effect for each individual, holding the value of *Poverty* for that individual fixed at its initial value. The NDE cannot be calculated without strong assumptions unless individual-level data are available.

- The *interventional direct effect* (or stochastic direct effect) of *Exposure* on *Risk* in a population is the difference in average values of *Risk* in the population between the final and initial levels of *Exposure,* holding the value of *Poverty* for each individual fixed at a random value drawn from the conditional frequency distribution of *Poverty* values in the population for that individual, given any information about individual exposure and covariates. This can often be calculated even when the NDE cannot.
- The *natural indirect effect* of a change in *Poverty* on *Risk* is the change in *Risk* from its value when *Exposure* is at its level for the initial value of *Poverty* to when *Exposure* is at its level for the final value of *Poverty*, holding *Poverty* fixed at its final value.
- The *indirect effect* of *Poverty* on *Risk* mediated by *Exposure* is 0.6 + 2**Exposure* units of risk reduced per unit reduction in *Poverty*.
- The *total effect* of *Poverty* on *Risk* (meaning the change in *Risk* if *Poverty* is changed and other explanatory variables – here, *Exposure* – respond to that change) is 0.4 + 0.6 + 2**Exposure* = 1 + 2**Exposure* units of risk reduced per unit reduction in *Poverty*.

Table 1. Mediation Estimand Definitions, Descriptions, and Assumptions

Estimand	Description	Identifying Assumptions in Addition to Positivity and Consistency
Controlled direct effect $E(Y_{a,m}) - E(Y_{a^*,m})$	Difference in the expected value of Y setting A to a versus a^* and in both cases setting M to m	1. No unmeasured confounding between A and Y $(A \perp Y_{a,m} \mid W)$. 2. No unmeasured confounding between M and Y $(M \perp Y_{a,m} \mid W, A)$.
Natural direct effect $E(Y_{a,M_{a^*}}) - E(Y_{a^*,M_{a^*}})$ Natural indirect effect $E(Y_{a,M_a}) - E(Y_{a,M_{a^*}})$	Difference in the expected value of Y setting A to a versus a^* and in both cases letting M be the value that it would naturally be under a^* Difference in the expected value of Y in both cases setting A to a and contrasting M under a versus a^*	1. No unmeasured confounding between A and Y $(A \perp Y_{a,m} \mid W)$. 2. No unmeasured confounding between M and Y $(M \perp Y_{a,m} \mid W, A)$. 3. No unmeasured confounding of $A - M$ $(A \perp M_a \mid W)$. 4. No measured or unmeasured posttreatment confounding of the $M - Y$ relationship $(M_{a^*} \perp Y_{a,m} \mid W)$. 5. Y_a is equivalent to Y_{a,M_a}.
Stochastic direct effect $E(Y_{a,G_{M\mid a^*,W}}) - E(Y_{a^*,G_{M\mid a^*,W}})$ Stochastic indirect effect $E(Y_{a,G_{M\mid a,W}}) - E(Y_{a,G_{M\mid a^*,W}})$	Difference in the population average of Y setting A to a versus a^* and in both cases drawing the value of M from a distribution of M conditional on $A = a^*$ and the individual's set of covariate values, W Difference in the population average of Y in both cases setting A to a and contrasting drawing the value of M from a distribution of M conditional on $A = a$ versus $A = a^*$ and the individual's set of covariate values, W	1. No unmeasured confounding between A and Y $(A \perp Y_{a,m} \mid W)$. 2. No unmeasured confounding between M and Y $(M \perp Y_{a,m} \mid W, A)$. 3. No unmeasured confounding of $A - M$ $(A \perp M_a \mid W)$.

Abbreviations: A, treatment; M, mediator; W, covariates; Y, outcome.

Source: Rudolph KE, Goin DE, Paksarian D, Crowder R, Merikangas KR, Stuart EA. Causal Mediation Analysis With Observational Data: Considerations and Illustration Examining Mechanisms Linking Neighborhood Poverty to Adolescent Substance Use. Am J Epidemiol. 2018 Dec 18. doi: 10.1093/aje/kwy248.

For regulatory risk assessment, the most essential distinction is that between *direct effects* of changes in exposure on changes in risk (with exposure being changed from a stated initial level to a stated final level and all other predictors being held fixed at specified levels) and *total effects*, in which covariates (such as copollutants and comorbidities) adjust realistically in response to changes in exposure.

These terms are often defined and applied specifically in the context of causal mediation analysis. Their usage in this context is illustrated in Table 1, taken from a recent article surveying current options for carrying out such analyses. In the preceding example, *Exposure* is a mediator of the effects of *Poverty* on *Risk*. The controlled direct effect of *Poverty* on *Risk* is found by holding *Exposure* fixed at a specified level and quantifying the change in *Risk* for a given change in *Poverty*. Concern about new methods and concepts not yet being vetted adequately for use in a ISA is perhaps better justified for parts of causal mediation analysis than for the rest of causal analysis, as causal mediation analysis is still being rapidly refined and improved. For current practical use in regulatory risk assessment to support policy analysis, the most useful concepts are probably controlled direct effects and interventional direct effects. For example, Naimi et al. (2014) argue that "The debate on the relevance of natural direct and indirect effects rests on whether one takes as a target of inference the mathematical object per se, or the change in the world that the mathematical object represents. We further note that **public health questions may be better served by estimating controlled direct effects**." (Naimi AI, Kaufman JS, MacLehose RF. Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. Int J Epidemiol. 2014 Oct;43(5):1656-61. doi: 10.1093/ije/dyu107).

Terminology aside, causal analysis addresses the following crucial scientific questions: *What do adverse effects depend on? How does the dependency of an adverse effect probability on exposure to PM vary with the levels of other variables on which it also depends (e.g., age, sex, income, smoking, temperature, humidity, comorbidities, etc.)?* This is the core scientific information needed to support well-informed deliberations and policy-making.

Appropriate use of concepts and terms from modern (post-1980) causal analysis and epidemiology removes the ambiguity from terms such as "effect" and "causal relationship" by specifying *which type* of effect or causal relationship (e.g., controlled direct, natural direct, indirect mediated by copollutants, total, etc.) is intended. This is essential information for accurately communicating scientific information about health effects of exposure reductions to policy makers and the public. For example, if reducing PM has a large beneficial *total* effect on reducing cardiovascular mortality risk among elderly people living in poverty, but has a negligible natural *direct* effect on this group (perhaps because of mediation by other pollutants), then *accurate risk communication about "the effect" on human health of a proposed reduction in PM requires clarity about which effect – the large total effect or the negligible direct effect – is being communicated.*

In addition to using standard epidemiological terms and definitions, the ISA should rely upon technically sound methods of data analysis appropriate for estimating the specific causal effects of interest. As noted by Richiardi et al. (2013),

> "In epidemiological studies it is often necessary to disentangle the pathways that link an exposure to an outcome. Typically the aim is **to identify the total effect of the exposure on the outcome, the effect of the exposure that acts through a given set of mediators of interest (indirect effect) and the effect of the exposure unexplained by those same mediators (direct effect)**. **The traditional approach to mediation analysis is based on adjusting for the mediator in standard regression models to estimate the direct effect**. However, several methodological papers have shown that under a number of circumstances **this traditional approach may produce flawed conclusions**. Through a better understanding of the causal structure of the variables involved in the analysis, with a formal definition of direct and indirect

effects in a counterfactual framework, **alternative analytical methods have been introduced to improve the validity and interpretation of mediation analysis**." (Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013 Oct;42(5):1511-9. doi: 10.1093/ije/dyt127. Emphases added.)

The ISA should critically evaluate the methods used in individual studies and should avoid accepting flawed conclusions while accepting conclusions based on application of appropriate analytical methods to identify specific causal effects of interest. In carrying out its own assessments, EPA might consider augmenting existing exposure simulation models such as EPA's Stochastic Human Exposure and Dose Simulation [SHEDS] and Air Pollutants Exposure [APEX] models or the Exposure Model for Individuals [EMI] to carry out the necessary calculations for quantifying total and direct effects of changes in PM exposures on changes in human health.

Technical references on different types of causal effects and how to estimate them from data and stated assumptions include the following (emphases added):

- Albert JM, Cho JI, Liu Y, Nelson S. (2018) Generalized causal mediation and path analysis: Extensions and practical considerations. Stat Methods Med Res. Jan 1:962280218776483. doi: 10.1177/0962280218776483. This article is accompanied by a free software package in R: https://cran.r-project.org/web/packages/gmediation/vignettes/gmediate.html.
- Albert JM, Geng C, Nelson S. Causal mediation analysis with a latent mediator. Biom J. 2016 May;58(3):535-48. doi: 10.1002/bimj.201400124. The authors explain how to estimate natural direct and indirect effects in the presence of an unobserved mediator, and provide sensitivity analyses for untested modeling assumptions: "Health researchers are often interested in assessing the direct effect of a treatment or exposure on an outcome variable, as well as its indirect (or mediation) effect through an intermediate variable (or mediator). For an outcome following a nonlinear model, the mediation formula may be used to estimate causally interpretable mediation effects. This method, like others, assumes that the mediator is observed. However, as is common in structural equations modeling**, we may wish to consider a latent (unobserved) mediator**. We follow a potential outcomes framework and assume a generalized structural equations model (GSEM). **We provide maximum-likelihood estimation of GSEM parameters using an approximate Monte Carlo EM algorithm, coupled with a mediation formula approach to estimate natural direct and indirect effects**."
- Baron RM, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research – conceptual, strategic, and statistical considerations, Journal of Personality and Social Psychology, Vol. 51(6), pp. 1173–1182.
- Cho SH, Huang YT. Mediation analysis with causally ordered mediators using Cox proportional hazards model. Stat Med. 2018 Dec 18. doi: 10.1002/sim.8058. "**Causal mediation analysis aims to investigate the mechanism linking an exposure and an outcome**. However… existing multi-mediator analyses for survival outcomes are either performed under special model specifications such as probit models or additive hazard models, or they assume a rare outcome. …**We develop a methodology under a counterfactual framework to identify path-specific effects (PSEs) of the exposure on the outcome through the mediator(s)** [and] derive closed-form expressions for PSEs on survival probabilities."

- Moerkerke B, Loeys T, Vansteelandt S. Structural equation modeling versus marginal structural modeling for assessing mediation in the presence of posttreatment confounding. Psychol Methods. 2015 Jun;20(2):204-20. doi: 10.1037/a0036368.
- Moreno-Betancur M, Carlin JB. Understanding Interventional Effects: A More Natural Approach to Mediation Analysis? Epidemiology. 2018 Sep;29(5):614-617. doi: 10.1097/EDE.0000000000000866. The authors warn that "The causal mediation literature has mainly focused on **'natural effects' as measures of mediation**, but these **have been criticized for their reliance on empirically unverifiable assumptions**. They are also impossible to estimate without additional untestable assumptions in the common situation of exposure-induced mediator-outcome confounding. **'Interventional effects' have been proposed as alternative measures that overcome these limitations… In contrast with natural effects, which are defined in terms of individual-level interventions, the definitions of interventional effects rely on population-level interventions**. This distinction underpins the previously described advantages of interventional effects, and reflects a shift from individual effects to more tangible population-average effects."
- Naimi AI. Invited commentary: boundless science--putting natural direct and indirect effects in a clearer empirical context. Am J Epidemiol. 2015 Jul 15;182(2):109-14. doi: 10.1093/aje/kwv060. This comment notes that "even if one accepts them as relevant clinical or public health quantities, and even under ideal conditions of no selection, information, or confounding bias, **natural direct and indirect effects will always be compatible with a range of possible values** for a given data set, and thus will not be (point-) identifiable" and that "Though it is shrouded in some controversy, use of natural direct and indirect effects is becoming more common in epidemiology. Reporting bounds for these effects would do much to quell the controversy and place them in a clearer empirical context."
- Pearl J. (2001) Direct and Indirect Effects. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA: Morgan Kaufmann, 411-420.
- Petersen ML, Sinisi SE, van der Laan MJ. (2006) Estimation of direct causal effects. Epidemiology. May; 17(3):276-84.
- Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol. 2013 Oct;42(5):1511-9. doi: 10.1093/ije/dyt127.
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 1992, 3(2):143-155. doi:10.1097/00001648-199203000-00013.
- Sheng-Hsuan L, VanderWeele T (2017). Interventional Approach for Path-Specific Effects, Journal of Causal Inference 5(1): 1-10. "The effect mediated by a certain combination of mediators, i. e. path-specific effect (PSE), is not always identifiable without making strong assumptions. In this paper, the authors propose a method, defining a randomly interventional analogue of PSE (rPSE), as an alternative approach for mechanism investigation. **This method is valid under assumptions of no unmeasured confounding and allows settings with mediators dependent on each other, interaction, and mediator-outcome confounders which are affected by exposure**. In addition, under linearity and no-interaction, our method has the same form of traditional path analysis for PSE. Furthermore, under single mediator without a mediator-outcome confounder affected by exposure, it also has the same form of the results of causal mediation analysis."
- Sobel ME (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*. 13: 290–312. doi:10.2307/270723

- Steen J, Vansteelandt S. (2018) Graphical models for mediation analysis. https://arxiv.org/pdf/1801.06069.pdf. This excellent exposition notes that "**Unlike identification of total effects, adjustment for confounding is insufficient for identification of path-specific effects** because their magnitude is also determined by the extent to which individuals who experience large exposure effects on the mediator, tend to experience relatively small or large mediator effects on the outcome. This chapter therefore provides an accessible review **of identification strategies under general nonparametric structural equation models (with possibly unmeasured variables)**, which rule out certain such dependencies. In particular, it is shown **which path-specific effects can be identified** under such models, and how this can be done."
- 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).
- Vansteelandt S, Daniel RM. Interventional Effects for Mediation Analysis with Multiple Mediators. Epidemiology. 2017 Mar;28(2):258-265. doi: 10.1097/EDE.0000000000000596. This paper discusses estimation of policy-relevant interventional direct effects and interventional indirect effects: "**In contrast to natural (in)direct effects, interventional (in)direct effects are policy-relevant**: they are relevant about a policy that involves fixing the mediator distribution, or shifting it to the extent that it is affected by the exposure."

Revise definitions of causal determination categories for clarity, correctness, and consistency.

The definitions of the five causal determination categories should be revised and clarified to resolve the following questions. The category descriptions as currently given (e.g., in Table P-2) appear to have several ambiguities, confusions, and contradictions that should be resolved so that the categories have clear meanings and can be used by different people independently to mean the same thing for purposes of communicating findings.

1. *Causality and preventability.* Does the "Causal" determination for a PM exposure-response relationship imply that reducing exposure would reduce response?
    a. Conversely, might the "Causal" category be an appropriate designation for a PM concentration-response (C-R) association even if it is known that reducing PM would *not* change the probability distribution of the response in the exposed population? If so, under what conditions would this be appropriate?
    b. More specifically, does a "Causal" determination imply that reducing exposure alone (without changing anything else such as socioeconomic, co-morbidity, co-pollutant, or weather variables) would necessarily reduce any or all of following: prevalence, incidence, average annual frequency per 100,000 capita-years, or age-specific hazard functions for the response in exposed populations?
    c. Might the "Causal" category be an appropriate designation for a PM concentration-response (C-R) association even if it is not known whether reducing PM would change the response (or the probability distribution of the response) in the exposed population?

d. *Direct vs. total causation.* Does the determination that a PM exposure-response relationship is causal imply that reducing exposure would reduce response even if other causally relevant factors (e.g., socioeconomic, co-morbidity, co-pollutant, or weather variables) were held fixed at their current values?
e. More generally, what are the empirically testable implications or predictions of a "Causal" designation?

In response to this question, Dr. Vandenberg (Letter from Vandenberg to Cox dated February 20, 2019) noted that "The draft PM ISA…conveys the available evidence on the relationship between exposures and response, including studies that indicate the occurrence and extent of reduction in responses observed with reductions in PM exposures. Such studies, often termed accountability studies, are evaluated and discussed within the ISA…if they fit within the scope of the ISA as detailed in the Preface, i.e., included a composite measure of PM, such as PM2.5 mass." This response does not address or answer any of questions (a)-(e). The *definitional* issue here is what the term "Causal" is intended to mean. A crucial question that remains unanswered of whether a causal determination that exposure to PM is causally related to an adverse health response implies that reducing PM exposure would reduce the risk of the adverse health response. Citing "studies that indicate the occurrence and extent of reduction in responses observed with reductions in PM exposures" does not answer this question, since reductions in PM and in responses can and do occur for reasons other than causality (e.g., because both PM and cardiovascular risks are declining over time, even if there is no indication that either causes the other, as occurred in the Dublin study). In addition, the Draft ISA omits 14 of the 15 studies tabulated by the Health Effects Institute (Table 1, "Overview of accountability studies funded by HEI" in the letter from Dan Greenbaum of HEI to Aaron Yeow of EPA dated February 21, 2019), suggesting that its consideration of evidence from high-quality accountability studies is not comprehensive. If necessary, the ISA should broaden its scope to include such studies.

2. *Strength of causal relationships and sizes of effects.* Does the determination that a PM exposure-response relationship is causal imply that reducing exposure by, say, 10 $\mu g/m^3$, must reduce response by at least a certain positive amount? In other words, is there any lower limit to how small a change in health effects caused by a given reduction in exposure can be to make "Causal" the appropriate determination? (To use an extreme example, if eliminating exposure completely were to lengthen the life expectancy of just one person by one trillionth of a second, but had no effect on anyone else, would that suffice to designate the C-R relationship as "Causal" for the population? If not, is there a minimum size of effect that must be achieved for the "Causal" label to be appropriate?) Is the five-point categorization intended to convey any information about effect sizes or strength of association?

This question remains unanswered.

3. *Homogeneity of causal relationships.* Does the category "Causal relationship" mean the same thing as "Causal relationship for 100% of the members of the exposed population"?
    a. Conversely, does the category "Causal relationship" mean the same thing as "Causal relationship for at least one member of the exposed population"?

b. Can an exposure-response or C-R association be causal for part of an exposed population (e.g., men over 70 years old with COPD) without being causal for other parts of the population (e.g., healthy women under 30)?
c. If so, is there a minimum size or fraction of the population for whom the C-R relationship must be causal in order to imply that "Causal" is the correct designation for that relationship in the exposed population as a whole?
d. Might evidence of a C-R relationship be causal for some subpopulations (e.g., COPD patients) but only "Suggestive of, but not sufficient to infer, a causal relationship" for other subpopulations? Why or why not?
e. If different causal determination categories apply to different subpopulations, how should the causal determination category for the population as a whole be determined from the causal determination categories of its subpopulations?

In response to this question, Dr. Vandenberg (Letter from Vandenberg to Cox dated February 20, 2019) writes that "Chapter 12 [of the ISA] identified, evaluates, and summarizes the evidence of factors that influence inter-individual and inter-population differences in health responses from exposures to PM." This response does not address any of questions (a)-(e). The main definitional issue here is not whether there are other factors that also influence health responses, but whether a causal determination that an exposure-response pattern is a "Causal relationship" implies that it is causal for all members of a population. That question remains unanswered.

4. *Are causal determination categories mutually exclusive and collectively exhaustive?*
   a. Is it possible for evidence to be both "Suggestive of, but not sufficient to infer, a causal relationship" and also "lnadequate to infer the presence of absence of a causal relationship"? Why or why not? (At present, I do not see why both descriptions might not apply simultaneously, or why both might not also be compatible with the "likely to be causal" category.)
   b. More generally, what prevents a body of evidence from being correctly described by more than one of these categories? The descriptions given in the framework for causality determinations described in the Preamble to the ISAs (U.S. EPA, 2015) and in Table P-2 appear to allow for considerable overlap between some of the five categories. For example, the "Causal" category includes as examples "observational studies that cannot be explained by plausible alternatives or that are supported by other lines of evidence (e.g., animal studies or mode of action information). Generally, the determination is based on multiple high-quality studies conducted by multiple research groups." The "Likely to be causal" category includes this: "animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited or no human data are available. Generally, the determination is based on multiple high-quality studies." Now, suppose that a body of evidence consists of
      i. Observational studies in humans that can plausibly be explained by plausible alternatives such as an unmeasured confounder or coincident historical trends; and
      ii. Supporting animal toxicological evidence from multiple studies from different laboratories that demonstrate species-specific effects based on multiple high-quality studies conducted by multiple research groups.

   Which category applies in this case? On the one hand, the evidence satisfies the description "observational studies that… are supported by other lines of evidence (e.g., animal studies or

mode of action information). Generally, the determination is based on multiple high-quality studies conducted by multiple research groups." That would indicate that it belongs to the "Causal" category. On the other hand, it also satisfies the description "animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited… human data are available. Generally, the determination is based on multiple high-quality studies." Thus, it seems it should also belong to the "Likely to be causal" category. To which category should such a body of evidence that is described by more than one be assigned, and on what basis?

c. In the example just given, the body of evidence consisted of observational data in humans that can plausibly be explained by alternatives such as an unmeasured confounder or coincident historical trends, together with animal data showing a species-specific response. Shouldn't this be categorized as "Inadequate to infer a causal relationship" rather than (or in addition to) being categorized as "Causal" and/or "Likely to be causal" as Table P-2 seems to require?
d. Two of the causal determination categories are "Likely to be a causal relationship" and "Not likely to be a causal relationship." Why doesn't at least one of these labels apply to each body of evidence?
e. The description for "Not likely to be a causal relationship" says "Evidence indicates there is no causal relationship with relevant pollutant exposures." What evidence would indicate that there is no causal relationship (rather than that there is no detected causal relationship)? This seems to require proving a negative.
f. Suppose that an initial body of evidence consists of animal toxicological evidence from a relatively few (but multiple) high-quality studies from different laboratories that demonstrate effects, but that no human data are available. This matches one of the cases described as "Likely to be a causal relationship" in Table P-2. Now suppose that two further supporting studies are added: a high-quality epidemiologic study that shows an association with a given health outcome; and a high-quality toxicological study that shows effects relevant to humans in an animal species. The evidence now matches one of the cases described as "Suggestive of, but not sufficient to infer, a causal relationship" in Table P-2. Should the addition of these two new supportive studies result in a downgrade of the evidence from its previous label of "Likely to be a causal relationship" to a new label of "Suggestive of, but not sufficient to infer, a causal relationship," to match the classification of case descriptions in Table P-2? Why or why not?

In response to this question, Dr. Vandenberg provided the following clear, responsive answer: "In regard to your questions 'Are causal determination categories mutually exclusive and collectively exhaustive", the simple answer is no." This answer implies that there is no objective way, even in principle, to assign a unique causal determination category to a body of evidence. Rather, participants in the labeling process for assigning causal determination labels to evidence must choose which label to apply when more than one (or none) is applicable, making the final result depend on something other than the evidence itself. This risks conflating non-fact-based opinions with scientific evidence in the causal determination process.

5. *Operational definitions of "adequacy" and "sufficiency".*
    a. What are the defining (operationally testable and independently verifiable) conditions that make evidence "sufficient to infer a causal relationship"? What is the operational definition of this category?
    b. Are there also defining conditions that make evidence "insufficient to infer a causal relationship?" If so, what are they?
    c. What are the defining conditions that make evidence "lnadequate to infer the presence or absence of a causal relationship"? What is the operational definition of this category?
    d. Are there also defining conditions that make evidence sufficient to infer absence of a causal relationship? If so, what are they?

In response to this question, Dr. Vandenberg's letter of February 20, 2019 replies that "The defining conditions are as indicated in the Preamble [and] Preface [in] Table P-2." However, Table P-2 simply says that evidence is "inadequate to infer a causal relationship" if "Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quantity, quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect." This is almost a circular definition: it provides no operational definition of what constitutes sufficient or insufficient "quantity, quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect." Without a clear operational definition of these terms, it is left up to participants to decide what evidence they will consider as "sufficient" or "insufficient" for establishing conclusions. There is no guarantee that consistent criteria will be applied for different studies, or for bodies of evidence that support different conclusions.

6. *Certainty of causal relationship category.* Does the category "Causal relationship" mean the same thing as "Causal relationship with 100% certainty, probability, or confidence"? If not, is there a threshold for certainty, probability, or confidence below which it would be inappropriate to call a relationship "causal"?

This question remains unanswered.

7. *Categorizing simple cases where all relevant information is known.* Consider an example in which it is known that that an observed C-R relationship is either causal (if there is no unmeasured confounder that explains it) or not (otherwise). There are no other relevant facts, considerations, or lines of evidence. The probability of such an unmeasured confounder has been bounded by data analysis of multiple past studies as being no greater than p, where p is a number between 0 and 1.
    a. For what values of p should the C-R relationship be categorized as "Causal"? Is there a smallest value of p (the probability that the relationship is not causal) that is required for the "Causal" label to be applicable?
    b. Similarly, suppose that different data analyses establish that the probability of an unmeasured confounder is no less than q, where q is a number between 0 and 1. For what values of q should the relationship *not* be considered causal?
    c. Are there values of p and q for which the causal determination category is ambiguous?
    d. If further research determines that the probability of an unmeasured confounder is in fact r (to two decimal places), where r is a number between 0 and 1, then for what values of r

is each of the five causal determination categories the correct description? For example, if $r = 0.5$, which causal determination category would be the correct one to use, and why, assuming that there are no other relevant uncertainties or facts?

In response to this question, Dr. Vandenberg's letter of February 20, 2019 states that "The EPA does not conduct hypothetical probability analyses in the ISA." However, the question does not ask for probability analyses; rather it asks for definition, e.g., "For what values of p should the C-R relationship be categorized as 'Causal'?". Answers do not require analysis, but conceptual clarity. These questions remain unanswered.

8. *Quantity of evidence needed for a causal determination.* Suppose that each of 10 independent studies (possibly including diverse types of evidence, e.g., epidemiological, toxicological, and clinical studies) concludes that the hypothesis of no causal relationship between C and R can be rejected with at least 95% statistical confidence. For simplicity, assume that this is the totality of the available evidence. (Thus, no studies have reached a different conclusion.) Would this constitute sufficient evidence to conclude that the C-R relationship should be classified as causal? Would 2 such studies be enough? In this simple setting, is there a minimum number of such studies that would be necessary and sufficient to warrant labeling the studied C-R relationship as "Causal" even though 100% certainty can never be achieved?

In response to this question, Dr. Vandenberg's letter of February 20, 2019 states that the EPA considers evidence spanning scientific disciplines using a well-established weight-of-evidence framework and that "The EPA does not count studies nor conduct hypothetical analyses in developing causality determinations." This does not answer the question of how, specifically, the weight-of-evidence framework would or should be applied in specific cases. For such applications to be based on articulated normative principles that can be independently applied by others to reach the same conclusions, it must be made clear what principles are used and how they would work in simple hypothetical cases. This has yet to be done. In addition, if "EPA does not… conduct hypothetical analyses in developing causality determinations," then the resulting determinations are unlikely to be valid, insofar as valid causal determination typically requires comparing differences in responses at different levels of exposures, at least some of which are counterfactual (i.e., hypothetical).

9. *Discordant evidence.* Suppose that 7 studies estimate a significant positive C-R relationship at the 95% confidence level (e.g., the 95% confidence interval for the relative risk is entirely to the right of 1), but another 3 studies estimate significant negative C-R relationships (95% confidence intervals entirely to the left of 1). Upon close scrutiny, all studies appear to have the same high quality and their conclusions appear to be equally sound and credible. If this were the only relevant evidence, then what conclusions about causal determination category, if any, should be drawn from such discordant evidence? If the numbers were changed (e.g., to 1000 studies reaching one conclusion and 2 reaching the opposite conclusion), how, if at all, should the resulting causal determination category change in response?

In response to this question, Dr. Vandenberg's letter of February 20, 2019 states that the EPA "considers the varying evidence" and "does not engage in hypothetical analyses." But the question is

about *how* varying evidence is (and should be) considered and used in making causal determinations. This question remains unanswered. To answer it in a principled way – one based on general rules that can be stated before considering any specific body of evidence and that can be independently applied so that different people can reach the same answers independently starting from the same evidence – it is necessary to state what the rules are with enough precision and clarity so that they can indeed be applied to hypothetical situations. (If the ways in which the rules are to be applied are decided only after looking at the particular evidence of interest, then they are not rules at all and no objective, independently verifiable procedure is being followed.) Simply "considering" evidence and then drawing conclusions in some unspecified way is not consistent with sound science or sound decision-making. For this reason, it is important to explain the principles used to handle realistically discordant evidence, and to explain them with enough clarity and generality so that they can be independently applied to a variety of bodies of evidence.

10. *Updating evidence categorizations.* Is it possible that a C-R relationship that is presently classified as causal might later be reclassified in light of additional evidence? Are there any restrictions on how likely this possibility must be in order for the "Causal" classification to be applied? For example, is a current designation of "Causal" for a relationship inconsistent with a judgment that there is a 90% probability that the relationship will be reclassified as "lnadequate to infer the presence of absence of a causal relationship" as soon as an accountability study now underway is concluded? What restrictions, if any, does a current designation of "Causal" imply for possibilities and probabilities of future reclassifications?

In response to this question, Dr. Vandenberg's letter of February 20, 2019 states that "The EPA does not engage in hypothetical analyses" such as these. Thus, these questions remain unanswered.

Several commentators, including some former members of CASAC, have voiced strong support for the continued use of the existing causal determination framework and categories in the Draft ISP without further changes. However, it is essential that definitions of key terms used to communicate major ISA findings should be clear enough so that all who use them understand what is being asserted when they are used – and, specifically, whether what is being asserted is that changes in NAAQS made to reduce PM exposures will materially reduce risk to human health or welfare with some degree of certainty. This essential clarity is presently missing. Communication of results must clearly differentiate between causal relationships that have no implications for whether reducing PM will reduce harm (e.g., associational, attributive, predictive, and unverified counterfactual/potential outcome causal relationships) and causal relationships that do have such implications (e.g., validated manipulative, structural, mechanistic, and but-for causal relationships). Frameworks and category definitions that do not inform policy makers about whether human health and welfare consequences would be materially changed by making different choices about NAAQS exposure levels (thereby providing manipulative causality information) fall short of communicating the crucial information required for scientifically well-informed decision-making.

Concerns expressed by some commentators, including some former members of CASAC, that modern methods of causal analysis have not been adequately vetted by the air pollution health effects community might be alleviated by noting that the cost of waiting for such vetting has been high (decades of effort spent documenting and estimating associations with no clear valid causal or policy implications, such as the original Dublin coal burning ban studies and many others in the US that

continue until this day) and that this community has generally eventually adopted mainstream methods that have been well developed and vetted years to decades earlier in more mature areas of science and applied statistics. As one example among many, the use of quasi-experimental (QE) designs and analyses for causal interpretations of observational data was introduced and well developed in social statistics in the 1960s (e.g., Campbell DT, Stanley JC (1963) *Experimental and Quasi-Experimental Designs for Research on Teaching*, in N. L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963). Half a century later, it was hailed as a very welcome and promising development for overcoming limitations of air pollution health effects studies as they were (and still are) currently being performed: "However, the path to the best available evidence about the benefits of reducing PM and other air pollutants lies in an increased focus on developing and using QE evidence," and QE techniques "provide an opportunity to improve understanding of the relation between human health and regulation of air pollution from particulates." (Dominici F, Greenstone M, Sunstein CR. Science and regulation. Particulate matter matters. Science. 2014 Apr 18;344(6181):257-9). These same insights could have been achieved decades earlier if the air pollution health effects research community had been swifter to adopt QE methods that had already been well vetted by more mature fields. Earlier adoption could have prevented time and effort spent on false policy-relevant causal conclusions drawn from the original Dublin coal burning ban studies and many others. Similarly, structural equation models for causal modeling have been extensively developed, vetted and applied in numerous disciplines since the 1950s. Path analysis is now a century old. A few recent air pollution health effects research for PM2.5 and other pollutants have started to take advantage of these and closely related ideas and methods (e.g., Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, Samet J, King G, Dominici F; HEI Health Review Committee. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. Res Rep Health Eff Inst. 2016 May;(187):5-49), but the adoption curve has been needlessly slow and is costly in terms of missed opportunities to produce more relevant and less ambiguous results on how changes in PM affect human health risks.

More generally, modern causal analysis methods have already been extensively developed, vetted, and applied in top health sciences and epidemiology journals for decades, as reflected in part in the references cited in these comments. Adopting these concepts, terms, and methods now can enable air pollution health effects researchers and policy makers to take advantage in the current NAAQS review cycle of the improved precision that they offer for explaining, quantifying, and communicating about health risks caused by air pollution and preventable by reducing exposures to air pollution. Policy makers and the public deserve such precision now in EPA's and CASAC's communications about health risks associated with air pollution.

Include and discuss more of the relevant literature on observed changes in health effects following changes in PM exposures, including well-conducted accountability studies and quasi-experiments.

The Draft ISA omits some relevant studies of observed health effects following changes in PM. The final ISA should include and discuss results of high-quality accountability studies, natural experiments, intervention studies, and causal mediation studies for PM health effects. Its conclusions should synthesize lessons learned from these studies. Specific studies that are not discussed in the Draft ISA but that appear to contain useful information – including some that appear to be discordant with conclusions in the Draft ISA – include the following (emphases added):

- Health Effects Institute (2013). Did the Irish Coal Bans Improve Air Quality and Health? HEI Update, Summer, 2013. This accountability study found that **substantial reductions in ambient particulate air pollution (by up to 70% and several dozen µg/m^3) in Ireland were not found to cause reductions in all-cause or cardiovascular mortality rates**, despite strong, consistent, coherent etc. historical associations between levels of PM in air and levels of all-cause and cardiovascular mortality due to coincident historical trends. See also Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, Kotlov T; HEI Health Review Committee. Effect of air pollution control on mortality and hospital admissions in Ireland. Res Rep Health Eff Inst. 2013 Jul;(176):3-109. (This article does not appear to correct for multiple testing bias.) Although the Draft ISA may have followed study selection criteria (per the guidance of previous CASACs) that exclude the Irish coal ban studies because they address black smoke (soot) air pollution, rather than PM2.5 *per se*, Ireland's EPA specifically stated that "An extension of the bituminous coal ban across Ireland would help in reducing the levels of PM2.5 in ambient air." Thus, estimated effects of the bans on subsequent mortality rate (mediated by reductions in PM/PM2.5) appear relevant. www.epa.ie/pubs/epasub/EPA%20Response%20on%20Smoky%20Coal%20Regs%20Consultation.pdf. The finding that large reductions in ambient PM air pollution did not cause detectable reductions in all-cause or cardiovascular mortality rates should be carefully discussed in the final ISA. It appears to be an unexpected finding, and hence potentially valuable for improving understanding and modeling of conditions under which PM reductions do or do not cause detectable changes in mortality rates.
- Health Effects Institute (2016). Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. See also Synopsis of Research Report 187: Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. This study found that "Contrary to expectations, their **analysis suggested a reduction, on average, in mortality even in areas where their analyses reported that PM10 was not causally affected**. The authors suggested that **the observed causal effect of nonattainment designation on mortality,** in the absence of a strong associative effect for PM10, **may be due to causal pathways other than the one involving reduction of PM10**. However, they suggested their results provide evidence that PM10 played a causal role in the reduction of hospitalization for respiratory disease, but again, not for cardiovascular disease. As the authors noted, **all of the estimates from these analyses were accompanied by substantial uncertainty, indicated by broad posterior 95% confidence intervals that included zero**. As a result, the HEI Health Review Committee thought the investigators generally overstated the average causal effects of nonattainment designation and the role of PM10 in this study." See also Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, Samet J, King G, Dominici F; HEI Health Review Committee. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. Res Rep Health Eff Inst. 2016 May;(187):5-49. This study may help to clarify or place bounds on the fraction of the observed PM10-mortality C-R function that is *not* causal (mediated by reduction in PM10). Similar non-PM mechanisms creating a non-causal fraction of the PM-mortality C-R association might also be relevant for PM2.5.
- Enstrom JE. Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. Inhal Toxicol. 2005 Dec 15;17(14):803-16. This association-based study concluded that "**For the initial period, 1973-1982, a small positive risk was found: RR was 1.04 (1.01-1.07) for a 10-microg/m(3) increase in PM(2.5). For the subsequent period, 1983-2002, this risk was no longer present**: RR was 1.00 (0.98-1.02). For the entire follow-up period, RR was 1.01 (0.99-1.03). The

RRs varied somewhat among major subgroups defined by sex, age, education level, smoking status, and health status. None of the subgroups that had significantly elevated RRs during 1973-1982 had significantly elevated RRs during 1983-2002. …These epidemiologic results do not support a current relationship between fine particulate pollution and total mortality in elderly Californians, but they do not rule out a small effect, particularly before 1983." Although it is a traditional association-based study, and therefore less relevant to determination of manipulative causality than intervention studies, this study does suggest the important possibility that C-R relationships have changed over time. This underscores the importance of using recent, relevant data on PM reductions and changes in public health where possible.

- 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. 2011; 106 (494):396-406. doi: 10.1198/jasa.2011.ap09392. This applies similar methods to Janes et al. (2007), *op cit*, to individual-level information on time of death and age in a population of 18.2 million patients in the Medicare Cohort Air Pollution Study (MCAPS) for 2000-2006. The authors conclude that "**Results based on the global coefficient indicate a large increase in the national life expectancy for reductions in the yearly national average of PM2.5. However, this coefficient based on national trends in PM2.5 and mortality is likely to be confounded** by other variables trending on the national level. Confounding of the local coefficient by unmeasured factors is less likely, although it cannot be ruled out. **Based on the local coefficient alone, we are not able to demonstrate any change in life expectancy for a reduction in PM2.5**."
- Haikerwal A, Akram M, Del Monaco A, Smith K, Sim MR, Meyer M, Tonkin AM, Abramson MJ, Dennekamp M. Impact of Fine Particulate Matter (PM2.5) Exposure During Wildfires on Cardiovascular Health Outcomes. J Am Heart Assoc. 2015 Jul 15;4(7). pii: e001653. doi: 10.1161/JAHA.114.001653. This study concluded that "**PM2.5 exposure was associated with increased risk of out-of-hospital cardiac arrests and IHD during the 2006-2007 wildfires in Victoria**. This evidence indicates that PM2.5 may act as a triggering factor for acute coronary events during wildfire episodes."
- Hutchinson JA, Vargo J, Milet M, French NHF, Billmire M, Johnson J, Hoshiko S. The San Diego 2007 wildfires and Medi-Cal emergency department presentations, inpatient hospitalizations, and outpatient visits: An observational study of smoke exposure periods and a bidirectional case-crossover analysis. PLoS Med. 2018 Jul 10;15(7):e1002601. doi: 10.1371/journal.pmed.1002601. This study found that "**Respiratory diagnoses, especially asthma, were elevated during the wildfires in the vulnerable population of Medi-Cal beneficiaries**. Wildfire-related healthcare utilization appeared to persist beyond the initial high-exposure period. Increased adverse health events were apparent even at mildly degraded AQI levels."
- Janes H, Dominici F, Zeger SL (2007). Trends in air pollution and mortality: an approach to the assessment of unmeasured confounding. Epidemiology. Jul;18(4):416-23. These authors applied a method for diagnosing confounding bias in a model with spatially and temporally varying exposure and health outcomes to test whether counties having steeper declines in PM2.5 also have steeper declines in mortality relative to their national trends. They found "that the exposure effect estimates are different at these 2 spatiotemporal scales, which raises concerns about confounding bias" and that, at the local scale**, "there is little evidence of an association between 12-month exposure to PM2.5 and mortality."**
- Merrifield A, Schindeler S, Jalaludin B, Smith W. Health effects of the September 2009 dust storm in Sydney, Australia: did emergency department visits and hospital admissions increase? Environ

Health. 2013 Apr 16;12:32. doi: 10.1186/1476-069X-12-32. This article concludes that "**The dust storm period was associated with large increases in asthma emergency department visits** (relative risk 1.23, 95% confidence interval 1.10-1.38, $p < 0.01$), and to a lesser extent, all emergency department visits (relative risk 1.04, 95% confidence interval 1.03-1.06, $p < 0.01$) and respiratory emergency department visits (relative risk 1.20, 95% confidence interval 1.15-1.26, $p < 0.01$). There was **no significant increase in cardiovascular emergency department visits ($p = 0.09$) or hospital admissions** for any reason. Age-specific analyses showed the dust storm was associated with **increases in all-cause and respiratory emergency department visits in the ≥65 year age group; the ≤5 year group had higher risks of all-cause, respiratory and asthma-related emergency department presentations**."

- You C, Lin DKJ, Young SS. PM2.5 and ozone, indicators of air quality, and acute deaths in California, 2004-2007. Regul Toxicol Pharmacol. 2018 Jul; 96:190-196. doi: 10.1016/j.yrtph.2018.05.012. The authors report that "**There is no statistically significant association between either ozone or $PM_{2.5}$ and acute human mortality"** in a large dataset for eight air basins in California for the years 2004-2007, after statistical adjustment for seasonal and weather effects. (The Draft ISA, p. 11-9, discusses other negative studies by Young et al. 2017 and Lanzinger et al. 2016.)
- Zhou M, He G, Fan M, Wang Z, Liu Y, Ma J, Ma Z, Liu J, Liu Y, Wang L, Liu Y. Smog episodes, fine particulate pollution and mortality in China. Environ Res. 2015 Jan;136:396-404. doi: 10.1016/j.envres.2014.09.038. This natural experiment study examined the impacts on mortality rates of prolonged and severe smog episodes (**PM2.5 hourly peak concentrations over 800 $\mu g/m^3$**) in China in 2013, finding that "Without any meteorological control, the smog episodes are positively and statistically significantly associated with mortality in 5 out of 7 districts/ counties. However, the findings are sensitive to the meteorological factors. **After controlling for temperature, humidity, dew point and wind, the statistical significance disappears in all urban districts**. In contrast, the **smog episodes are consistently and statistically significantly associated with higher total mortality and mortality from cardiovascular/respiratory diseases in the two rural counties**." Such findings suggest substantial geographic heterogeneity in estimated PM2.5-mortality associations.
- Zu K, Tao G, Long C, Goodman J, Valberg P. Long-range fine particulate matter from the 2002 Quebec forest fires and daily mortality in Greater Boston and New York City. Air Qual Atmos Health. 2016; 9:213-221. This natural experiment study concluded that "**substantial short-term elevation in PM2.5 concentrations from forest fire smoke were not followed by increased daily mortality in Greater Boston or New York City**." Although the Preface to the Draft ISA notes that "Studies that conduct an assessment of the PM effect from a source-based mixture (e.g., wood smoke, diesel exhaust, gasoline exhaust, etc.) are only included if they use filtration (e.g., a particle trap) or another approach to differentiate between effects due to the mixture and effects due to the particles alone," natural experiments involving long-range transport of PM from fires may also be useful in showing effects of elevated PM2.5 levels on human health, insofar as other components settle or volatilize out during long-range transport.

Some of these studies present information and reach conclusions that appears to be discordant with the assumption in many other studies that observed PM-mortality associations correspond to (manipulative) causal relationships. Acknowledging, understanding, and resolving such seemingly discordant evidence and discussing it in the ISA may help to refine and improve scientific understanding of the conditions under which human health benefits are materially increased by reducing PM exposure and conditions

under which they are not (possibly due to important roles for temperature, PM composition, copollutants, etc. that are not yet well understood). This may help to refine and improve the predictive validity of the scientific theories and models currently used to predict health effects of PM reductions. For example, a possible explanation for the lack of observed reductions in mortality risks following the large reductions in PM exposures in Ireland might be that *coal is burned primarily in the winter*, and PM effects on health in winter are minimal. As noted by Stafoggia et al. (2008) for PM10 and mortality, "**Season and temperature levels strongly modified the PM10–mortality association**: for a 10-μg/m3 variation in PM10, a 2.54% increase in risk of death in summer (95% confidence interval: 1.31, 3.78) compared with 0.20% (95% confidence interval: −0.08, 0.49) in winter. … The authors found much higher PM10 effects on mortality during warmer days." (Stafoggia M, Schwartz J, Forastiere F, Perucci CA; SISTI Group. Does temperature modify the association between air pollution and mortality? A multicity case-crossover analysis in Italy. Am J Epidemiol. 2008 Jun 15;167(12):1476-85. doi: 10.1093/aje/kwn074.) If this is the explanation, then it would highlight the crucial importance of conditioning on season and temperature in all C-R functions used to predict public health effects of reducing PM exposures. Such a refined C-R model might have more successfully predicted the outcomes of the Ireland coal burning bans.

The ISA should also be meticulous in reporting negative results accurately. For example,

- Puett et al. (2011) state that "Among this cohort of men with high socioeconomic status living in the midwestern and northeastern United States, **the results did not support an association of chronic PM exposures with all-cause mortality and cardiovascular outcomes** in models with time-varying covariates. Whether these findings suggest sex differences in susceptibility or the protective impact of healthier lifestyles and higher socioeconomic status requires additional investigation"(Puett RC, Hart JE, Suh H, Mittleman M, Laden F Particulate matter exposures, mortality, and cardiovascular disease in the health professionals follow-up study. Environ Health Perspect. 2011 Aug;119(8):1130-5. doi: 10.1289/ehp.1002921). The Draft ISA (p. 6-146) includes this study in its summary as follows: "The remaining North American studies, which examined populations of men, or both men and women, **generally report positive associations** between long-term PM2.5 exposure and MI, although the width of the confidence intervals varies between studies. Puett et al. (2011) conducted a prospective analysis of the Health Professionals Follow-up Study (HPFS), which consists of male medical professionals reporting an association of 1.08 (95%CI: 0.90, 1.28). This association was largely unchanged after adjustment for PM10−2.5 (Puett et al., 2011)." This summary may not fully convey that Puett et al. found that "the results did not support an association of chronic PM exposures with all-cause mortality and cardiovascular outcomes."
- The Draft ISA summarizes a study by Madrigano et al. (2013) as follows: "In an incident case control analysis of confirmed acute MI **Madrigano et al. (2013) reported a stronger association** [OR: 1.21 (95%CI: 1.00, 1.38)] between long-term exposure to PM2.5 and acute MI. This study derived exposure metrics to distinguish regional PM2.5 from local traffic-related PM2.5 sources of exposure, and found the association with regional PM2.5 was not attenuated in a copollutant model containing local traffic-related PM2.5." This does not fully convey the following aspect of the author's own assessment: "In contrast to our previous analysis specifically examining traffic particles, **we only found a weak association between our measure of local PM2.5 pollution and occurrence of AMI**" (non-significant association with OR: 1.04 (95%CI: 0.96, 1.11) for GEE models with exchangeable correlation within census block group).

- Hartiala et al. (2016) concluded that "Exposure **to higher PM2.5 levels was also significantly associated with increased risk of incident myocardial infarction** (hazard ratio 1.33, 95% CI 1.02-1.73, P=0.03) **but not stroke or all-cause mortality**. …Exposure to PM2.5 increased the likelihood of having severe coronary artery disease and the risk of incident myocardial infarction among patients undergoing elective cardiac evaluation. These results suggest that ambient air pollution exposure may be a modifiable risk factor for risk of myocardial infarction in a highly susceptible patient population." The Draft ISA treats this information by noting correctly that "In another study, Hartiala et al. (2016) reported an association of long-term exposure to PM2.5 with confirmed MI among those undergoing cardiac evaluation at a clinic in Ohio." But it does not mention Hartiala et al.'s negative finding that exposure to higher PM2.5 levels was not significantly associated with all-cause mortality.
- The Draft ISA (p. 11-91) says of the ESCAPE project that "**Previous analyses of the ESCAPE cohort observed associations between long-term PM2.5 exposure and CVD mortality**. The results presented by Wolf et al. (2015) are consistent with these associations." It does not mention the more recent review by Lipfert (2017), which states that "**No significant associations were reported for cardiovascular mortality**" (Lipfert FW. A critical review of the ESCAPE project for estimating long-term health effects of air pollution. Environ Int. 2017 Feb;99:87-96. doi: 10.1016/j.envint.2016.11.028).

The final ISA should avoid any appearance of distorting, simplifying, or selecting results from the literature to reduce the salience of conflicting information.

At the public meeting on the PM Draft ISA in December, in response to a question about the omission of seemingly relevant negative study results, EPA mentioned that previous CASACs encouraged use of study selection criteria that would exclude the Dublin coal burning ban accountability study. However, understanding the totality of relevant scientific evidence about human health effects caused by reducing PM exposures requires considering negative as well as positive results. The ISA should include and discuss the results of the Dublin accountability study and other high-quality negative studies of changes in PM and changes in health effects. Comments or concerns about individual study quality should be noted in discussions or summary tables on individual study quality, as discussed next, but *the ISA should carefully consider and discuss the body of negative study results such as those listed above.*

At the same time, the ISA should not uncritically accept results based on poor-quality or speculative quasi-experimental studies. For example, Schwartz et al. (2015), cited by the Draft ISA (p. 11-13), state that their propensity score and instrumental variable analyses rest on untestable assumptions (e.g. Cox LA and Goodman JE (2016) Re: "Estimating causal associations of fine particles with daily deaths in Boston" American Journal of Epidemiology 183(6): 593, https://doi.org/10.1093/aje/kww023). This study does not control for confounding by daily temperature extremes and humidity in the weeks preceding death, but only for same-day and previous-day temperatures. As discussed below, this is not nearly long enough to control for likely confounding of PM mortality C-R associations by lagged temperatures. For example, Yang et al. (2012) report in a study of distributed lags for temperatures affecting mortality that the "Hot effect was immediate and limited to the first 5 days… Cold effect persisted for approximately 12 days, (Yang J, Ou CQ, Ding Y, Zhou YX, Chen PY. (2012) Daily temperature and mortality: a study of distributed lag non-linear effect and effect modification in Guangzhou. Environ Health. Sep 14;11:63. doi: 10.1186/1476-069X-11-63.) Well-conducted quasi-experimental studies should use relevant comparison groups, control for obvious potential confounders

such as lagged temperatures, test for and correct for unobserved confounders (for which there are actually many well-developed techniques, as discussed later), and avoid relying on unverified assumptions, e.g., by using invariant causal prediction (ICP) models for which invariance assumptions have been verified in multiple data sets (e.g., Heinze-Deml C, Peters J, Meinshausen N. 2017. Invariant causal prediction for nonlinear models. https://arxiv.org/pdf/1706.08576.pdf).

Present explicit, verifiable derivations of all conclusions. Provide explicit criteria for how individual studies and evidence are systematically reviewed, selected, evaluated, interpreted causally, combined or synthesized, resolved when they conflict, and summarized in the ISA. Explain the explicit objective, independently verifiable criteria and methods for:

- ***Including or excluding individual studies***
- ***Evaluating the internal and external validity and technical soundness of each study's conclusions***
- ***Evaluating, combining, and synthesizing and summarizing results across studies (including resolving any conflicting results); and***
- ***Interpreting and validating their implications for causal determinations.***

Although the Draft ISA makes some useful remarks about the limitations of certain studies (e.g., Mirabelli et al. 2016), it does not provide independent and systematic critical assessments of the internal or external validity of most of the studies and conclusions that it cites, nor does it systematically assess their methodological soundness and document the results of these systematic assessments. The final ISA should present such systematic evaluations of individual study methods and conclusions for methodological soundness and validity. The Draft ISA leaves unclear exactly how studies were selected, why some apparently valuable ones were not, what makes evidence "sufficient to conclude" something, and what principles were or should be followed in presenting and integrating conflicting evidence. The final ISA should address each of these points. It should be thorough in critically assessing the internal and external validity of the study conclusions that it presents and synthesizes. It should provide explicit, objective, transparent (i.e., clear and independently verifiable) criteria and methods for carrying out each of the following steps:

- *Study selection:* For each published study that presents information about health and welfare effects of PM, determine whether the study meets the ISA's criteria for inclusion (e.g., based on explicit criteria for quality and relevance of study design, analysis, interpretation, and validation).
- *Individual study evaluation:* Explicitly assess each included study on each criterion used to determine individual study quality. (The ISA might use a table similar to the one shown by Dr. Goodman in public comments to display study evaluation results.) Explicitly assess the internal and external validity of each study's conclusions.
- *Study result combination:* For any set of studies presenting information (possibly discordant) about health effects of PM, determine the summary statements that are warranted by the studies.
- *Causal determination interpretation of entire body of evidence:* For any set of studies presenting information (possibly discordant) about health effects of PM, determine the causal determinations that are warranted by the studies. In doing so, as previously discussed, the ISA should specify the types of causal effects (e.g., direct vs. total vs. mediated) that are being discussed.

The criteria and methods for each step should be presented in sufficient operational detail so that different investigators working independently, given only the ISA's descriptions of the criteria and a set of studies to apply them to, can reach identical conclusions on the results of each of these steps. This will help to make clear whether stated conclusions are justified by independently reproducible derivations from the factual evidence presented. It will also encourage more critical and thorough engagement with, and evaluation of, the factual basis and technical contents of the individual studies underlying ISA conclusions.

"Evidence" consisting of merely repeating selected published results and conclusions from studies with inappropriate designs or analyses or with conclusions that depend on unverified or mistaken assumptions or models (e.g., with estimated exposures treated as true exposures, model uncertainty not quantified, effects of unobserved confounders and latent variables not tested for and quantified, etc.) is not necessarily valid evidence. It should not be presented or summarized as if it were known to be valid. Conclusions from the literature should not be quoted, cited, or used without systematic independent review and documentation of the quality of the methods used to produce them (Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, Ratcliffe JM, Kraft AD, Schünemann HJ, Schwingl P, Walker TD, Thayer KA, Lunn RM. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. Environ Int. 2016 Jul-Aug;92-93:617-29. doi: 10.1016/j.envint.2016.01.005). Perhaps most importantly, conclusions with unknown internal and external validity should not be cited as facts. As stated by Campbell and Stanley (1963, p.5):

> "***Internal validity*** **is the basic minimum without which any experiment is uninterpretable:** Did in fact the experimental treatments [here, changes in PM pollution levels] make a difference in this specific experimental instance? *External validity* asks the question of *generalizability*: To what populations, settings, treatment variables, and measurement variables can this effect be generalized? Both types of criteria are obviously important, even though they are frequently at odds in that features increasing one may jeopardize the other. While *internal validity* is the *sine qua non*, and while the question of *external validity*, like the question of inductive inference, is never completely answerable, the selection of designs strong in both types of validity is obviously our ideal." (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).

A noteworthy technical advance is that the question of external validity *can* now be answered completely under certain conditions. As explained by Pearl and Bareinboim (2014),

> "The generalizability of empirical findings to new environments, settings or populations, often called 'external validity,' is essential in most scientific explorations. This paper treats a particular problem of generalizability, called 'transportability,' defined as a license to transfer causal effects learned in experimental studies to a new population, in which only observational studies can be conducted. [It presents] **procedures for deciding, prior to observing any data, whether causal effects in the target population can be inferred from experimental findings in the study population**. **When the answer is affirmative, the procedures identify what experimental and observational findings need be obtained from the two populations**, and how they can be combined to ensure bias-free transport." (Pearl J, Bareinboim E (2014). External validity: from do-calculus to transportability across populations. Statistical Science

29(4): 579–595 DOI: 10.1214/14-STS486, https://ftp.cs.ucla.edu/pub/stat_ser/r400-reprint.pdf) This line of research has recently been updated to "demonstrate how transportability analysis can guide the transfer of knowledge among non-experimental studies to minimize re-measurement cost and improve prediction power. We further provide a causally principled definition of 'surrogate endpoint' and show that the theory of transportability can assist the identification of valid surrogates in a complex network of cause-effect relationships." (Pearl J, Bareinboim E (2018), Transportability across studies: A formal approach https://ftp.cs.ucla.edu/pub/stat_ser/r372.pdf).

Likewise, Bareinboim and Pearl (2014) address the problem of "transferring causal knowledge collected in several heterogeneous domains to a target domain in which only passive observations and limited experimental data can be collected." They present a "necessary and sufficient condition for deciding… whether causal effects in the target domain are estimable from the information available." (Bareinboim E, Pearl J (2014) Transportability from Multiple Environments with Limited Experiments: Completeness Results. In Advances in Neural Information Processing Systems 27 (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence and K. Weinberger, eds.). Curran Associates, Inc., 280–288. https://papers.nips.cc/paper/5536-transportability-from-multiple-environments-with-limited-experiments-completeness-results.pdf). These ideas have proved useful in epidemiology (Infante-Rivard C, Cusson A. Reflection on modern methods: selection bias - A review of recent developments. Int J Epidemiol. 2018 Oct 1;47(5):1714-1722. doi: 10.1093/ije/dyy138.) Thus, some useful conditions for establishing external validity are now well understood. c*Before accepting causal conclusions from the literature, or drawing new causal conclusions from the synthesis of results in included studies, the ISA should explicitly evaluate and discuss the extents to which both the internal validity and the external validity of each causal conclusion have been established.*

To avoid conflating science and policy judgments, the ISA should also define and explain the criteria and methods used to decide which studies to include and exclude, how to combine their results in preparing summary and synthesis statements, and how to derive implications and conclusions about causal determinations from the entire body of evidence. The definitions of criteria and methods should be sufficiently clear, objective, and operational that the decisions they imply are independently reproducible and verifiable. They should leave no room for possible implicit policy decisions in the study selection, evaluation, evidence synthesis, or conclusion-drawing steps. (For example, a tendency to exclude relevant high-quality studies unless they support a certain policy position, or to include lower-quality study results when they do support a policy position but not otherwise, would be clearly inappropriate.) The ISA should explain and then execute the criteria and methods for the above steps with enough specificity, objectivity, and independent verifiability to make it obvious that no such mixing of policy with science is possible.

Recommended criteria to consider for each of these steps include the following.

Study inclusion criteria

Include studies that address changes in health effects caused by changes in PM exposures.

- Include well-designed and well-conducted accountability studies, natural experiments, intervention studies, and other quasi-experiments (QEs). Such studies are especially valuable for learning empirically how changes in PM affect human health under real-world conditions.
    - Studies that do not include data on actual changes in exposures and subsequent actual changes in responses should not be included, as they lack information needed for valid causal inference (see e.g., discussion on pages 6-7 of 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).
    - Specifically, studies that only examine static associations between levels of exposures and corresponding levels of responses should be excluded. Such data do not permit valid inferences about how or whether changes in exposure would cause responses to change. (See discussion of the "Static-group comparison" design on p. 12 of Campbell and Stanley (1963).)
- The specific *causal effects* assessed in each included study should be clearly stated. This requires specifying the following components:
    - What *changes in exposure* were considered (when, where, over what time interval)?
    - What *observed changes in effects or responses* were considered, e.g., what changes in specific health or welfare endpoints in what population(s) over what intervals?
    - What *specific causal relationship(s)* between changes in exposure and changes in effects were assessed? (If the answer is unknown, this should be stated.) As previously discussed, specific causal relationships include the following: controlled direct, natural direct, indirect, total, and mediated effects of PM exposure on risks of adverse effects.
    - How were the changes in exposures and responses operationally defined and measured?
    - How were measurement error, missing data, and unobserved variables accounted for?
- Included studies should have designs and analyses that allow *threats to internal validity* of causal conclusions (e.g., from history, maturation, testing, instrumentation, regression to the mean, selection, experimental mortality, and interactions of threats) to be refuted, e.g., through the use of appropriate comparison groups and tests of exchangeability and conditional independence assumptions (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).

Study exclusion criteria

Studies with any of the following deficiencies can be excluded from further consideration unless they can be corrected *ex post* during the ISA review, critical evaluation, and synthesis:

1. *Study design, data collected, or analyses performed do not control for obvious potential confounders or selection biases that could plausibly explain the study results.*
2. *Study does not distinguish between true exposure values and estimated exposure values in analyzing and presenting information.*
3. *Study design, data collected, or analyses performed do not permit threats to internal validity to be tested and refuted.*
4. *Study design, data collected, or analyses performed do not allow external validity to be established and correct generalizations to target populations to be made.*
5. *Conclusions are sensitive to unstated, untested, unverified, or mistaken assumptions.*

6. *Study design or data only address association and do not permit valid inferences about (manipulative) causation.*
7. *Animal experiment or in vitro study identifies changes caused by PM exposures in test systems, but does not show that these changes are relevant to (or likely to cause) adverse effects in humans under real-world conditions.*

Each of these considerations is discussed next.

Study design, data collected, or analyses performed do not control for obvious potential confounders or selection biases that could plausibly explain the study results.

Important obvious potential confounders that are often omitted include daily low and high temperatures and humidity over the weeks preceding a death or illness. Both hot and cold temperature extremes in the days to weeks preceding a health response can confound and/or modify PM C-R associations. The importance of lagged temperatures for various health effects associated with PM is well documented in many studies and in many different countries.

- For example, Yang et al. (2012) report that "**Hot effect was immediate and limited to the first 5 days, with an overall increase of 15.46% (95% confidence interval: 10.05% to 20.87%) in mortality risk** comparing the 99th and the 90th percentile temperature. **Cold effect persisted for approximately 12 days, with a 20.39% (11.78% to 29.01%) increase in risk** comparing the first and the 10th percentile temperature. **The effects were especially remarkable for cardiovascular and respiratory mortality**. **The effects of both hot and cold temperatures were greater among the elderly**. Females suffered more from hot-associated mortality than males. We also found significant effect modification by educational attainment and occupation class. CONCLUSIONS: There are significant mortality effects of hot and cold temperatures in Guangzhou. **The elderly, females and subjects with low socioeconomic status have been identified as especially vulnerable to the effect of ambient temperatures**." (Yang J, Ou CQ, Ding Y, Zhou YX, Chen PY. (2012) Daily temperature and mortality: a study of distributed lag non-linear effect and effect modification in Guangzhou. Environ Health. Sep 14;11:63. doi: 10.1186/1476-069X-11-63.)

Despite the importance of temperature on various time scales as modifiers and confounders of PM C-R associations, however, very few studies control for both short-run (same-day and recent-day) extreme temperatures, intermediate (weeks to month and season) and long-run average temperatures in quantifying PM C-R associations. For example, Wang et al. (2016) state that "**Many studies have reported the associations between long-term exposure to PM2.5 and increased risk of death. However, to our knowledge, none has** used a causal modeling approach or **controlled for long-term temperature exposure**. … The mean summer temperature and the mean winter temperature in a census tract significantly modified the effects of long-term exposure to PM2.5 on mortality. We observed a higher percentage increase in mortality associated with PM2.5 in census tracts with more blacks, lower home value, or lower median income…. We consistently found that an increase in the mean winter temperature was associated with an increase in the effects of PM2.5 on mortality. … **Although temperature may be the strongest confounder between PM2.5 and mortality, the change over time in other variables such as the employment rate may also confound the relationship**." (Wang Y, Kloog I, Coull BA, Kosheleva A, Zanobetti A, Schwartz JD. Estimating Causal Effects of Long-Term PM2.5 Exposure on Mortality in New Jersey. Environ Health Perspect. 2016 Aug;124(8):1182-8. doi:

10.1289/ehp.1409671.) Thus, it appears that *most of the literature does not control for potential confounding of PM C-R associations by temperatures on both long and short time scales*, leaving the correct causal interpretation of reported C-R associations in such studies objectively undetermined by the data collected.

Following are some examples of the many studies cited in the Draft ISA that do not control for confounding by temperature or humidity:

- Lipsett MJ, Ostro BD, Reynolds P, Goldberg D, Hertz A, Jerrett M, Smith DF, Garcia C, Chang ET, Bernstein L. Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort. Am J Respir Crit Care Med. 2011 Oct 1;184(7):828-35. doi: 10.1164/rccm.201012-2082OC.
- Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. Estimating the Causal Effect of Low Levels of Fine Particulate Matter on Hospitalization. Epidemiology. 2017 Sep;28(5):627-634. doi: 10.1097/EDE.0000000000000690.
- Mirabeli et al. (2015), which is cited a dozen times in the Draft ISA, specifically states that "Our study population included 18 participants with asthma and 21 participants without asthma, 17 and 19 of whom, respectively, completed two study commutes each. Despite measurements provided by each participant at up to 10 time points, **the relatively small number of observations in our analysis, and missing data for several measures limit our ability to conduct additional analysis to explore the roles of** body mass index, medication use, general health status, automobile characteristics, in-vehicle ventilation, **temperature, humidity, rainfall, season, multipollutant exposures or other factors** that may influence susceptibility to in-vehicle air pollutants." (Mirabelli MC, Golan R, Greenwald R, et al. Modification of Traffic-related Respiratory Response by Asthma Control in a Population of Car Commuters. *Epidemiology*. 2015;26(4):546-55.)
- Pinault L, Tjepkema M, Crouse DL, Weichenthal S, van Donkelaar A, Martin RV, Brauer M, Chen H, Burnett RT. Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. Environ Health. 2016 Feb 11;15:18. doi: 10.1186/s12940-016-0111-6.
- Perez L et al. (2015) Air pollution and atherosclerosis: a cross-sectional analysis of four European cohort studies in the ESCAPE study. Environ Health Perspect. 2015 Jun;123(6):597-605. doi: 10.1289/ehp.1307711. (See Draft ISA Table 6-67.)
- Pun VC, Hart JE, Kabrhel C, Camargo CA Jr, Baccarelli AA, Laden F. 2015. Prospective study of ambient particulate matter exposure and risk of pulmonary embolism in the Nurses' Health Study cohort. Environ Health Perspect 123:1265–1270; http://dx.doi.org/10.1289/ehp.1408927. (See Draft ISA Table 6-69.)
- Thurston GD, Ahn J, Cromar KR, et al. Ambient Particulate Matter Air Pollution Exposure and Mortality in the NIH-AARP Diet and Health Cohort. *Environ Health Perspect*. 2015; 124(4):484-90. The Draft ISA cites this study in Tables 5-28, 6-52 and Figures 5-34, 5-35, 6-19, 6-21.

C-R associations and curves estimated in studies that do not control for confounding by same-day and lagged daily temperature extremes and humidity and for confounding by longer-term (e.g., seasonal) average temperatures, should not be presented in the ISA as "evidence" for adverse health effects of PM, unless it is shown that uncontrolled confounding does not provide a plausible alternative explanation. Otherwise, *no valid causal conclusions about C-R relationships can be drawn from studies that do not*

control for plausible confounding (Campbell and Stanley 1963, op cit). Lagged values of daily temperature extremes out to at least 2 weeks should be included to control for confounding, based on observations that PM levels are strongly autocorrelated and that effects of cold temperatures persist for at least 12 days (Yang J, Ou CQ, Ding Y, Zhou YX, Chen PY. (2012) Daily temperature and mortality: A study of distributed lag non-linear effect and effect modification in Guangzhou. Environ Health. Sep 14;11:63. doi: 10.1186/1476-069X-11-63.)

More generally, omitted confounders arise if not enough lagged values of explanatory variables are included and the omitted lagged values affect both exposure and response variables. Many studies only control for a few recent days of temperature and humidity variables, or for temperatures averaged over some time window. Such studies fail to control for confounding by daily temperature extremes at longer lags. The following are examples (among many) of studies that are cited and used in the Draft ISA that consider only a few days of temperature and/or humidity, or only longer-term averages, without showing that temperatures on earlier days do not confound the reported C-R associations:

- Belleudi et al. (2010) included no lagged values for temperature or humidity; thus, confounding by temperatures from the previous day(s) was not controlled. The reference is: Belleudi V, Faustini A, Stafoggia M, Cattani G, Marconi A, Perucci CA, Forastiere F. Impact of fine and ultrafine particles on emergency hospital admissions for cardiac and respiratory diseases. Epidemiology. 2010 May;21(3):414-23. doi: 10.1097/EDE.0b013e3181d5c021. The Draft ISA cites this study multiple times, e.g., in Tables 5-8, 5-10, 5-42, 5-43; Figures 3-13, 5-6, 5-7.
- Powell et al. (2015) used "smooth functions of the current day's temperature and the mean of the previous 3 days' temperatures, both of which used 6 degrees of freedom; smooth functions of the current day's dew point temperature and the mean of the previous 3 days' dew point temperatures, both of which used 3 degrees of freedom." Using 3-day averages does not show that the results are not confounded by temperature extremes over the previous 2 weeks. The reference is: Powell H, Krall JR, Wang Y, Bell ML, Peng RD. Ambient Coarse Particulate Matter and Hospital Admissions in the Medicare Cohort Air Pollution Study, 1999-2010. Environ Health Perspect. 2015 Nov;123(11):1152-8. doi: 10.1289/ehp.1408720. The Draft ISA cites this study multiple times (e.g., Tables 5-11, 5-32, 5-34, 6-58, 6-62, Figures 5-42, 5-44).
- Mirabelli et al. (2016) associate asthma symptoms with county-wide PM2.5 levels but average daily maximum temperatures over the entire preceding two weeks, making it impossible to determine whether confounding by lagged daily minimum and maximum temperatures explains the association. The reference is: Mirabelli MC, Vaidyanathan A, Flanders WD, Qin X, Garbe P. Outdoor PM2.5, Ambient Air Temperature, and Asthma Symptoms in the Past 14 Days among Adults with Active Asthma. Environ Health Perspect. 2016 Dec;124(12):1882-1890. The Draft ISA cites this study but correctly notes that "However, this study is limited by its cross-sectional design, and residual confounding may arise from the 14-day PM2.5 averaging time and lack of consideration of confounding by community-level SES."
- Stafoggia et al. (2013), in a study of PM-associated hospitalizations cited by the Draft ISA in Figures 5-8, 5-23, 5-44, 5-45, Tables 5-11, 5-34, and over 20 times in the text, states that "We controlled for the effect of temperature by modeling high and low temperatures separately. **For high temperatures we calculated the average temperature on the current and previous day (lag 0–1)** and fit a natural spline with 3 df on the lagged variable only for days on which the lag 0–1 temperature was higher than the median annual temperature for the city as a whole. Similarly, **we adjusted for low temperatures by fitting a natural spline with 2 df for the average temperature on previous 6**

days (lag 1–6) only for days on which the lag 1–6 temperature was below the median annual value for the city (Chiusolo et al. 2011). This method accounts for differences in the lag structures and effects of cold and warm temperatures on hospitalizations while reducing the correlation between the two spline terms. We also performed an analysis that adjusted for potentially prolonged effects of warm temperatures on hospitalizations by **replacing the lag 0–1 temperature term from the base model with the lag 0–6 average**." (Stafoggia M, Samoli E, Alessandrini E, et al. Short-term associations between fine and coarse particulate matter and hospitalizations in Southern Europe: results from the MED-PARTICLES project. *Environ Health Perspect*. 2013;121(9):1026-33.) This procedure does not address confounding by extreme temperatures in the days (or weeks) prior to hospitalization. Averaging temperatures over a 6-day window obscures the effects of lagged extreme temperatures (e.g., daily highs or lows), which have been found to be important in other studies (e.g., Wang et al. 2015, discussed below).

Studies that report C-R associations without controlling for lagged daily temperature extremes with lags out to at least 2 weeks, and for which confounding by lagged temperatures and humidity are not refuted as a plausible explanation for the reported C-R associations, have no clear valid causal interpretations for their reported C-R associations. Whether "enough" lagged values have been included in modeling C-R functions can be determined from data by verifying that the response variable is conditionally independent of the excluded lagged values of predictors, given the included ones.

References for temperature as a confounder and modifier of PM C-R functions include the following:

- Fang X, Fang B, Wang C, Xia T, Bottai M, Fang F, Cao Y. Relationship between fine particulate matter, weather condition and daily non-accidental mortality in Shanghai, China: A Bayesian approach. PLoS One. 2017 Nov 9;12(11):e0187933. doi: 10.1371/journal.pone.0187933. This study concludes that **"The effect of PM2.5 on non-accidental mortality differed under specific extreme weather conditions** and SWTs [synoptic weather types, e.g., hot dry, warm humid, etc.]. Environmental policies and actions should take into account the interrelationship between the two hazardous exposures."
- Gouder et al. (2013) state that "Previous results from our study show that **emergency department (ED) visits for acute asthma** exhibit seasonality in Malta. Visits were positively correlated with wind speed and precipitation and negatively correlated with humidity, barometric pressure and temperature. … Regression analysis **showed temperature to be the best predictor for ED visi**ts ($p < 0.05$). Conclusion: **High wind speeds and temperature are associated with elevated air pollutant levels**. Precipitation, humidity and pressure seem to be independent triggers for acute asthma. Increased vigilance during such periods may avoid exacerbations." (Gouder C, Gerada E et cl. (2013), Are air pollutants confounders in relation to weather variables as triggers for acute asthma? European Respiratory Journal Sep 2013, 42 (Suppl 57) P771)
- Kim H, Bell ML, Lee JT Does a lag-structure of temperature confound air pollution-lag-response relation? Simulation and application in 7 major cities, Korea (1998-2013). Environ Res. 2017 Nov;159:531-538. doi: 10.1016/j.envres.2017.08.047. The authors state that "**Temperature must be controlled when estimating the associations of short-term exposure to air pollution and mortality**." In Korea, for PM10 and mortality, "Controlling for temperature as distributed lags for 21 days provided 0.25% (95% CI: 0.1, 0.4) increase in the risk of all-cause mortality."
- Kioumourtzoglou MA, Schwartz J, James P, Dominici F, Zanobetti A. PM2.5 and Mortality in 207 US Cities: Modification by Temperature and City Characteristics. Epidemiology. 2016

Mar;27(2):221-7. doi: 10.1097/EDE.0000000000000422. The authors reported that "We observed a **higher association between long-term PM2.5 exposure and mortality in warmer cities**." Furthermore, we observed increasing estimates with increasing obesity rates, %residents and families in poverty, %black residents and %population without a high school degree, **and lower effects with increasing median household income** and %white residents."

- Schnell JL, Prather MJ. Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America. Proc Natl Acad Sci U S A. 2017 Mar 14;114(11):2854-2859. doi: 10.1073/pnas.1614453114. "There is evidence indicating that **combined pollution extremes and heat waves are such synergistic stressors** (i.e., impact modifiers), and that combined extremes produce disproportionately greater adverse health impacts."
- Scortichini M, De Sario M, de'Donato FK, Davoli M, Michelozzi P, Stafoggia M. Short-Term Effects of Heat on Mortality and Effect Modification by Air Pollution in 25 Italian Cities. Int J Environ Res Public Health. 2018 Aug 17;15(8). pii: E1771. doi: 10.3390/ijerph15081771. This study again demonstrates geographic heterogeneity of C-R functions as well as strong temperature-PM10 interactions in affecting mortality: "Evidence on the health effects of extreme temperatures and air pollution is copious. However few studies focused on their interaction. The aim of this study is to evaluate daily **PM10 and ozone as potential effect modifiers of the relationship between temperature and natural mortality in 25 Italian cities**. … Differential temperature-mortality risks by air pollutants were found. **For PM10, estimates ranged from 3.9% (low PM10) to 14.1% (high PM10) in the North, from 3.6% to 24.4% in the Center, and from 7.5% to 21.6% in the South**. Temperature-related mortality was similarly modified by ozone in northern and central Italy, while no effect modification was observed in the South."
- Stafoggia et al. (2008) report that "**Season and temperature levels strongly modified the PM10–mortality association**: for a 10-μg/m3 variation in PM10, a 2.54% increase in risk of death in summer (95% confidence interval: 1.31, 3.78) compared with 0.20% (95% confidence interval: −0.08, 0.49) in winter. Analysis of the interaction between PM10 and temperature within temperature strata resulted in positive but, in most cases, nonstatistically significant coefficients. The authors found **much higher PM10 effects on mortality during warmer days**." (Stafoggia M, Schwartz J, Forastiere F, Perucci CA; SISTI Group. Does temperature modify the association between air pollution and mortality? A multicity case-crossover analysis in Italy. Am J Epidemiol. 2008 Jun 15;167(12):1476-85. doi: 10.1093/aje/kwn074.)
- Wang et al.(2015) found strong effects and substantial geographic heterogeneity of lagged extreme temperatures on cardiovascular mortality: "For all cause-specific cardiovascular mortality, Beijing had stronger cold and hot effects than those in Shanghai. The **cold effects on cause-specific cardiovascular mortality reached the strongest at lag 0-27**, while the **hot effects reached the strongest at lag 0-14**. The **effects of extremely low and high temperatures differed by mortality types in the two cities**. Hypertensive disease in Beijing was particularly susceptible to both extremely high and low temperatures; while for Shanghai, people with ischemic heart disease showed the greatest relative risk (RRs = 1.16, 95% CI: 1.03, 1.34) to extremely low temperature. CONCLUSION: **People with hypertensive disease were particularly susceptible to extremely low and high temperatures in Beijing. People with ischemic heart disease in Shanghai showed greater susceptibility to extremely cold days**." Wang X, Li G, Liu L, Westerdahl D, Jin X, Pan X. Effects of Extreme Temperatures on Cause-Specific Cardiovascular Mortality in China. Int J Environ Res Public Health. 2015 Dec 21;12(12):16136-56. doi: 10.3390/ijerph121215042.

- Willers SM, Jonker MF, Klok L, Keuken MP, Odink J, van den Elshout S, Sabel CE, Mackenbach JP, Burdorf A. High resolution exposure modelling of heat and air pollution and the impact on mortality. Environ Int. 2016 Apr-May;89-90:102-9. doi: 10.1016/j.envint.2016.01.013. This case-crossover study found that "**Significant interaction between maximum air temperature (Tamax) and PM10 was observed. During "summer smog" days (Tamax>25°C and PM10>50µg/m(3)), the mortality risk at lag 2 was 7% higher compared to the reference** (Tamax 15°C and PM10 15µg/m(3)). Persons above age 85 living alone were at highest risk. CONCLUSION: **We found significant synergistic effects of high temperatures and air pollution on mortality. Single living elderly were the most vulnerable group.** Due to spatial differences in temperature and air pollution, **mortality risks varied substantially between neighbourhoods, with a difference up to 7%**."

The Draft ISA (p. 11-24) states that "to date studies conducted within the U.S. have not provided evidence of a modification of the PM2.5-mortality association by temperature." This appears to be contradicted by findings such as those in Kioumourtzoglou et al. (2016), which concluded that "living in cities with high temperatures and low socio economic status (SES) is associated with higher effect estimates." (Kioumourtzoglou MA, Schwartz J, James P, Dominici F, Zanobetti A. PM2.5 and Mortality in 207 US Cities: Modification by Temperature and City Characteristics. Epidemiology. 2016 Mar;27(2):221-7. doi: 10.1097/EDE.0000000000000422.) Moroever, Cox et al. (2012) found that "a significant, approximately linear, statistical C-R association exists in simple statistical models" between PM2.5 and mortality rates, but that conditioning on daily temperature and month of year eliminated this association 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. doi: 10.2203/dose-response.12-034.)

In addition to temperature and humidity, many other confounders should be controlled for. Longer-term studies (e.g., of mortality or lung cancer associated with PM) often omit socioeconomic variables such as income and residential and occupational locations. Long-term trends in unobserved variables have also been suggested as important sources of confounding for PM-mortality C-R functions.

- Strak et al. (2017) state that "Cohorts based on administrative data have size advantages over individual cohorts in investigating air pollution risks, but often lack in-depth information on individual risk factors related to lifestyle. If there is a correlation between lifestyle and air pollution, **omitted lifestyle variables may result in biased air pollution risk estimates.** Correlations between lifestyle and air pollution can be induced by socio-economic status affecting both lifestyle and air pollution exposure. … **Current smoking and alcohol consumption were generally positively associated with air pollution**. **Physical activity and overweight were negatively associated with air pollution**. … Despite the small associations between air pollution and smoking intensity, **indirect adjustment resulted in considerable changes of air pollution risk estimates for cardiovascular and especially lung cancer mortality**. CONCLUSIONS: Individual lifestyle-related risk factors were weakly associated with long-term exposure to air pollution in the Netherlands**. Indirect adjustment for missing lifestyle factors in administrative data cohort studies may substantially affect air pollution mortality risk estimates**. … For most lifestyle-related risk factors, **unhealthy lifestyle was associated with higher air pollution exposure** in our survey data. … indirect adjustment for missing lifestyle factors may either increase or decrease observed air pollution effect estimates. … the **results of indirect adjustments of air pollution**

effect estimates in administrative cohorts are study-specific." (Strak M, Janssen N, Beelen R, Schmitz O, Karssenberg D, Houthuijs D, van den Brink C, Dijst M, Brunekreef B, Hoek G. Associations between lifestyle and air pollution exposure: Potential for confounding in large administrative data cohorts. Environ Res. 2017 Jul;156:364-373. doi: 10.1016/j.envres.2017.03.050.)

Confounders can usually be controlled for in multiple ways by conditioning on appropriate adjustment sets, provided that the study collects the needed data (e.g., Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2016 Dec 1;45(6):1887-1894). *Studies that do not collect the data or perform the analyses needed to control for confounding and collider biases that could plausibly explain their results should not be used to draw causal conclusions.*

Study does not distinguish between true exposure values and estimated exposure values in analyzing and presenting information.

In general, studies that treat estimated exposures as true exposures and that ignore exposure estimation errors (or assume without justification that they bias risk estimates downward) do not support valid inferences about the shape of the C-R curve for PM2.5. This is because realistic exposure estimation errors for PM2.5 are large enough to substantially distort the shape of the estimated C-R function, e.g., by making even threshold C-R relationships appear to be linear no-threshold. As noted by Sheppard et al. (2012), "**Exposure measurement error is a challenge in epidemiology because inference about health effects can be incorrect** when the measured or predicted exposure used in the analysis is different from the underlying true exposure. **Air pollution epidemiology rarely if ever uses personal measurements of exposure** for reasons of cost and feasibility. … Exposure assessment for epidemiology should be evaluated in the context of the health effect estimation goal**. It is important to design the exposure assessment to capture the underlying exposure variability** for the pollutants of interest, obtain exposure data that are directly relevant to the study population (e.g., representative of residence locations), and ensure there are sufficient exposure data to support good predictions." (Sheppard L, Burnett RT, Szpiro AA, Kim SY, Jerrett M, Pope CA 3rd, Brunekreef B. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Health. 2012 Jun;5(2):203-216.) *Studies that do not address exposure measurement and estimation errors should not be used or cited as "evidence"* but should be excluded, unless they can be retroactively reanalyzed and corrected to model the effects of realistic exposure estimation errors.

One of many examples of such studies, Lepeule et al. (2012), states that "**Including more recent observations with PM2.5 exposures down to 8 µg/m3**, we continued to find a **statistically significant association between chronic exposure to PM2.5 and all-cause and cardiovascular mortality**. … **The concentration–response relationship was linear without any threshold,** even at exposure levels below the U.S. annual 15-µg/m3 standard (U.S. EPA 1997). Taken together with the results of a previous reanalysis of the Harvard Six Cities study (Krewski et al. 2005b), there is evidence for a robust association between chronic PM2.5 exposure and early mortality." (Lepeule J, Laden F, Dockery D, Schwartz J. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environ Health Perspect*. 2012;120(7):965-70.) This statement neglects the crucial distinction between *true* PM2.5 exposures, which are unknown for people who died, and *estimated* average PM2.5 exposures. The study analysis treats the estimated exposure values as if they

were accurately measured values with zero measurement error variance. Yet, the study's description of the exposure estimation process makes clear that PM2.5 concentrations for individuals were not measured, but were simply imputed based on year and city: "Annual PM2.5 concentration was assigned for each participant until death or censoring. PM2.5 concentration was measured in the participant's city by a centrally located monitor from 1979 to 1986–1988, depending on the city (Dockery et al. 1993). **Therefore, the study has no spatial contrast on the within-city scale. PM2.5 concentrations for the years before monitoring started were assumed to be equal to the earliest monitored year**." A correct analysis would require modeling exposure estimation errors and using appropriate errors-in-variables techniques (e.g., Mallick R, Fung K, Krewski D. Adjusting for measurement error in the Cox proportional hazards regression model. J Cancer Epidemiol Prev. 2002;7(4):155-64). Because Lepeule et al. ignored exposure estimation errors, their conclusion that "The concentration–response relationship was linear without any threshold" is unwarranted, insofar as (a) They did not observe the shape of the true concentration-response relationship and whether it was linear or had thresholds; and (b) PM C-R relationships with realistic errors in estimated exposures typically appear to be linear without any threshold even when the true C-R relationship has a sharp threshold (e.g., Cox LAT. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ Res. 2018 Jul;164:636-646. doi: 10.1016/j.envres.2018.03.038).

The Draft ISA (p. 6-208) uncritically propagates the false and misleading suggestion by Lepeule et al. that they had "recent observations with PM2.5 exposures down to 8 $\mu g/m^3$." What they actually had instead an "annual PM2.5 concentration [that] was assigned for each participant" based on city and year. This assigned value was almost certainly wrong for each individual participant, as it ignored all inter-individual differences in exposures within a city. No observed or measured PM2.5 exposure values were available for any of the individuals who died. Similarly, the Draft ISA mistakenly states that "A number of the concentration-response analyses include concentration ranges ≤ 12 μg/m3. For example, Lepeule et al. (2012) observed a linear, no-threshold concentration-response relationship for cardiovascular mortality in the most recent analysis of the Harvard Six Cities study, with confidence in the relationship down to a concentration of 8 μg/m3." This is a mistaken characterization of what was observed, since the concentration-response relationship was not actually observed (all individual concentration data were missing), as just described. Correct statements would be that (a) The concentration ranges experienced by individuals who died are unknown; and (b) The true concentration-response relationship was not observed and could not be confidently estimated from the available data (since no concentration data were available for any of the individual participants who died).

Other studies acknowledge that exposure measurement error is an issue in Lepeule et al. and similar studies, but incorrectly assert that it is expected to exert a downward bias on risk estimates. For example, Shi et al. (2016) note that "[There] is spatial variability in PM2.5 concentrations within cities that time series studies generally do not take into account, **which can introduce exposure measurement error** (Laden et al. 2006; Lepeule et al. 2012). Chronic effects studies began using comparisons across cities of mortality experiences of cohorts living in various communities and the monitored air pollutant concentrations in those communities (Dockery et al. 1993; Pope et al. 1995). **Again, these studies suffered from exposure error due to failure to capture within-city spatial variability in exposure**. Because the geographic exposure gradient is the exposure contrast in these studies, the **failure to capture within-city contrasts leads to classical measurement error with expected downward bias**." (Shi L, Zanobetti A, Kloog I, Coull BA, Koutrakis P, Melly SJ, Schwartz JD. Low-Concentration PM2.5 and Mortality: Estimating Acute and Chronic Effects in a Population-Based Study. Environ Health

Perspect. 2016 Jan;124(1):46-52. doi: 10.1289/ehp.1409111). In reality, in multivariate models, the direction of bias is unknown, but in models with threshold C-R functions it is upward for estimated exposures below the threshold and downward for estimated exposures above the threshold.

Many other studies cited in the Draft ISA also misrepresent estimated exposures as if they were true exposures, leading to false statements about what has been found. For example:

- Crouse et al. (2012) state that "In this large national cohort of nonimmigrant Canadians, **mortality was associated with long-term exposure to PM2.5. Associations were observed with exposures to PM2.5 at concentrations that were predominantly lower (mean, 8.7 µg/m3; interquartile range, 6.2 µg/m3) than those reported previously**." (Crouse DL, Peters PA, van Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, Khan S, Atari DO, Jerrett M, Pope CA, Brauer M, Brook JR, Martin RV, Stieb D, Burnett RT. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. Environ Health Perspect. 2012 May;120(5):708-14. doi: 10.1289/ehp.1104049.) Again, true long-term exposures to PM2.5 were unmeasured and are unknown for any of the individuals who died; associations of mortality with true long-term exposures were not quantified or observed. No observations were made of any deaths at true concentrations that were known to be below those reported previously. Instead, the authors "assigned estimates of exposure to ambient PM2.5 derived from satellite observations to a cohort of 2.1 million Canadian adults" and then analyzed these assigned estimates *as if* they were true (measured without error) exposure values, much as in the Lepeule et al. study.
- Di et al. (2017) state that "**In the entire Medicare population, there was significant evidence of adverse effects related to exposure to PM2.5 and ozone at concentrations below current national standards**. This effect was most pronounced among self-identified racial minorities and people with low income."(Di Q, Wang Y, Zanobetti A, et al. Air pollution and mortality in the Medicare population. N Engl J Med 2017;376:2513-22. DOI: 10.1056/NEJMoa1702747, https://www.nejm.org/doi/full/10.1056/NEJMoa1702747). Again, this is a false and misleading statement, insofar as it suggests that adverse effects were observed "at concentrations below current national standards." In reality, the supplement to the article shows that the exposure concentrations experienced by individuals who experienced adverse effects were not measured. Instead, individual exposure concentrations were guessed at, or imputed, using techniques such as this: "We also acquired daily 1 km × 1 km gridded air pollution levels (PM2.5 and ozone) from previously developed and validated air pollution prediction models. We obtained ZIP code-level variables by taking inverse-distance averages of the four nearest grid cells to the ZIP code's centroid and then computed the annual averages for temperature, humidity, and PM2.5, and the warm-season (from April 1 to September 30) average for ozone. …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." https://www.nejm.org/doi/suppl/10.1056/NEJMoa1702747/suppl_file/nejmoa1702747_appendix.pdf). In other words, *no actual measurements or observations of PM2.5 concentrations for any individual with an adverse health effect were made* in this study. Consequently, the claim that "In the entire Medicare population, there was significant evidence of adverse effects related to exposure to PM2.5 and ozone at concentrations below current national standards" would be more accurately described as "In the entire Medicare population, there was significant evidence of adverse effects

related to estimated exposure to PM2.5 and ozone at estimated concentrations below current national standards, but the accuracy of theestimates for individuals with adverse effects is unknown."

- Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. Estimating the Causal Effect of Low Levels of Fine Particulate Matter on Hospitalization. Epidemiology. 2017 Sep;28(5):627-634. doi: 10.1097/EDE.0000000000000690. This study claims to find substantial effects of changes in PM2.5 at concentrations below 12 $\mu g/m^3$ on changes in hospitalization rates, even though no actual changes in exposures were made, no changes in hospitalization rates were observed, and no actual exposure measurements of exposures for hospitalized patients were available.

References on effects of exposure estimation errors include the following:

- Basagaña X, Aguilera I, Rivera M, Agis D, Foraster M, Marrugat J, Elosua R, Künzli N. Measurement error in epidemiologic studies of air pollution based on land-use regression models. Am J Epidemiol. 2013 Oct 15;178(8):1342-6. doi: 10.1093/aje/kwt127.
- Cox LAT. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ Res. 2018 Jul;164:636-646. doi: 10.1016/j.envres.2018.03.038
- Rhomberg LR, Chandalia JK, Long CM, Goodman JE. (2011) Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit Rev Toxicol. Sep;41(8):651-71. doi: 10.3109/10408444.2011.563420.
- Vlaanderen J, Portengen L, Chadeau-Hyam M, Szpiro A, Gehring U, Brunekreef B, Hoek G, Vermeulen (2018) Error in air pollution exposure model determinants and bias in health estimates. R. J Expo Sci Environ Epidemiol. 2018 Jun 8. doi: 10.1038/s41370-018-0045-x.

Study design, data collected, or analyses performed do not permit threats to internal validity to be tested and refuted.

Only studies with sound designs that permit plausible threats to internal validity to be tested and refuted should be accepted as providing useful evidence for making causality determinations (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). Observational studies that lack sound quasi-experimental designs do not permit valid causal inferences to be drawn and should not be included as sources of evidence in the ISA. Even if a sound design is used, available data may not suffice to refute plausible threats to internal validity of conclusions so that, again, no valid causal inferences can be drawn. If this is the case, then the study should be excluded from use in causal determination. The ISA should clearly document which threats to internal validity have been tested and refuted, via what tests and with what confidence levels. The ISA should exclude conclusions from observational studies from use in making causal determinations if one or more plausible threats to the internal validity of their conclusions have not been refuted (e.g., threats from history, maturation, testing, instrumentation, regression to the mean, selection, experimental mortality, and interactions of threats). Possible design flaws that could prevent refutation of threats to internal validity include the following:

- Appropriate comparison groups are not used; study design not valid for causal inference
- Hold-out samples or validation samples not used to validate conclusions

- Negative controls not used to validate conclusions (e.g., by showing that exposure-response relationship is stronger for endpoints such as cardiovascular or respiratory mortalities or morbidities than for auto accidents)

Study design, data collected, or analyses performed do not allow external validity to be established and correct generalizations to target populations to be made.

Most PM health effects studies that estimate PM C-R functions do not specify the conditions under which the estimated C-R function holds, such as causally relevant weather conditions (including daily temperature extremes and humidity) over the past 2 weeks, income and SES variables, and other location-specific covariates that modify the PM C-R function. However, weather and location-specific variables are now known to strongly affect C-R associations between PM exposure and health effects, as documented in the previous discussion of temperature effects. For example,

- Kioumourtzoglou et al. (2016) op. cit. "observed a higher association between long-term PM2.5 exposure and mortality in warmer cities… and lower effects with increasing median household income and %white residents."
- Fang et al. (2017) op. cit. found that "The effect of PM2.5 on non-accidental mortality differed under specific extreme weather conditions and SWTs [synoptic weather types].

In the absence of conditioning information specifying the weather conditions, socioeconomic conditions, and other location-specific conditions that affect PM C-R functions, it is unclear how or whether a C-R function estimated from the conditions in a study can be applied elsewhere. (This is the previously discussed problem of "transportability," generalization, or external validity.) Appropriately designed and analyzed studies and data sets can be used to solve this "transportability" problem, as previously noted (e.g., Pearl J, Bareinboim E (2018), Transportability across studies: A formal approach https://ftp.cs.ucla.edu/pub/stat_ser/r372.pdf). Conclusions from studies that do not provide the transportability conditions needed for valid generalization cannot be applied with confidence to conditions that differ from those in the original studies in causally relevant ways.

Study conclusions are sensitive to unstated, untested, unverified, or mistaken assumptions.

Many studies cited by the Draft ISA present conclusions that depend on unverified assumptions such as the following:

- Assumption of no unobserved confounders. This assumption is often made but not tested (and, some authors assert, wrongly, that it is untestable. Methods for testing it are discussed later.)
- "Positivity" assumption that all members of the exposed population can receive any level of exposure is made but not tested.
- Counterfactual assumption that an observed change in health or welfare effects would not have occurred had it not been for a preceding change in exposure. Again, this assumption is sometimes made without being tested; the previously discussed Dublin study is an example.)
- Exchangeability assumption that a comparison group is exchangeable with a treatment group. This should be tested, e.g., by showing that group membership is conditionally independent of all other variables, and hence unpredictable from them, given exposure histories. (If this screening test fails,

invariant causal prediction (ICP) tests and transportability formulas can be used to allow comparisons even without exchangeability.)

The ISA should list the critical unverified assumptions on which each study's stated results and conclusions depend and assess whether the assumptions are consistent with data. For example, Makar et al. (2017) state that "Finally, no unmeasured confounding implies that our full set of available covariates (p=122) is adequate to adjust for residual confounding. This assumption is not testable, but we argue that it is unlikely that there exists covariates that are uncorrelated with the p=122 observed covariates and that can lead to confounding bias." (Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. Estimating the causal effect of low levels of fine particulate matter on hospitalization. Epidemiology. 2017 Sep;28(5):627-634. doi: 10.1097/EDE.0000000000000690.) In critically evaluating this study, the ISA should note that it did not include weather variables such as same-day and lagged temperatures and humidity. It is plausible that these omitted weather variables are unmeasured strong confounders not significantly correlated with the 122 covariates (mainly income and other SES, age and other demographics, and medical history variables) considered by Makar et al. The hypothesis that these unmeasured confounders do not explain away the reported associations in this study could be tested by adding them (temperature data are readily available, and they are only "unmeasured" in this study in the sense that the authors did not include them) and see whether the reported PM C-R associations are significantly reduced or eliminated after conditioning on them.

More generally, *model-based conclusions that depend on specified models (e.g., Poisson regression models, Cox Proportional Hazard models, conditional logistic regression models, etc.) should always be accompanied by appropriate regression model diagnostics and by sensitivity analyses and characterizations of model uncertainty.* Many air pollution health effects research papers do not take these steps. Nonetheless, uncertainty characterization is a critical component of risk assessment (https://www.nap.edu/read/12972/chapter/10). Free software is available for performing regression diagnostics if authors make their data available, e.g.,

- https://www.statmethods.net/stats/rdiagnostics.html
- Zhang Z. Residuals and regression diagnostics: focusing on logistic regression. Ann Transl Med. 2016 May;4(10):195. doi: 10.21037/atm.2016.03.36.
- http://www.sthda.com/english/articles/36-classification-methods-essentials/148-logistic-regression-assumptions-and-diagnostics-in-r/
- https://cran.r-project.org/web/packages/LogisticDx/LogisticDx.pdf
- http://www.sthda.com/english/wiki/cox-model-assumptions

Non-parametric model ensembles (e.g., random forests) can be used to reduce dependence of results on specific modeling choices and assumptions and to characterize remaining model uncertainty in conclusions about causal impacts (e.g., natural direct effects or total effects) of exposures on health outcomes. In light of the ready availability of methods and software for performing model diagnostics and characterizing model uncertainties and sensitivities of conclusions to untested assumptions, *EPA might exclude from further consideration the results of studies that do not provide these steps* and that do not provide data to enable others to carry them out. Conclusions that have not been shown to be independent of (or robust to) unverified assumptions and modeling choices and assumptions are not necessarily valid or suitable for use in informing policy deliberations and decisions.

Study design or data only address association and do not permit valid inferences about (manipulative) causal effects of changes in exposure on changing risks.

Studies that do not contain data on actual changes in exposures and subsequent changes in effects should be excluded from use in making causal determinations, as they lack the essential data needed to draw valid causal inferences about whether changing exposures changes effects (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). For example, a study by Makar et al. (2017), cited on p. 6-185 of the Draft ISA (Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. Estimating the Causal Effect of Low Levels of Fine Particulate Matter on Hospitalization. Epidemiology. 2017 Sep;28(5):627-634. doi: 10.1097/EDE.0000000000000690) concludes that "changes in exposure to $PM_{2.5}$, even at levels always below the standards, leads to significant increases in hospital admissions for all-cause, cardiovascular and respiratory diseases." However, the study presents no data or analyses of actual (real-world) changes in $PM_{2.5}$ or subsequent observed changes in hospital admissions: its conclusions about effects caused by changes in exposure are based entirely on unverified modeling assumptions. Some of these are implausible, such as that there are no unmeasured confounders (which the authors describe as an untestable assumption), even though the study does not measure or correct for obvious potential confounders such as temperature. Such a study has no known relevance or validity for estimating the effects caused by real-world changes in PM exposures. It should not be used in the ISA to support causal conclusions. The Draft ISA concludes that "In summary, these studies generally support an effect of long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes." In reality, however, such studies offer no support for any conclusions about real-world health effects of changes in PM.

Some investigators argue that static C-R associations can be used to establish causality and to predict how changing exposures would change effects, even without observing real-world changes, by making untestable assumptions. For example, Schwartz et al. (2018) state that "Causal modeling contrasts the results of two potential outcomes: what would have been observed had the entire population been exposed to exposure *a*, vs. observations made had they all been exposed to *a′*. At most, one potential outcome is observed, and various methods provide legitimate surrogates for the unobserved potential outcome under certain assumptions, some of which are untestable" (Schwartz J, Fong K, Zanobetti A (2018) A national multicity analysis of the causal effect of local pollution, NO2, and PM2.5 on mortality. Environmental Health Perspectives. doi.org/10.1289/EHP2732 https://ehp.niehs.nih.gov/doi/10.1289/EHP2732). Arguably, *unverified predictions of harm based on untestable assumptions are not science,* insofar as science depends on *testable* (and potentially falsifiable) theories and predictions that can be compared to data and verified or refuted based on agreement with data. In practice, the untestable-assumptions approach to air pollution health risk assessment leads to causal attributions of unknown validity, as the models needed to correctly predict unobserved potential outcomes are seldom known, and causal impacts estimated using them are often sensitive to model specification errors and uncertainties and to the unknown validity of the assumptions (e.g., Lundin M (2011), *Sensitivity Analysis of Untestable Assumptions in Causal Inference*. www.diva-portal.org/smash/get/diva2:412501/FULLTEXT01.pdf.) Therefore, studies that do not use data on actual changes in exposures and changes in effects should not be used to support conclusions about real-world changes in health caused by changes in air pollution. On the other hand, data on actual changes in exposures and changes in effects in different studies, together with data on causally relevant individual and ecological covariates, can readily be used to make, test, validate, and refine causal models by

applying the principle that the conditional probability of a health effect, given the values of its direct causes in a data set (e.g., PM2.5 exposure and weather, demographic and SES, smoking, and co-morbidity variables), should be constant ("invariant) across studies and policy interventions (Peters J, Buhlmann P, Meinshausen N (2016) Causal inference using invariant prediction: identification and confidence intervals). This is a testable prediction that can potentially be refuted (falsified) by data and that is unlikely to hold by chance alone in multiple diverse settings. Hence it provides a useful foundation for scientific risk assessment of the human health effects caused by exposure and other causes.

Animal experiment or in vitro study identifies changes caused by PM exposures in test systems, but does not show that these changes are relevant to (or likely to cause) adverse effects in humans under real-world conditions.

- For a hypothesized causal chain A → B → C → D, where A is exposure, D is an adverse effect, and B and C are intermediate variables (e.g., biomarkers), studies that establish one or several of the links (e.g., A → B and/or B → C) do *not* constitute valid evidence supporting the hypothesis that changes in A cause changes in D unless and until it is validated that the entire chain transmits effects from A to D. That is, discovering that the link from C to D does not exist, or that its arrow points from D to C (where arrows signify that changes in the quantity at the tail cause changes in the probability distribution of the quantity at the head) would make the evidence that changes in A cause changes in B and that changes in B cause changes in C irrelevant, rather than supportive of the causal hypothesis that changes in A cause changes in D. Likewise, establishing changes caused by PM when it is not known that they will propagate to cause harm stops short of providing evidence that reducing PM will reduce harm.
- Finding that certain changes in PM exposures cause a specific change (such as "increased ROS production by activated alveolar macrophages") in a test system and that a change with that description also increases the risk of an adverse health effect (e.g., COPD or lung cancer) in humans does *not* by itself constitute valid evidence that the changes in PM cause increased risk of the adverse health effect. The reason is that the same description may apply to both healthful and harmful changes (e.g., ROS production may increase as part of a reversible homeostatic response or as part of an irreversible pathogenic response, and the description of the change does not by itself distinguish between these very different contexts). To demonstrate an exposure-related increase in risk, it is necessary and sufficient to show that the identified changes caused by PM exposures in test systems propagate along known causal pathways through causal biological networks in humans to cause changes in adverse effects. This can be done by validating that the changes in exposure cause intermediate changes that, in turn, cause increased risk of adverse consequences. This end-to-end causal connectivity must be established before valid causal inferences can be drawn that increases in exposure increase risk of adverse consequences.

Evaluating Individual Studies

A Suggested Checklist of Methodological Issues for Evaluating Studies and their Conclusions

For each study included in evaluating, synthesizing, and stating conclusions about the policy-relevant science of health effects caused by PM exposures, the ISA should critically evaluate the *internal validity* of the study's conclusions (do they follow from the study design and data analysis presented?) and the

external validity of its conclusions (have they been appropriately generalized and caveated for applications beyond the specific conditions of study?) For a comprehensive evaluation, the ISA should also systematically report how well each study has tested and corrected for each of the following potential threats to valid conclusions:

- Confounding by weather variables
- Unmeasured confounders
- Residual confounding
- Other unmeasured (latent) variables and selection biases
- Errors in estimates and measurements of exposures and covariates.
- Model uncertainty and dependence of conclusions on unverified assumptions
- Multiple testing bias and modeling of time-varying C-R models
- Interactions and dependencies among explanatory variables
- Interindividual heterogeneity
- Generalization of study results

Each of these is discussed next, and selected technical references are provided on constructive methods for dealing with each.

1. **Confounding by weather variables.**
    a. **Short-term confounding by temperature.** Were minimum and maximum daily temperatures with lags out to at least 2 weeks before the occurrence of an adverse health effect considered as potential confounders of short-term C-R associations? Were omissions of lagged temperature extremes and means justified by conditional independence tests showing that they had no detectable effect on the C-R function being estimated?
    b. **Short-term confounding by humidity.** Was daily humidity with lags out to 2 weeks considered and were the lagged values of humidity that were excluded justified by conditional independence tests showing that they had no detectable effect on the C-R function being estimated?
    c. **Longer-term confounding by temperature.** Were longer-term (seasonal and annual average) temperatures controlled for as potential confounders?

2. **Unmeasured confounders.** Did the study use appropriate designs and analyses to correct for effects of unmeasured confounders? References on how to test for and control for effects of unmeasured confounders include the following.
    - Best N, Hansell AL. Geographic variations in risk: adjusting for unmeasured confounders through joint modeling of multiple diseases. Epidemiology. 2009 May;20(3):400-10. doi: 10.1097/EDE.0b013e31819d90f9. This paper proposes that "Joint modeling of multiple diseases can be used to investigate geographic variations in risk. **These models reveal patterns that are adjusted for the effects of shared area-level risk factors for which no direct data are available**."
    - Carnegie NB, Harada M, Hill JL. (2016) Assessing sensitivity to unmeasured confounding using a simulated potential confounder. Journal of Research on Educational Effectiveness 9(3) https://www.tandfonline.com/doi/full/10.1080/19345747.2015.1078862. "Attempts to infer

causality using nonexperimental studies are vulnerable to violations of the assumption that requires, colloquially speaking, that all confounders have been measured. Rather than abandoning the goal of causal inference using nonexperimental data, methods that allow researchers to explore sensitivity of their inferences to violations of this assumption can act as a middle ground. … T**his article presents a set of graphical and numeric tools to investigate the sensitivity of causal estimates in nonexperimental studies to the presence of an unmeasured confounder**…. Our approach to assessing sensitivity to an unmeasured confounder has two primary advantages over similar approaches (with two sensitivity parameters) that are currently implemented in standard software. First, it can be applied to both continuous and binary treatment variables. Second, our method for binary treatment variables allows the researcher to specify three possible estimands (average treatment effect, effect of the treatment on the treated, effect of the treatment on the controls). All tools described in this article are available in an R package called treatSens."

- Ding P, VanderWeele TJ. Sensitivity Analysis Without Assumptions. Epidemiology. 2016 May;27(3):368-77. doi: 10.1097/EDE.0000000000000457. Erratum in: Epidemiology. 2018 May;29(3):e19. "A crucial task in causal inference with observational studies is to **assess the sensitivity of causal conclusions with respect to unmeasured confounding**. In sensitivity analysis, because one is assessing the sensitivity of conclusions to the assumption of no unmeasured confounding, additional untestable assumptions may often seem undesirable and suspect to researchers. **We have introduced a new joint bounding factor that allows researchers to conduct sensitivity analysis without assumptions, that is, we provide an inequality, which is applicable without any assumptions, such that the sensitivity analysis parameters must satisfy the inequality if an unmeasured confounder is to explain away the observed effect estimate or reduce it to a particular level**. We can obtain a conservative estimate of the true causal effect by dividing the observed relative risk by the bounding factor; the method does not assume a single binary confounder or no exposure–confounder interaction on the outcome."
- Dorie V, Harada M, Carnegie NB, Hill J. A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. Stat Med. 2016 Sep 10;35(20):3453-70. doi: 10.1002/sim.6973. The main contribution of this paper is that "When estimating causal effects, **unmeasured confounding and model misspecification are both potential sources of bias. We propose a method to simultaneously address both issues** in the form of a semi-parametric sensitivity analysis. In particular, our approach incorporates Bayesian Additive Regression Trees [BART] into a two-parameter sensitivity analysis strategy that assesses sensitivity of posterior distributions of treatment effects to choices of sensitivity parameters. This results in an easily interpretable framework for testing for the impact of an unmeasured confounder that also limits the number of modeling assumptions. … More often than not, it is impractical to implement randomized experiments to address many of the most interesting causal questions. The alternative approach of using observational studies to draw causal conclusions requires structural as well as functional assumptions. These **structural assumptions are typically not trivially plausible, which motivates analysis of the sensitivity of causal estimates** drawn from observational studies to violations of these assumptions, in particular of ignorability. … However, these goals can be more difficult to achieve if one is forced to rely on parametric models, as the potential for model misspecification introduces its own biases. **We sidestep this issue by allowing for a nonparametric fit** of the relationship between the outcome and the observed covariates via the BART algorithm. This approach appears to be competitive with

existing approaches when no nonlinear confounding exists and to outperform these approaches in the presence of nonlinear confounding. Moreover, **the procedure has been integrated into the treatSens package for the R programming language** available, on the Comprehensive R Archive Network."

- Genbäck M, de Luna X Causal inference accounting for unobserved confounding after outcome regression and doubly robust estimation. Biometrics. 2018 Nov 14. doi: 10.1111/biom.13001. "Causal inference with observational data can be performed under an assumption of no unobserved confounders (unconfoundedness assumption). There is, however, seldom clear subject-matter or empirical evidence for such an assumption. **We therefore develop uncertainty intervals for average causal effects** based on outcome regression estimators and doubly robust estimators, **which provide inference taking into account both sampling variability and uncertainty due to unobserved confounders**. In contrast with sampling variation, uncertainty due to unobserved confounding does not decrease with increasing sample size. The intervals introduced are obtained by modeling the treatment assignment mechanism and its correlation with the outcome given the observed confounders, allowing us to derive the bias of the estimators due to unobserved confounders. We are thus also able to contrast the size of the bias due to violation of the unconfoundedness assumption, with bias due to misspecification of the models used to explain potential outcomes."
- Groenwold RH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. J Clin Epidemiol. 2009 Jan;62(1):22-8. doi: 10.1016/j.jclinepi.2008.02.011. The authors argue that "Methods to quantify unobserved confounding can be categorized in methods with and without prior knowledge of the effect estimate. Without prior knowledge of the effect estimate, **unobserved confounding can be quantified using different types of sensitivity analysis**. When prior knowledge is available, the size of unobserved confounding can be estimated directly by comparison with prior knowledge. CONCLUSION: **Unobserved confounding should be addressed in a quantitative way to value the inferences of nonrandomized intervention studies**."
- Kasza J, Wolfe R, Schuster T. Assessing the impact of unmeasured confounding for binary outcomes using confounding functions. Int J Epidemiol. 2017 Aug 1;46(4):1303-1311. doi: 10.1093/ije/dyx023. "**A critical assumption of causal inference is that of no unmeasured confounding**: for estimated exposure effects to have valid causal interpretations, a sufficient set of predictors of exposure and outcome must be adequately measured and correctly included in the respective inference model(s). **In an observational study setting, this assumption will often be unsatisfied**, and the potential impact of unmeasured confounding on effect estimates should be investigated. **The confounding function approach allows the impact of unmeasured confounding on estimates to be assessed, where unmeasured confounding may be due to unmeasured confounders and/or biases such as collider bias or information bias**. Although this approach is easy to implement and pertains to the sum of all bias, its use has not been widespread, and discussion has typically been limited to continuous outcomes. In this paper, we consider confounding functions for use with binary outcomes and illustrate the approach with an example. We note that confounding function choice encodes assumptions about effect modification: some choices encode the belief that the true causal effect differs across exposure groups, whereas others imply that any difference between the true causal parameter and the estimate is entirely due to imbalanced risks between exposure groups. **The confounding function approach is a useful method for assessing the impact of unmeasured confounding,**

in particular when alternative approaches, e.g. external adjustment or instrumental variable approaches, cannot be applied."

- Marra, G., Radice, R. & Missiroli S. (2014). Testing the hypothesis of absence of unobserved confounding in semiparametric bivariate probit models. Comput Stat 29: 715. https://doi.org/10.1007/s00180-013-0458-x
- Pearl J. An introduction to causal inference. Int J Biostat. 2010 Feb 26;6(2):Article 7. doi: 10.2202/1557-4679.1203. See especially Section 3 on "Coping with unmeasured confounders."
- Sanderson E, Macdonald-Wallis C, Davey Smith G. Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration. Int J Epidemiol. 2018 Apr 1;47(2):587-596. doi: 10.1093/ije/dyx213. "**Negative control exposure studies are increasingly being used in epidemiological studies to strengthen causal inference regarding an exposure-outcome association when unobserved confounding is thought to be present**. Negative control exposure studies contrast the magnitude of association of the negative control, which has no causal effect on the outcome but is associated with the unmeasured confounders in the same way as the exposure, with the magnitude of the association of the exposure with the outcome. A markedly larger effect of the exposure on the outcome than the negative control on the outcome strengthens inference that the exposure has a causal effect on the outcome. .. Measurement error is common in the variables used in epidemiological studies; these results show that negative control exposure studies cannot be used to precisely determine the size of the effect of the exposure variable, or adequately adjust for unobserved confounding; however, they can be used as part of a body of evidence to aid inference as to whether a causal effect of the exposure on the outcome is present." … Due to the unmeasured confounding that is inherent in studies in which it is necessary to use a negative control, **the estimates of regression coefficients are always expected to reflect the confounded association rather than the causal relationship**. In the analysis above, we have shown that measurement error in the exposure and negative control will add a bias which may increase or decrease the difference between the estimated coefficient and the causal relationship … The results we have found mean **we cannot give a general statement about the direction of any bias caused by measurement error in a negative control exposure study**."
- Streeter AJ, Lin NX, Crathorne L, Haasova M, Hyde C, Melzer D, Henley WE. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. J Clin Epidemiol. 2017 Jul;87:23-34. doi: 10.1016/j.jclinepi.2017.04.022. "Motivated by recent calls to use electronic health records for research, **we reviewed the application and development of methods for addressing the bias from unmeasured confounding in longitudinal data**. … Among the 121 studies included for review, **84 used instrumental variable analysis (IVA), of which 36 used lagged or historical instruments. Difference-in-differences (DiD) and fixed effects (FE) models were found in 29 studies**. Five of these combined IVA with DiD or FE to try to mitigate for time-dependent confounding. **Other less frequently used methods included prior event rate ratio adjustment, regression discontinuity nested within pre-post studies, propensity score calibration, perturbation analysis, and negative control outcomes**."
- Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. Am J Epidemiol. 2014 Mar 1;179(5):633-40. doi: 10.1093/aje/kwt303. "Unobserved confounding can seldom be ruled out with certainty in nonexperimental studies. **Negative controls are sometimes used in epidemiologic practice to detect the presence of**

unobserved confounding. … In this paper, **we go beyond the use of control outcomes to detect possible unobserved confounding and propose to use control outcomes in a simple but formal counterfactual-based approach to correct causal effect estimates for bias due to unobserved confounding**."

- VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. Annu Rev Public Health. 2016;37:17-32. doi: 10.1146/annurev-publhealth-032315-021402.
- Zhang Z, Uddin MJ, Cheng J, Huang T. Instrumental variable analysis in the presence of unmeasured confounding. Ann Transl Med. 2018 May;6(10):182. doi: 10.21037/atm.2018.03.37.

3. **Residual confounding.** Were effects of residual confounding appropriately quantified, e.g., using bounds and sensitivity analyses? For example, if a seasonal indicator was used, was confounding by daily temperatures within the same season also controlled for? If a seasonal or annual average temperature was used, was confounding by daily temperatures within the same season or year also controlled for? In practice, controlling for residual confounding Is not merely a matter of methodological rigor, but is essential, since biases introduced by model selection and residual confounding are plausibly of the same approximate size as estimated health impacts of air pollution (Lumley T, Sheppard L (2000) Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses. Environmetrics 11(6): 705-717. Special Issue: Statistical Analysis of Particulate Matter Air Pollution.)
   - Chen K, Wolf K, Hampel R, *et al* OP VII – 2 Does temperature confounding control influence the modifying effect of air temperature in ozone-mortality associations? *Occup Environ Med* 2018;**75:**A14. (Similar methods can be applied to PM2.5.)
   - Flanders WD, Strickland MJ, Klein M. A new method for partial correction of residual confounding in time-series and other observational studies. Am J Epidemiol. 2017 May 15;185(10):941-949. doi: 10.1093/aje/kwx013.
   - Groenwold RH, Klungel OH, Altman DG, van der Graaf Y, Hoes AW, Moons KG; PROTECT WP2 Adjustment for continuous confounders: an example of how to prevent residual confounding. CMAJ. 2013 Mar 19;185(5):401-6. doi: 10.1503/cmaj.120592. Epub 2013 Feb 11.
   - Halonen JI, Blangiardo M, Toledano MB, Fecht D, Gulliver J, Ghosh R, Anderson HR, Beevers SD, Dajnak D, Kelly FJ, Wilkinson P, Tonne C. Is long-term exposure to traffic pollution associated with mortality? A small-area study in London. Environ Pollut. 2016 Jan;208(Pt A):25-32. doi: 10.1016/j.envpol.2015.06.036. (Suggests the potential practical importance of uncontrolled confounding in C-R estimates, albeit for exhaust-related PM2.5.)

4. **Other unmeasured (latent) variables and collider biases.** Did the study use appropriate designs and methods to test for and correct for effects of unmeasured variables? For example, did it test and use invariance properties for causal dependencies, finite mixture distribution models, causal graph criteria or other techniques to quantify or bound the effects of latent variables on the PM C-R function? Did the study avoid collider bias by not conditioning on (e.g. stratifying on) common effects? Did the analysis use appropriate adjustment sets to estimate causal effects without bias?
   - Bobb JF, Claus Henn B, Valeri L, Coull BA. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. Environ Health. 2018 Aug 20;17(1):67. doi: 10.1186/s12940-018-0413-y.
   - Hu ZG, Wong CM, Thach TQ, Lam TH, Hedley AJ. Binary latent variable modelling and its application in the study of air pollution in Hong Kong. Stat Med. 2004 Feb 28;23(4):667-84.)

- Ma Z, Li D, Zhan S, Sun F, Xu C, Wang Y, Yang X. Analysis of risk factors of metabolic syndrome using a structural equation model: a cohort study. Endocrine. 2018 Aug 21. doi: 10.1007/s12020-018-1718-x.
- Pearl J. An introduction to causal inference. Int J Biostat. 2010 Feb 26;6(2):Article 7. doi: 10.2202/1557-4679.1203
- Shook-Sa BE, Chen DG, Zhou H. Using structural equation modeling to assess the links between tobacco smoke exposure, volatile organic compounds, and respiratory function for adolescents aged 6 to 18 in the United States. Int J Environ Res Public Health. 2017 Sep 25;14(10). pii: E1112. doi: 10.3390/ijerph14101112.
- Salway R, Lee D, Shaddick G, Walker S. Bayesian latent variable modelling in studies of air pollution and health. Stat Med. 2010 Nov 20;29(26):2732-42. doi: 10.1002/sim.4039
- Strand M, Sillau S, Grunwald GK, Rabinovitch N. Regression calibration with instrumental variables for longitudinal models with interaction terms, and application to air pollution studies. Environmetrics. 2015 Sep;26(6):393-405.
- Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2016 Dec 1;45(6):1887-1894

5. **Errors in estimates** and measurements of exposures and covariates. Did the study use appropriate errors-in-variables methods or other techniques to correct for differences between true and estimated exposure values and between true and estimated values of other variables? Did it quantify (or bound) the magnitudes and effects of errors in exposure estimates, e.g., using sensitivity analyses and uncertainty analyses?
   - Alexeeff SE, Carroll RJ, Coull B. Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. Biostatistics. 2016 Apr;17(2):377-89. doi: 10.1093/biostatistics/kxv048.
   - Baxter LK, Wright RJ, Paciorek CJ, Laden F, Suh HH, Levy JI. 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. doi: 10.1038/jes.2009.5.
   - Keller JP, Chang HH, Strickland MJ, Szpiro AA. Measurement error correction for predicted spatiotemporal air pollution exposures. Epidemiology. 2017 May;28(3):338-345. doi: 10.1097/EDE.0000000000000623.
   - Samoli E, Butland BK. Incorporating measurement error from modeled air pollution exposures into epidemiological analyses. Curr Environ Health Rep. 2017 Dec;4(4):472-480. doi: 10.1007/s40572-017-0160-1.
   - Silva R. (2016) Comments on "Causal inference using invariant prediction: identification and confidence intervals" by Peters, Buhlmann and Meinshausen. http://www.homepages.ucl.ac.uk/~ucgtrbd/papers/comment_peters.pdf

6. **Model uncertainty and dependence of conclusions on unverified assumptions.** Were conclusions (e.g., about the shapes of C-R functions) shown to hold with high confidence in the absence of unverified modeling assumptions, e.g., by using non-parametric model ensembles (such as partial dependence plots and individual conditional expectation plots) and sensitivity analyses of the dependence of conclusions on any remaining unverified conclusions? If regression models were used, were regression diagnostics and results of model specification tests presented? Important

advances for dealing with model uncertainty (e.g., initially unknown form of the C-R function) without relying on unverified modeling assumptions use non-parametric methods and nonparametric model ensembles (e.g., random forest). The following references discuss and illustrate these and other techniques and demonstrate their practical importance for various pollutants, covariates, and health effects.

- Pannullo F, Lee D, Waclawski E, Leyland AH. How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using Bayesian model averaging. Spatiotemporal Epidemiol. 2016 Aug;18:53-62. doi: 10.1016/j.sste.2016.04.001.
- Fang X, Li R, Kan H, Bottai M, Fang F, Cao Y. Bayesian model averaging method for evaluating associations between air pollution and respiratory mortality: a time-series study. BMJ Open. 2016 Aug 16;6(8):e011487. doi: 10.1136/bmjopen-2016-011487.
- Smith AE, Glasgow G. Integrated uncertainty analysis for ambient pollutant health risk assessment: a case study of ozone mortality risk. Risk Anal. 2018 Jan;38(1):163-176. doi: 10.1111/risa.12828.
- Cox LA Jr.(2018). Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. Crit Rev Toxicol. Nov 15:1-31. doi: 10.1080/10408444.2018.1518404

7. **Multiple testing bias and modeling of time-varying C-R models, effects and interactions** among variables. Were interactions and statistical dependences among variables for various lags quantified and displayed, e.g., using methods based on C&RT trees or dynamic Bayesian networks? Was multiple testing bias accounted for (e.g., in identifying effects of PM on health in some seasons but not others, or in some subpopulations but not others, or under some weather conditions but not others)?
   - Gass K, Klein M, Sarnat SE, Winquist A, Darrow LA, Flanders WD, Chang HH, Mulholland JA, Tolbert PE, Strickland MJ. Associations between ambient air pollutant mixtures and pediatric asthma emergency department visits in three cities: a classification and regression tree approach. Environ Health. 2015 Jun 27;14:58. doi: 10.1186/s12940-015-0044-5. The current Draft ISA mentions this study on p. 5-115. A result of the study is that "**No single mixture emerged as the most harmful**. Instead, the rate ratios for the mixtures suggest that all three pollutants drive the health association, and that the rate plateaus in the mixtures with the highest concentrations. In contrast, the results from the comparison model are dominated by an association with ozone and suggest that the rate increases with concentration. …Examination of the differences between the C&RT and comparison model results suggests that **the two approaches for modeling multipollutant exposures lead to different conclusions** regarding the roles of individual pollutants. In the comparison model, **joint effects are driven by O3 concentration**." (Emphases added.) It is not clear that these points are well captured by the ISA's summary of results from this and other studies, which reads as follows: "In summary, the studies that examined multipollutant mixtures that include PM2.5 indicate that mixtures encompassing days with high PM2.5 concentrations are often those mixtures with the highest risk estimates. Additionally, when comparing single-pollutant PM2.5 results with those based on mixtures, the risk estimate associated with the mixture is relatively similar and, in some cases, larger than that observed for PM2.5."
   - Castner J, Guo L, Yin Y. Ambient air pollution and emergency department visits for asthma in Erie County, New York 2007-2012. Int Arch Occup Environ Health. 2018 Feb;91(2):205-214. doi: 10.1007/s00420-017-1270-7.

- Li G, Sun J, Jayasinghe R, Pan X, Zhou M, Wang X, Cai Y, Sadler R, Shaw G. Temperature Modifies the Effects of Particulate Matter on Non-Accidental Mortality: A Comparative Study of Beijing, China and Brisbane, Australia. Public Health Research p-ISSN: 2167-7263 e-ISSN: 2167-7247 2012; 2(2): 21-27 doi:10.5923/j.phr.20120202.04
- Szyszkowicz M, Kousha T. Emergency department visits for asthma in relation to the Air Quality Health Index: a case-crossover study in Windsor, Canada. Can J Public Health. 2014 Jul 31;105(5):e336-41. (The current Draft ISA mentions this study on p. 5-115.)
- Zeng Q, Li G, Cui Y, Jiang G, Pan X. Estimating Temperature-Mortality Exposure-Response Relationships and Optimum Ambient Temperature at the Multi-City Level of China. Int J Environ Res Public Health. 2016 Mar 3;13(3). doi: 10.3390/ijerph13030279.)

8. **Modeling of interactions and dependencies among explanatory variables and between explanatory and risk variables.**
    a. Were dependencies among exposure and other causes of responses or health effects modeled explicitly so that direct, indirect, total, and other causal effects of exposure on risk (or of C on R in C-R models) could be isolated and displayed (e.g., using partial dependence plots)?
    b. Were formal tests performed for identifiability of the (manipulative causal) PM C-R functions from available data, and the results reported? Were confounding effects of socioeconomic gradients adequately modeled? (Milojevic A et al.. Socioeconomic and urban-rural differentials in exposure to air pollution and mortality burden in England. Environ Health. 2017 Oct 6;16(1):104. doi: 10.1186/s12940-017-0314-5.)
    c. Were interactions among air pollution and other explanatory variables (such as noise, green space, income, and activity level) quantified and modeled so that the effects of air pollution could be distinguished from the effects of other variables? (Cole-Hunter T et al. Estimated effects of air pollution and space-time-activity on cardiopulmonary outcomes in healthy adults: A repeated measures study. Environ Int. 2018 Feb;111:247-259. doi: 10.1016/j.envint.2017.11.024.)

    Relevant references include the following:
- Causal graph models, directed acyclic graph (DAGs) methods, structural equations models (SEMs) and related methods now provide excellent techniques for quantifying dependencies among explanatory variables (Cox LA Jr.(2018). Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. Crit Rev Toxicol. Nov 15:1-31. doi: 10.1080/10408444.2018.1518404).
- Pearl J. An introduction to causal inference. Int J Biostat. 2010 Feb 26;6(2):Article 7. doi: 10.2202/1557-4679.1203.

9. **Interindividual heterogeneity** in C-R functions. Was interindividual heterogeneity in C-R functions quantified and visualized, e.g., using finite mixture distribution models, latent variable analysis, or individual conditional expectation (ICE) plots?
- Susan Athey, Guido Imbens Recursive partitioning for heterogeneous causal effects Proc Natl Acad Sci U S A. 2016 Jul 5; 113(27): 7353–7360.
- Kim C, Daniels M, Li Y, Milbury K, Cohen L. A Bayesian semiparametric latent variable approach to causal mediation. Stat Med. 2018 Mar 30;37(7):1149-1161. doi: 10.1002/sim.7572

- Li X, Xie S, McColgan P, Tabrizi SJ, Scahill RI, Zeng D, Wang Y. Learning subject-specific directed acyclic graphs with mixed effects structural equation models from observational data. Front Genet. 2018 Oct 2;9:430. doi: 10.3389/fgene.2018.00430).
- https://cran.r-project.org/web/packages/ICEbox/ICEbox.pdf

10. **Generalizaton of study results**. Were transportability tests and formulas used to appropriately generalize study results? Section 1.5.3, p. 1-49 of the Draft ISA states that "conducting C-R and threshold analyses is challenging due to the (1) limited range of available concentration levels (i.e., sparse data at the low and high end); (2) **heterogeneity of (at-risk) populations** (between cities); and (3) influence of measurement error." Important advances since 2009 in methods for valid extrapolation of C-R analyses that adjust for heterogeneity of at-risk populations between locations and that help to overcome some of the challenges of limited ranges of data include greatly improved theories and algorithms for transportability and transport formulas for generalizing study results. Relevant technical references include the following.
    - Bareinboim E, Pearl J. Causal transportability with limited experiments. In Proceedings of the 27th AAAI Conference on Artificial Intelligence, pp. 95-101, 2013. http://ftp.cs.ucla.edu/pub/stat_ser/r408.pdf
    - Hernán MA, Vanderweele T. On compound treatments and transportability of causal inference. *Epidemiology*. 2011; 22:368.
    - Lee S, Honavar V. (2013) m-Transportability: Transportability of a causal effect from multiple environments. Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence. www.aaai.org/ocs/index.php/AAAI/AAAI13/paper/viewFile/6303/7210
    - Schwartz S, Gatto NM, Campbell UB. Transportabilty and causal generalization. Epidemiology: Sep 2011 22(5): 745-6

Throughout the ISA, conclusions from cited studies should not be presented as evidence until their internal and external validity and technical soundness have been carefully, critically, independently, and systematically evaluated and documented as part of the ISA process. Unwarranted, unsound, and unverified conclusions are prevalent in the literature on health effects caused by PM2.5 exposures, and investigators often misrepresent their own work. Examples are discussed next. Common misrepresentations include the following:

- Referring to estimated, guessed-at, or imputed exposures as if they were accurately measured observed values;
- Presenting results of unverified modeling assumptions as if they were empirical findings; and
- Describing hypothetical differences in responses between hypothetical exposure scenarios as if they were observed changes.

Therefore it is important for the ISA not to passively repeat and summarize conclusions or accept them at face value, but rather to actively engage in critical evaluation and synthesis. The above checklist may help to quickly assess the methodological soundness of different studies and whether their conclusions are trustworthy or might instead result from unaddressed issues on this list.

Examples of Critical Evaluation of Individual Studies

We previously proposed the following screening criteria for excluding individual studies from further consideration when their design or analysis precludes valid causal interpretation of their results:

- *Study design, data, or analyses do not control for obvious potential confounders or selection biases that could plausibly explain the study results.*
- *Study does not distinguish between true exposure values and estimated exposure values in analyzing and presenting information.*
- *Study design, data collected, or analyses performed do not permit threats to internal validity to be tested and refuted.*
- *Study design, data collected, or analyses performed do not allow external validity to be established and correct generalizations to target populations to be made.*
- *Conclusions are sensitive to unstated, untested, unverified, or mistaken assumptions.*
- *Study design or data only address association and do not permit valid inferences about (manipulative) causation.*
- *Animal experiment or in vitro study identifies changes caused by PM exposures in test systems, but does not show that these changes are relevant to (or likely to cause) adverse effects in humans under real-world conditions.*

Each study that is included, e.g., because it passes these screening tests, should be evaluated systematically for how well it addresses each of the following methodological issues:

1. Confounding by weather variables
2. Unmeasured confounders
3. Residual confounding
4. Other unmeasured (latent) variables and collider bias
5. Errors in estimates and measurements of exposures and covariates.
6. Model uncertainty and dependence of conclusions on unverified assumptions
7. Multiple testing bias and modeling of time-varying C-R models
8. Interactions and dependencies among explanatory variables
9. Interindividual heterogeneity
10. Generalizaton of study results

For example, a study that controls for obvious confounders such as temperature might do so more or less well, depending on how it deals with lagged values and whether it tested for residual confounding.

How to apply these considerations to evaluate individual studies can be illustrated by the following comments on studies in Table 6-47 of the Draft ISA (p. 6-186), "Characteristics of the studies examining the association between long-term PM2.5 exposures and cardiovascular diseases." The Draft ISA concludes that "In summary, these studies generally support an effect of long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes," but it does not provide a systematic critical assessment of the individual studies. Such a systematic critical review using previously discussed criteria might make the following points.

Table 6-47 Characteristics of the studies examining the association between long-term $PM_{2.5}$ exposures and cardiovascular diseases.

Study	Study Population	Exposure Assessment	Concentration µg/m³	Outcome	Copollutants Examined
Miller et al. (2007) 36 metro areas, U.S. Prospective cohort $PM_{2.5}$: 2000 Follow-up: 1994/98-2002	WHI observational cohort N = 65,893 Median follow-up: 6 yrs	Annual avg of closest monitor (2000) within 10 km of monitor	Median 13.4 IQR 11.6-18.3	CVD event (MI, revascularization, stroke, death from CHD, CBVD) Medical record review by physician adjudicators	Copollutant model: NR Copollutant correlations: NR
†(Chi et al., 2016a) 36 metro areas, U.S. Prospective cohort $PM_{2.5}$: 2000 Follow-up: 1994/98-2005	WHI observational cohort Post-menopausal women 50-79 yrs N = 51,754 Mean follow-up 7.6 yrs	Annual avg (2000) kriging interpolation to estimate concentration at residential address C-V R^2 = 0.88 (Sampson et al., 2013)	Mean: 12.7 (SD: 2.9) IQR: 4.1	CVD Event (MI, stroke, death from CHD or CBVD)	Copollutant model: NR Copollutant correlations: NR
†Makar et al. (2017) Prospective cohort $PM_{2.5}$: 2000-2010 Outcome: 2002-2010	Medicare N = 32,119 MCBS survey participants 65+yrs	Spatiotemporal model incorporating satellite observations of AOD over a 1 x 1 km grid for entire US C-V R^2 = 0.84	Full Cohort Mean: 12 IQR: 3.41 Low pollution cohort Mean: 10.18 IQR: 2.46	Circulatory system HA ICD9: 390-459	Copollutant model: NR Copollutant correlations: NR

Avg = average, CVD = cardiovascular disease, CHD = coronary heart disease, CBVD = cerebrovascular disease, C-V = cross validation, hospital admissions = hospital admission, ICD = International Classification of Disease, MCBS = Medicare current beneficiary survey, MI = myocardial infarction, N, n = number of subjects, NR = not reported, WHI = Women's Health Initiative.

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

- Miller et al. (2007) (www.nejm.org/doi/full/10.1056/NEJMoa054409) report that "**Each increase of 10 µg per cubic meter was associated with a 24% increase in the risk of a cardiovascular event** (hazard ratio, 1.24; 95% confidence interval [CI], 1.09 to 1.41) **and a 76% increase in the risk of death from cardiovascular disease** (hazard ratio, 1.76; 95% CI, 1.25 to 2.47). For cardiovascular events, the between-city effect appeared to be smaller than the within-city effect. The risk of cerebrovascular events was also associated with increased levels of PM2.5 (hazard ratio, 1.35; 95% CI, 1.08 to 1.68). **Long-term exposure to fine particulate air pollution is associated with the incidence of cardiovascular disease and death among postmenopausal women**. **Exposure differences within cities are associated with the risk of cardiovascular disease**. …Aspects of our analytic approach also **reduce the concern over confounding, such as our examination of the between-city and within-city components of exposure**. **We controlled for the factors that vary from city to city** (e.g., imperfectly measured subject characteristics, the composition or toxicity of particulate matter, and particle infiltration) in the analysis, which included a city indicator variable. By investigating many potential covariates, and **by including both within-city and between-city exposures, we provided confirmation of the observed association between long-term exposure to air pollution and cardiovascular disease**. The role of socioeconomic status has received attention in air-pollution epidemiology. Beyond controlling for educational level and household income, our results were not sensitive to further adjustment for occupation or Census-derived measures of income, wealth, or poverty on the basis of ZIP Code. **Neither educational level nor household income significantly modified the relationship between air pollution and cardiovascular disease**, although there was a trend toward greater effects among those with less education."

Although this discussion makes clear that the design and analysis of this study clearly have many strengths, applying the above checklists draws attention to the following limitations:

  - *Obvious confounders not controlled:* The study does not explicitly control for confounding by temperature (or humidity). It is therefore left unclear to what extent the reported associations are caused by confounding by same-day and lagged temperatures and/or humidity. Adding data to show that temperature variations within cities cannot explain the variations in PM-associated mortality rates (if indeed that is the case) could strengthen the results by refuting the threat to internal validity from uncontrolled confounding by temperature.
  - *Study design:* "*Study only addresses association and does not permit valid inferences about (manipulative) causation."* No actual changes in exposures or corresponding changes in responses were observed and analyzed in this study.

- Chi et al. (2016a) (www.ncbi.nlm.nih.gov/pmc/articles/PMC5132637/) state that "**5 µg/m3 higher exposure to PM2.5 was associated with a 13% increased risk of cardiovascular event** [hazard ratio (HR) 1.13; 95% confidence interval (CI): 1.02, 1.26]. Adjustment for SES factors did not meaningfully affect the risk estimate. Higher risk estimates were observed among participants living in low-SES neighborhoods. The most and least disadvantaged quartiles of the NSES score had HRs of 1.39 (95% CI: 1.21, 1.61) and 0.90 (95% CI: 0.72, 1.07), respectively. Conclusions: Women with lower NSES [neighborhood-level SES] may be more susceptible to air pollution-related health effects. **The association between air pollution and cardiovascular disease was not explained by confounding from individual-level SES or NSES**." Limitations of this study that are apparent upon systematic review include the following:

  - *Obvious confounders not controlled:* This study did not control for confounding by temperature (or humidity). Thus, it does not pass the first proposed screening criterion: "*Study design, data collected, or analyses performed do not control for obvious potential confounders or selection biases that could plausibly explain the study results.*"
  - *Study design:* "*Study only addresses association and does not permit valid inferences about (manipulative) causation."*
  - *Omitted confounders:* The authors note that "An individual-level measure of wealth was not available in this data set."
  - *Exposure estimation errors ignored:* Exposures were not measured accurately for any individual in the study. Instead, "For each address, the point-specific annual average PM2.5 concentration was predicted using U.S. Environmental Protection Agency's (EPA) Air Quality System (AQS) and Interagency Monitoring of Protected Visual Environments (IMPROVE) monitoring data for the year 2000 and used to represent ambient PM2.5 concentrations at that address over the entire follow-up." This procedure ignores the fact that individuals with cardiovascular events are more likely to have higher-than-estimated exposures (if exposure is a risk factor for these adverse responses) than individuals who did not. But the stated conclusions, such as ""5 µg/m3 higher exposure to PM2.5 was associated with a 13% increased risk of cardiovascular event" ignore this distinction between real and estimated exposure levels. Errors in exposure estimates are simply ignored. This violates the proposed screening criterion "*Study does not distinguish between true exposure values and estimated exposure values in analyzing and presenting information."*

- *Generalization of study results (external validity):* No transportability conditions are given to allow generalization beyond the specific study population and conditions in the study. As noted by Ahmed et al. (2017) (www.ncbi.nlm.nih.gov/pmc/articles/PMC5226706/), "**the findings of this report can be generalized only to a subset of the population who are female, white, postmenopausal, over 50 years old, and free of CVD at baseline**. In order for the findings to apply to the general population, potential differences related to factors such as sex and race would need to be considered. .. **its applicability to current populations is limited due to the data collection timeline**. Participants were initially enrolled between 1993 and 1998. This enrollment period occurred before a crucial turning point in health care, specifically the 1999 release of the first woman-specific clinical recommendations by the American Heart Association (Lewis et al. 2009; Mosca et al. 2011; Ski et al. 2014). Since the release of the recommendations, there have been major improvements and changes in risk factor awareness, prevention, and treatment of CVD in women." This illustrates the proposed screening criterion "*Study design, data collected, or analyses performed do not allow external validity to be established and correct generalizations to target populations to be made."*

- Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. Estimating the causal effect of low levels of fine particulate matter on hospitalization. Epidemiology. 2017 Sep;28(5):627-634. doi: 10.1097/EDE.0000000000000690. This study states that "To protect public health and welfare against the dangers of air pollution, the U.S. Environmental Protection Agency (EPA) establishes National Ambient Air Quality Standards (NAAQS). In response to mounting evidence demonstrating the harmful effects of exposure to fine particulate matter, in 2012 the EPA enacted more stringent NAAQS for fine particulate matter (PM2.5). … Using a nationally representative sample of Medicare enrollees, **we found that changes in exposure to $PM_{2.5}$, even at levels always below the standards, leads to significant increases in hospital admissions for all-cause, cardiovascular and respiratory diseases**. The **robustness of our results to inclusion of many additional individual level potential confounders adds validity** to studies of air pollution that rely entirely on administrative data." The ISA summarizes the results as follows (p. 6-185): "In an analysis of data from Medicare recipients across the U.S. Makar et al. (2017) examined the association of 2-year PM2.5 concentrations with hospital admissions for diseases of the circulatory system among those with annual average concentrations less than 12 μg/m3. Authors found an increase in circulatory system hospital admissions [HR: 1.06 (95%CI: 1.02, 1.09), cutpoint of μg/m3 and [HR: 1.18 (95% CI 1.10, 1.27) cutpoint of 8 μg/m3]. Positive associations between long-term exposure to PM2.5 and cardiovascular disease were reported in cross-sectional studies (Feng and Yang, 2012; Johnson and 27 Parker, 2009). **In summary, these studies generally support an effect of long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes**."
  Systematic review identifies limitations of this study, including the following:
  - *Obvious confounders not controlled:* The study does not control for confounding by temperature or humidity. This illustrates the first proposed screening criterion for excluding studies: "*Study design, data, or analyses do not control for obvious potential confounders or selection biases that could plausibly explain the study results.*"
  - *Exposure estimation errors:* The authors state that "We then estimate each individual's exposure to PM2.5 by averaging PM2.5 levels across space (from the 1km x1km grid to ZIP code of residence) and across time (for the 2 years prior to the reference date). …In previous work, we reported a ten-fold cross-validation of $R^2 = 0.84$ for daily measurements, at the monitoring sites, for the period 2000 to 2012, and for the entire continental US. This indicates high correlation

between predicted and monitored PM2.5. This correlation is anticipated to be even higher when we aggregate these values across time (day to year) and across space (1kmx 1km grid cells to ZIP code)." In other words, individual exposures were not measured. Yet, conclusions are phrased in terms of actual exposure concentrations, not estimated ones, as in "We found that changes in exposure to PM2.5, even at levels always below the standards, leads to significant increases in hospital admissions for all-cause, cardiovascular and respiratory diseases." This conflation of real and estimated exposure values, while ignoring errors in estimates, illustrates the proposed study exclusion criterion "*Study does not distinguish between true exposure values and estimated exposure values in analyzing and presenting information."* Citing a high correlation between predicted and measured values at monitoring sites is irrelevant for answering the question of how large exposure errors are for individuals. For example, if true values were always 5 times higher than estimated values, the correlation between them would be $R^2 = 1.00$, but this is an irrelevant statistic for quantifying the extent of exposure estimation error.

- *No changes observed*. Although the authors state that "we found that changes in exposure to PM2.5, even at levels always below the standards, leads to significant increases in hospital admissions for all-cause, cardiovascular and respiratory diseases," this is misleading Not only were true exposures of patients in this study not measured, and hence not known to have been "always below the standards," but also the "changes" that are said to have been "found" are entirely hypothetical. No real-world changes were observed or analyzed. Hence, no causal relationship was observed or established between changes in exposure and changes in response. This illustrates the criterion "*Study design or data only address association and do not permit valid inferences about (manipulative) causation.*
- *Conclusions are sensitive to unstated, untested, unverified, or mistaken assumptions.* The authors state that "**Our study uses inverse probability weighting (IPW), enabling us to estimate 'causal' effects**. The results are consistent with existing literature on the adverse health effects of long-term exposure to PM2.5. **We found robust evidence that increasing long-term exposure to PM2.5 (two years average) from levels lower than 12 µg/m3 to levels higher than 12 µg/m3 causally increases all-cause admissions and circulatory admission hazard rates**; and among individuals with exposure levels below 12 µg/m3, exposure to PM2.5 levels above 8 µg/m3 increases all-cause, circulatory and respiratory admission hazard rates." However, the "robust evidence" referred to consists solely of imagining what might happen under different conditions, using untested assumptions. No actual increases in PM2.5 or actual resulting increases in hazard rates were observed and analyzed. The authors explain: "**Throughout, we will be relying on three key assumptions** necessary for making [our] causal statements: the stable unit treatment value assumption (SUTVA), positivity, and the assumption of no unmeasured confounding. … **Finally, no unmeasured confounding implies that our full set of available covariates (p=122) is adequate to adjust for residual confounding. This assumption is not testable**, but we argue that it is unlikely that there exists covariates that are uncorrelated with the p=122 observed covariates and that can lead to confounding bias." Applying systematic critical review criteria can identify limitations of these methods, assertions, and conclusions. For example,
  - The study did not include measurements of temperature or other weather variables (or close surrogates for them). Thus, its untested assumption of "no unmeasured confounding" is implausible. Many studies have found lagged temperatures to be important confounders of PM C-R associations.

- The claim that the assumption of no unmeasured confounding is "not testable," which is repeated in multiple publications cited by the Draft ISA, is incorrect. First, it is clear that weather variables were not measured in this study: no further testing is required to identify these unmeasured confounders. Second, many methods are in fact available for testing and controlling for unmeasured confounding. These include the following:
  - ***Hypothesis testing*** (e.g., Marra, G., Radice, R. & Missiroli S. (2014). Testing the hypothesis of absence of unobserved confounding in semiparametric bivariate probit models. Comput Stat 29: 715. https://doi.org/10.1007/s00180-013-0458-x);
  - ***Confounding functions*** (Kasza J, Wolfe R, Schuster T. Assessing the impact of unmeasured confounding for binary outcomes using confounding functions. Int J Epidemiol. 2017 Aug 1;46(4):1303-1311. doi: 10.1093/ije/dyx023). For supporting R software, see Blackwell M (2018), *causalsens: Sensitivity Analysis for Causal Effects*. https://cran.r-project.org/web/packages/causalsens/vignettes/causalsens.pdf
  - ***Control outcome calibration*** (Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. Am J Epidemiol. 2014 Mar 1;179(5):633-40. doi: 10.1093/aje/kwt303.
  - ***Double negative control*** **(**Miao W, Tchetgen Tchetgen E. (2018) Confounding bridge approach for double negative control inference on causal effects. https://arxiv.org/pdf/1808.04945.pdf) A special case is using future PM levels as negative control exposures to test and reduce bias from unmeasured confounding.
  - ***Doubly robust estimation methods*** (Genbäck M, de Luna X Causal inference accounting for unobserved confounding after outcome regression and doubly robust estimation. Biometrics. 2018 Nov 14. doi: 10.1111/biom.13001);
  - ***Graph methods*** (Pearl J. An introduction to causal inference. Int J Biostat. 2010 Feb 26;6(2):Article 7. doi: 10.2202/1557-4679.1203)
  - ***Instrumental variables***, under certain conditions (Zhang Z, Uddin MJ, Cheng J, Huang T. Instrumental variable analysis in the presence of unmeasured confounding. Ann Transl Med. 2018 May;6(10):182. doi: 10.21037/atm.2018.03.37).
  - ***Joint modeling of multiple diseases*** (Best N, Hansell AL. Geographic variations in risk: adjusting for unmeasured confounders through joint modeling of multiple diseases. Epidemiology. 2009 May;20(3):400-10. doi: 10.1097/EDE.0b013e31819d90f9);
  - ***Negative controls*** (Sanderson E, Macdonald-Wallis C, Davey Smith G. Negative control exposure studies in the presence of measurement error: implications for attempted effect estimate calibration. Int J Epidemiol. 2018 Apr 1;47(2):587-596. doi: 10.1093/ije/dyx213)
  - ***Sensitivity analysis*** (e.g., Groenwold RH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies.; Ding P, VanderWeele TJ. Sensitivity Analysis Without Assumptions. Epidemiology. 2016 May;27(3):368-77. doi: 10.1097/EDE.0000000000000457.)
  - ***Simulation*** (Carnegie NB, Harada M, Hill JL. (2016) Assessing sensitivity to unmeasured confounding using a simulated potential confounder. Journal of Research on Educational Effectiveness 9(3) https://www.tandfonline.com/doi/full/10.1080/19345747.2015.1078862.)

- IPW is often unreliable and biased in practice. Relevant references include the following:
  - Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012 Feb;21(1):31-54. doi: 10.1177/0962280210386207.
  - Austin PC, Stuart EA. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. Stat Methods Med Res. 2017 Aug;26(4):1654-1670. doi: 10.1177/0962280215584401. This paper mentiones that "**Propensity score methods were found to result in biased estimation** of conditional hazard ratios."
  - Shu D, Yi GY. Weighted causal inference methods with mismeasured covariates and misclassified outcomes. Stat Med. 2019 Jan 4. doi: 10.1002/sim.8073. This paper notes that "**Inverse probability weighting (IPW) estimation has been widely used in causal inference. Its validity relies on the important condition that the variables are precisely measured**. This condition, however, is often violated, which distorts the IPW method and thus **yields biased results**." In this study, as discussed above, variables were not precisely measured.

Since studies such as those in Table 6-47 both ignore measurement errors and also omit important confounders, it is worth noting that these two sources of bias can interact. As warned by Fewell et al. (2007), "**Measurement error in confounders will lead to residual confounding**, but this is not a straightforward issue, and it is not clear in which direction the bias will point. Unmeasured confounders further complicate matters.. **With plausible assumptions, effect sizes of the magnitude frequently reported in observational epidemiologic studies can be generated by residual and/or unmeasured confounding alone**." Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol. 2007 Sep 15;166(6):646-55.

These examples for the three studies cited in Table 6-47 suggest that applying systematic review criteria to individual studies can be valuable in identifying important limitations and caveats that should be included in summarizing and interpreting their conclusions. As noted by Dr. Goodman in public comments during the December PM Draft ISA review public meeting, systematically evaluating each study on each of numerous well-specified criteria and documenting the results is not very burdensome, and such systematic reviews can produce highly valuable information for assessing the extent to which study conclusions are sound and independently verifiable. The burden of systematic review is further reduced if studies that clearly meet criteria for exclusion are excluded without further evaluation. For example, simply excluding studies that fail to control for temperature as an important confounder would eliminate many of the studies in Table 11-5 (North American epidemiologic studies of long-term exposure to PM2.5 and mortality) and other tables and figures summarizing evidence the Draft ISA. Further excluding studies that ignore exposure measurement error, or that make claims about how changing PM2.5 exposure would affect mortality or morbidity without analyzing any data on actual changes in PM2.5 exposure or ensuing changes in mortality or morbidity, would eliminate most (possibly all) of these numerous studies. This is not because the methodological bar is being set high. Controlling for temperature and other obvious confounders, applying appropriate errors-in-variables methods, and studying real-world changes in exposures and responses have all been done many times in other areas of epidemiology, but appear not to have been high priorities for many air pollution health

effects studies such as those in Table 11-5. If EPA makes clear that only studies that meet at least such minimal methodological standards will be considered for use as "evidence" in NAAQS cycles, air pollution health effects investigators may put more emphasis on conducting studies from which sound conclusions can be drawn about how changing exposure affects human health and welfare.

Interpreting, Combining, Synthesizing, Reconciling, and Summarizing Individual Studies

The ISA should specify the rules or criteria used to interpret evidence from each study, combine evidence across studies, reconcile conflicting information, and summarize results. The Draft ISA interprets evidence from studies with uncontrolled confounding, ignored exposure measurement or estimation errors, untested assumptions, and conflicting results as supporting conclusions about causal determinations and linearity of C-R functions without providing clearly stated scientific grounds to justify such interpretations. For example, it states that "In an analysis of data from Medicare recipients across the U.S. Makar et al. (2017) examined the association of 2-year PM2.5 concentrations with hospital admissions for diseases of the circulatory system among those with annual average concentrations less than 12 μg/m3. Authors found an increase in circulatory system hospital admissions [HR: 1.06 (95%CI: 1.02, 1.09), cutpoint of 12 μg/m3 and [HR: 1.18 (95% CI 1.10, 1.27) cutpoint of 8 μg/m3]. Positive associations between long-term exposure to PM2.5 and cardiovascular disease were reported in cross-sectional studies (Feng and Yang, 2012; Johnson and Parker, 2009). **In summary, these studies generally support an effect of long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes**." However, *none of these studies controlled for confounding by temperature*. Thus, a different interpretation of the same evidence would be that "In summary, these studies generally support an effect of uncontrolled confounders associated with long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes." No objective basis is provided for choosing between these rival causal interpretations, or for concluding that "these studies generally support an effect of long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes" instead of "these studies generally support an effect of uncontrolled confounders on a variety of pooled cardiovascular outcomes." If there is an objective basis for accepting one of these interpretations and rejecting the other, it should be explicitly stated in the ISA. If not, then the multiple possible alternative interpretations should be presented, rather than selecting one particular interpretation ("an effect of long-term exposure PM2.5 on a variety of pooled cardiovascular outcomes") and ignoring others (e.g., that the reported associations represent effects of uncontrolled confounding).

Similarly, the Draft ISA repeatedly emphasizes the consistency of evidence supporting positive C-R associations, but without providing clear rules or criteria for independently deriving or verifying this conclusion. It uses the phrase "consistent, positive associations" in multiple places. For example, the Executive Summary offers the following summaries:

- For asthma, the Executive Summary (p. ES-9) states that "**The consistent, positive associations observed for asthma and COPD emergency department visits and hospital admissions** are further supported by evidence of increased symptoms and medication use in response to short-term PM2.5 exposure."
- For respiratory mortalities, the Executive Summary (p. ES-10) states that "Evidence of **consistent, positive associations between PM2.5 and respiratory mortality** demonstrate a continuum of respiratory-related effects."

- For total mortality, the Executive Summary (p. ES-13) states that "Recent multicity studies conducted in the U.S., Canada, Europe, and Asia in combination with the single- and multicity studies evaluated in the 2009 PM ISA continue to provide evidence of **consistent, positive associations between short-term PM2.5 exposure and total mortality**."
- The Executive Summary (p. ES-20) also states that "There are many recent epidemiologic studies conducted in diverse geographic locations, encompassing different population demographics, and using a variety of exposure assignment techniques, that continue to report **consistent positive associations between short- and long-term PM2.5 exposure and respiratory and cardiovascular effects and mortality**. This evidence continues to support the large body of previously published epidemiologic studies reporting positive PM2.5 associations with respiratory and cardiovascular effects and mortality and in some cases strengthens and extends the evidence base for other health effects."

Since the studies cited to support these conclusions do not fully control for important confounders such as lagged temperature extremes, it might be reasonable to expect consistent positive associations between PM2.5 and a variety of weather-associated effects such as cardiovascular and respiratory mortality and morbidity. However, the following articles describe patterns different from the consistent positive associations emphasized in the Draft ISA:

- Cortzez Lugo et al. (2015) state that "the few studies that have been published on adults with asthma, on individuals with and without chronic respiratory symptoms, and on adults with chronic obstructive pulmonary disease (COPD) **show inconsistent results of the effects of air pollution**." Cortez-Lugo M, Ramírez-Aguilar M, Pérez-Padilla R, Sansores-Martínez R, Ramírez-Venegas A, Barraza-Villarreal A. Effect of Personal Exposure to PM2.5 on Respiratory Health in a Mexican Panel of Patients with COPD. Int J Environ Res Public Health. 2015 Aug 28;12(9):10635-47. doi: 10.3390/ijerph120910635.
- Fan et al. (2015) note that "Although the relationship between asthma and exposure to fine particulate matter (PM2.5) has been frequently measured, **reported conclusions have not been consistent**." (Fan J, Li S, Fan C, Bai Z, Yang K. The impact of PM2.5 on asthma emergency department visits: a systematic review and meta-analysis. Environ Sci Pollut Res Int. 2016 Jan;23(1):843-50. doi: 10.1007/s11356-015-5321-x.) These authors conclude "that ambient PM2.5 has an adverse impact on asthma ED visits after short-term exposure and that children are a high-risk population when PM2.5 concentrations are high, particularly in warm seasons, during which measures should be taken to prevent PM2.5." Thus, the point here is not that PM2.5 is not associated with asthma and ED visits, but that there is inconsistency in the literature that should be discussed in the ISA.
- Hartiala et al. (2016) concluded that "Exposure **to higher PM2.5 levels was also significantly associated with increased risk of incident myocardial infarction** (hazard ratio 1.33, 95% CI 1.02-1.73, P=0.03) **but not stroke or all-cause mortality**."
- Lipfert (2017) finds that "**No significant associations were reported for cardiovascular mortality**" (Lipfert FW. A critical review of the ESCAPE project for estimating long-term health effects of air pollution. Environ Int. 2017 Feb;99:87-96. doi: 10.1016/j.envint.2016.11.028).
- Janes et al.(2007) found that, at the local scale**, "there is little evidence of an association between 12-month exposure to PM2.5 and mortality." (**Janes H, Dominici F, Zeger SL (2007). Trends in

air pollution and mortality: an approach to the assessment of unmeasured confounding. Epidemiology. Jul;18(4):416-23.)

- Puett et al. (2011) state that "Among this cohort… **the results did not support an association of chronic PM exposures with all-cause mortality and cardiovascular outcomes** in models with time-varying covariates." (Puett RC, Hart JE, Suh H, Mittleman M, Laden F Particulate matter exposures, mortality, and cardiovascular disease in the health professionals follow-up study. Environ Health Perspect. 2011 Aug;119(8):1130-5. doi: 10.1289/ehp.1002921).
- Wang et al. (2018) note that "the ESCAPE Project showed that **PM10, but not PM2.5, had a statistically significant association with pneumonia incidence in early children**" (Wang J, Chen S, Zhu M, Miao C, Song Y, et al. (2018) Particulate Matter and Respiratory Diseases: How Far Have We Gone?. J Pulm Respir Med 8: 465. doi: 10.4172/2161-105X.1000465)
- You et al. (2018) report that "**There is no statistically significant association between either ozone or $PM_{2.5}$ and acute human mortality"** in a large dataset for eight air basins in California for the years 2004-2007, after statistical adjustment for seasonal and weather effects. You C, Lin DKJ, Young SS. PM2.5 and ozone, indicators of air quality, and acute deaths in California, 2004-2007. Regul Toxicol Pharmacol. 2018 Jul; 96:190-196. doi: 10.1016/j.yrtph.2018.05.012
- Zu et al. (2016) state that "**substantial short-term elevation in PM2.5 concentrations from forest fire smoke were not followed by increased daily mortality in Greater Boston or New York City**."

The Draft ISA also describes examples of inconsistent and discordant evidence. For example, page 11-9 of the Draft ISA mentions that "Additionally, in contrast to Ostro et al. (2006), a recent study by Young et al. (2017**) did not provide any evidence of an association between short-term PM2.5 exposure and mortality when examining eight air basins in California.**"

The Draft ISA does not explain the rules or criteria used to assign a label such as "consistent, positive associations," or alternative labels, to bodies of evidence. The final ISA should do so. It is important to acknowledge substantial discordant evidence when it exists because it can be used constructively to inform *accurate generalizations* about exposure-associated risks by clarifying the conditions under which conclusions hold. For example, C-R associations found only in some locations or at certain times of year or in some populations but not others may reveal combinations of conditions that are required for the associations to hold. More generally, the conditional probability of an adverse response occurring in an exposed individual in a given interval of time typically depends on the values of multiple direct causes. In addition to exposure, factors such as age and sex, income and education, same-day and lagged temperature extremes, and obesity and medical history have all been identified as important interacting causes of health responses associated with PM2.5. The dependency of conditional probabilities of response on the direct causes of the response in one or more data sets can be described by a table or model showing these conditional probabilities for different combinations of the values of the direct causes. (In modern causal analysis, this information is usually summarized in a conditional probability table (CPT) or a conditional probability model or model ensemble such as a random forest.) Studies that reach different conclusions about C-R associations between exposure concentrations and conditional probabilities of adverse responses may do so because other variables are different in the populations studied. Seeking to explain observed differences in C-R associations by a single ("invariant") conditional probability table or model that holds in all settings provides a flexible approach for combining multiple diverse sources of causal evidence. Formal methods and algorithms that can help identify such universal explanations for apparently discordant C-R associations based on values of other

variables are being developed (e.g., Heinze-Deml C, Peters J, Meinshausen N. (2017). Invariant causal prediction for nonlinear models. https://arxiv.org/pdf/1706.08576.pdf); Triantafillou and Tsamardinos I (2015) Constraint-based causal discovery from multiple interventions over overlapping variable sets Journal of Machine Learning Research 16: 2147-2205; Tillman RE, Eberhardt F. (2014) Learning causal structure from multiple data sets with similar variable sets. Behaviormetrika 41(1): 41-64). However, even informal approaches that collect data on a variety of relevant factors such as age, sex, education, socioeconomic status, occupation class, and extreme daily temperatures can be used to understand how C-R functions are modified by these factors (e.g., Yang J, Ou CQ, Ding Y, Zhou YX, Chen PY. (2012) Daily temperature and mortality: a study of distributed lag non-linear effect and effect modification in Guangzhou. Environ Health. Sep 14;11:63. doi: 10.1186/1476-069X-11-63).

Many published articles on air pollution health effects state over-generalized conclusions that do not follow from data alone, but that depend crucially on unverified assumptions that might well be untrue. For example, a recent article leads with the following three unwarranted generalizations (Vodonos A, Awad YA, Schwartz J. The concentration-response between long-term PM2.5 exposure and mortality; A meta-regression approach. Environ Res. 2018 Oct;166:677-689. doi: 10.1016/j.envres.2018.06.021):

- "PM2.5-mortality effect is significant below 10 μg/m^3 and above 20 μg/m^3." This claim rests on misinterpreting estimated exposure levels as if they were true exposure levels and on ignoring important unmeasured confounders such as lagged temperatures.
- "Better exposure estimates result in higher effect size estimates." This generalization assumes that the true C-R function has no threshold – an assumption that is commonly made but that has not been demonstrated in studies that do not implicitly assume it. (The many studies that report absence of a threshold typically ignore measurement error – perhaps justifiable for a linear no-threshold model, but not for a model with a threshold (e.g., Cox LAT. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ Res. 2018 Jul;164:636-646. doi: 10.1016/j.envres.2018.03.038). In effect, this assumes the conclusion.)
- "More control for SES results in higher effect size estimates." This contradicts many studies in which controlling for SES variables income reduces or eliminates the estimated effect of PM2.5 (e.g., "Our findings suggest that living in cities with high temperatures and **low socioeconomic status (SES) is associated with higher effect estimates**" (Kioumourtzoglou MA, Schwartz J, James P, Dominici F, Zanobetti A. PM2.5 and Mortality in 207 US Cities: Modification by Temperature and City Characteristics. Epidemiology. 2016 Mar;27(2):221-7. doi: 10.1097/EDE.0000000000000422); "The hazard ratio (HR) for death was 1.021 (95% confidence interval: 1.019, 1.022) per 1 μg m increase in annual PM2.5. …It was **higher in neighborhoods with lower SES** or higher urbanicity. The HR increased with mean summer temperature." (Wang Y, Shi L, Lee M, Liu P, Di Q, Zanobetti A, Schwartz JD. Long-term Exposure to PM2.5 and Mortality Among Older Adults in the Southeastern US. Epidemiology. 2017 Mar;28(2):207-214. doi: 10.1097/EDE.0000000000000614.)

This study concludes that "This meta-analysis provides strong evidence for the adverse effect of PM2.5 on mortality, that studies with poorer exposure have lower effect size estimates, that more control for SES increases effect size estimates, and that significant effects are seen below 10 μg/m^3. The concentration-response function produced here can be further applied in the global health risk assessment of air particulate matter." However, none of these conclusions appears to be justified, for the reasons just mentioned. (The idea that a single "concentration-response function… can be further

applied in the global health risk assessment of air particulate matter" neglects the fact that significantly different C-R functions hold in different places, subpopulations, seasons, and years.) The prevalence of such sweeping but unjustified generalizations in the research literature on PM2.5 health effects makes it especially necessary and valuable for the ISA to provide careful independent critical review and synthesis that recognizes and seeks to reconcile the discrepancies among stated conclusions and generalizations in the literature.

The Draft ISA does not emphasize using discordant data and conflicting conclusions to improve the external validity of its models and findings. For example, page 11-9 of the Draft ISA states that "Additionally, in contrast to Ostro et al. (2006), a recent study by Young et al. (2017) did not provide any evidence of an association between short-term PM2.5 exposure and mortality when examining eight air basins in California. **The difference in results between these two studies could be attributed to: (1) the larger spatial domain over which exposure was assigned in Young et al. (2017), i.e., an air basin (encompassing multiple counties), compared to Ostro et al. (2006), i.e., a single county; (2) the use of only the highest monitor on each day to assign exposure Young et al. (2017) versus the averaging of all monitors over the spatial domain examined Ostro et al. (2006); and (3) the statistical models used in both studies.**" However, the discrepancy is not further analyzed or resolved. It is important to understand whether the difference in results is indeed due to differences in statistical models used and, if so, to determine which modeling approach (if either) produces correct results. Ostro et al. state that "**We used Poisson multiple regression models** incorporating natural or penalized splines to control for covariates that could affect daily counts of mortality, including time, seasonality, temperature, humidity, and day of the week. **We used meta-analyses using random-effects models to pool the observations in all nine counties**. The analysis revealed **associations of PM2.5 levels with several mortality categories**." Since these authors pooled observations across all counties, the Draft ISA's potential explanation (1) ("the larger spatial domain over which exposure was assigned in Young et al.") is probably not correct. It has previously been found that use of Poisson regression can induce significant positive C-R associations (regression coefficients) even in the absence of a causal C-R relationship, due to model specification errors that make PM2.5 useful for improving predictions of mortality by serving as a partial surrogate for omitted lagged temperatures (Cox LA (2017) Do causal concentration–response functions exist? A critical review of associational and causal relations between fine particulate matter and mortality, Critical Reviews in Toxicology, 47:7, 609-637, DOI: 10.1080/10408444.2017.1311838). If the difference in findings between Ostro et al. and Young et al. is indeed due to differences in choice of modeling methods (Poisson regression or time series analysis, respectively), then the ISA should do one of the following:

- *Explain which modeling approach is preferred and why.* For example, the Ostro et al. approach does not present results of sensitivity analyses, regression diagnostics, or model specification tests that indicate that its selection of a Poisson regression model is appropriate for the data. Thus, model specification error is an untested possible explanation for its reported positive C-R associations. By contrast, the Young et al. approach uses time series analysis and presents results of sensitivity analyses to argue that its findings are robust to modeling choices. These considerations might suggest that the Young et al. results are less dependent on untested modeling assumptions.
- *Explain why neither approach is preferred, and state that the conclusions reached depend on the modeling approach selected*. In this case, all equally credible conclusions should be reflected in the ISA summaries and conclusions and in the Executive Summary and Chapter 1.

- *Apply both approaches to the data and perform sensitivity analyses and regression diagnostics*. Select a preferred approach if these additional analyses make clear that one is more appropriate than the other for the data.

The Draft ISA does none of these. Instead it notes that the conclusions reached might depend on the modeling approach selected, and then selects results consistent with Ostro et al. approach to report in summary statements in the Excecutive Summary and Chapter 1 (e.g., "more recently published scientific evidence reaffirms and further strengthens that there is a '*causal relationship*' between both short- and long-term PM2.5 exposure and total mortality. These causality determinations are based on the consistency of findings across a large body of epidemiologic studies and coherence…"). No basis is given for choosing the Ostro et al. conclusions but not the Young et al. conclusions. *The final ISA should apply explicitly stated criteria to select results to include or exclude in its summaries, conclusions, and causal determinations. When results in the literature disagree, the ISA should apply explicitly stated, independently reproducible methods to determine how conflicts should be resolved* – or, if they are not resolved, how the conflicting results should be summarized, synthesized, and reported in the ISA's conclusions and causal determinations.

Validating Conclusions

The final ISA should assure that all of its conclusions are supported by explicit, independently verifiable derivations from stated premises (facts, data, and assumptions). Empirically testable predictions implied by the conclusions and assumptions, and the extent to which they have been tested and verified, should be discussed. The extent to which alternative explanations and interpretations of the same facts and data have been tested and refuted should be discussed, and results of tests of these alternative explanations should be provided.

Specific Comments on Preface and Executive Summary

The Preface, Executive Summary, and Chapter 1 are mostly well written. They describe the aspirations and summarize the major conclusions from the rest of the Draft ISA. However, the content being summarized, presented more fully in subsequent chapters of the Draft ISA, stops short of providing essential scientific information about health and welfare effects that could be reduced or prevented by reducing particulate matter (PM) pollution levels. Instead, consistent with guidance received from previous CASACs, the Draft ISA focuses on assigning causal determination category labels to selected evidence based on the subjective judgments of those involved. These labels have no clear implications for effects of changes in the current NAAQS on human health or welfare. They make no empirically testable or falsifiable predictions about what changes, if any, are needed to reduce current risks and thereby protect human health and welfare from adverse effects of exposure. In this sense, they do not provide scientific information. They merely reflect, via labels with no clear scientific meanings, the consensus opinions of those involved about what should be considered further for possibly regulatory action. But the missing scientific information on changes in public health produced by changes in exposures is needed for well-informed policy deliberations and decision-making. It is crucial for determining what changes, if any, in current NAAQS are needed to protect public health with an adequate margin of safety. Therefore, these comments first discuss recommendations for adding to the scope and contents of the Draft ISA to restore strong, empirically-based, scientific information to the development and contents of the final ISA.

Association vs. Causation and Protection of Human Health and Welfare

The Draft ISA states on page ES-1 that it "is a comprehensive evaluation and synthesis of policy-relevant science aimed at characterizing exposures to ambient particulate matter (PM), and health and welfare effects associated with these exposures." To be most useful for informing risk management and policy deliberations and decisions, however, the ISA should focus not simply on "effects *associated with* these exposures" (emphasis added), but on effects *caused by* exposures, and *preventable or reducible* by reducing exposures. Normative principles of rational decision-making require that decisions be informed about how alternative choices would change outcomes (or the probabilities of outcomes, when uncertainty is important) (Clemen RT. *Making Hard Decisions: An Introduction to Decision Analysis*. Wadsworth Publishing, Belmont, CA 1996 and subsequent editions). Enabling legislation further requires the NAAQS review process to identify whether changes in current standards are required to protect public health with an adequate margin of safety. This requires being informed about whether and how changes in standards will change public health outcomes or risks. Associations do not provide this essential policy-relevant scientific information (Pearl J, (2009) Causal inference in statistics: An overview. Statistics Surveys 3: 96-146). The scope and emphasis of the ISA should therefore be updated to assure that it provides policy-relevant scientific information by identifying and characterizing harmful effects that can be prevented or reduced by reducing PM exposures. This scientific information about preventable harm caused by PM exposures will also be of fundamental importance for the Risk and Exposure Analysis (REA) and Policy Analysis (PA) efforts later in this review cycle.

Comments on specific sentences in the Preface and Executive Summary follow. These comments overlap with the general points already discussed for the Draft ISA more generally.

Preface

Page P-11: "*To address these questions and update the scientific judgments in the 2009 PM ISA (*U.S. EPA, 2009a*), this ISA aims to:*

• Assess whether new information (since the last PM NAAQS review) further informs the relationship between exposure to PM and specific health and nonecological welfare effects?
• Inform whether the current indicators (i.e., PM2.5 for fine particles and PM10 for thoracic coarse particles), averaging times (e.g., 24-hour average, annual average), and levels of the PM NAAQS are appropriate?

In addressing policy-relevant questions, this ISA aims to characterize the independent health and welfare effects of PM, specifically PM2.5"

- The term "the relationship," as used here and throughout the draft ISA in contexts such as "the relationship between exposure to PM and specific health and nonecological welfare effects," is ambiguous. There are many quantitative relationships between exposure to PM and specific effects, including the following:
    - *Descriptive* relationships such as ratios of mean effects to mean exposure concentrations; ratios of differences in mean effects levels to differences in mean exposure concentrations; and regression coefficients for lines drawn through the mean values of estimated measures of exposures and effects

- Various measures of statistical *association* (e.g., how much more frequently do exposure and effects tend to occur together than would be expected to by chance alone?)
- Measures of statistical *information* (does knowledge of exposure help to predict effects better (e.g., with smaller mean squared prediction errors) than they could be predicted otherwise, and, if so, by how much?)
- Measures of statistical *explanation* (how much of the variance in observed effects is explained by differences in exposures?), such as Pearson's or Spearman's rank correlations
- Measures of various types of *causation* (e.g., by *how much* would changing exposure change effects (manipulative causation))
- Measures of different types of causal *effects* (e.g., controlled direct, natural direct, indirect, mediated, and total effects).

Throughout the ISA, every reference to a "relationship" between exposure and response or PM concentration and response should clearly state *which* relationship, specifically, is being referred to.

P-12: "Table P-2 provides a description of each of the five causality determinations and the types of scientific evidence that is [sic] considered for each category for both health and welfare effects."

- To support scientifically well-informed risk management deliberation and decision-making, it is essential to augment these qualitative (category) determinations with corresponding *quantitative* determinations of the fraction of each adverse effect caused by exposure that would be prevented if exposure were reduced or eliminated.
- If this fraction is uncertain for a particular effect, then its probability distribution should be estimated.
- If this fraction depends on other factors (e.g., sex, age, income, education, ethnicity, co-exposures, co-morbidities, recent daily temperatures, etc.) then the conditional probability distribution for its value given the values of other variables on which it depends should be estimated. Technical methods for characterizing the dependence of adverse health effects on exposures in the presence of other causal factors include partial dependence plots and conditional probability tables or models in causal graph models or Bayesian networks and influence diagrams. Relevant technical references for these methods include the following:
  - Cox LA Jr. (2018). Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. Crit Rev Toxicol. 2018 Nov 15:1-31
  - Howard RA. Decision analysis: Practice and promise. Management Science, Vol. 34, No. 6. (Jun., 1988), pp. 679-695. Stable URL: http://links.jstor.org/sici?sici=0025-1909%28198806%2934%3A6%3C679%3ADAPAP%3E2.0.CO%3B2-M.
  - Greenwell BM. (2017) pdp: An R Package for Constructing Partial Dependence Plots. The R Journal. Jun 9(1): 421-436. ISSN 2073-485).
- Table P-2 offers as a definition for its "Causal relationship" category determination that "Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures. That is, the pollutant has been shown to result in effects in studies in which chance, confounding, and other biases could be ruled out with reasonable confidence." But an observed association between exposure and effects can be partly due to confounding, biases, coincident historical trends, and other non-causal factors and partly due to manipulative causation. An association should not be classified as wholly "causal" or wholly not in such cases. Rather the *fraction* of effects that would be prevented by reducing or eliminating exposure (the manipulative causal fraction) should be estimated.

P-18 "Causal relationship: the pollutant has been shown to result in health and welfare 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."

- The meaning of "result in" and criteria for determining whether a pollutant "has been shown to result in effects" should be clearly defined.
- The definition given for "result in" should be applicable to the realistic case in which PM exposure and other factors that are correlated with PM exposure, including sociodemographic and weather variables, jointly cause or contribute to health effects. For example, if the simple regression model

  E(RISK | EXPOSURE, POVERTY) = 0.01*EXPOSURE*POVERTY + 0.5*POVERTY

  were found to describe data from several different studies with relevant exposures accurately and was not found to be inaccurate for any study, and if chance, confounding, and biases could be ruled out with reasonable confidence, would this provide an adequate basis to conclude that "the pollutant has been shown to result in health and welfare effects at relevant exposures?" Why or why not? If the answer is no, what else would have to be considered to make such a determination? The ISA should address these conceptual and definitional issues in sufficient clarity and detail so that different scientists independently applying them to the same data and studies can independently reach the same conclusion.
- To support scientifically well-informed policy deliberations and decisions, the ISA should develop, state, and use definitions of the following core concepts and terms:
  - "causal relationship"
  - "result in"
  - "the relationship" between exposure and response
  - "concentration-response relationship."
- These and other terms could be listed and defined in a technical glossary in the final PM ISA, along with definitions of more refined terms, such as different types of causal relationships and causal effects that have been defined and distinguished in the epidemiological and risk analysis literature.
- All definitions should meet the *clarity test* often used in decision analysis (Howard RA (1988). Decision analysis: Practice and promise. *Management Science*, 34(6):679-695.)
- Several commentators have offered written public comments that express a high degree of comfort and satisfaction with previous practices and that note to the evolution and improvement of the causal determination framework in Table P-2 over the years with the help of previous CASAC committees. These commentators may see little or no need to clarify key concepts and definitions as recommended here. However, normative principles of decision analysis for supporting responsible science-informed decisions and policy deliberations require such clarity. Ambiguous, unstated, or conflicting definitions of these key concepts are not adequate to support scientifically well-informed decisions. Admittedly, informality and lack of clarity in core definitions and concepts may facilitate consensus-building and political or psychological comfort with resulting statements (especially about causality and effect) despite – or because of – their unclear meanings. But they are inadequate for sound scientific work and for scientifically well-informed deliberation and decision-making based on understanding of how changes in NAAQS are likely to change health outcomes. Therefore, clear definitions should be stated. This may require some new conceptual work to precisely define various types of "relationships" and "effects" when multiple causally relevant factors interact in jointly increasing the probability or frequency of undesirable effects.

Executive Summary

*ES-1: "**Purpose and Scope of the Integrated Science Assessment***
This Integrated Science Assessment (ISA) is a comprehensive evaluation and synthesis of policy-relevant science aimed at characterizing exposures to ambient particulate matter (PM), and health and welfare effects associated with these exposures."

- Consider replacing "is a comprehensive evaluation" with "seeks to provide a comprehensive evaluation" or similar language to indicate that this is a goal for the ISA, not a declaration that it has yet been accomplished.
- To provide a comprehensive evaluation and synthesis of policy-relevant science of health effects caused by PM exposures, the scope of the ISA should be expanded to discuss results and implications of *accountability studies* for the effects of observed changes in PM levels on observed health effect. Relevant references for accountability studies include the following:
  - Boogaard H, van Erp AM, Walker KD, Shaikh R. (2017) Accountability Studies on Air Pollution and Health: the HEI Experience. Curr Environ Health Rep. Dec;4(4):514-522. doi: 10.1007/s40572-017-0161-0.
  - Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, Samet J, King G, Dominici F; HEI Health Review Committee. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations. Res Rep Health Eff Inst. 2016 May;(187):5-49.
  - Henneman LR, Liu C, Mulholland JA, Russell AG. (2017) Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. J Air Waste Manag Assoc. Feb;67(2):144-172. doi: 10.1080/10962247.2016.1242518.)
  - Health Effects Institute (HEI). 2013. Did the Irish Coal Bans Improve Air Quality and Health? HEI Update, Summer, 2013. http://www.healtheffects.org/system/files/UpdateSummer2013.pdf.

  The new discussion should address the implications of the Irish Coal Bans accountability study for the following issues:
  - *C-R relationships for PM in different locations.* In Ireland, reducing ambient particulate air pollution by up to 70% and several dozen $\mu g/m^3$ was not found to cause reductions in all-cause or cardiovascular mortality rates despite strong, consistent, coherent etc. associations between levels of PM in air and levels of all-cause and cardiovascular mortality. Are (manipulative causal) C-R relationships for PM in Ireland expected to be different from those in the US? Why or why not?
  - *Testing and validation of causal determination methods.* Before the accountability study was done, would the methods used in the ISA to make causal determinations for health effects of PM exposures have determined that PM was a cause of increased all-cause and cardiovascular mortality risk in Ireland? Why or why not?
    Refinement of causal determination methods. Are any refinements needed to the causal determination methods used in previous PM ISAs to adequately account for the results of recent accountability studies?

ES-6: "In summary, exposure error tends to produce underestimation of health effects in epidemiologic studies of PM exposure, although bias in either direction can occur."

- Please add citations or explanations for this claim. It is not true in general. Instead, estimation errors typically lead to over-estimates of low-dose risks and under-estimates of high-dose risks if the true manipulative causal C-R function has a threshold or threshold-like nonlinearity. These two errors can cause the estimated C-R function to flatten and appear linear even if the true C-R function has a well-defined threshold (e.g., Cox LAT. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ Res. 2018 Jul;164:636-646. doi: 10.1016/j.envres.2018.03.038).

ES-9: "As in the 2009 PM ISA, the current ISA concludes there is a "likely to be causal relationship" between short-term PM2.5 exposure and respiratory effects (Section 5.1)."

- The ISA should augment this qualitative determination with a quantitative discussion of what is currently known about the fraction of short-term respiratory effects that could be prevented by reducing or eliminating PM2.5 exposure.

ES-16: "An examination of the C-R relationship between short- and long-term PM2.5 exposure and health effects can inform both the shape of the C-R curve and whether there is a threshold (i.e., concentration level) below which there is no evidence of an effect of PM2.5 on health."

- This is not usually true when there is substantial estimation error for the concentrations to which individuals are exposed, as is the case for PM, and specifically for PM2.5 studies. Examining a C-R relationship estimated from data with individual exposure estimates containing unmodeled estimation errors does not in general reveal the shape of the true (error-free) C-R curve or whether it has a threshold (i.e., concentration level) below which exposure does not affect health. The draft ISA's discussion of concentration-response (C-R) relationships should be revised throughout to address effects of exposure estimation and measurement errors on estimated C-R functions and on uncertainty about the shapes of true C-R functions. Technical references on the effects of exposure estimation errors on estimated shapes of C-R functions include the following:
    - Rhomberg LR, Chandalia JK, Long CM, Goodman JE. (2011) Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit Rev Toxicol. Sep;41(8):651-71. doi: 10.3109/10408444.2011.563420.
    - Cox LAT. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ Res. 2018 Jul;164:636-646. doi: 10.1016/j.envres.2018.03.038
- The definition of "the C-R relationship" should be clearly stated using standard epidemiological terms such as controlled direct effect, natural direct effect, mediated effect, total effect, etc. There are many C-R relationships, and it is important to be clear about which one(s) are being discussed. Without such a clear specification, it appears that the draft ISA uses the same term, "the C-R relationship," to refer to both natural direct effects and total effects, and perhaps also some controlled direct effects; these should be separate curves. Technical references on different types of effects and how to estimate them include the following:

- Pearl J. (2001) Direct and Indirect Effects. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA: Morgan Kaufmann, 411-420.
- 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).

- It is not clear that any single C-R relationship exists that applies to different areas of the United States (e.g., both west coast and east coast cities). The ISA should address whether a single C-R relationship exists before estimating and applying such an assumed relationship to estimate changes in health risks caused by changes in PM concentrations. The hypothesis that a single manipulative causal C-R relationship exists can be tested using C-R data from different studies by testing whether the property of invariant causal prediction (ICP) holds across the studies. Relevant technical references include the following for testing whether ICP holds across multiple studies include the following:
  - Cox LA Jr. (2018). Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. Crit Rev Toxicol. 2018 Nov 15:1-31. The discussion of external consistency is particularly relevant.
  - Heinze-Deml C, Peters J, Meinshausen N. 2017. Invariant causal prediction for nonlinear models. https://arxiv.org/pdf/1706.08576.pdf
  - Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society Series B, 2016 78(5):947-1012

p. ES-20: "Epidemiologic studies that conducted copollutant analyses show that associations remain relatively unchanged when adjusting for gaseous pollutants and other particle size fractions (e.g., PM10−2.5), addressing a key uncertainty identified in the 2009 PM ISA."

- The ISA should address whether natural direct, controlled direct, and total manipulative causal effects of PM exposures on health outcome probabilities also remain relatively unchanged after adjusting for gaseous pollutants and other particle size fractions
- The ISA should also address whether natural direct, controlled direct, and total causal effects of PM exposures on health outcome probabilities remain relatively unchanged after adjusting for other risk factors such as sociodemographic factors and daily minimum and maximum temperatures over the two weeks preceding the adverse health effects(s) of interest?

p. ES-21: "Evidence continues to support a linear, no-threshold concentration-response relationship, but with less certainty in the shape of the curve at lower concentrations (i.e., below about 8 µg/m3)."

- The ISA should clearly distinguish throughout between *true* exposure concentrations, which are usually unknown, and *estimated* exposure concentrations, which usually contain estimation or measurement errors.
- Wherever the terms "exposure" or "concentration" are used in the ISA, it should be made clear whether the exposures and concentrations referred to are *actual* (true) or *estimated* values. For epidemiological studies, the answer is usually that they are estimated (often with large errors and uncertainties). Much of the epidemiological literature on air pollution health effects, including for PM2.5 specifically, conflates actual and estimated values and ignores errors in estimates. The ISA should be meticulous in avoiding this conflation.
- Current evidence does not support a linear no-threshold (LNT) relationship between *true* concentration and response probability, but only a LNT relationship between *estimated* concentration and response probability. It is now known that even a sharp threshold in the true C-R function is compatible with LNT for the estimated C-R relationship, so evidence supporting LNT for the estimated C-R function does not constitute evidence supporting LNT for the true C-R function. Technical references include the following:
    - Rhomberg LR, Chandalia JK, Long CM, Goodman JE. (2011) Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit Rev Toxicol. Sep;41(8):651-71. doi: 10.3109/10408444.2011.563420.
    - Cox LAT. Effects of exposure estimation errors on estimated exposure-response relations for PM2.5. Environ Res. 2018 Jul;164:636-646. doi: 10.1016/j.envres.2018.03.038
- This conclusion that "*Evidence continues to support a linear, no-threshold concentration-response relationship*" should be revisited and updated if necessary after the draft ISA's discussion of concentration-response (C-R) relationships is revised throughout to address effects of exposure concentration estimation and measurement errors on estimated C-R functions and on uncertainty about their true shapes.
- The ISA's discussion of evidence about LNT assumptions should be updated to address recent advances in understanding of biological mechanisms of PM-induced lung inflammation, such as the role of the NLRP3 inflammasome.
    - Cevallos VM, Díaz V, Sirois CM. Particulate matter air pollution from the city of Quito, Ecuador, activates inflammatory signaling pathways in vitro. Innate Immun. 2017 May;23(4):392-400. doi: 10.1177/1753425917699864.
    - Du X, Jiang S, Zeng X, Zhang J, Pan K, Zhou J, Xie Y, Kan H, Song W, Sun Q, Zhao J. Air pollution is associated with the development of atherosclerosis via the cooperation of CD36 and NLRP3 inflammasome in ApoE-/- mice. Toxicol Lett. 2018 Jun 15;290:123-132. doi: 10.1016/j.toxlet.2018.03.022.
    - Xin L, Che B, Zhai B, Luo Q, Zhang C, Wang J, Wang S, Fan G, Liu Z, Feng J, Zhang Z. 1,25-Dihydroxy Vitamin D3 Attenuates the Oxidative Stress-Mediated Inflammation Induced by PM2.5 via the p38/NF-κB/NLRP3 Pathway. Inflammation. 2018 Nov 14. doi: 10.1007/s10753-018-0928-y.
    - Xu F, Qiu X, Hu X, Shang Y, Pardo M, Fang Y, Wang J, Rudich Y, Zhu T. Effects on IL-1β signaling activation induced by water and organic extracts of fine particulate matter

(PM2.5) in vitro. Environ Pollut. 2018 Jun;237:592-600. doi: 10.1016/j.envpol.2018.02.086.
- Zheng R, Tao L, Jian H, Chang Y, Cheng Y, Feng Y, Zhang H. NLRP3 inflammasome activation and lung fibrosis caused by airborne fine particulate matter. Ecotoxicol Environ Saf. 2018 Nov 15;163:612-619. doi: 10.1016/j.ecoenv.2018.07.076.

Dr. Mark Frampton

Major Comments

Need to re-appoint the CASAC PM review panel. Prior to the release of this draft PM ISA, and without consulting CASAC, EPA disbanded the expert PM review panel that had been previously appointed to assist CASAC in this important review. Over the past 30 years, NAAQS document reviews by CASAC have been assisted by expert review panels that supplement and expand the scientific expertise brought to bear. The seven chartered CASAC members by themselves do not have the breadth and depth of knowledge or expertise in many areas that are necessary to adequately advise the EPA, and to meet the statutory requirement for a thorough and accurate review.

For example, among the current seven chartered CASAC members, there are no experts in epidemiology, and there is inadequate expertise in health outcomes for which there is new evidence linking to PM, including reproductive and developmental outcomes, and neurobehavioral effects.

In order to provide the needed expertise in the review process, EPA should immediately re-appoint the PM review panel, and convene an additional CASAC public meeting to review and discuss the panel's comments, before CASAC finalizes its advice on the current draft ISA.

Study selection and quality assessment are described in general in the ISA Preamble, and more specifically for the PM ISA in Appendix 1. That Appendix indicates that studies are not necessarily excluded from consideration based on quality assessment. What is missing from the current draft PM ISA, as well as the preamble and Appendix 1, is a description of how quality assessments are used in the review process. The text of the ISA occasionally provides quality-related commentary in the text and/or tables, but this seems to be left up to the individual author of that section. The ISA would benefit from a clear description of the process for considering the study quality assessments in the development of the ISA. In addition, it would be helpful to provide answers to the following: Are the quality reviews performed by the section author(s) themselves, or independently? Are there written quality assessments for each study that are available to the author(s)? While the methods for assessing quality appear appropriate, there is currently a gap between study quality assessment and its application in the ISA preparation and subsequent risk assessment process.

Possible pulmonary vascular effects of PM, and cardiopulmonary interactions. In general, the background sections of chapters 5 and 6 ignore the importance of inter-relationships between respiratory and cardiac function. The mechanistic figures showing potential pathways for PM pulmonary and cardiovascular effects should be modified to reflect these considerations. Acute PM-related effects on LV ischemia or function, or effects on pulmonary artery pressure, could present as respiratory effects, with dyspnea. This is especially true for COPD, where many patients have co-existing cardiac disease and/or pulmonary arterial hypertension, and acute exacerbations often have a major cardiac contribution.

Pulmonary vascular effects are a likely pathway, in addition to inflammation and translocation, for both acute and long-term PM effects. Pulmonary hypertension and right sided heart failure are briefly discussed in section 6.2.5, under long-term effects, but there is additional evidence for pulmonary vascular and right heart effects that is not discussed. Also, the findings of the study cited dealing with

diastolic dysfunction (Ohlwein et al., 2016) is related, because RV dysfunction can worsen LV diastolic dysfunction by encroachment on the LV, with impaired filling. This is a pathway leading to clinical findings of acute heart failure, but with preservation of LV systolic function. This is a very common occurence in COPD patients, and a major contributor to exacerbations. There is epidemiological, clinical, and toxicological evidence to support a pathway of pulmonary vascular effects for PM. Only one of the following studies (Aaron et al.) was cited in the ISA, in the context of heart failure in general.

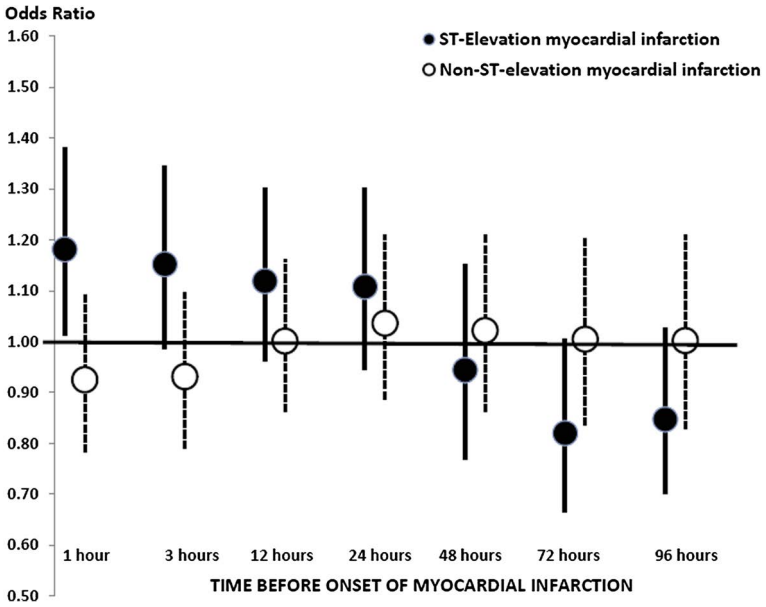
1. Aaron CP, Chervona Y, Kawut SM, Roux AV, Shen M, Bluemke DA, et al. Particulate Matter Exposure and Cardiopulmonary Differences in the Multi-Ethnic Study of Atherosclerosis. Environ Health Perspect. 2016;124(8):1166-73. *Long-term PM2.5 exposures were associated with greater RV mass and RV mass/ end-diastolic volume ratio conditional on the LV.*a
2. Grunig G, Marsh LM, Esmaeil N, Jackson K, Gordon T, Reibman J, et al. Perspective: ambient air pollution: inflammatory response and effects on the lung's vasculature. Pulm Circ. 2014;4(1):25-35.
3. Leary PJ, Kaufman JD, Barr RG, Bluemke DA, Curl CL, Hough CL, et al. Traffic-related Air Pollution and the Right Ventricle. The Multi-ethnic Study of Atherosclerosis. Am J Respir Crit Care Med. 2014;189(9):1093-100. *NO2 as marker of TRAP linked with increased RV mass. Accompanying editorial.*
4. Liu J, Ye X, Ji D, Zhou X, Qiu C, Liu W, et al. Diesel exhaust inhalation exposure induces pulmonary arterial hypertension in mice. Environ Pollut. 2018;237:747-55.
5. Park SH, Chen WC, Esmaeil N, Lucas B, Marsh LM, Reibman J, et al. Interleukin 13- and interleukin 17A-induced pulmonary hypertension phenotype due to inhalation of antigen and fine particles from air pollution. Pulm Circ. 2014;4(4):654-68.
6. Rich DQ, Freudenberger RS, Ohman-Strickland P, Cho Y, Kipen HM. Right heart pressure increases after acute increases in ambient particulate concentration. Environ Health Perspect. 2008;116(9):1167-71. *Panel study of patients with heart failure. Acute increase in PA pressure with PM2.5.*
7. Wauters A, Vicenzi M, De Becker B, Riga JP, Esmaeilzadeh F, Faoro V, et al. At high cardiac output, diesel exhaust exposure increases pulmonary vascular resistance and decreases distensibility of pulmonary resistive vessels. Am J Physiol Heart Circ Physiol. 2015;309(12):H2137-44. Human clinical study of diesel exhaust using echocardiography, with exercise testing and hypoxia.

Page 5-6, line 5. “Activation of sensory nerves in the respiratory tract can trigger local reflex responses resulting in lung irritation.” “Lung irritation” lacks specificity, and may have different meanings for different people. The more accurate term is “airway irritant response” which refers to this whole sensory-mediated process, not just its result. Suggest replacing lung irritation in this sentence with “lung function decrements and airway inflammation”. Elsewhere would replace “lung irritation” with “airway irritant response”.

Page 6-14, line 18. “There were generally consistent results across recent studies looking specifically at MI, and registry studies, which are likely to reduce outcome misclassification, report evidence of positive associations with MI subtypes.” This sentence seems somewhat at odds with the first paragraph on this page, which indicates inconsistencies, especially in the European studies. The interpretation should be further clarified, with justification for disregarding the negative European studies.

Chapter 6, Figure 6-1. The potentially important role for NO and endothelins in PM effects on vascular function are not adequately covered in the figures or the mechanistic paragraphs. There is evidence that PM may act through both, with reduction in NO bioavailability and increased production of endothelins by a variety of cells. There is also the possibility that translocated particles or their components may directly injure the vascular endothelium.

Figure 6-2, page 6-13. The depiction of the associations with MI in the Gardner study appear to be incorrect. The ISA Figure shows minimal associations with very broad CIs, but the figure (below) and data from the paper show a significant effect on STEMI with a 1 hr lag.



Section 6.2.6 is "Cardiac Electrophysiology and Arrhythmia", and section 6.2.11 is "Heart Rate (HR) and Heart Rate Variability (HRV)". But cardiac electrophysiology encompasses HRV, and all are measured using ECG. These sections could be combined, retaining the electrophysiology and arrhythmia heading. Having widely separated sections with closely related outcomes is confusing. Similar comments apply to sections 6.1.4 and 6.1.10.

Section 6.2.8, Peripheral Vascular Disease (PVD), Venous Thromboembolism, Pulmonary Embolism. The diagnosis "peripheral vascular disease" generally refers to disease in the peripheral arterial system, rather than venous disease. The discussion in this section is limited to venous thromboembolism, and does not address arterial PVD, so this term should be removed from the title. In any case, PVD should not be lumped together with venous TE disease; they have different etiologies, pathophysiology, and treatments.

Page 6-176, line 3. "A study of newborns in Massachusetts found elevated SBP with higher PM2.5 averages over the 30-, but not 60- or 90-day periods before birth (van Rossem et al., 2015) while trimester specific associations between PM2.5 and increased SBP increased but confidence intervals were wide…" This sentence is run-on and needs clarification. The description should make clear that the exposure estimates were during the 90 days before birth, but the BP measurements were 30 hours after birth.

Page 6-196, line 31. In the description of the Wilker 2014 study, the ISA states, "Only hyperemic flow velocity was additionally associated with PM2.5 [-1.80 % change (95%CI: -3.45, -0.15)] These effects are relatively large given that normal ranges are between 5-10% (Järhult et al., 2009). This description of the findings of this study is incorrect. The normal range for FMD% is 5-10. Hyperemic flow velocity is expressed in the units of cm/s, not %. Also, it is not clear where the "-1.80% change" comes from. The Wilker 2014 abstract states: "An inter-quartile range difference in PM2.5 (1.99 mug/m(3)) was associated with -0.16% (95% confidence interval [CI] -0.27%, -0.05%) lower flow-mediated dilation% and -0.72 (95% CI -1.38, -0.06) cm/s lower hyperemic flow velocity%."

Page 6-283, line 26. "Weichenthal et al. (2014a) reported positive associations between 2-hour averages of NCs with SBP measurements taken 3 hours post-exposure, but associations with SBP were null." This sentence is contradictory and needs clarification. Associations of UFP with SBP were not significant in this study.

Chapter 7, Metabolic Effects. A better distinction needs to be made between the potential metabolic effects of PM, and metabolic abnormalities as markers of susceptibility to CV effects of PM. These two issues are inappropriately thrown together here. Metabolic effects could include increased insulin resistance, blood glucose, hemoglobin A1c, and incidence of diabetes. Alternatively, having diabetes, obesity, or metabolic syndrome could render increased susceptibility to the CV effects of PM exposure. These are separate, important questions. However, the latter should not be described as "metabolic effects", but considered with other susceptibility factors.

The first paragraph on page 7-18 describes a potential pathway for metabolic effects involving the hypothalamus of the brain, and this important pathway is represented by a blue box in Figure 7-2. However, there is no relationship shown between any of the mechanisms, represented by the green boxes. Given the strong evidence for UFP translocation from the nasal mucosa to the brain (discussed elsewhere in the document), this is a likely pathway for brain effects and should be represented in the figure.

Section 7.1.3, Other Indicators of Metabolic Function, should be re-thought and re-organized. The subheading topics of systemic inflammation and blood pressure have already been reviewed as outcomes, and it is redundant to revisit them here. It is enough for the background to point out the interplay of inflammation in metabolic effects and in obesity, as well as hypertension as a clinical component of the metabolic syndrome, and reference the previous sections.

It is unclear what is meant by "peripheral inflammation" in Figure 7-2 and the accompanying text, and would recommend not using this term. In reading section 7.1.3 (see below), peripheral inflammation seems to be referring to increased inflammation in adipose tissue in various organs, which could have important implications for obesity and metabolic responses. This should be stated more clearly. In addition, this evidence would support a pathway that differs from the current pathways in the figures, suggesting that PM exposure may lead to focal or organ-specific inflammation/oxidative stress, that could be mediated by translocated PM or their components.

Figure 7-2 is incomplete in several aspects. The text describes a potential pathway for metabolic effects involving the brain, but this important pathway is not represented in Figure 7-2. UFP have been

shown in animal models to translocate to the brain via the olfactory nerves. This is a pathway different from the ANS effects of irritant nerve stimulation in the respiratory tract.

The biological plausibility section and Figure 7-2 **do not adequately address the distinctions and differences between type 1 and type 2 diabetes**.

Section 7.2.10, Metabolic Disease Mortality. The title of this section, and some of the text, are a bit misleading. People don't often die of "metabolic disease" (although there are certainly deaths from diabetic ketoacidosis). Their metabolic conditions increase risk for mortality from a variety of causes, from cardiovascular deaths to pneumonia and other infections. The Pope 2014 paper described in this section looks at cardiovascular mortality, and examines whether metabolic disease such as diabetes, contribute to the PM risk for CV mortality. This issue fits best in Chapter 12.

Chapter 8. UFP short-term nervous system effects. Page 8-82. The Liu et al 2017 human clinical study is mistakenly characterized as showing an effect on the HPA axis. Again, the p-value was <0.1, not <0.05. From abstract of the Liu study: "Ultra fine CAP was not significantly associated with changes in any blood and urinary neural biomarkers examined."

UFP long-term effects. 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 most of the animal studies that provide coherence were done by a single group in a single location.

Figure 9-1. Vascular effects could mediate erectile dysfunction, independent of classic systemic inflammation. Progression of atheromatous disease is a possible long term intermediary. Add Vascular/endothelial effects to first set of blue boxes.

Page 9-4. "Inhalation of PM2.5 can result in translocation of particles or soluble factors from the lungs (see Chapter 5) which then can increase respiratory tract inflammation…" The sequence is likely wrong here. Particles in contact with airway epithelium initiate airway inflammation, in part via chemokine production by the epithelial cells. That takes a few hours to develop, while transport of particles likely starts before airway inflammation is fully developed. The translocated particles that enter the pulmonary capillary bed are quickly transported to the left heart and then the systemic circulation. This sentence seems to make the assumption that translocation causes pulmonary inflammation, which does not accurately represent the pathophysiology.

Table 9-6, page 9-29. For the Kloog et al. 2012 study, the last column indicates the effect estimate is 1.03, with 95% CIs of 0.54, 0.63. These values are incorrect. According to the Abstract and Table 3 of the publication, the odds ratio of premature birth was 1.06 (1.01 to 1.13).

Figure 9-1. Vascular effects could mediate erectile dysfunction, independent of classic systemic inflammation. Progression of atheromatous disease is a possible long term intermediary. Add Vascular/endothelial effects to first set of blue boxes.

Page 9-4. "Inhalation of PM2.5 can result in translocation of particles or soluble factors from the lungs (see Chapter 5) which then can increase respiratory tract inflammation…". The sequence is likely wrong

here. Particles in contact with airway epithelium initiate airway inflammation, in part via chemokine production by the epithelial cells. That takes a few hours to develop, while transport of particles likely starts before airway inflammation is fully developed. The translocated particles that enter the pulmonary capillary bed are quickly transported to the left heart and then the systemic circulation. This sentence seems to make the assumption that translocation causes pulmonary inflammation, which does not capture the pathophysiology.

Page 9-6, line 23. The Tallon et al. study is described as showing "…positive associations between exposure to annual PM2.5 concentrations and erectile dysfunction in men aged 57−85 years (OR: 1.26; 95% CI: 0.81, 1.96)." Although the OR is positive, the 95% CI includes 1, so the findings are not statistically significant. Highlighting this in the ISA as a positive study, without further qualification, is misleading.

The words fecundity and fecundability are used interchangeably in this section. Suggest changing the latter to the former wherever it occurs.

Figure 10-2 does not accurately reflect the likely pathways for lung cancer. The current emphasis in the figure is on transport of particles and systemic or brain effects. However, the most relevant pathway is direct effects of PM or its components on the airway epithelium. While airway inflammation may be involved, direct mutagenic, genotoxic, and epigenetic effects on the airway epithelium are likely more important. Systemic inflammation and particle translocation away from the lung are not relevant for lung cancer.

Page 10-9. "…an in vivo study by Sato et al. (2003) reported increased DNA adducts in lung, liver, and nasal mucosal tissues after inhalation exposure to urban roadside air. Because this study evaluated effects of exposure to a mixture of PM and gases, it does not inform the current ISA, which identifies the hazard for effects after exposures to only the PM component of complex mixtures…". Virtually all epi studies involve exposures to mixtures of PM and gases, and yet they can and do inform the PM ISA. The issue with the study in this case is not the exposure to a mixture, but that PM concentrations in the roadside air were not quantified. One could therefore argue that this study should not be included in the ISA since it does not meet the screening criteria stated in the Preface, page P-14, indicating the focus is on studies that "…(1) include a composite measure of PM or (2) characterize PM and apply some approach to assess the direct effect of PM when the exposure of interest is a source-based mixture…"). . At the very least, the preceding statement should be edited to clarify the limitation.

Also, page 10-35 line 3. Similar statement, no measurement of PM. Would not mention this study, or at least correct the statement.

Page 10-49 line 27. This statement is incorrect: "Specifically, an assessment of adenocarcinoma, the only subtype that develops in nonsmokers…" Adenocarcinoma is not the only type of lung cancer that occurs in nonsmokers. "Only" should be changed to "predominant". The same applies to page 10-53, line 11, and page 10-74, line 29.

Chapter 12

This chapter delineates the approach to considering the evidence for at-risk groups and populations. It is an improvement on the approach taken in the 2009 ISA, and clearly presents the rationale and evidence base for the conclusions. The 4-level grading of the conclusions is logical and reasonable, and parallels the approach taken for causality determinations.

Minor/Editorial Comments

Page P-18. The bulleted list of definitions of causal relationships on this page could be removed, as it duplicates information in Table P-2, page P-12.

Table 5-30, page 5-232. "Mild to moderate individuals with asthma" should be "Individuals with mild to moderate asthma".

Page ES-13, footnote 31. "Whole PM exposures" is a poor terminology for "exposures that contain both PM and gaseous pollutants". "Whole atmosphere" may be more descriptive.

There are several places in the ISA where "Section 0" is referenced. Presumably this is a placeholder that needs to be corrected/completed. Examples: Page 5-5, line 13; page 5-8, line 13; Table 5-49, page 5-310.

Variable Figure quality. See Fig. 5-4, page 5-25.

Some CHE studies have failed to find BP elevations.

Page 167, line 3. The study being referenced is missing here. Judging from the text, it seems to be Aaron et al., 2016.

Page 6-16, line 20. Provide the reference referred to here.

Section 6.1.5, Page 6-41. There should be a concluding sentence to the first paragraph indicating that there are new studies since the 2009 review.

Page 6-56, line 2. The reference should be "Gong Jr. et al.".

Page 6-56, line 31. "…although it was noted that assessing changes in blood pressure in the HF group is difficult given beta-blocker use." Assessing the changes is not difficult; the problem is that beta-blocker use may blunt the effect.

Page 6-60, line 18. "…animal toxicological studies that provide biological plausibility for these associations by demonstrating changes in hemodynamics (e.g., an increase in coagulation factors) following short-term PM2.5 exposure…" "Hemodynamics" refers to blood circulation, including blood flow, pressure, and rheology, not levels of coagulation factors or other soluble blood components.
Page 6-79, line 14. HFn needs to be defined.

Page 6-91, line 5. "…increase the potential for an embolism." The major concern is the increased potential for thrombus formation obstructing blood flow, especially in diseased coronary arteries. That is the most common cause of acute MI.

Page 6-148, line 11. This sentence is incomplete and unclear.

The long-term CV effects sections contain frequent missing words, incomplete sentences, and grammatical errors. The document as a whole would benefit from a careful editorial review.

Page 6-176, line 3. "A study of newborns in Massachusetts found elevated SBP with higher PM2.5 averages over the 30-, but not 60- or 90-day periods before birth (van Rossem et al., 2015) while trimester specific associations between PM2.5 and increased SBP increased but confidence intervals were wide…" This sentence is run-on and needs clarification. The description should make clear that the exposure estimates were during the 90 days before birth, but the BP measurements were 30 hours after birth.

Section 6.2.8, Peripheral Vascular Disease (PVD), Venous Thromboembolism, Pulmonary Embolism. The diagnosis "peripheral vascular disease" generally refers to disease in the peripheral arterial system, rather than venous disease. The discussion in this section is limited to venous thromboembolism, and does not address arterial PVD, so this term should be removed from the title. PVD should not be lumped together with venous TE disease; they have different etiologies, pathophysiology, and treatments.

Last sentence of 6.5.3. "However, relative to control animals, a toxicological study did not find an increase in markers consistent with cardiac damage following short-term exposure to PM10-2.5." PM10-2.5 is meant to be UFP here?

Page 7-12, line 2. "However, effects may be transient, so the upstream consequences are somewhat uncertain." Many of the outcomes discussed in this document are transient; it is not clear why the emphasis here. Also, it is not clear what is meant by "upstream" consequences. Did the author mean "downstream"? Perhaps "clinical consequences" would be more clear.

NFκβ should be NFκB (the English letter B instead of the Greek letter β).

"IVF" needs to be defined at its first appearance in the document, which is on page 9-5.

Dr. Sabine Lange

General note for these comments: a reference list can be found at the bottom of this document for those studies that are not referenced in the PM ISA.

General Comments

Literature Review and Study Quality:

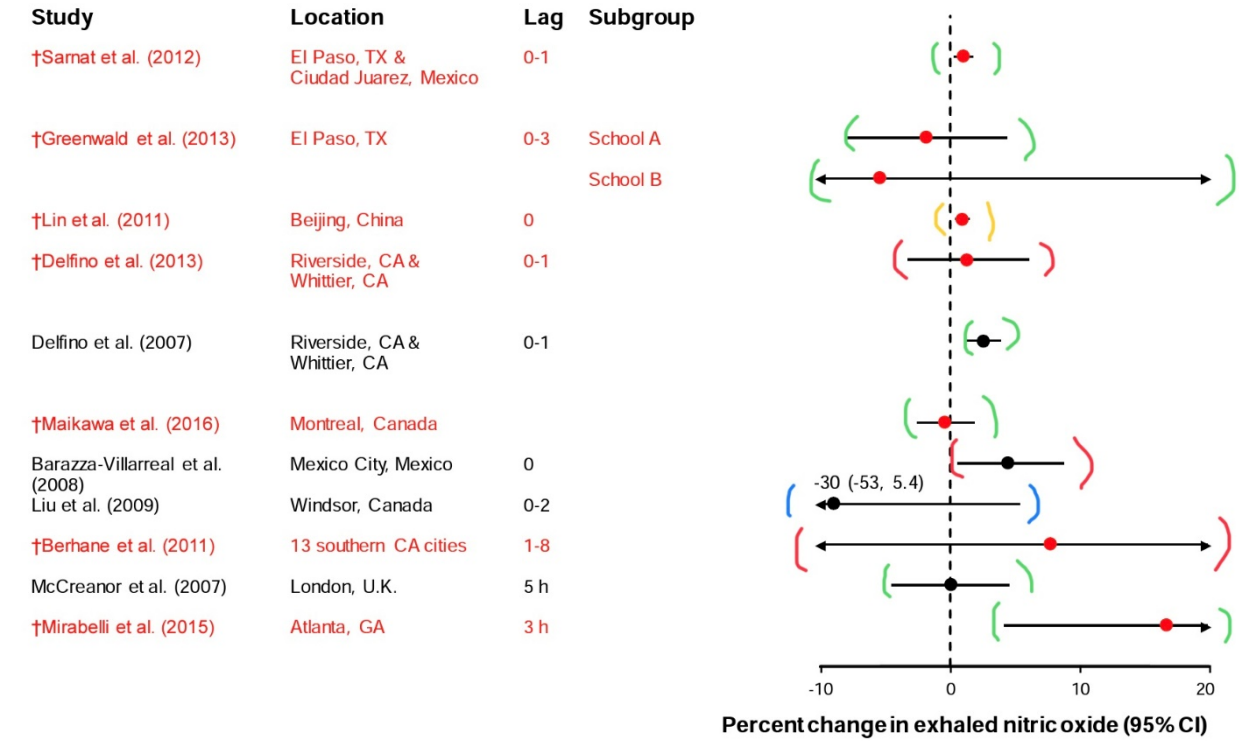
- The EPA needs to provide specific details about how studies were chosen for this review. In the absence of this information it is very difficult to determine whether a comprehensive, unbiased review has been completed.
- The EPA states that other recent studies are not the focus of this evaluation because they did not address uncertainties and limitations in the evidence previously identified. This suggests that the EPA only included studies that address uncertainties – does that mean only those with better methods, that consider copollutants and other confounders, etc were included? Looking at the listed studies, that doesn't seem to be the case. Which studies weren't included? Referencing 14000 studies in the HERO database is not helpful for the reader to determine which studies were not discussed in the ISA.
- Similarly, a detailed explanation of how study quality criteria was applied to the reviewed studies should be described. These study quality criteria also need to be included in the discussion of the study results in the health effects sections, so that appropriate conclusions can be drawn that consider how the study was conducted.
- The quality of measured outcomes needs to be discussed in the health effects chapters, because not all measured outcomes are equally reliable.
- The EPA should better explain when and why studies of Asian air pollutants are included in the analysis. These studies are conducted in environments with PM concentrations much higher than in the US, and with a much different combination of constituents that makes the comparability to US populations difficult to determine.
- "Uncertainties" should be a column in each of the data tables, laying out the potential concerns with each study. This makes it easy for reviewers to see what gaps still need to be filled in the literature and helps them appraise whether those gaps may be substantial.

Evidence Integration:

- There should be a discussion at the beginning of this document about how the EPA addresses the combination of positive and null or negative studies. For example, in the biological plausibility sections, is a single paper showing some effect on an end-point of interest enough to conclude that the pathway is plausible? What if there are a lot of studies not showing that effect?
- Noting whether a change in a biomarker or a subclinical effect in a pathway is adverse or is a substantial change (in terms of disease states) would be helpful for distinguishing important changes from PM exposure. For example, what is the significance of the change in glomerular filtration rate associated with long-term PM2.5 concentrations (pg 6-180)?
- The EPA should hypothesis-test its conclusions. For example, if PM2.5 concentrations are causally related to total mortality, you might expect that PM2.5 only actually impacts some subset of mortality types. If this is the case, then that mortality should have a larger more

significant association with PM2.5 than total mortality. Similarly, if all these effects are occurring at the same concentrations, then you would expect the milder effects to be more common and more likely to show an association than the more serious effects (e.g. HA or mortality). You would also expect that long-term effects would occur at lower concentrations and would show stronger effects than short-term, because of the cumulative exposure (assuming that PM2.5 has an impact via cumulative exposure). In addition, EPA notes that increasing variability in exposure or outcome estimates can bias the health effect estimate towards the mean. Therefore, one would hypothesize that studies with better exposure or outcome measures would have higher, more precise estimates than studies with poorer exposure or outcome estimates. I completed a simple analysis of this type of hypothesis testing using data in several of the presented forest plots:

- Figure 1 – this figure is a copy of Figure 5-5 from this ISA (associations between short-term PM2.5 exposures and exhaled nitric oxide in asthmatics), with marks around those effect estimates with more precise (green) or less precise (red) estimates of exposure; or with the highest (yellow) or lowest (blue) PM2.5 concentrations. No clear patterns are discernible demonstrating higher effect estimates with more precise exposure estimates or with higher PM2.5 concentrations.
- Figure 2 - this figure is a copy of Figure 5-8 from this ISA (associations between short-term PM2.5 exposures and respiratory-related HA and ED visits), with marks around those effect estimates with more precise (green) or less precise (red) estimates of exposure. The effect estimates generated with more precise exposure estimates tend to be lower, but with narrower confidence intervals, than those with less precise exposure estimates.
- Figure 3 – this figure is a copy of Figure 11-1 from this ISA (associations between short-term PM2.5 exposures and all-cause mortality), with marks around those effect estimates with more precise (green) or less precise (red) estimates of exposure; or with the highest (yellow) or lowest (blue) PM2.5 concentrations. Concentration does not seem to impact the association. Mortality associated with studies using more precise effect estimates shows generally higher, although with similar widths of confidence intervals, risks for health effects.
- Figure 4 – this figure shows the health effect estimates from forest plots in the short-term PM2.5 and respiratory effects section, with the range of effects representing the range of central estimates from each study. This does not show a clear pattern of increasing risk of health effects with decreasing effect severity.

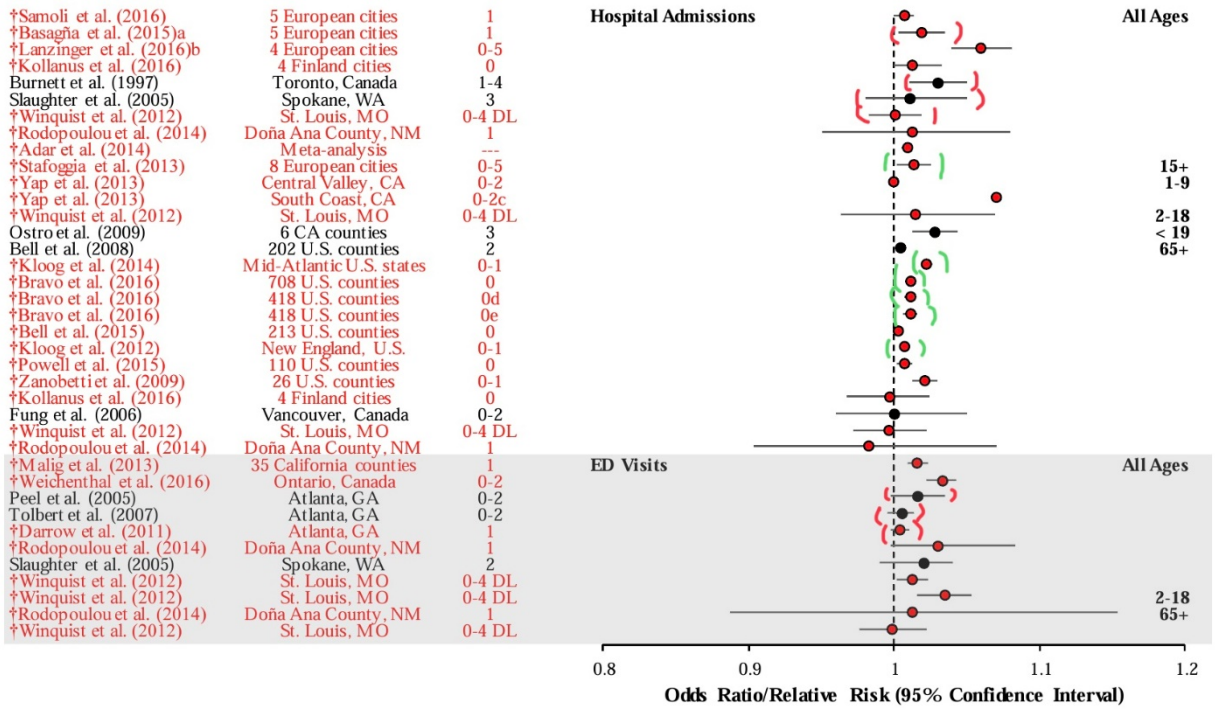


CI = confidence interval.

Note: Studies in red with a dagger are recent studies. Studies in black were included in the 2009 PM ISA. Effect estimates are standardized to a 10 μg/m^3 increase in 24-hour average $PM_{2.5}$. Lag times reported in days. Corresponding quantitative results are reported in Supplemental Material (U.S. EPA, 2018).

Figure 5-5 Summary of associations between short-term $PM_{2.5}$ exposures and exhaled nitric oxide in populations with asthma.

Figure 1. Copy of Figure 5-5 from this ISA (associations between short-term PM2.5 exposures and exhaled nitric oxide in asthmatics), with marks around those effect estimates with more precise (green) or less precise (red) estimates of exposure; or with the highest (yellow) or lowest (blue) PM2.5 concentrations.

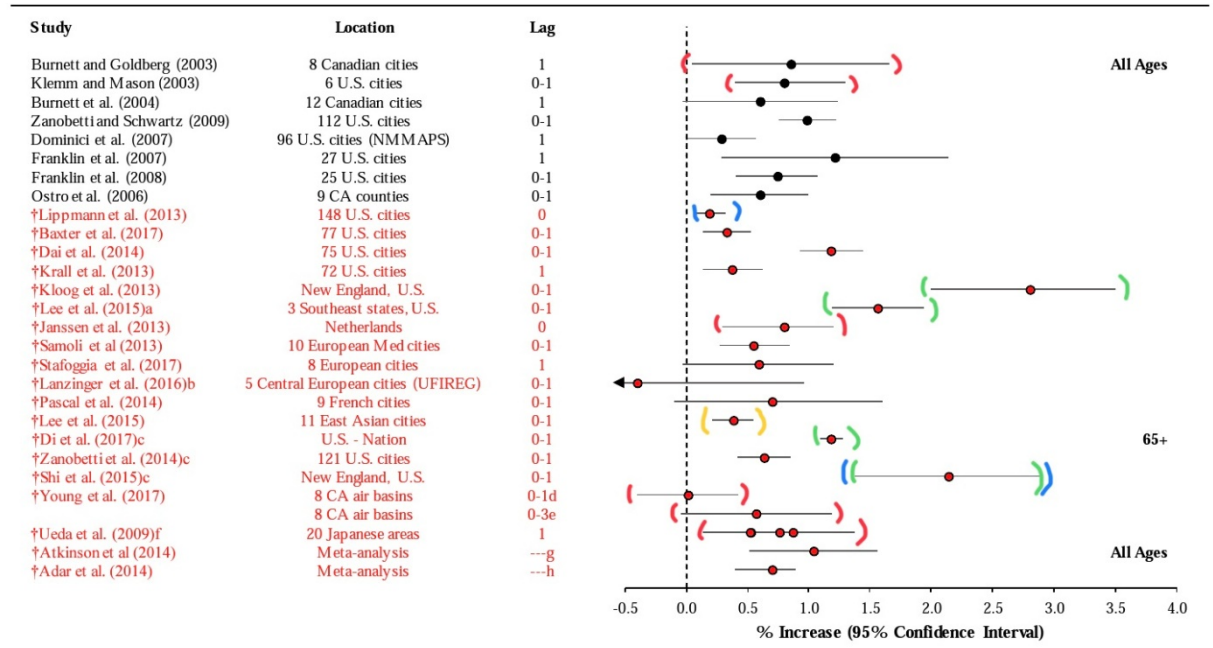


DL = distributed lag.

Note: †Studies published since the 2009 PM ISA. Black text: U.S. and Canadian studies included in the 2009 PM ISA. a = five European cities as part of the MED-PARTICLES project; b = only four of the five cities had $PM_{2.5}$ data; c = quantitative data for confidence intervals not reported, but above the null; d = monitoring data result; e = downscaler CMAQ, only counties and days with monitoring data. Corresponding quantitative results are reported in Supplemental Material (U.S. EPA, 2018).

Figure 5-8 Summary of associations from studies of short-term $PM_{2.5}$ exposure and respiratory-related hospital admission and emergency department (ED) visits for a 10 µg/m³ increase in 24-hour average $PM_{2.5}$ concentrations

Figure 2. Copy of Figure 5-8 from this ISA (associations between short-term PM2.5 exposures and respiratory-related HA and ED visits), with marks around those effect estimates with more precise (green) or less precise (red) estimates of exposure.



NMMAPS = National Morbidity, Mortality, and Air Pollution Study; UFIREG = Ultrafine Particles—an evidence based contribution to the development of regional and European environmental and health policy.

[a]Results are from modeled $PM_{2.5}$ analysis, analysis focusing on measured $PM_{2.5}$ reported 1.21% (95% CI: 0.94, 1.47).

[b]Only four of the five cities measured $PM_{2.5}$.

[c]Shi et al. (2015) and Zanobetti et al. (2014b) only had data for all-cause mortality including accidental mortalities.

[d]Main model used in Young et al. (2017) included current and average of 3 previous days daily maximum temperature, daily minimum temperature, and maximum daily relative humidity.

[e]Sensitivity analysis in Young et al. (2017) focusing on only the San Fransisco Bay air basin, dropping out the maximum daily relative humidity term, where the shortest duration of lag days examined was 0−3 days.

[f]Ueda et al. (2009) presented results for three different modeling approaches, which are presented here: GAM, GLM, and case-crossover.

[g]Atkinson et al. (2014) primarily focused on single-day lag results.

[h]Adar et al. (2014) focused on single-day lag results, specifically lag 0, 1, or 2.

Note: †Studies published since the 2009 PM ISA. Black circles = U.S. and Canadian multicity studies evaluated in the 2004 PM AQCD and 2009 PM ISA. Red circles = Multicity studies and meta-analyses published since the completion of the 2009 PM ISA.

Corresponding quantitative results are reported in the Supplemental Material for this chapter, see (U.S. EPA, 2018a).

Figure 11-1 Summary of associations between short-term $PM_{2.5}$ exposure and

Figure 3. Copy of Figure 11-1 from this ISA (associations between short-term PM2.5 exposures and all-cause mortality), with marks around those effect estimates with more precise (green) or less precise (red) estimates of exposure; or with the highest (yellow) or lowest (blue) PM2.5 concentrations.

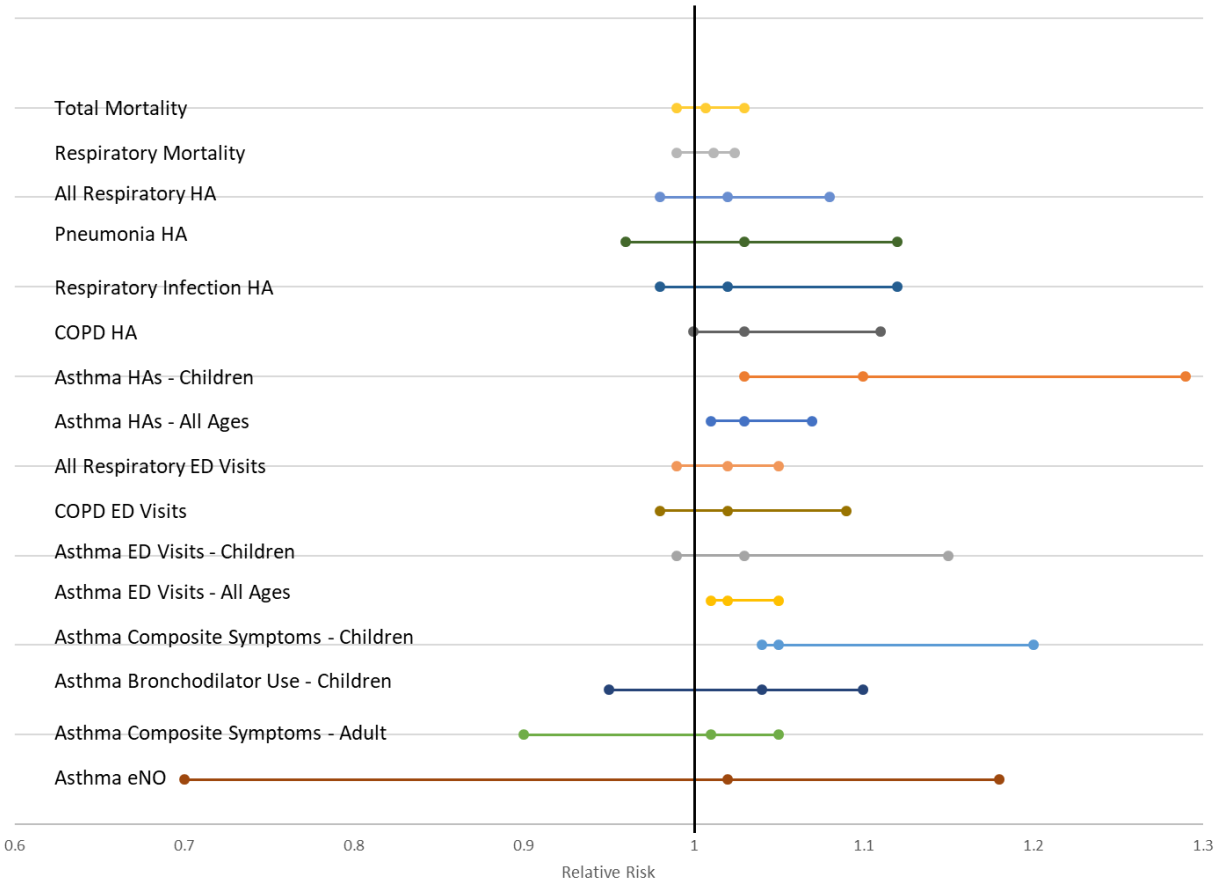


Figure 4. Health effect estimates from forest plots in the short-term PM2.5 and respiratory effects section, with the range of effects representing the range of central estimates from each study. The effects are in the order of decreasing severity from top to bottom.

Confounding:

- The EPA should specifically look at studies that investigate confounding by known other causes of the effects of interest, such as allergens for asthma.
- This document needs more discussion in the epidemiology sections about confounders considered in the different studies and the evidence of their impact on associations.
- EPA states throughout this document that you can't reliably do a copollutant analysis if the copollutant is well correlated with the primary pollutant. But it is the correlation that makes the copollutant a potential confounder (if they aren't correlated, then it can't be a confounder, by definition). As noted in the exposure section there is certainly a problem with determining which of two correlated pollutants is actually causal (whichever is measured more precisely will have the effect attributed to it (Carrothers and Evans, 2000; Fewell et al., 2007; Lipfert and Wyzga, 1996) but that seems like something to directly address and develop an answer for.
- Genetics contributes to many diseases, including cardiovascular disease, lung cancer, and asthma. In addition, it is not unusual for families to live in the same city or neighborhood. Therefore, family history is a potential confounder for PM effects (because it may be related to both the exposure and the effect, while not being a part of the possible causal pathway) that the EPA should consider when looking at incidence datasets.

Measurement Error, Statistics, and Concentration-Response:

- Measurement error and variability in epidemiology studies linearizes a non-linear relationship, and supra-linearizes a linear relationship (Rhomberg et al., 2011). Measurement error can also obscure a threshold (Brauer et al., 2002; Cox, 2018; Lipfert and Wyzga, 1996; Watt et al., 1995; Yoshimura, 1990). Therefore, epidemiology studies with known measurement error (and other types of error, such as outcome error) should not be used to determine the actual shape of the C-R curve. I discuss this point more in my comments on the exposure chapter. Animal and human controlled exposure studies should be used to determine a likely threshold – there is quite a lot of data for this.
- Concentration needs to be considered whenever a result is discussed. For example, the EPA states that mortality evidence provides coherence for a continuum of effects, without ever considering the concentrations at which these effects occur. Similarly, concentration plays a part in biological plausibility – is it plausible, that both mild effects and mortality occur at the same low concentrations? This should be explicitly discussed in this document.

Editorial Comments:

- Most of the figures are hard to read, because they are low resolution.
- There is often reference made to Section 0 – there is no section 0.
- Most of the chapters require some copy-editing, particularly chapters 4 and 9.
- Much of the verbiage in the biological plausibility sections is recycled from one section to another – if the pathways are the same, these sections should be consolidated. Also, all the biological plausibility sections state that their intention is to show "how" PM causes the health effect. Why is "how" in quotation marks? This makes it sound like the EPA doesn't take this important consideration seriously.
- The tables should specify what type of effect estimate was generated for a particular study (e.g. OR, HR, RR, % increase, etc).

- This document needs a list of figures and tables.
- At the beginning of every subsection on a particular health effect there should be a summary of EPA's conclusions about that health effect based on the new data. As it stands, some of the sections have these summaries, and others do not.
- What is the pattern for studies listed in the tables? E.g. studies in Tale 5-1 aren't alphabetical, by year, or by age, or by exposure type. These should be arranged in some way to make a particular study easier for the reader to find, or easier for patterns to be discerned.
- Different chapters have inconsistent organization – CNS has biomarkers of effects first, then diseases, whereas the opposite is true with respiratory and CVD. The organization should be consistent between chapters.
- When including monitoring information in tables in this ISA, there should be inclusion of information about the monitoring sample schedule (e.g. 1 in 3-day, 1 in 6-day, a combination) – this can impact the information that can be gleaned from the study, and potentially the exposure measurement error. Similarly, any data interpolation that is done in a study (particularly the long-term studies) to estimate PM2.5 concentrations should be included in the tables. Also, the model fit if exposure was modeled and was presented by the authors (and a note if the model fit was not presented by the authors).
- EPA should not present negative associations that aren't statistically significant as "near null" (e.g. pg 11-72), and positive associations that aren't statistically significant as "positive".
- In general, the EPA should state which estimate is chosen to present in the graphs, if there are multiple estimates presented in a study.

Chapter 3: Exposure to Ambient Particulate Matter

- The EPA should provide more discussion of the personal-exposure measurement literature, including some key systematic reviews published in 2010 (Avery et al., 2010a, 2010b) that described the variability in personal-ambient relationships, and stated that "The wide range in estimated correlations between personal and ambient PM2.5, as well as the associations with participant, study and environment characteristics, suggest that the potential for exposure misclassification can be substantial." This should be further discussed in this document and used to better inform interpretation of studies that assume a relationship between ambient and personal PM2.5 concentrations. The systematic review guidelines for TSCA (US EPA, 2018) lists study quality criteria for epidemiology studies (amongst others). They state as a criterion for deeming a study unacceptable (and therefore for removal from the review) "There is evidence of substantial exposure misclassification that would significantly alter results." This needs to be seriously considered for studies that use ambient monitors as surrogates for personal exposure.
- There is considerable evidence in the literature that exposure measurement error (and likely other types of error) in epidemiology studies can generate shapes of C-R curves that do not accurately represent the underlying C-R function (Cox, 2018; Yoshimura, 1990). For example, from Rhomberg et al. 2011: "Overall, because of the prevalence of exposure measurement error in epidemiology data and lack of reliable error-mitigating techniques, conclusions about the linearity of the exposure-response curve must be examined carefully and treated with some skepticism."

- "Lipfert and Wyzga (1996) found that for a true PM10 threshold of up to 150 µg/m3, an underlying "hockey-stick" risk model would appear consistent with linear (no-threshold) models in the presence of independent variable error."
    - From Lipfert and Wyzga: "If the variables that we are forced to work with (from fixed ambient monitors) already contain a lot of exposure error, no amount of analysis of this type can provide a remedy since the error cannot be removed. In such situations (which may include most of the PM studies), even sophisticated statistical analysis cannot impart real meaning to the data." (Lipfert and Wyzga, 1996)
  - From Watt (1995), "Using the same data and parameters from Lioy et al., (1990) as Lipfert and Wyzga (1996) but a slightly different computational approach in which individual exposures were assumed to be lognormally distributed around the central/ambient exposure, they also showed that error can mask a true threshold function." (Lioy et al., 1990; Watt et al., 1995)
  - From Brauer 2002: "From both sets of figures, it is evident that when surrogate measures (ambient concentrations) are not highly correlated with personal exposures, a threshold can be masked at the population level even if there is a clear, common threshold at the individual level (individual risk function). Furthermore, even if the threshold is not completely masked, it is likely to be biased." (Brauer et al., 2002)
- Therefore, epidemiology studies that have known, substantial error in the exposure estimates, should not be used to determine the shape of the C-R function between PM and health effects.
- For the general conclusion that exposure measurement error biases towards the null, evidence suggests that this is only true if the following assumptions hold: 1) the concentration-response is linear (Fuller, 1987); 2) the measured concentrations are a good surrogate for ambient concentrations (not a valid assumption for PM, as noted above); 3) it is a single-pollutant model; and 4) the differences between measured concentrations and personal concentrations are constant (Zeger 2000). If any of these assumptions are false, then the estimated effect of PM on health could either be an under-estimate of the true effect, or it could reveal an effect that is not due to ozone concentrations. Many studies have shown that exposure error can in fact have complicated effects on the health effect estimate that are not captured by the generalization that there is bias towards the null (Cefalu and Dominici, 2014; Goldman et al., 2011; Jurek et al., 2008, 2005; McGuinn et al., 2017).

Methodological Considerations:

- The EPA cites Pope 2009 and Zanobetti & Schwartz 2009 in their argument that ambient monitors should continue to be used as exposure surrogates. One of these arguments is "The ambient monitor approach is the least data intensive approach among all exposure concentration estimation methods because it only requires data from a single monitor to represent exposures to a large area (on the order of 100 km2)." Are the authors and the EPA suggesting that one monitor per 100 square kilometers is adequate to capture exposure estimates for all the individuals in that area? This is completely inconsistent with the personal exposure-ambient correlations, which show great variability in personal-ambient correlations, even using ambient monitors that are much closer than 100 km2 (Avery 2010a, b).

Exposure Assessment and Interpretation of Epidemiologic Study Results:

- The EPA states that "If this occurs, the health effect related to PM exposure would be underestimated or potentially not detected. Positive correlation between PM and the copollutant and between the exposure measurement errors of PM and the copollutant can add more negative bias to the PM health effect estimate. Spatial variability of concentration differs among the particle size spectrum, and this may cause more exposure measurement error in PM10−2.5 or UFP compared with PM2.5 (Section 3.4.2.2). Hence, if PM2.5 is measured with less error than copollutants, it is likely that the effect will be attributed to PM2.5". This means that in copollutant models whichever pollutant is measure with the least error is most likely to be ascribed the positive effect, as has been demonstrated by several groups (Carrothers and Evans, 2000; Fewell et al., 2007; Lipfert and Wyzga, 1996). This makes interpreting copollutant models quite tricky and requires considerations of exposure measurement error for each component. The EPA should specifically address this point when discussing copollutant models in the health effects chapters.

Chapter 4: Dosimetry of Particulate Matter

- This chapter provides very useful and up-to-date information about PM dosimetry in humans and model organisms. This information is crucial for interpreting doses caused by PM exposure, and should be an integral part of the interpretation of health effects studies in Chapters 5-11. One way to make dosimetry more interpretable in the health effects chapters would be to include the exposure concentrations at which different patterns were observed. For example, the Miller 2017 study exposed people to particle numbers that are about 1000-times higher than ambient (compare 4.15 x 10E6/cm3 of ~4 nm particles, with the Stanier 2004 study measuring 5.6 x 10E3 particles/cm3 for 3-10 nm particles cited on pg 2-32 of this ISA).
- The summary for this section captures the relevant conclusions from this chapter. However, the sentence "New dosimetric information shows that PM10 overestimates the size of particles likely to enter the human lung." is somewhat confusing and could be reworded. I recommend using wording based on the summary in Chapter 1: PM10 uses a 50% cut-point at 10 μm, which provides a conservative (protective) overestimate of particles that reach the thoracic compartment of the lung.

Structure and Function of the Respiratory Tract:

- To streamline this chapter so that it only includes necessary information, the paragraph reviewing the history of scientific views on post-natal alveolar development could be removed.

Ventilation Distribution:

- This section discusses ventilation distribution in dogs and horses, including pregnant Shetland ponies, and then concludes: "Thus, the position in which rats are exposed may influence the regional delivery and deposition of inhaled aerosols." Extrapolation of ventilation distribution to rats should probably be restricted to data that is more similar to rats (dogs and horses may be quite different), and if these species are relevant to rats, that should be explained.

Thoracic and Respirable Particles:

- I appreciate the inclusion of the information that PM10 was chosen as a cutpoint to over-represent the true penetration of particles into the thoracic region; and that penetration of 1 um particles into the human lower RT is more affected by route of breathing than by age, sex, activity level, or breathing pattern. This is very helpful for interpretation of human health effect studies and for assessment of at-risk populations.
- The discussion of translocation of insoluble versus soluble components provides good information. A great add-on to this would be providing some information about how big a contribution is made by soluble particles to total particles, because these particles could have a more direct or obvious linkage to the systemic effects than insoluble particle translocation (which occurs at a very low frequency).

Deposition Patterns:

- This section notes that inertial impaction at carinal ridges can result in concentrations at those locations that are hundreds to thousands of times higher. There may be very local responses at these sites of deposition that aren't captured by whole-lung washes. Would a reaction at these small sites be enough to mediate the health effects seen in toxicology and epidemiology studies?

Factors Modulating Deposition:

- Physical Activity - The last paragraph on page 4-26 provides important information about the impact of exercise on lung deposition of different sized particles. The clarity of this information summary would be improved by making a clear distinction between lung deposition fraction and total lung deposition. It seems that while lung deposition fraction may not change during exercise, because the individuals are breathing more air, more total particles are deposited in the lungs.

Particle Clearance:

- Interspecies clearance and retention – The information about the differences in particle clearance between animals and humans is very useful for the extrapolation of health effect results between species. It would be helpful if this section also included information about what concentrations and dose-rates in rats cause the inhibition of mucocilliary clearance, because that would provide information about what effects in rats may be more or less relevant to humans.

Particle Translocation:

- This section provides information that is used by the EPA to support biological plausibility in further chapters of this ISA. With this in mind there are several points that are worth emphasizing (perhaps summarizing at the beginning of this chapter) that will help readers apply this information to potential pathways of effect:
    - Translocation of particles < 200 nm may occur along the axon to the olfactory bulb, although there is little data for this in humans. Because of a reduction in the foramina area in humans with age, there may be a decrease in this pathway in older humans.
    - Translocation to the olfactory bulb has been demonstrated in animals with Mn from welding fumes. However, other poorly soluble metals in the fumes did not translocate to the olfactory bulb, suggesting that translocation may be component specific, or is due to soluble particles.

    - It has been difficult to determine how much translocation outside of the respiratory tract is caused by movement of insoluble particles, versus movement of soluble particle components. Because this translocation of particles to the blood and nervous system may contribute to health effects, it would be very valuable if the EPA provided some information about what fraction of particles are soluble, and perhaps what types of soluble components could cause certain types of health effects.
    - While olfactory particle translocation may happen, the human data is quite uncertain (the human autopsy studies published in 2010 and 2013 by Calderón-Garcidueñas et al. do not provide definitive evidence because of problems with proper controls and determining the source of the UFPs found in brain tissue). However, even if there is translocation it is likely to be a very tiny fraction of particles, as estimated by Garcia et al. 2015, with only 0.001% of 20 nm particles being deposited on the human olfactory mucosa.
    - From Miller et al. 2017, about 0.03% of gold nanoparticles seem to have translocated from the respiratory tract to the blood in humans.
- In reference to the note made by EPA on page 4-59 that the "absolute numbers of particles reaching the olfactory bulb over time can be considerable (Figure 4-7).", more data should be provided to support this conclusion. Figure 4-7 provides the dose-rate of UFPs to the olfactory epithelium, which doesn't provide much explanation about the absolute number of particles reaching the olfactory bulb, or whether it reaches a threshold that could be labeled as "considerable".
- Many studies test the translocation of particles into the blood or olfactory apparatus using a label, typically a radio label. However, some of the label may be attached to a soluble subset of the tested chemical, causing a signal that may not be due to the insoluble particle. Therefore, when discussing these types of studies, it would be of value to the reader if the EPA explicitly discussed that study's control for solubility, and whether the signal from the non-respiratory tissues was shown to be particle-bound (e.g. the discussion of the Geiser et al. (2005) results on page 4-59). The information provided can be similar to how EPA addressed the 99mTc-labeled particles discussed on page 4-60.
- Because Miller et al. 2017 is a central paper in this analysis, a discussion of solubility and potential ingestion of the gold particles should be provided. There should also be consideration of the doses used in Miller 2017 compared to ambient concentrations.
- There is a discussion in this chapter about the potential translocation of particles to the fetus. It is not accurate to portray the available data on fetal translocation of particles as providing "biological plausibility of effects during pregnancy", for two reasons. 1) The only information available for fetal translocation is from oral or intravenous animal studies that are not relevant to inhalation exposures (as evidenced by data showing that the extra-pulmonary distribution of particles from inhalation is different compared to IV or oral administration). 2) The administered doses generate systemic particle concentrations that are orders of magnitude higher than would be attained via inhalation. Altogether these studies do not provide biological plausibility of particle translocation to the fetus at relevant exposures via inhalation in humans.

Factors Modulating Particle Clearance:

- The discussion of particle overload in rats is very helpful for the extrapolation of rat data to humans, particularly for chronic effects of PM exposure. Providing data on the doses or dose-rates at which this occurs would aid further in applying this information to rat-human

extrapolation. Are any of the doses or dose-rates used in the rat studies reviewed in the ISA likely to cause particle overload?

Summary:

- EPA states in the summary that "The fraction of nanoparticles translocating from the peripheral lung into circulation is generally low (less than a fraction of a percent) for larger nanoparticles (18−80 nm) but can approach several percent for extremely small particles (1.4−2.8 nm)." EPA should note here that while several percent of extremely small particles may translocate into the peripheral circulation in rodent studies with exposure by lung installation, there is no evidence that this much translocation occurs with exposure to even very small particles (4-5 nm) in humans.
- Similarly, when discussing results showing fetal translocation of particles, EPA should state that this was using oral or IV particle administration and cannot be extrapolated to human inhalation exposures at relevant ambient concentrations.

Chapter 5: Respiratory Effects

<u>Short-Term Effects of PM2.5</u>

Biological Plausibility:

- In this discussion of biological plausibility, the EPA does a good job of laying out the potential paths that particles may take to influence the respiratory system. In the summary EPA states that "Immune system responses due to the presence of particles in the interstitial space may contribute to respiratory health effects." What is the impact in the respiratory tract of particle translocation to the interstitial space and then movement to the lymph nodes? This is a method of removal, but is there direct evidence of its adversity? If particles activate the immune responses in the local lymph nodes, this could be added to the biological plausibility summary.
- The beginning of the biological plausibility pathway for all PM-induced health effects in this ISA is respiratory tract inflammation, oxidative stress, and injury. In this section EPA states that strong evidence for these effects is provided by human and animal studies. However, the human studies, which have exposed individuals at higher-than-ambient concentrations with exercise (healthy as well as vulnerable populations) show very little evidence for this (data summarized in the following bullet points).
- Inflammation in human controlled exposure studies of fine CAPs:
    - Ghio 2003, Gong 2003, Gong 2004, and Huang 2012 did not show an increase in inflammatory cell infiltration or soluble inflammatory mediators after exposure to fine CAPS from different locations, and Gong 2005 & Holgate 2003 did not show an increase in infiltration of immune cells. These studies were conducted on people who were healthy, asthmatic, elderly, or had COPD, and at concentrations up to 178 ug/m3. (Ghio et al., 2003; Holgate et al., 2003)
    - Ghio 2000 did show an increase in neutrophil infiltration with PM2.5 exposure, but no change in soluble inflammatory mediators. Urch 2010 showed an increase in soluble IL-6 at 3 hours after PM2.5 exposure in people exposed to concentrations higher than 100 ug/m3, but no change in inflammatory cell infiltration, and no change in soluble

inflammatory markers when PM2.5 exposure was combined with 120 ppb ozone exposure.
    - Altogether, the evidence for PM2.5-induced respiratory inflammation in human controlled exposure studies is inconsistent and largely negative.
- Lung injury in human controlled exposure studies of fine CAPs:
    - Ghio 2000 showed a decrease in bronchial total protein, and Gong 2003 and 2005 found a decrease in sputum total or epithelial cells. This is the opposite direction of adversity, with lung damage usually manifesting in an increase in total protein or cells. Behbod 2013, Gong 2004, Huang 2012, and Urch 2010 showed no increase in pulmonary damage markers with fine CAPs exposure.
    - Altogether these studies show evidence of a lack of pulmonary damage with higher than ambient PM2.5 concentrations in multiple populations.
- Similarly, this section says that there is evidence of lung function changes in humans, but actually most studies don't show this effect. No adverse pulmonary function effects of exposure to CAPs were shown in Brauner 2007, Ghio 2000, Gong 2003, Gong 2004, Huang 2012, Lay 2001, Sivagangabalan 2011, or Urch 2010, who used exposures up to 206 ug/m3 and in healthy younger and older subjects, and subjects with asthma or COPD. (Bräuner et al., 2007; Lay et al., 2001). A few studies showed some lung function effects – Gong 2005 observed a decrease in FEF25-75 in healthy older adults with PM2.5 exposure, but not with PM2.5+400 ppb NO2, nor in individuals with COPD. Hazucha 2013 observed a decrease in FEV1 in smokers and ex-smokers. Altogether this is not compelling evidence that PM2.5 causes lung function deficits. The EPA should explicitly state how they interpret the occasional positive study in light of many negative studies.
- The EPA conjectures that activation of the autonomic nervous system (ANS) causes the respiratory effects, but they should include a discussion about whether the lack of FEV1 responses is consistent with an ANS or airway irritant response (usually an irritant/neural response in the airways triggers a decrease in FEV1, as with ozone).
- The EPA states that PM2.5 caused changes in SaO2, FEV1, and tidal volume in human and animal studies with COPD. However, there was very little response in human studies (some evidence that CAPs cause less responsiveness in COPD people or maybe improvement in animals with chronic bronchitis (CB)). Saldiva 2002 found less neutrophil density in CB rats with PM2.5 exposure, and Clarke 1999 found increases in tidal volume (TV) and PEF (decreases would be adverse). The results of these studies need to be integrated into EPA's conclusions.
- Immune responses are cited as occurring subsequent to respiratory tract inflammation and oxidative stress and are blocked by anti-oxidants. However, this evidence comes from Whitekus 2002, a diesel exhaust particle (DEP) study that only found effects with OVA treatment + 600 ug/m3 DEP and not with 2000 ug/m3 DEP alone. The OVA system is cited as being similar to human asthma, but a few sentences about how they are similar and different, and to what degree the severities are similar would be helpful (e.g. is this like severe asthma, or mild asthma?).
- For the pathway activation of sensory nerves, the EPA notes that the previous ISA and this one demonstrated changes in respiratory rate and lung volumes (i.e. rapid shallow breathing). However, the Clarke 1999 paper demonstrated an increase in TV with CAPs in healthy or CB animals, and no increase in breathing rate. There is data that lung irritant responses are mediated by the vagus nerve and the parasympathetic nervous system. However, isn't this the opposite

direction from the heart rate variability (HRV) responses cited in the next chapter? This needs to be addressed.

- The EPA notes that Ghelfi 2008 found involvement of the TRP sensory nerve receptors in response to PM2.5 exposure, because TRP antagonists blocked PM2.5-mediated effects. This study lacks important study details (could be addressed with a study quality evaluation), but its data suggests parasympathetic ANS activation. However, Chiarella 2014 is also cited, and this study demonstrates activation of the sympathetic ANS with increased norepinephrine. The EPA need to consider these study results in their biological plausibility pathways.

Asthma Exacerbations:

- HA & ED Visits – the EPA states on pg 5-7 that there is controlled human exposure and animal toxicology evidence for short-term PM2.5-induced allergic inflammation. However, I don't think that there is any human evidence for this, and there are no citations offered for this statement. If there is evidence, it should be cited, as well as discussed in the biological plausibility section. For animal toxicological studies, the Harkema 2009 and Wagner 2012 papers did not show evidence of CAPs alone inducing allergic inflammation, and only showed an enhancement of OVA-induced inflammation at higher CAPs concentrations from Detroit (not Grand Rapids).
    - Figure 5-2 – the confidence intervals are not presented for Yap 2013.
    - Most of the ED visit effect estimates are not statistically significant – how does this affect EPA's determinations about the conclusions from this data?
    - There needs to be discussion of other types of confounders besides copollutants, such as the aeroallergens presented in Hebern & Cakmak (2015). EPA states that there is evidence of seasonality, but it isn't clear if this represents PM having different effects based on the season, or if it is just unexplained heterogeneity in the study results.
    - Several of the studies in Figure 5-3 have different years than the corresponding entry in Table 5-1.
- Asthma Respiratory Symptoms and Medication Use – most of the studies presented in Figure 5-4 show positive results, but almost none are statistically significant. This section states that the EPA has increased confidence in the results because recent evidence shows associations with PM2.5 measured outside of children's schools. However, the Spira-Cohen (2011) study showed that personal exposures were not associated with symptoms, and that outside concentrations were not good surrogates of personal exposure. How does this support the EPA's conclusions about the effects of PM2.5 on asthma?
- Lung Function Changes in Asthmatics – the EPA states that lung function changes in asthmatics were only evaluated in epidemiology studies, but Gong 2004 and Urch 2010 (both human controlled exposure studies) investigated effects of CAP exposures in asthmatics and found no effects on lung function. No summary figure or information is provided for the epidemiology studies, making study results difficult to interpret without looking up all the papers individually. From looking at the studies, it is clear that the Spira-Cohen (2011) results weren't statistically significant, and the Delfino (2008) paper only had associations with lag 0 for 1 or 8 hr max and it wasn't clear if the authors made sure that the maximum concentration occurred before the asthma lung function effect. Smargiassi (2014) found no effect of personal PM2.5 exposure on an array of lung function effects. So there is considerable variability in the findings of these studies that the EPA needs to address. Many of the studies summarized in Figure 5-4 are not present in the associated Table 5-2.

- Subclinical Asthmatic Effects – EPA focuses on a relationship with eNO (Figure 5-5). There is an even distribution of negative, null, and positive effect estimates. However, there are also CAPs studies in asthmatics that have shown little or no effect of exposure to PM2.5 on lung function, inflammation, or damage (Gong 2003, Urch 2010). The Harkema (2009) paper does not show independent effects of 600 ug/m3 CAPs on pulmonary endpoints but does show that it can enhance OVA-induced bronchopneumonia. This did not happen with the animals exposed to 356 ug/m3 CAPs (8 Hrs per day for 3 days), demonstrating a threshold of effects. Wagner 2012 also showed that Grand Rapids CAPS at 600 ug/m3 actually diminished OVA-induced effects, showing a constituent-importance, and possibly a counter-intuitive protective effect. These results need to be considered when drawing conclusions in this section.

COPD Exacerbation:

- HAs and ED Visits – A new meta-analysis is cited as providing positive (statistically significant) evidence of an association between PM2.5 and COPD exacerbation (Li 2016). Li (2016) conducted a systematic review and meta-analysis of 18 studies from North and South America (1 study), Europe, and Asia (3 studies). The heterogeneity statistic for PM2.5 and COPD HAs was 88%, meaning that these studies are very heterogenous and probably shouldn't have been meta-analyzed (a study quality assessment would have picked this up). The mortality estimates were heavily weighted by the Santiago Chile study, and Taiwan studies weight the HA estimate.
- COPD Lung Function Changes – Ebelt (2006) and Trenga (2007) from the last review are cited as having found lung function effects in people with COPD. Both of these studies only showed positive associations with ambient monitors, and lesser non-significant associations with personal exposure. This isn't consistent with poorer exposure estimates biasing towards the null, or with biological plausibility of PM2.5 exposure causing the lung function effects. Two new American studies show no or inconsistent effects; the only significant effects presented are in studies in Mexico City or Asia. Controlled exposure studies that are cited (Gong 2004 and 2005) are noted as having found decreases in oxygen saturation in adults with COPD, but no changes in lung function. There were also no symptoms, and no evidence of pulmonary inflammation or damage in those studies that should be considered here.
- Subclinical COPD effects - The only new epidemiology information provided is from Chinese studies which have unknown applicability to US exposure. Two panel studies were cited as evidence of changes in eNO associated with PM2.5. One of these, Jansen 2005, showed no statistically significant associations with fixed site monitors, and the association with indoor PM2.5 (which will be closer to personal), was null. EPA notes that Gong 2005 found a decrease in columnar epithelial cells with PM2.5 (more pronounced in healthy people than people with COPD). But this was a decrease in *sputum* epithelial cells – an increase in sputum epithelial cells suggests damage, but the authors couldn't explain the decrease in cells.

Respiratory mortality:

- New studies seem to provide inconsistent evidence that there is an association between short-term exposure to PM2.5 and respiratory mortality: Figure 11-2 presents 7 new effect estimates from the US and Europe with a similar range of magnitude, but variable effects (3 statistically significantly positive, 2 non-statistically significantly positive, 2 non-statistically significantly negative). EPA noted that there was limited coherence with human controlled exposure and

animal studies in the previous ISA, and presumably this is still the case. How does this affect the EPA's conclusions?

Policy-Relevant Considerations:

- Policy-relevant considerations are included sporadically in the individual sections as well. This section is not a separate component of the other end-points or exposures in this ISA, so it is not clear why the EPA put a separate section for policy relevant considerations here.
- EPA notes that epidemiologic studies often conduct analyses to determine whether the observed effects are due to chance, bias, or confounding. However, there is essentially no discussion of chance in this chapter, despite its inclusion on this list. There is also very little discussion of confounding that is not due to copollutants. These topics need to be included much more thoroughly by EPA.
- Copollutant Confounding - In the figures the open and closed symbols need to be defined. Sometimes this section refers to a discussion of copollutant confounding in the relevant health effects section, and that section refers to this policy-relevant section (e.g. subclinical asthmatic effects section). This section needs to be consolidated and clarified.
- Model Specification - EPA notes that degrees of freedom for temporal trends and weather variables mostly don't affect the results, and therefore that there is reduced uncertainty from model mis-specification. However, it seems like there are far more modeling options than just df for temporal trends and lags for weather variables. Two studies (Strickland 2010 and Sarnat 2015) look at the effect estimates for lag -1 day as a control. EPA notes that the results of the base model were similar to those for lag -1 day, but because the associations for lag -1 day were smaller (1.03 compared to 1.05 for lag 0-2; and 1.02 compared to 1.04 for lag 0-2), that potential confounders were adequately controlled in the model. But both show positive results, one borderline stat sig, for a lag -1. If those results were for lag 1 EPA would have considered them to be indicative of a positive association. This type of result needs to be considered when EPA takes any positive result as indicative of an effect. In general, Sarnat 2015 only saw associations between PM2.5 and ED visits for asthma/wheeze, not pneumonia, COPD, general respiratory disease, general CVD, IHD, arrhythmia, or CHD.
- Lag Structure - EPA notes that lag structure can be informative about whether PM2.5 has immediate, delayed, or prolonged effects. However, effects at inconsistent lags can place doubt on the veracity of the results. For example, Strickland 2010 showed early lag effects, and Kim 2012 showed 4-12-day lag effects. In addition, lag 0 is problematic because it is not entirely clear if the effect occurs before or after the exposure. EPA notes a somewhat delayed lag of 0-5 days for respiratory morbidity, but a shorter lag of 0-2 for respiratory mortality – it seems that the opposite would be true – that more immediate concentrations would lead to less severe effects, while it would take more exposure to cause mortality.
- Season Effects – EPA notes that the 2009 ISA found that some associations between PM2.5 and respiratory effects were stronger in the warm months, and for others they were stronger in the colder months. Newer data that EPA presents demonstrates seasonal heterogeneity, with no clear pattern. How is this heterogeneity interpreted? Before the EPA interprets this is showing a complex relationship between PM2.5 and season, they should consider whether the results are due to bias, chance or confounding, and if the heterogeneity is in fact suggesting a lack of causal association between PM2.5 and the relevant health effect. Only once these concerns have been addressed should the EPA investigate what these seasonal patterns may be indicating about how

or under what circumstances PM2.5 causes a health effect. Studies of aeroallergens (e.g. Hebern and Cakmak 2015) suggest that these could be important potential confounders and are a reminder that copollutants aren't the only confounder.

- Temperature – What would the EPA expect the effect of temperature (or season) to be on PM2.5-mediated respiratory effects? More effects in the cold season because of burning as a source of heat, or during warm season because of interactions with allergens? EPA should generate some mechanism-based hypotheses for these effects, and then see if the data matches. For example, if you think that sulfate is an etiologic factor, then look at places with higher sulfate in certain seasons (even if it is not measured for the study, emissions inventory data can be used), and see if that is when the effect is greater.
- Concentration-Response and Threshold Analyses – EPA presents Figure 5-21 from Silverman and Ito (2010), and states that the authors found that the non-linear model wasn't any better at fitting the data than the linear model. However, Figure 5-21 presents a distinctly non-linear shaped curve. Evidence presented from Gleason 2014, showing positive estimates in the 5th quintile of PM2.5, no association in the 3rd and 4th quintiles, and the largest association in the 2nd quintile suggests a lack of evidence of any association. EPA concludes that there is some evidence for linearity, and some evidence for non-linearity. There is no discussion of MOA or results from experimental studies that would inform this decision.

<u>Long-Term Respiratory Effects of PM2.5</u>

Biological Plausibility:

- It would be helpful if the EPA could clarify how changes in the renin-angiotensin system could impact long-term respiratory function. In addition, since there is only a single study with information on this endpoint, a shorter discussion could be devoted to it.
- EPA states that there is evidence of Th2 immunity from Kim 2016a, however the cited Deiuliis 2012 suggested Th1 and not Th2, so the EPA should clarify why the Kim 2016a study provides more definitive evidence of the activated immune pathway. In addition, Kim 2016a used Penh as a marker of airway hyperresponsiveness (AHR), but this is generally acknowledged to be a poor marker for AHR in animals (Bates and Irvin, 2003).
- EPA cites studies showing increased oxidative stress, injury, inflammation, and morphologic changes in the nasal mucosa (Guo 2017 is cited twice). However, the relevance of nasal changes in rats and mice (obligate nasal breathers) to human effects is not clear and should be directly addressed.
- It would generally be useful to have concentrations and exposure conditions provided, with appropriate dosimetric adjustments to allow the reader to understand the comparison to human effects and doses.

Lung Function and Development:

- EPA states that PM2.5 effects on lung function and development are supported by several iterations of the Children's Health Study in California. They note that associations are supported by a multicity cohort in Taiwan, although the concentrations are considerably higher. Since they are reviewing these higher concentration Asian studies, it would be good if the EPA noted if the associations were stronger or the effect estimates larger, as one would hypothesize (based on Figure 5-28, this doesn't seem to be the case). EPA states that pre-adolescent effects are

uncertain (positive effects in a Chinese cohort, but not in the European PIAMA study). In general, it seems that the only real evidence comes from the CHS cohort, which is just studied over and over again.
- Table 5-19 shows that there are moderate to high correlations with many copollutants, which makes interpretation of the CHS study results problematic.
- EPA notes that they expect to have only low-to-moderate spatial heterogeneity for the CHS study, so that means that there is unlikely to be major exposure measurement error. This statement assumes that spatial heterogeneity is the only source of exposure measurement error, and in general needs to be better supported.
- There seem to be animal toxicology studies only from Beijing and Sao Paolo looking at developmental effects of PM2.5. They show some evidence of effects, but interpretation is tricky because of the exposure locations.

Development of Asthma:
- Several longitudinal studies look at the relationship between new asthma and PM2.5. Many of them show positive associations, but most are not statistically significant (Gehring 2010 is non-statistically significant when adjusting for study region, and Gehring 2015, Yang 2016, McIntyre 2014 for ever asthma, McConnell 2010, Clarke 2010, and Nishimura 2013 aren't statistically significant). EPA states that studies generally provide support for an association between asthma prevalence and PM2.5, though not all studies – Fuertes 2013b and Akinbami 2010. Why aren't these last two studies shown in the summary Figure 5-30? Altogether this doesn't seem to be compelling evidence of an association between new asthma and PM2.5.

<u>Long-Term Respiratory Effects of PM10-2.5</u>
- Table 5-30 and 5-31 are the same.

<u>Short-Term Respiratory Effects of UFPs</u>
- Biological Plausibility – For respiratory tract inflammation, Frampton 2004, Frampton 2006, and Gong 2008 didn't show any increase in immune cells or soluble inflammatory mediators after UFP exposure (with very high number concentrations, in the 10^5 to 10^6 range), and Samet 2009 saw no immune cell infiltration, and only an increase in IL-8 at 0 hrs after exposure. Altogether this is not convincing of a respiratory tract inflammatory effect. In the inflammatory section when discussing evidence the EPA should be sure to note which species is being discussed (humans, rats, etc). (Frampton et al., 2006, 2004)
- There is a lot of discussion in the biological plausibility section about how UFPs can penetrate more deeply into the lungs than fine or coarse PM, and maybe can translocate into the blood, but at some point in this section the EPA should address why the UFP results show less evidence of health effects than for PM2.5.
- There are few studies for short-term respiratory effects of UFPs, with essentially no statistically significant results for any of the analyzed endpoints (asthma HAs and ED visits, inflammation or pulmonary fxn changes in controlled human exposure studies, COPD exacerbation, respiratory infection, total respiratory HA or ED visits, healthy human controlled exposure studies). EPA should be clearer as to why this data merits a "suggestive" causality determination.
- Respiratory mortality incorrectly references Table 11-9, which is for PM10-2.5, not UFPs – should be Table 11-13.

Long-Term Respiratory Effects of UFPs

- EPA states that a paucity of data prevent the description of biological pathways that may underlie long-term respiratory effects of UFPs, and then they drew a diagram of those pathways (Figure 5-50). Is this a mistake, or was this pathway really drawn based on very little data?
- Data present from Tyler 2016, Araujo 2008, Reed 2008, and Tanaka 2013a does not support pulmonary inflammation as a pathway, because there was no pulmonary inflammation observed.
- No association was demonstrated between UFPs and respiratory mortality in Ostro 2015.
- This section incorrectly discusses PM10-2.5 in the summary and causal determination.

Chapter 6: Cardiovascular Effects

Short-Term PM2.5 Exposure and Cardiovascular Effects

Biological Plausibility:

- The first pathway starts with respiratory inflammation, leading to systemic inflammation. However, there was poor evidence of respiratory inflammation in CHE studies, and animal toxicology studies only showed respiratory inflammation when coupled with a strong allergen such as OVA, or SO2-damage to induce chronic bronchitis (discussed in comments on Chapter 5). Where does the systemic inflammation come from, in the absence of respiratory inflammation?
- Evidence for systemic inflammation in CHE studies shows some studies with increasing blood immune cells (neutrophils, monocytes, etc), but the types of cells are inconsistent, and other studies don't show increases. Behbod 2013 – shows increased leukocytes and neutrophils, but links these to endotoxin, not CAPs, and no increase in soluble inflammatory markers. Urch 2010 showed an increase in blood IL-6 at 3 hours after exposure to CAPs >100 ug/m3, but not with CAPs>100 + 120 ppb ozone; Brook 2009 – at 148.5 ug/m3 showed increased WBCs and neutrophils, but not soluble inflammatory mediators; Gong 2004 saw increased basophils in blood at 4 hrs after exposure to 167 ug/m3 in healthy older adults, but not in older adults with COPD. Studies that did not find changes in blood immune cells and/or soluble inflammatory mediators: Bellavia 2013, Devlin 2003 (or one of this group of studies), Gong 2003, Hazucha 2013, Hemmingsen 2015a, Huang 2012. Ghio 2003 saw a decrease in total blood leukocytes. Brauner 2008 saw no change in soluble inflammatory markers.
- Budinger 2011 is cited as evidence that PM induces inflammation in the lung, which increases systemic thrombosis. This group exposed mice to Chicago CAPs at 88.5 ug/m3 for 8 hrs per day for 3 days and saw a 2.5-fold increase in IL-6 mRNA, as well as TNF-alpha. These mice also had a 2.5-fold increase in blood thrombin-antithrombin complexes and adipose PAI-1. They also intra-tracheally installed 200 ug of CAPs, and saw lung injury, decreases in clotting time (not dependent on PAI-1), and much higher increases in IL-6. This shows that increases in PAI-1 aren't necessary for changes in clotting time, and that the changes in the clotting parameters are not PM (or inflammation) dose-dependent. Unfortunately, the authors did not look at lung injury or clotting time in the inhalation-exposed mice. This evidence does not support the hypothesis that inflammation in the respiratory tract is related to clotting effects, because when substantially

more inflammation was induced (by intratracheal installation instead of inhalation), there was no further increase in blood thrombin-antithrombin complexes.

- EPA also references Xu 2013 to demonstrate an increase in systemic inflammatory mediators (although Xu 2012, discussed in the respiratory section, did not see this increase). Xu 2013 exposed mice to 143.8 ug/m3 Columbus OH CAPs for 6 hrs/day, 5 days per week for 5, 14, or 21 days. Adhesive leukocytes were increased at 14 days only and rolling leukocytes at 21 days only. Serum cytokine MCP-1 was increased at 5 days only, with no change in IL-6, IL-10, IL-12, TNF-alpha, or IFN-gamma. IL-6 expression was increased in epididymal fat at 5 days only. An increase in activated macrophages was seen in bronchial sections (but not BALF) with no change in neutrophils. The authors incorrectly conclude that there is a neutrophilic response. This is not convincing of a systemic inflammatory response.
- EPA state that there are CHE, epidemiology, and animal toxicology studies showing changes in thrombotic measures after PM exposure. The CHE studies they cite are Lucking 2011, Ghio 2000 and 2003, and Gong 2003 (incorrectly labeled as Jr. et al.). Ghio 2000 and 2003 both saw increased fibrinogen, but Gong 2003 saw a decrease in Factor VII, Gong 2004 and Huang 2012 saw no change in clotting factors, and Mills 2008 saw an increase in platelets. Lucking 2011 is a diesel exhaust paper. EPA cites Lucking 2011 to state that these increases in prothrombotic factors can increase thrombosis, but filtering particles out of the diesel exhaust in that study did not decrease the size of ex vivo thrombotic plaques.
- For the animal toxicology studies supporting changes in thrombotic measures, the EPA cites Kodavanti 2005, who studied the total results from 6 one-day CAPs exposures and 7 2-day CAPs exposures. There were no biological effects (pulmonary or systemic) in the SH rats exposed for 4 hrs to 1172-1765 ug/m3 CAPs. Two-day exposures (4 hrs each) to 144-2758 ug/m3 caused variable responses. No breathing parameters were different in the WKY rats, but the SH rats had decreased breathing frequency. The WKY rats had a decrease in total and macrophage cells in BAL, and an increase in neutrophils, but not on the day with the highest concentrations. No BALF changes were observed in the SH rats. There were increases in GGT (a damage marker) and fibrinogen in the SH rats, but not on the days with the highest PM. There was no correlation between any response and the PM mass, but some correlation with metals concentrations. This paper does not provide convincing evidence of PM2.5-induced inflammation or thrombotic effects.
- EPA cites changes in vascular function and blood pressure demonstrated in CHE, epidemiology, and animal studies. For the CHE fine CAPs studies investigating BP:
    - Bellavia 2013 showed increased SBP with exposure to 242 ug/m3 Toronto CAPs; Brook 2009 showed increased DBP (not SBP) with exposure to 148 ug/m3 Toronto CAPs; and Sivagangabalan 2011 showed increased DBP (not SBP) with exposure to 154 ug/m3 Toronto CAPs.
    - No effects on BP were seen with fine CAPs exposure in Brauner 2008, Brook 2002, Devlin 2003, Gong 2003, 2004, or 2005, Hemmingsen 2015a, Huang 2012, or Mills 2008. These studies exposed individuals who were healthy, elderly, overweight, with COPD, asthma, or CHD, to PM2.5 CAPs concentrations up to 207 ug/m3. (Brook et al., 2002)
    - Altogether, these CHE studies don't appear to provide convincing evidence of an effect of CAPs on BP.

- EPA suggests that one of the pathways of effects of PM2.5 on the CV system is via activation of the sympathetic arm of the autonomic nervous system (ANS). However, this is the opposite arm of the suggested ANS pathway from the respiratory section.
- EPA notes that there can be resulting heart conduction abnormalities from PM2.5 exposure, as evidenced in CHE, epidemiology, and animal tox studies. However, the EPA does not reference or discuss the results from Langrish 2014 who showed with 12,500 hours of ECG recordings that there was no association between PM concentration and arrhythmia in CHE studies. (Langrish et al., 2014)
- One aspect of biological plausibility that doesn't seem to be considered here is that all of these pathways have to be activated in a single person for there to be movement form the initial exposure to the apical event. Therefore, citing one study for one aspect of the pathway, then another study for another part of the pathway does not prove that the whole pathway could happen in one person. There are studies that look at many of the steps in a single analysis, and the total results from these studies should be discussed, not single independent results. For example:
    - Ghio 2000 and the other studies that published results from this dataset (Devlin 2003, Harder 2001, Holgate 2003) measured respiratory effects, and looked at systemic inflammation, BP effects and HRV. While increased pulmonary neutrophilia was observed, there was no increase in systemic inflammatory markers, and no change in BP or HRV. This study exposed healthy individuals to on average 120 ug/m3 Chapel Hill fine CAPs. (Harder et al., 2001)
    - Gong 2003 exposed healthy and asthmatic individuals to 141 ug/m3 LA fine CAPs and observed no respiratory or systemic inflammation, and a decrease in heart rate and an increase in the high frequency HRV, both of which are indicative of parasympathetic activation. The lack of pulmonary or systemic inflammation suggests that this is not mediated by inflammation.
    - Huang 2012 exposed healthy individuals to 90 ug/m3 Chapel Hill fine CAPs and did not observe any signs of pulmonary or systemic inflammation, nor changes in pulmonary function, clotting factors, BP, HR, or HRV.

Ischemic Heart Disease and Myocardial Infarction:

- This section notes a diminishment in concern about exposure measurement error from the last review because of better exposure modeling. However, there is no discussion about concern for copollutant confounding from the last review, although I don't see from the study summaries that the studies looked at copollutant confounding. Has this concern been addressed?
- Section 6.1.2.1 (ED visits and HAs) concludes by saying that recent studies "continue to provide evidence for positive associations between short-term PM2.5 exposure and IHD ED visits and HA." However, in the paragraph before it there was discussion of one study with a positive but not statistically significant result (Bell 2015), one with a positive statistically significant result (Kloog 2014), one with associations only in NYC but not the rest of the state (Hsu 2017), one with associations in 2 of 7 states (Talbot 2014), one with a negative association (Milojevic 2014), and two single city studies with opposite results (Kim 2012, Sarnat 2015). How does this add up to continuing to provide positive estimates? And where is the discussion of copollutant confounding?

- MI studies are even less consistent, although EPA concludes again that there is a generally positive association. Figure 6-2 shows studies on both sides of the "no-effect" line, and the majority aren't statistically significant. "Although not all studies observed positive associations, overall, recent administrative studies continue to provide evidence of a positive association between PM2.5 and MI, particularly for immediate lag periods (see Section 6.2)." What is your criteria for evidence of a positive association? This statement is also not supported by EPA's note that the MI registry-based studies, which have less outcome error than the administrative studies, show even less consistent results. If there is a real association, you would expect that studies with better outcome assessment would show more consistent, cleaner associations.
- Final statement: "Consistent, positive associations across multicity and single-city studies continue to provide strong evidence for the relationship between short-term PM2.5 and IHD that is unlikely to be driven by chance or systematic bias." However, only measurement error was even discussed in this section, although these studies didn't directly assess it. If systematic bias (e.g. from copollutants) is not addressed, then it can't be ruled out. Also, there weren't consistent, positive associations, as noted in the previous text, there were many null and negative associations.
- Studies of ST-Segment Depression: It is not clear why the CHE studies are not discussed in this section. ST-segment changes have been measured in several studies.

Heart Failure and Impaired Heart Function:

- Two of the studies in Table 6-3, published in 2015, are not marked as being published since the last ISA.
- Although the EPA again concludes that there is consistent positive evidence, similar heterogeneities in presented results are observed as were noted in the IHD section – some positive, some null, some negative. In addition, there was no discussion of copollutants or other biases.
- CHE: The EPA cite Vieira 2016 that shows decreased pulse O2 (a surrogate marker of ventricular stroke volume) in CHF patients with exposure to 325 ug/m3 DE for 6 minutes with submaximal exercise, but not when the particles were filtered out. This study is missing important details, like how the subjects were exposed, or when measurements were taken. This is a great example of the importance of study quality criteria.

Cardiac Electrophysiology, Arrhythmia, and Cardiac Arrest:

- This section concludes that there are inconsistent results for arrhythmia HAs and ED visits, which is consistent with the data showing positive, negative, and null results.
- Associations between PM2.5 and arrhythmia in panel studies was fairly inconclusive in the 2009 ISA, but there are more studies with less severe arrhythmia (e.g. atrial) in this ISA that the EPA considers as showing largely positive associations. As before, when looking at the details the studies show mostly positive effect estimates, but often are not statistically significant, and some are null or negative.
- Conclusion for conduction abnormalities: "Although evidence from recent studies is inconclusive, taken together these studies indicate a potential for cardiac depolarization and repolarization disturbances by PM2.5. These disturbances may increase the risk for malignant ventricular arrhythmias that could result in cardiac arrest." I think concluding that there is a

"potential for polarization disturbances" is a better, more nuanced conclusion than stating something like "the data are generally positive".

- CHE Studies – EPA cites Gong 2000 and 2003 showing no effect on conduction, but Gong 2004 showing effects only in healthy adults (not COPD), and then Tong 2012, Kusha 2012, and Sivangangabalan 2011 as showing evidence of conduction changes. Not mentioned as showing no effect are Devlin 2003, Huang 2012, and a thorough review by Langrish 2014.
- Summary: "Most studies found at least some indication of conduction abnormalities as measured by ECG." How do you interpret the fact that many found no changes, and that the changes that were found were in different indicators?

Cerebrovascular Disease and Stroke:

- EPA notes that "Older age, female sex, smoking, obesity and prior stroke are known risk factors for stroke and should be considered in epidemiologic analysis." Therefore, these should be explicitly discussed in the following sections.

Blood Pressure and Hypertension:

- EPA notes in their summary that the epidemiological study results are inconsistent, but animal and human studies show some impacts of PM2.5 on blood pressure. I discuss the inconsistent human results for this endpoint in the biological plausibility section.
- HA and ED visit studies, and panel studies for BP, have had mixed and inconsistent results. Quasi-experimental studies generally did not show associations between PM2.5 and BP. However, EPA found that panel studies of older populations, particularly in nursing homes or assisted living facilities showed more consistent associations. This makes some sense – you would expect older people, or those recovering from cardiac arrests, to be more sensitive.
- CHE studies – in the previous review there were inconsistent effects on BP. The EPA suggested then that longer follow up may be required to see changes in BP. However, several of the more recent CHE studies have shown BP changes only during, and not after, exposure. Not included in the studies that don't show changes in BP are: Brauner 2008, Brooks 2002, Devlin 2003, Fahkri 2009, Gong 2003, 2004, 2005, Huang 2012, and Mills 2008. (Fakhri et al., 2009)

Peripheral Vascular Disease, Venous Thromboembolism, Pulmonary Embolism

- EPA considers the evidence for a connection between PM2.5 and PVD to be uncertain, despite evidence form CHE and animal toxicological studies showing changes in thrombotic factors. They present somewhat-consistent results in ED visit and HA studies.
- Why is there no discussion of the animal and human data?

Combined CV-Related Events

- EPA concluded from the 2009 ISA that there is strong evidence of associations between PM2.5 and total CVD HAs or ED visits, and that the more recent evidence adds to that conclusion. Does it make sense that most of the separate diseases have inconsistent evidence, but the total diseases have stronger evidence? Is this an indication that sample size is driving the association?
- Effect estimates for HA and ED visits are quite small (1.01-1.1), many are not statistically significant, and some are null or negative. Some of the estimates in Figure 6-6 may be missing error bars.

Cardiovascular Mortality

- EPA concluded in their 2009 ISA that there were consistent positive associations between CV mortality and PM2.5, with a 0.47-0.94% increase in mortality per 10 ug/m3 increase in PM2.5. They note that more recent studies of total CV mortality are consistent with this conclusion, but that cause-specific mortality results are less consistent. That makes sense with the morbidity outcomes outlined above but doesn't provide a causal pathway. Most of the new studies showed positive effect estimates, but many weren't statistically significant.
- The ISA 2009 also concluded that there was coherence with CHE and animal tox studies, but this ISA doesn't say how the EPA arrived at that conclusion. Was mortality observed in the animal toxicological studies? There was no conclusion about CHE or animal toxicological studies from current data.

Heart Rate and Heart Rate Variability

- EPA summarizes that there is additional evidence across disciplines that PM2.5 can impact HRV, but limited/inconsistent evidence that it can impact HR.
- Panel studies - The studies that EPA reports seem to show inconsistent results – some with no change in one HRV marker, but a change in another, and then the opposite in another study. EPA considers this to generally show that PM2.5 can lead to changes in HRV, but it is difficult to draw any conclusions beyond that. Many of the effect estimates are not statistically significant, which is often not discussed.
- CHE – EPA notes a few recent studies that don't show any change in HR, and some that do. From my notes, ones that show changes in HR are Gong 2003 (decreased HR with PM2.5), but no change with Brook 2009, Fahkri 2009, Gong 2005, Huang 2012, Mills 2008, Sivangangabalan 2011, Urch 2005. EPA states that HR increased with CAPs in Gong 2003, but it actually decreased. Changes in HRV are noted in several studies, but EPA states that the lack of HRV effects found in Huang 2012 may be reflective of the lower exposure concentrations (89.5 ug/m3) – suggests that there might be a threshold.
- EPA states that Brook 2009 showed reductions in time and frequency domain measurements of HRV. The paper states that "The changes in BAD (Table 4), BP, heart rate, and HRV measures (Tables S1 and S2) did not differ across the 4 different exposure conditions when measured at any time point outside the chamber", and from the supplemental tables there doesn't seem to be changes during exposure either. The EPA should revise their conclusions about this paper that clearly didn't show effects in HRV.
- CHE Conclusions: "Considered as a whole, the CHE studies discussed above provide some evidence of a change in HRV following PM2.5 CAP exposure, but not following exposure to DE." There are several studies that don't show an effect, not just the DE study.
- Animal Tox studies – They present some evidence that studies show some changes in HR, but it is inconsistent in direction. The HRV data is also inconsistent and in different directions – the EPA presents this as showing a pattern with diet or season, but it just seems to show heterogeneity and inconsistency in responses.

Systemic Inflammation and Oxidative Stress

- EPA notes that the evidence for inflammatory changes with PM2.5 remains limited, because some studies show increases in inflammatory mediators, while others don't. They note that there

are a few more animal toxicology studies showing increases in oxidative stress markers. The EPA should consider how this conclusion impacts the biological plausibility pathway.

- CHE – EPA said that in the 2009 ISA there was essentially no evidence of systemic inflammation. For this ISA, they cite Behbod 2013 as having increased leukocytes at 0 hr, but not 3 hrs post-exposure. For the Urch 2010 study the increase in IL-6 was seen with the higher PM exposure, but not with PM + O3. EPA presents a considerable number of studies that do not show any effects. Despite this, they conclude "Overall, the evidence presented above is inconsistent. This is not unexpected however, given the variability in design and subjects across these studies (Table 6-24). Thus, it can still be concluded that the studies presented above provide limited evidence that short-term exposure to PM2.5 can result in an increase in inflammation. Moreover, these results also provide evidence that the amount of endotoxin present in PM2.5 exposure appreciably contributes to inflammatory potential." This is not justified by the data. EPA also did not find any evidence of increased oxidants in blood or urine but noted that different endpoints may have different results.
- Animal studies – as with the CHE studies there are inconsistent results, but EPA says there is evidence of some effects, and the study design can significantly impact the results.

Coagulation:

- EPA concludes that despite limited and inconsistent evidence in CHE and epidemiological studies, animal studies showing increased clotting factors in genetic mouse models, but not in rats, means that there is evidence of PM-induced clot formation. This data is not consistent with this conclusion – just mouse results in direct contrast to negative human data.
- CHE – For previous studies, EPA notes that Gong 2003 did not find any change in fibrinogen, or vWF or Factor VII. In fact, they observed a decrease in Factor VII. From the new studies there seems to be very little evidence of effect, and Tong 2015 shows evidence of an anti-thrombotic effect (as with Gong 2003). Again, EPA concludes that while the evidence is inconsistent, because of differences in subjects and study design that there is some evidence of PM2.5 promoting pro-thrombotic changes. What aspects of the study design and subjects do you suspect of causing the inconsistency? There are enough studies that EPA should be able to narrow down a particular culprit.
- Animal Tox – EPA cites studies that do not show any effect of PM2.5 pro-thrombotic effect in rats. In mice they cite Budinger 2011 and Chiarella 2014 as showing evidence of PM2.5-mediated pro-thrombotic effects. However, most of those studies are done with PM intratracheal installation, which induces far more damage and inflammation than the 3-day CAPs exposure. Interestingly, in Budinger 2011 despite the far greater damage and IL-6 expression with installation versus inhalation, the increase in plasma TAT complexes is almost identical, which doesn't speak to a dose-response or an inflammatory precursor. That most of the data comes from installation should be reflected in the EPA's discussion and Table 6-28.

Endothelial Dysfunction and Arterial Stiffness:

- CHE – EPA presents results from several studies that show vascular responses to PM2.5, but the responses aren't entirely consistent, nor are they clearly presented. Brook 2009 did not see a significant difference compared to FA control. Several studies saw increases in VEGF, which is a marker of vasodilation and is not consistent with the suspected direction of effects (vasoconstriction). This should be discussed before EPA comes to its conclusion of evidence that

PM2.5 affects vascular function and that there is evidence for an increase in endothelial dysfunction markers in blood and urine. Also not mentioned is Mills 2008 who did not show any vascular dysfunction in older adults who were healthy or had CHD.

- Animal Tox – EPA reports consistent evidence of PM2.5 effects on vascular function, mostly based on ex vivo assays. They also report two studies that show a decrease in circulating endothelial progenitor cells, but don't make the implications of this decrease clear. Which direction of change is adverse, and is it consistent with the direction of effects in other studies?

Policy-Relevant Considerations:

- EPA focuses on copollutant confounding, temperature and season, and lag effects. For the respiratory section there were also considerations of model specification, averaging time, and shape of the C-R function. Were these not addressed in any of the CV studies? If so, that should be discussed at the beginning of this section.
- Temperature and Season- the 2009 ISA concluded that there was variability of PM2.5 associations with CV effects by season, and recent studies have continued to show that. This demonstrates heterogeneity in effect estimates that is not simply explained by sources. This section refers several times to Figure 6-6, which is not a figure showing seasonal effects, it shows ED visits and HAs for CV-related effects.
- Lagged Effects – EPA discusses how studies show effects at different lag periods, depending on the endpoint, or even within endpoints. Generally stronger effects are seen at lag 0 and 1, but some studies have shown delayed or prolonged effects. As I stated in the respiratory section, it seems that these results need to be considered not only for what they say about the timing of PM effects, but whether there are effects at all given the heterogeneity of results.

PM2.5 Components and Sources:

- EPA generally concludes that studies that evaluated sources and species of PM showed inconsistent results. Does it make sense that there is only a "clear" association with PM mass, and not any of the components or sources (which would make more sense toxicologically?).
- This section doesn't seem to assess results from CHE studies, although several have done component analysis compared to total mass. Those should be included, particularly because there is far less question about the causal inferences for the observed effects.
- Sources – EPA states that there is some evidence for associations between traffic PM, and wildfires, and CVD HA. Again, there is a lack of consideration of CAPs CHE and animal studies that can provide some source information.
- EPA presents results from Chen 2010 which also describes the NPACT study, and notes that the mice were exposed for 6 months from May-Sept 2007. Although this is what the authors reported, it should be noted that May-Sept is only 5 months. EPA should also note that while this is a chronic study, the authors looked at acute effects during the day. There does not seem to be any discussion of the fact that opposite results were obtained at the two New York sites.
- Rohr et al present results for a 13-day exposure in Detroit in summer and winter to 518 ug/m3 ad 357 ug/m3 CAPs respectively. Most 8-hr HR and HRV effects weren't affected, but there were some elements associated with increased HRV and decreased HR. An explanation of the opposite response shown here compared to what is the direction of adversity would be helpful. A lot of these nuances aren't presented in the corresponding HRV section in the main text, which would provide more information for interpreting the weight of evidence. EPA also presents a

summary analysis of the results but should also include a better explanation about what the results mean, and what the uncertainty in this integration method is. It should also be noted that the HEI review committee were skeptical of the authors' analysis and it wasn't clear whether components or concentrations impacted the different results in different areas. Was there any evidence of increasing effects over time with PM exposure?

Summary and Causality Determination:

- This whole section emphasizes only the positive results (and often isolated, positive results from CHE and animal studies), making the literature seem cohesive and consistent, when in fact a much more nuanced and far less consistent picture is provided by the detailed data analysis. It doesn't seem like the CVD data is much stronger than the respiratory data – so why the difference in causality determinations?
- EPA notes that there is a coherence in the results from different endpoints, demonstrating the plausibility of an effect of PM2.5 on CVD effects. However, there is no discussion of concentration of effect, heterogeneity, or the type of hypothesis-testing that I recommend, which involves looking at the associations for patterns that would be expected of exposure to a toxicant in the population (e.g. increases in confidence in effect estimates that are more effect-specific, and that are less severe). The opposite actually appears to be true – the more refined and less severe the effect, the less likely it is to be positively significantly associated, making these effects seem more like statistical artifacts. There is also no discussion in this section of bias, chance, or confounding, which could be impacting the effect estimates.
- Respiratory effects are broken down into sub-effects for the causal determination. Why is that not done here? There are a number of CV effects that do not have good supporting data, and these appear to be lumped together and not distinguished from those with more compelling data.

<u>Long-Term PM2.5 Exposure and Cardiovascular Effects</u>

- The long-term PM2.5 CVD section needs to be carefully copy-edited.

Summary:

- The previous determination was causal for long-term PM2.5 exposure and CV effects, with the strongest evidence coming from cohort studies associating PM2.5 with CVD mortality. I would hypothesize that the strongest evidence would come from milder or more specific effects, not from mortality.
- Why does this document focuses on epidemiology studies conducted in areas with PM concentrations less than 20 ug/m3? (i.e. why 20 ug/m3 specifically?)

Ischemic Heart Disease and Myocardial Infarction:

- EPA presents 11 effect estimates for long-term PM2.5 and IHD or MI – only one is statistically significant. Of the 6 US studies, 4 have substantial amounts of data from before the PM2.5 monitoring network was established. A meta-analysis of 11 European cohorts found a non-statistically significant HR of 1.13, and an almost statistically significant HR of 1.19 for PM2.5 concentrations <15 ug/m3. A separate study (Hoffmann 2015) that used physician-confirmed diagnoses did not report an association between PM and coronary events (but did with stroke) after considering noise and other cofactors. There is text missing at the sentence at the end of the paragraph, which ends with Koton (2013).

- The summary paragraph at the end again emphasizes the positive results, even though throughout this section the EPA has emphasized the inconsistent results. They also note that there is little information about copollutant confounding, and that copollutant correlations were generally moderate-to-high.

Cerebrovascular Disease and Stroke:

- Several of the epidemiological studies did not have time concordance between PM2.5 measurements and outcome assessment (e.g. Lipsett 2011 – followup 1995-2000, PM2.5 – 1999-2005). How does this impact study interpretation, given the requirement of the exposure to come before the effect in a causality determination? Of the 9 effect estimates presented, only one is maybe statistically significant (hard to tell – the Hoffmann 2015 study doesn't have an effect estimate, only an interval. If it is off the scale, perhaps the numeric value of the estimate can be put on the graph?). EPA notes that the Hartiala 2016 study showed an effect with wide CI, that was attenuated with consideration of various potential confounders (smoking, obesity, etc.). It seems that the unadjusted estimate is presented in Figure 6-18, but that the adjusted estimate should be presented.
- The Figures provide a column specifying the years for each study, but should also provide information about which years are being specified – years where the PM2.5 concentration was measured, or years of followup/health effect analysis? This is important because years of pollutant measured and outcome assessment aren't always the same.

Atherosclerosis:

- What is the significance of observing changes in DNA methylation in circulating monocytes in the MESA-AIR cohort (Chi 2016b)? No explanation or justification of this result is offered. Also, the study results should be presented in the table, or in a figure.

Heart Failure and Impaired Heart Function:

- Epidemiology studies – EPA notes several studies that show positive associations with CHF or HF and PM2.5, although not necessarily statistically significant. The cross-sectional study, To 2015, is not in Table 6-39.
- There were no positive statistically significant associations with various indices of CHF in the MESA-air study (one for right ventricle mass was positive). Another cross-sectional study (SALVIA) found positive associations between some CHF metrics and PM2.5. Again, effect estimates should be provided in the table or in a separate figure. The data seems insufficient and debatable to me, but EPA concludes that there is evidence of a possible relationship between PM2.5 and CHF and HF.
- Animal Tox – EPA presents evidence from multiple animal studies (although not consistent in all) of effects of CAPs exposure on cardiac wall thickness and heart function. No information is presented about whether any of the exposed animals experienced CHF or HF or died from the exposure. EPA also summarizes several studies that expose animals in utero and found cardiac changes. EPA says that "Tanwar et al. (2017) demonstrated that prenatal exposure alone was sufficient to produce heart failure in adulthood" from exposure to 74 ug/m3 Ohio State CAPs for 6 hrs/day throughout pregnancy. Did these animals experience heart failure (i.e. did their hearts fail)? Also, Tanwar 2017 is listed twice in Table 6-40.

Cardiac Electrophysiology and Arrhythmia:

- Summary – The EPA states that current animal tox evidence is still lacking, although it seems that a lot of animal studies did ECG Analysis, and likely presented some information about cardiac depolarization and repolarization.

Blood Pressure and Hypertension:

- Some studies showed associations between long-term PM concentrations and some BP metrics, but not all studies. The changes were small – about 1 mm Hg. This seems like it is well-within the margin of error of BP measurements, so how significant is this result?
- Hypertension – EPA concludes that there is generally a positive association between long-term PM2.5 and hypertension, particularly supported by the Ontario hypertension study which has better outcome assessment. There are cross-sectional studies listed as being supportive, but no further discussion – why aren't these studies listed in Table 6-44?
- Gestational hypertension and preeclampsia – epidemiology studies generally present inconsistent results. EPA noted that meta-analyses of PM2.5 and preeclampsia showed positive effects but had high heterogeneity scores so it may have been inappropriate to combine studies. This is good to note, and the EPA should strive to make sure that they assess this for all meta-analyses they include in assessments (e.g. Li 2016).
- Renal Function – one epidemiology study observed an association between PM2.5 and reduced glomerular filtration rate. EPA should state whether the change observed was substantive and would be associated with adverse effects.
- Animal tox – studies in rodents showed increased BP with longer exposure to PM2.5 (85-375 ug/m3) as well as changes in the renin-angiotensin system. One study at 85 ug/m3 for 9 months (Wold 2012) showed increased blood pressure and decreases in pulse pressure (the difference between SBP and DBP). What is the expected adverse effect direction for changes in pulse pressure?

Cardiovascular Mortality:

- Pope 2014 and Turner 2016 extended the ACS followup and showed associations between long-term PM2.5 and mortality from HF, cardiac arrest, CVD, and hypertensive disorder. EPA discusses the CanCHEC study, but the first set of results has no reference. The CanCHEC studies generally found associations with PM2.5 and IHD, diabetes, and MI, although one study (Weicenthal 2016a) showed that for people living within 5 km of a group monitor there was a null association (not consistent with the assumption that there is a bias towards the null with less precise exposure assessments).
- Why is Weicenthal 2016a not included in Figure 6-19?
- Why is there no table providing details about these studies? Are they in the mortality chapter? This chapter says that more detail is provided in Section 6.2.10, but this is Section 6.2.10.
- In their summary EPA mentions the large European cohort meta-analysis Beelen 2014 study that showed no association of PM2.5 with CVD mortality except a positive but non-statistically significant association with CBVD. Why wasn't this discussed earlier, or included in the Figure? It doesn't fit with the discussion before it, which states that European studies generally show positive associations.

Heart Rate and Heart Rate Variability:

- The only epidemiology data is from the MESA panel study (Park 2010), which shows non-statistically significant negative associations between 30-day or 60-day PM2.5 and rMSSD or SDNN, with higher effects in people with MetS. Does EPA's lack of consideration of statistical significance apply to panel studies as well?
- Animal Tox – The NPACT studies showed increases in HR in the early days in Manhattan but decreases in HR in the early days in Tuxedo, with no changes in in other study cities (Lippman 2013a). No changes were observed in HRV with chronic exposure to CAPs from any location. Wold 2012 also showed an increase in HR with a 9-month exposure to 85 ug/m3. EPA concludes that there is some evidence of increased HR with long-term PM2.5 (although very inconsistent) but not of increased HRV. Without evidence of changes in HRV, how does this impact the ANS pathway in the biological plausibility section?

Systemic Inflammation and Oxidative Stress:

- Epidemiology studies – some studies showed associations between long-term PM2.5 and increases in CRP, whereas others did not, including the MESA study and the ESCAPE cohort. MESA did show a small increase in circulating IL-6, and it is not clear whether this was observed in other studies. Why is there no table or figure of results? There is a table for the animal studies.
- Animal Tox – EPA presents variable results for inflammatory markers in different animal studies. They note that while these results appear inconsistent, because it is difficult to compare inflammatory markers across studies because of differences in timing and design, this provides information for PM2.5-induced inflammation. Why is the default interpretation of a variable marker automatically on the side of showing an association? Why doesn't this just show that there is a lot of background variability in inflammatory markers that may not have any significance? Or perhaps it shows a threshold or other pattern of response?

Coagulation:

- Summary: several recent studies show that long-term PM2.5 can impact fibrinogen, D-dimer, and platelet count. Why isn't this combined with the thrombosis section? How can you tell which direction of change is associated with pro-thrombosis? It seems that an increase in plasminogen could mean that more is being made in preparation for more clot formation, or that there is less fibrin formation. Similarly with D-dimer.
- Epidemiology studies – most summarized studies showed no effect of long-term PM2.5 on fibrinogen, and those that saw statistically significant effects were in opposite directions. Cross-sectional studies, and a meta-analysis of the ESCAPE cohort showed null effects. This does not support EPA's statement that recent studies show impacts on fibrinogen, D-dimer, and platelet count.

Impaired Vascular Function and Arterial Stiffness:

- Epidemiology studies – the MESA Air and Framingham offspring studies showed small statistically significant negative associations with FMD, but no changes with BAD or several other markers. There was also a small statistically significant negative association with hyperemic flow velocity in the Framingham study. These studies showed changes of 0.5-1.8%, which the EPA says are large given that normal ranges are usually 5-10%. What does this mean?

Normal ranges of normal function or variability? That doesn't make a 0.5% change seem large. Tallon 2017 showed associations with erectile dysfunction, which EPA says may be associated with vascular function. I would recommend removing this, unless a stronger connection is shown. Several studies showed no association of PM2.5 with arterial stiffness. There should be an evidence summary table for this section.

Copollutant Confounding

- Is there a reason that the long-term exposure section doesn't have a separate policy-relevant considerations section, but there is one in the short-term exposure section?
- The beginning of this section notes "A change in the PM2.5 risk estimates, after adjustment for copollutants, may indicate the potential for confounding." How much of a change? This seems very subjective. A statistically significant change?
- EPA states that there are more studies looking at copollutant confounding for mortality, and fewer for morbidity, but those that are available generally show that the effect estimates are unchanged when copollutants are considered. Can you make a final conclusion based on limited data? I would agree that the studies shown in Figure 6-20 mostly show no effects of copollutant confounding, but these also mostly don't show effects of PM2.5 either. There is a lack of labeling on the figure – are the filled circles the ones without copollutants?

Shape of the C-R Function

- Summary: some studies have suggested largely linear concentration-response functions, but in general there is a paucity of information, and cut-point analysis from other studies suggest non-linear C-R curves.
- Morbidity Studies - Kaufman 2016 used the MESA-AIR CAC data to look at the C-R function and found a somewhat supra-linear shape, whereas Dorans 2016 with the Framingham cohort showed a very odd C-R function shape. Cesaroni 2014 found similar HRs below and above a 15 ug/m3 cut-point for the ESCAPE cohort, and Chen 2014 showed an exponential-like C-R function shape. This combination of data makes it difficult to draw any conclusions about the shape of the C-R function. As noted in the general section, there is evidence that variability and error in epidemiology study estimates prevents one from determining the appropriate shape of the curve (Rhomberg 2011).
- Mortality – EPA concludes that most studies support a linear no-threshold response between long-term PM2.5 and CVD mortality, including studies with concentrations <12 ug/m3. Crouse 2012 showed higher risks for IHD mortality at concentrations <10 ug/m3 (although no departure from linearity) – similar in Jerret 2016 and Weicenthal 2014. Two studies by Pope (2009, 2011) showed that the risks at low PM concentrations were higher than the risks at higher concentrations associated with smoking, and so there could be a supra-linear relationship. EPA concludes that "This indicates the importance of considering the cause of death when characterizing the concentration-response relationship between long-term PM2.5 exposure and cardiovascular mortality." While I agree that all types of death shouldn't be lumped together, how does this conclusion follow from the statement about supra-linearity and cigarette smoke? Also, do the Pope analyses really make sense – that the risks from ambient PM are more than the risks from smoking? Shouldn't this call into question the PM results?
- Why isn't animal study data used to produce/inform the C-R function? At least it can be used for comparison to the epidemiology study results.

PM2.5 Components and Sources

- EPA states that Wolf 2015 showed positive associations with PM2.5 components in the ESCAPE cohort – which components? Information in this section is disjointed and needs to be better organized (or put back into the individual sections). From Figure 6-28 there seems to be more statistically significant associations with BC than with PM2.5 – what can be concluded from this?
- Regional Heterogeneity – EPA summarizes some data from studies showing evidence of regional heterogeneity in effect estimates. What conclusions do you draw from this? Are there any obvious areas that always (or never) show associations that can be used to draw conclusions? Why is there no regional heterogeneity section for short-term exposure, where there is more data about it?
- Animal Tox studies on components and sources – EPA discusses the Campen 2014 study at length, which exposed animals for 50 days to motor vehicle exhaust, particle-filtered exhaust, or sulfate, ammonium nitrate, or road dust particles at 300 ug/m3. Mostly there was very little biological response, with some evidence of vascular changes. Rohr et al 2011 showed inconsistent source associations between winter and summer CAPs exposures in Detroit. Conclusions from the animal tox studies?

Summary and Causal Determinations:

- This section does not discuss any of the negative evidence and EPA's own "inconclusive" determinations for different endpoints.

<u>Short-Term PM10-2.5 Exposure and Cardiovascular Effects</u>

Biological Plausibility:

- EPA references section 5.2 to show that exposure to PM10-2.5 can cause respiratory tract inflammation. But section 5.2 is for long-term PM2.5 exposure. Section 5.3.1 is the PM10-2.5 biological plausibility section, and the evidence there for respiratory inflammation from coarse particle exposure in CHE studies are inconsistent.
- EPA cites Behbod 2013 as showing a potential increase systemic inflammation with PM10-2.5 exposure, even though there was no increase in soluble inflammatory mediators in the blood or respiratory tract – how does this inform the entire pathway? They also cite Graff 2009 as showing evidence of hemostasis effects because of a decrease in tPA, but don't mention that there was no change in platelets, Factor VII or IX, fibrinogen, PAI-1, vWF, protein C, prothrombin, plasminogen, or D-dimers. All of these tests make the tPA result seem like it might be spurious.
- For modulation of ANS, EPA cites Brooks 2014 for showing changes in HR and HRV with coarse particle exposure. However, these results were not concentration-responsive by regression analysis (subjects were exposed to a wide range of concentrations), and there was no pre-exposure measure, just a beginning of exposure measure for comparison. Byrd 2016 didn't show an effect on HRV. Very small changes in BP were seen in the referenced study, and different components of BP. Zhong 2015 is cited as showing BP changes, but it shows that endotoxin mediates this effect.

Systemic Inflammation and Oxidative Stress:

- EPA notes that the findings for systemic oxidative stress have been inconsistent, but given the transient nature of the effects, this is to be expected. This time they didn't go on to conclude that because of the transient effects, the few studies that showed effects must be showing a real effect.

Long-Term PM10-2.5 Exposure and Cardiovascular Effects

- Biological plausibility – EPA says that there is a plausible pathway connecting long-term exposure to PM10-2.5 and apical events, but then it offers almost no such data.
- The summary on pg 6-272 talks about the inconsistency in the epi study results, the attenuation with PM2.5 copollutant analysis, the poor measurement method (subtraction), and the lack of biological plausibility information. How does this earn a "suggestive of causality" designation?

Short-Term UFP Exposure and Cardiovascular Effects

Biological Plausibility:

- The first step in one of the pathways is respiratory inflammation, but there is very little evidence in numerous human studies that this actually occurs (see comments on Chapter 5).
- Liu 2015a and Devlin 2014 are cited as showing increased systemic inflammation (but not cited for respiratory inflammation – didn't look for it or didn't see it?). From the other studies there seems to be a decrease in inflammatory cells and mediators in Frampton 2004, and no change in either for Gong 2008 or Samet 2009.
- EPA also cites Devlin 2014 as showing evidence for altered vascular function and hemostasis. However, Frampton 2004 and 2006, and Gong 2008 showed no increase in clotting factors (Frampton 2004 showed decreases) or increased expression of vascular adhesion molecules (Frampton 2004 showed decreases). Gong 2008 showed no changes in vascular function. The only evidence from these studies was an increase in D-dimers in Samet 2009 (but no change in platelets, fibrinogen, factor VII or IX, vWF, PAI-1, tPA, or plasminogen).
- Samet 2009 and Devlin 2014 are cited as showing changes in HRV. But Samet 2009 showed an increase in HF, which is usually associated with increased parasympathetic activity and is not in the direction that EPA suggests is activated by PM.
- EPA cites epidemiology panel studies as showing increases in BP, while not citing the human studies that show no changes in baroreflex (Frampton 2004 and 2006, and Gong 2008). EPA cites Samet 2009 and Devlin 2014 as showing evidence of conduction abnormalities and arrhythmia but didn't cite the Langrish 2014 review that showed no arrhythmia effects with 12,500 hours of ECG recordings in many types of PM exposure studies.

Health Effects

- EPA concludes that overall epidemiology studies do not support an association between short-term UFP exposure and IHD or MI. EPA discusses the paucity of panel studies showing ST-depression with short-term PM2.5 but should also note that there are two CHE studies (Frampton 2003, Gong 2008) who didn't find ST segment changes.

- HF and impaired HF – in the summary, EPA refers to findings from a PM10-2.5 study, not a UFP study. Why are there tables with study information for the tox studies, but not for the epidemiology studies?
- Arrhythmia, Cardiac Arrest, Electrophysiology – The summary states that the 2009 ISA reviewed one epidemiology study of arrhythmia ED visits or HAs, and then in the epidemiology study section they say that there were no epidemiology studies of arrhythmia and ED visits or HAs. Which is it? This section references CHE studies showing a decrease in QT interval from Samet 2009 (this seems to be in the opposite direction of adversity), and an increase in GSTM1-null individuals in Devlin 2014. No mention of Frampton 2004 who also showed a decrease in QT, or Langrish who showed no change.
- CVD Mortality – there are only studies in Europe and China, and they suffer from serious exposure estimation problems. They show some positive effects, but most are not statistically significant. These studies are listed in Chapter 11 – EPA says table 11-9, but that is for PM10-2.5, not UFP.
- Generally, conclude that the evidence is suggestive by pointing out the few positive association studies. This isn't consistent with the details in individual sections which is largely inconsistent.

Chapter 7: Metabolic Effects

Short-Term PM2.5 and Metabolic Effects

- What kind of metabolic effects could be caused by short-term exposure to PM2.5? These are usually only adverse or only develop in the long-term. There should be some discussion of this in the document.

Biological Plausibility:

- EPA suggests that the activation of the ANS system by PM2.5 will cause an increase in output of norepinephrine (measured in an animal study, although not increased in at least one human study that looked at norepinephrine – Graff 2009) and might also increase the output of glucose (not measured). There needs to be a better connection between a transient increase in glucose (which happens every day when you eat), and changes in homeostasis and disease. (Graff et al., 2009)
- EPA cites Kim 2015 as evidence of effects of PM2.5 on liver function in humans – it should be noted that this is an epidemiology panel study, not a CHE. In general, the evidence of respiratory and systemic inflammation is poor with short-term exposure to PM2.5 (discussed in previous comments), and so does not provide great support for metabolic effects. Also, as noted above, the species and circumstances of the exposure data should be part of the discussion of biological plausibility.

Glucose and Insulin Homeostasis:

- As noted elsewhere, it would be good to discuss the significance of changes in biomarkers – for example, a 0.8 ug/dL change in glucose levels – what is the clinical significance of that?
- Animal tox – some data indicating changes in metabolic parameters, but not consistent. There were also 5 papers in Table 7-4 that were not discussed in the text – is there a reason for that?

These papers seem to be discussed in the following sections – the following sections should have references to this table.

Other Indicators of Metabolic Function:

- Inflammation – see comments in CVD chapter on systemic inflammation.
- Liver Effects – EPA cites a few epidemiology studies that show a change in CRP, and one animal study that does not. There is no mention of the human controlled exposure studies that show no change in CRP with exposure to fine CAPs: Mills 2008, Huang 2012, Graff 2009, Gong 2008, Ghio 2003, Brooks 2009, Brauner 2008, Behbod 2013. In sections such as this that probably weren't included in the last ISA, how do you deal with studies that were previously published with potentially relevant markers, such as CRP?
- Blood Lipids – Not mentioned in the CHE studies section is Huang 2012, who showed no change in a lipid profile with Chapel Hill CAPs exposure, except for an increase in blood HDL (opposite direction of adversity).
- Blood pressure – EPA notes that there is limited evidence for changes in blood pressure with PM2.5 exposure (somewhat less confidence in this endpoint than was noted in the CVD chapter). Would one expect a change in the ANS without a change in HR and BP? My comments on CHE BP studies can be found in reference to chapter 6.

<u>Long-Term PM2.5 Exposure and Metabolic Effects</u>

- Glucose and Insulin Homeostasis – Presented are a number of changes in various glucose and insulin homeostasis measures, but it would be good if these could be placed in the context of normal levels or normal changes in those metrics, or changes that would be considered adverse. For Figure 7-7, it would be helpful if the exposure concentrations and times were added to the labels, to give the reader the chance to see if the responses were concentration or exposure-time dependent. Animal studies are in Table 7-8, not 7-7.
- Other Metabolic Effects – many animal study results are presented here. As before, it would be valuable to include concentrations (or, better, modeling to a human equivalent concentration) in the discussion so as to put this in the context of human exposures. Also, some discussion of what the animals were exposed to would help – CAPs? Where were the CAPs generated? It seems that most of the studies were done in Columbus OH, which makes it difficult to translate study results to other locations or pollutant mixtures.
- Type I Diabetes – the two epidemiology studies that are compared seem to use different PM measures – Beyerlein 2015 uses PM2.5, and Rosenbauer 2016 uses PM10, supposedly as an attempt to repeat Beyerlein. Perhaps Rosenbauer 2016 isn't relevant for this section.
- Mortality – The summary in this section is supposed to refer to a figure that summarizes the findings from the ACS and CanCHEC cohorts, but Figure 7-12 provides a C-R function, not a summary of effect estimates. The table where the summary information can be found from these studies should be referenced here. Also, if Chapter 11 has a summary figure, that should be referenced. If not, there should be a summary forest plot in this section.

Long-Term PM10-2.5 and Metabolic Effects

- Biological Plausibility – as noted in previous sections in this chapter, the type of study and study population being discussed in this section should be noted. For example, Wolf 2016 should be labeled as a cohort study.
- The causal determination is suggestive of causality, based on one epi study that showed a non-statistically significant positive association. This is not adequate data for this determination.

Short-Term UFP and Metabolic Effects

- EPA only cited one study in this section, showing some effects of 28-day exposure in a longitudinal epidemiology study. They didn't cite Samet 2009, who studied the effects of blood lipids with UF CAP exposure and saw decreases in TG and VLDL at 0 hrs after exposure.

Chapter 8: The Nervous System

Short-Term PM2.5 Exposure and Nervous System Effects

Biological Plausibility:

- The EPA asserts that PM2.5 exposure activates lung irritant receptors, which cause lung irritant responses and also signal to the ANS, causing HRV. However, lung/respiratory irritant responses are mild to non-existent in human studies (measured by symptoms, FEV1 changes, etc – I discuss the CHE studies in the respiratory chapter comments), so the data isn't consistent with this pathway. Is there evidence that you can activate the TRP sensory nerves and have systemic effects in the absence of local irritant effects?
- Further, the respiratory chapter presented evidence of activation of the parasympathetic ANS, whereas this chapter discusses activation of the sympathetic ANS. Can they be activated differentially? Is this a time-dependent occurrence? This discrepancy needs to be addressed (it is also relevant to the biological plausibility section in the CVD chapter).
- As noted before, there is little supportive data that shows respiratory tract inflammation, making this an unlikely upstream component. EPA notes that two studies show brain inflammation in the absence of respiratory or systemic inflammation, using this as evidence to surmise that PM may act directly on the brain (Tyler 2016, Bos 2012). This may be true, but it also contradicts the EPA's presented inflammation pathway (that requires respiratory inflammation as an initial step). Bos 2012 seems to show a decrease in inflammatory mediators in the olfactory bulb, not an increase.
- EPA states that brain inflammation leads to particle uptake in the brain, citing Ljubimova 2013 – This paper doesn't show particles in the brain.
- The Fonken 2011 study showed some changes in behavior and cognition with 10-month Columbus CAPs exposure (95 ug/m3 6 hr per day 5 days per week). This paper is cited as a reference for the statement "Brain inflammation may be due to peripheral immune activation (Fonken et al., 2011)". It is not clear where the evidence is for peripheral immune activation in this study. It is also a chronic study cited in the acute section.

- EPA states that a CHE shows evidence of an impact of PM.5 on the blood-brain barrier (Liu 2017). This seems to be the only human study cited in this section – there is more human data for the earlier steps in the pathway, although it is often not supportive (but still informative). Liu 2017 doesn't show an increase in the BBB-related proteins S100B or UCHL1 with exposure to any of coarse, fine, or UFP CAPs, and therefore does not demonstrate a change in the BBB.

Activation of the SNS and Hypothalmus-Pituitary-Adrenal Stress Axis:

- CHE – the only study cited by EPA is Liu 2017, which shows no effects of PM2.5 exposure on the SNS or the HPA stress axis. I do appreciate that EPA cited this study and were upfront about the results.
- Animal Tox – EPA cites Balasubramian 2013 as showing effects of PM2.5 exposure on the SNS and HPA after 1 day of exposure, but not after 3 days of exposure. Firstly, this is mis-represented in the text ("increased levels of norepinephrine in the paraventricular nucleus of the hypothalamus 1 day ($p < 0.05$), but not 3 days, after exposure") – the way this is written suggests that the effect wasn't seen 3 days after exposure (which would suggest a reversal), as opposed to the lack of effects after 3 days versus 1 day of exposure, which suggests adaptation to the stimulus. Secondly, EPA should offer some analysis of why this pattern is seen (or offer the authors' analysis of it).

Brain Inflammation and Oxidative Stress:

- CHE – EPA again cites Liu 2017, this time for an increase in blood ubiquitin C-terminal hydrolase L1, which they said is "related to blood-brain barrier integrity", and that BBB integrity impacts brain inflammation. Was the level of the bio marker enough/in a range to determine that this was an adverse effect? In addition, the study does not show a statistically significant increase in this marker – the p value is >0.05.
- Animal Tox – the results summarized from the Bos 2012 study seem to suggest that in traffic-air exposed C57/Bl6 mice there was an up regulation in inflammatory genes in the hippocampus, but a down-regulation of inflammatory genes in the olfactory bulb, where presumably PM concentrations would be higher. How is this result explained? Did the authors look at protein expression, in addition to gene expression? Tyler 2016 shows a down-regulation of inflammatory markers in the hippocampus of WT C57/Bl6 mice but increases in ApoE-/- mice exposed to Chicago CAPs. Ljubimova 2013 saw no gene or protein expression changes in F344 rats exposed to Riverside CA air for 2 weeks (longer than the other studies). Bos 2012 had lower concentrations than Tyler, but for a longer duration. May also be the exposure conditions – for Bos 2012, the mice were actually in a roadway tunnel. How does a dampening of the response with increased exposure inform the biological plausibility pathways?

Disease of the Nervous System and Depression:

- Epidemiology studies show limited and somewhat conflicting associations. As in the other chapters, there should be some sort of summary at the end of each section, with a preliminary conclusion about the evidence presented in that section.

Summary and Causal Conclusions:

- It is not clear why the causal conclusion for this section is "suggestive". There is some animal evidence, but no attempt is made to convert this to relevant human concentrations, or to discuss the evidence of differences in effects in rodents versus humans in dosimetry and deposition. In addition, there is evidence of an adaptive response, and no discussion of whether there was an overt, adverse effect of the gene expression and norepinephrine changes in the animals.

Long-Term PM2.5 Exposure and NS Effects

Biological Plausibility:

- In Figure 8-2, EPA draws a line from RAS activation to SNS activation. However, they state that there is animal evidence that PM2.5 activates the SNS, which then impacts blood pressure (a RAS-impacted pathway), so which is the proper directionality between these pathways? Also, there is a lot of emphasis put on one study showing an impact of PM exposure on the RAS.

Brain Inflammation and Oxidative Stress:

- Describing the specifics of the animal study results shows inconsistencies in the findings between different studies, even with similar types of exposures (Columbus OH CAPs). Do these studies demonstrate inflammation by means other than mRNA expression? Perhaps protein expression or changes in immune cells? The results are restricted to Columbus OH CAPs, except for a study on resuspended DEP that showed somewhat different results. The only rat study showed no effect – which of these factors is mediating the effects seen?

Neurodegenerative Diseases:

- Parkinson's Disease in epidemiology studies showed some positive associations, but mostly they weren't statistically significant, and they were lower and not statistically significant in the studies with better outcome assessment. Are these studies for PD incidence? Hospital admissions studies are particularly problematic for this endpoint, because it is hard to determine the cause of the HA was (noted in the Ozone workshop discussion).

Neurodevelopmental Effects:

- The type of effect estimate is not always labeled in this section.
- ASD – somewhat consistent positive associations, but often not statistically significant. What confounders were considered in these studies? The only animal study cited is Klocke 2017, who show various brain morphological effects from GD 0.5-16.5 exposure for 6 hr/day to 93 ug/m3 NY CAPs. What is stated in the biological plausibility section, and should also be stated here, is that there was no evidence of cognitive or behavioral effects in this study.

Summary and Causal Determination:

- The conclusion that there is likely to be an effect on the CNS is not supported by the epidemiology studies that show largely null and inconsistent results. It may be supported by the animal studies, but the appropriate dose modeling to compare to doses experienced in humans has not been done, and most of the animal studies that provide coherence were done by a single group in a single location.

Short-Term Exposure to PM10-2.5 and NS Effects

- Section 8.3 has the wrong section referenced in its summary section.

Biological Plausibility:

- the EPA cites an animal study showing coarse CAPs causing decrements in lung function as potential evidence of ANS activation. However, human CHE studies show no effect of coarse PM exposure on lung function (Gong 2004, Graff 2009). The EPA states that Liu 2017 supports an association between coarse PM and activation of the HPA axis. Liu is again referenced for showing changes in the blood brain barrier – this study does not show changes in mRNA expression of the genes associated with BBB dysfunction, just endotoxin, not PM coarse mass. This section notes that the rodent effects are likely nasal, but don't go all the way to using the available dosimetry data to extrapolate about whether this would be relevant for humans.
- The Fonken 2011 study showed some changes in behavior and cognition with 10-month Columbus CAPs exposure (95 ug/m3 6 hr per day 5 days per week). This paper is cited as a reference for the statement "Brain inflammation may be due to peripheral immune activation (Fonken et al., 2011)". It is not clear where the evidence is for peripheral immune activation in this study. It is also a chronic study cited in the acute section.

Activation of the SNS and Hypothalmus-Pituitary-Adrenal Stress Axis:

- Liu 2017 is the only paper cited here, and they note the $P<0.1$ effects of changing biomarkers of BBB integrity – this was not a significant response, and there were plenty of samples for statistical power for this endpoint. Also, EPA should explain or address the fact that although urine cortisol went up, blood cortisol went down – what does that say about activation of the HPA axis?

Long-Term Exposure to PM10-2.5 and NS Effects

- Section 8.4 has some wrong section references in its summary section.
- Biological Plausibility – This section states that there is not enough data to be sure of the biological plausibility pathways. If this is the case, then the EPA should not draw a biological plausibility pathway figure, which misleads the reader into thinking that there is good data to support the pathway.
- Brain inflammation – what are the Arc and Rac genes, and what do their gene products do? Do they promote inflammation?
- Cognitive and behavioral effects in adults – the text references Table 8-24 for study-specific information, but it should reference Table 8-25.
- Causality – A better explanation should be provided for the suggestive causal determination. There is basically no data for mechanism or biological plausibility, and there are inconsistent epidemiology study results that don't account for copollutants.

Short-Term Exposure to UFPs and NS Effects

Biological Plausibility:

- The EPA cites Maher 2016 as showing that magnetite UFP, likely from combustion sources, is found in the brain. However, there is no evidence from that paper about where that UFP came from, or controls for people who would have been exposed to more or less UFP.
- Again, changes in lung function as an upstream indicator of ANS pathway activation is not very convincing, because of the lack of lung function changes observed in many CHE studies (Frampton 2004, Gong 2008, Samet 2009). The reference shouldn't be for Jr. 2008, it should be for Gong 2008, and there was no significant change in FEV1 observed with time (just a trend).
- Again, Fonken 2011 is referenced for peripheral immune activation (not shown in the paper), and Ljubimova 2013 is referenced for uptake of particles into the brain (not shown in the paper).
- The Chang 2016 findings of inflammation, oxidative stress, and apoptosis in the olfactory epithelium may indeed by mediating the neuronal effects. But given the differences in rodent and human dosimetry, would this response be expected in humans at relevant ambient concentrations?

Activation of the SNS and Hypothalmus-Pituitary-Adrenal Stress Axis:

- Again, there was not a stat-sig change in urinary excretion of VMA in UFP-exposed humans in the Liu 2017 study.

Brain Inflammation:

- As noted above, the Cheng 2016 study shows many olfactory effects of aerosolized UFP exposure, but is this a rodent specific effect?
- It should be noted when referencing the Tyler 2016 study that although they had a CAPS exposure that included more gases, the non-added gas exposure still had substantial concentrations of CO and NO2 (ppm levels).

Causality:

- There is a lack of human evidence, and an inconsistency in animal evidence, that doesn't support a suggestive of causality conclusion.

Long-Term Exposure to UFPs and NS Effects

- Again, the citation of Maher 2016 shows no evidence of where the UFP came from, if it has any effects, or if it is mitigated by changes in external UFP concentrations.
- Biological Plausibility section is 8.6.1, not 8.1.1.

Biological Plausibility:

- Evidence is applied inconsistently for supporting ANS and RAS between different chapters – in metabolism, the RAS was said to activate ANS (although evidence of the reverse was presented), and in this one, even though RAS activation is argued, no ANS activation is included in the chart.
- The Fonken 2011 and Ljubimova 2013 are again presented as providing information that isn't in these papers.

- The summary states that animal toxicology and CHE studies contributed to information about upstream and downstream events, but I didn't see any CHE studies cited, nor are they typically used to inform long-term effects.
- Because the information base is totally animal studies, then considerations of dosimetry, particularly in the olfactory compartment, must be considered.

Brain Inflammation:

- The Tyler 2016 study should be interpreted with caution because there were significant gas concentrations in all the particle exposures. Also, Ljubimova did a 3- and 10-month exposure, which should be cited in this section, not the two-week exposure.

Cognitive and Behavioral Effects:

- Were the behavioral effects seen in the animal studies consistent with the types of morphological changes seen with UFP exposure?

Neurodevelopmental Effects:

- Were the schools matched for other criteria besides SES to control for confounding? Were individual confounders considered?
- Animal Tox – it seems that Davis 2013 studied a lot of endpoints with prenatal exposure to UFPs and found very few changes. Are these significantly associated with any of the other pathways discussed in other parts of this section to show coherence of results? Or are they likely to be caused by the number of statistical tests? While a number of papers are cited showing neurodevelopmental effects of UFP, there were actually only two studies (with many publications about them), and they didn't show entirely consistent results.

Causality:

- A likely to be causal conclusion is not substantiated by the evidence. There are no supportive human studies, and there is no attempt made to show that the rodent effects aren't due to a difference between rodents and humans from the perspective of dosimetry and the part of the respiratory tract that is expected to be affected by UFP exposure.
- The neurodevelopmental data isn't extensive – it is just two studies but subdivided into half a dozen papers.

Chapter 9: Reproductive and Developmental Effects

<u>PM2.5 and Reproductive and Developmental Effects</u>

Biological Plausibility:

- The introductory section repeats 3 times the information that insoluble and soluble particles from PM can translocate into systemic circulation. This needs to be cleaned up. It also relies on pulmonary and systemic inflammation at the beginning of the pathway, which has only been rarely demonstrated in CHE and animal studies.
- There needs to be some connecting evidence showing that, if present, systemic inflammation can cause all these reproductive effects.

- It seems like all the female evidence comes from one study – Veras 2009. Similarly, for testicular and sperm effects, only one study is really cited – Pires 2011.
- "Together, these mechanisms provide plausible pathways by which inhalation of PM2.5 could progress from the initial events noted above to altered fertility, fecundity, and reproduction." In the paragraph above this statement, it wasn't mechanisms that were presented, it was endpoints – saying that PM2.5 exposure affected sperm motility and increased time for successful mating does not tell you anything about how it happened – there is almost no discussion of how in this section.
- The very last paragraph in this section just repeats the same information over again several times, without providing any more detail about how PM2.5 could affect reproduction besides "inflammation and oxidative stress".

Male Reproduction:

- There should be a summary table providing details and results for the epidemiology studies that are presented in this section.

Female Reproduction:

- Epidemiology evidence – EPA states that "Gametes (i.e., ova and sperm) may receive higher exposures while outside of the human body, as occurs with assisted reproduction." How are the gametes getting this exposure? They would be almost exclusively in closed containers and in protected environments to ensure sterility and viability. In the summary section, there should also be an acknowledgement of the studies that showed no effects. Is there a summary table of these papers?
- Animal Tox – This endpoint should not be characterized as being supported by multiple studies – just two papers from the same group in Brazil, and not supported by the more recent study.

Pregnancy and Birth Outcomes:

- Biological Plausibility – The first step in the pathway, systemic inflammation (actually the second step, they didn't address how systemic inflammation would happen), is demonstrated by evidence of a change in CRP in an epidemiology study, and in a CHE study. However, the Devlin 2014 study is a UFP CAP study, and the ISA doesn't reference all of the CHE studies that don't find a change in CRP (which is basically all of them – see my earlier comments on PM2.5 and CRP). A discussion of maternal toxicity versus direct fetal toxicity is important for this section.
- Maternal Health During Pregnancy – EPA cites a number of studies, but these should be summarized in a figure and in a table that is available for review (the supplementary information seems to be only available behind the HERO firewall, and not to the general public).
- In the fetal growth section, SGA and IUGR are defined, and the differences between them are noted (SGA is a small neonate, whereas IUGR is actually considered to be abnormal growth). But in the next sentence, EPA states that these terms are used interchangeably. Should they be? This should be clarified. In the paragraph describing the challenges of determining the effects of air pollution on adverse birth outcomes, the same text is repeated twice. These types of considerations, such as confounding and exposure, should be discussed when addressing the original studies. For the tox studies, one of the studies in the table didn't provide exposure concentrations (Gorr 2014). Also, there were more responses in the Blum 2017 study with

exposure to Sterling Forest CAPs, than in the Klocke 2017 study with the same type of exposure, but at a lower concentration – perhaps a threshold effect? This type of information should be discussed in this document. Was gestational time considered in the birth weight studies?
- Pre-term birth – the introduction notes that the mechanisms of pre-term birth are unknown, with multiple potential causes. What kinds of confounders were considered in these studies, and how were the outcomes assessed?
- Birth defects – table of study information?

Developmental Effects:
- It would make more sense if the summaries of developmental effects from other chapters followed the sequence of the other chapters – I.e. respiratory, cardiovascular, neurological, etc.
- For the neurodevelopmental section, the text states that the epidemiological data does not provide consistent evidence of positive associations. Then they say that this informs and contributes to the conclusion that there is likely to be a causal relationship. It does not seem like inconsistent evidence would contribute to that conclusion. It also says that these studies provide evidence that long-term PM2.5 exposure contributes to developmental effects – but it doesn't, because the data is inconsistent. I think this is a problem of copy-and-paste from text in other sections. For the animal tox section, when discussing the Klocke 2017 study, there should be reference to the fact that the study authors did not find any effect on behavioral or cognitive endpoints.

Summary and Causality Determination
- I agree with the "suggestive" causality determination, although it is not clear to me why this endpoint received a "suggestive" determination, and others (such as long-term PM2.5 and CNS effects) received a "likely" determination.

<u>UFP and Reproductive and Developmental Effects</u>

Female Reproduction and Fertility:
- Table 9-13 incorrectly characterizes the doses the animals received for Li 2009 – they should be ug/m3, not g/m3.
- The lack of effects seen with low-dose DE in Li 2012 suggests that there may be a threshold for this effect.

Neurodevelopmental Effects:
- In the brain morphology section, the same information about ventriculomegaly being related to ASD is repeated twice in the same paragraph.

Chapter 10: Cancer

Introduction:
- The Smith (2016) 10 characteristics of carcinogens represent necessary but not sufficient conditions for carcinogenesis. While it is true that carcinogens display one or more than one of

these characteristics, there are many chemicals that have those characteristics but are not carcinogens.

PM2.5 Exposure and Cancer

Biological Plausibility:

- This section discusses and emphasizes the Ames assay for mutagenicity, but also notes the drawback of the assays being conducted in bacterial cells. However, there are mammalian mutagenicity systems (particularly the in vivo assays), and EPA should emphasize results from those studies.
- The discussion of the hallmarks of cancer is very vague. Also, while changes in telomerase activity is mentioned here, no data is provided to support changed telomerase activity.
- It is interesting and very informative to note that no animal toxicology studies have shown direct carcinogenicity from PM2.5 exposure.

Genotoxicity:

- It is alluded to in this section, but EPA should more clearly state that the mutagenicity studies done in vitro are hazard identifications only, and don't provide clear information about what might happen in vivo. A good example is provided by the mutation in the Salmonella strains to allow larger molecules across the cell wall – this would not be the case in an in vivo model.
- Where is the summary information for these studies?
- In the epidemiology studies, the Ma 2015 study is discussed as rating DNA damage as <40% or >40%. What does this mean? That 40% of the DNA contained damage? What kind of damage? If 40% of the DNA is damaged, then the cell is dead – the study authors may have not been excluding non-live cells. If EPA can't answer these questions, they shouldn't include the study in their discussion (the importance of study quality criteria).
- The summary fails to discuss the import of negative study results.

Epigenetic Effects:

- The Soberanes 2012 study showed increased methylation of the MMP2 promoter, but that would decrease the potential for tumor invasion, it would not increase tumorigenicity.
- When discussing genetic or epigenetic changes in non-in vitro studies, the EPA should specify what tissue was being tested.
- What do the NAS Panni 2016 results, presented as changes in "1, 1, or 10 CpG sites" mean? That kind of resolution is far beyond the realistic ability of these assays to measure, not to mention the inter- and intra-human variability. In fact, it would be useful to discuss the normal human variability in these biomarkers. Were they the same CpG sites changed in different people, or all over the genome, or could the authors even tell?
- Changes in expression of mir-21 are presented from Borgie 2015b. Is mir21 considered a tumor suppressive or promoting mir?

Lung Cancer Incidence, Mortality, and Survival:

- "ecological study design, estimation of PM2.5 concentrations for entire study duration from concentrations of other pollutants using conversion factors, and inadequate control for potential confounders are not the focus of this section. These studies are available at:

https://hero.epa.gov/hero/particulate-matter." What is meant by inadequate control for potential confounders? Because non-copollutant confounders are discussed so little in this document, it is difficult to tell what EPA would consider to be adequate or inadequate control for confounders. In addition, there are plenty of studies referenced in this ISA that have ecological designs – why are they not usable in the cancer section, but are usable in other sections?

- In Table 10-4, the exposure assessment for Lepeule 2012 is stated as the average of the US EPA monitors for 1986-2009. This isn't the case – the authors used the EPA PM2.5 monitors for 1999-2009, but PM10 monitoring data with a conversion factor for 1986-1999.
- The meta-analyses at the bottom of Figure 10-3 should be included in the table (or in a separate table) with the relevant details.
- The effect estimates presented in Figure 10-3 are highly variable, with some negative estimates, and many that aren't statistically significant. From the perspective of exposure assessment, increasing the sophistication of the exposure assessment doesn't demonstrably impact the pattern of effect estimates as one would expect if there was bias towards the null with exposure error.
- With a better exposure assessment, one would expect to have higher effect estimates and narrower CIs, but this is not what is observed with Jerret 2013 and Thurston 2013, compared to Krewski 2009 (particularly comparing Krewksi and Thurston, who looked at the same population). Similarly, Turner 2016 only looked at never-smokers, and if this is considered to be a more susceptible sub population, then it should have a higher, more precise estimate, which it does not (Table 10-5). Related to this, why are never smokers the group with higher incidence? Wouldn't you expect (based on biological plausibility) for it to be the smokers – I.e. because they are receiving a higher lung dose and are pre-initiated (like the urethane animal study)? And why is there such a big difference between Turner 2014's estimate for non-smokers (1.26) and 2016's estimate (1.04)? I understand that one is comparing high versus low (presumably PM concentrations), but do these estimates correspond to one another when converted to a per 5 ug/m3 PM2.5 basis? It seems like they don't because the high versus low difference is less than 5 ug/m3.
- Table 10-5 says that Turner 2011 is the full cohort, but it should be never-smokers (appropriately specified in the text). Table 10-4 should also provide information about the sub-population analyzed, if there is one.
- For some of the ACS studies, the estimated HR for lung cancer increases when using a concentration estimation at a later time point (1999-2000), compared to the earlier time point (1979-1983). EPA should discuss how this could easily be the case for statistical reasons, and not because the risk is increasing in more recent times (it is because if you have the same number of deaths, but lower concentrations, then it makes the deaths per unit concentration look higher).
- Information about the H6C and CCHS cohorts presents the lag period as the 1-3 years of PM2.5 concentrations prior to cancer mortality. However, for standard cancer analyses for an agent that is considered to be carcinogenic, the cumulative exposure is the typical metric. By 1-3 years prior to death, the patient already has the disease, and likely is already undergoing treatment. So, is the EPA/study authors suggesting that PM2.5 exposure enhances mortality from lung cancer, but not its incidence? I understand that EPA is restricted to what the study authors have done, but there could be an emphasis on any studies that considered metrics that are consistent with the biological plausibility argument, such as cumulative exposure. One of the "unacceptable" study criteria from the TSCA systematic review data quality criteria is "Exposures clearly fell outside of relevant exposure window for the outcome of interest." This seems relevant here.

- Related to the last point, Gradient presented information about temporality of PM2.5 exposure measurement and lung cancer mortality in Figure 5.1 of their comments on this ISA, which showed that many of the studies investigated exposures that were measured concurrently, *or after* the lung cancer mortality.
- With Wong 2016 conducted in Hong Kong (much higher concentrations of PM2.5), the effect estimate was about the same as the cohorts in the US and Europe – this is not consistent with the concept of a concentration-response.
- EPA notes at the beginning of this section that they don't include studies with "inadequate control for potential confounders". It seems that the Hart 2011 study, which does not consider smoking at all (a VERY important potential confounder for lung cancer), should not have been included in this section. It would be deemed unacceptable using the TSCA systematic review guidelines: "Confounding was demonstrated and was not appropriately adjusted for in the final analyses (NTP, 2015a)." Smoking should always be considered when looking at lung cancer.
- For the incidence studies, it is even more important to look at the estimated time of exposure, with longer exposure times for PM2.5 being more important. The idea of a one-month average of PM2.5 contributing to lung cancer (from the Gharibvand 2016 study) is completely unsupported by what we know about cancer development. Cancer is known to take decades to develop (look at the smoking data for comparison), and the idea of a one-month exposure to PM2.5 contributing to lung cancer is highly unrealistic.
- Figure 10-4, showing effect estimates from different types of exposure estimates (and the discussion in the text) shows that there is no difference with more precise exposure estimates (the prediction has been a bias away from the mean) and no difference in CIs. This paper also showed effects only in former smokers, not never smokers, which is inconsistent with the mortality data. However, this is more consistent with the idea of dose-response and PM2.5 having a larger impact on people who have been pre-initiated by exposure to cigarette smoke.
- C-R Function Shape – EPA cites Pope 2011, who combined smoking and SHS data with predicted PM2.5 concentrations from various studies and plotted them. A few concerns – what is the length of assumed exposure here (as noted above, longer exposure is usually necessary for cancer formation); also, the authors produce HR's, and then seemingly equate them to RRs, and plot the other results as RRs. These are NOT the same thing (Hernán, 2013; Stare and Maucort-Boulch, 2016; Sutradhar and Austin, 2017). The authors also do this for CVD mortality, showing a much steeper D-R curve with ambient PM2.5 than smoking. Why does EPA/the authors think that gram for gram, PM2.5 is worse than smoking? Is there any basis for this? Perhaps both should consider the substantially different amounts of confidence in the exposure metric? As before, the variability and error in these studies can prevent the identification of the appropriate shape of the C-R function, and the presence of a threshold.
- Cancer Survival – What confounders were considered in Xu (2013) and Eckel (2016)? Many factors can impact cancer survival that might be different between two cities, particularly with such variation as would be expected between LA and Honolulu, for example.

Summary and Causal Determination:

- The data that PM2.5 components can be mutagenic in vitro, and that one study showed enhanced urethane-induced tumors, does not complete a pathway of biological plausibility of PM2.5 causing cancer. And it doesn't provide a plausibility pathway for mortality associated with relatively short-term PM2.5 exposure – months to several years. Those are two different effects

and should be separated out, because some genotoxicity caused by PM2.5 in the few years before death from cancer aren't contributing to the formation of that cancer.
- Positive associations in never-smokers is cited several times, but there is no discussion of whether this makes sense – wouldn't an association in smokers, with their pre-initiated lung cells, make more sense?
- There should be a discussion of statistical significance (**chance**, bias, and confounding). And there are many other confounders that were not discussed in this section that could contribute to lung cancer mortality, particularly for this short a time-span.
- This lack of coherence does not suggest a likely causal relationship.

PM10-2.5 Exposure and Cancer

Biological Plausibility:
- Why is brain inflammation particularly emphasized in this figure (also for PM2.5 and UFP)? Is there substantial evidence of brain cancer? Or is there a pathway of systemic dysregulation thought to be caused by brain inflammation? If so, this should be discussed in this section.

Genotoxicity:
- For the Wessels 2010 study, the location from which the particles were collected should be included in the discussion. The single CHE study used as evidence here did not show a statistically significant change in 8-oxo-dG ($p<0.1$ is not $p<0.05$). The epidemiology study results were also not statistically significant.

UFP Exposure and Cancer

- Genotoxicity: there is a discussion of the results from Hemmingsen 2015, but that study focuses on PM2.5, and it seems that a lot of the mass concentration is higher than the 100 nm cut-off for UFP.

Chapter 11: Mortality

Short-Term PM2.5 Exposure and Total Mortality

- The EPA states at the beginning of this chapter that "As detailed in the Preface, the focus of this section is on the evaluation of recently published studies that directly address policy-relevant issues, i.e., those studies where mean 24-hour average concentrations are less than 20 µg/m3 across all cities or where at least half of the cities have mean 24-hour average concentrations less than 20 µg/m3." Why 20 ug/m3? The 24-hr standard is 35 ug/m3. This statement references the Preface for the 2009 ISA – however, important information relevant to this ISA should be included in this ISA, even if just in summary form. This is also true of details about source apportionment in Section 11.1.11.2
- The EPA also states at the beginning of the chapter that "The following sections provide a brief overview of the consistent, positive associations observed in recent studies of mortality and

short-term PM2.5 exposures," Does this mean that the EPA did not include any studies that didn't provide consistent, positive associations? If not, then this should be restated.
- This same paragraph has a reference error (reference not found).
- The organization of this section, with a table at the beginning with all the relevant studies, is a better method than listing them over and over again in different sections, as is done in other chapters.

Biological Plausibility:
- The EPA states in this section that "However, the evidence for how the initial events and subsequent endpoints could lead to the observed increases in respiratory ED visits and hospital admissions, for particularly chronic obstructive pulmonary disease (COPD) and asthma, is limited. Collectively, the progression demonstrated in the available evidence for cardiovascular morbidity (and to a lesser extent, respiratory morbidity) supports potential biological pathways by which short-term PM2.5 exposures could result in mortality." This seems to be a contradictory statement – the first sentence says that evidence is limited for how initial events can lead to HAs and ED visits, and the second states that the progression supports biological pathways. This needs to be clarified, and the plausibility of low concentrations of PM causing mortality after a short exposure (0-1 days) needs to be considered.
- The EPA should include a discussion of animal studies where mortality was an endpoint after PM2.5 exposure.

Total Mortality in All Year Analyses:
- When discussing the Lanzinger 2016 study, is the EPA suggesting that the study size is not large enough to discern a pattern between PM2.5 short-term exposure and mortality? Because only having 2 years of data won't affect short-term mortality associations (unlike long-term). Having the sample size listed in the summary Table 11-1 would help with this distinction.
- In the discussion of causal analytics papers, the EPA should include a discussion of whether the authors addressed the SUTVA assumptions.

Cause-Specific Mortality:
- Is it logical that respiratory mortality has higher effect estimates than CVD, but the evidence for morbidity is weaker?

Non-Copollutant Confounding of PM2.5 Effects:
- The EPA states that "Recent multicity studies that assess the potential for copollutant confounding of the PM2.5-mortality relationship are limited to Europe and Asia" However, they then discuss Di et al. 2017, which is based in the US. It seems from Figure 11-3 that there is some attenuation of the risk estimate with copollutants, and some become non-statistically significant.
- None of the Sacks 2012 results (looking at different model specs for temp) were statistically significant. This should be included in the analysis/discussion of the results.
- Seasonal analyses show heterogeneity of results, even with similar study types and when the studies were only in the US. The EPA should offer some sort of explanation for this heterogeneity and lack of consistent pattern.

- For temperature patterns, the pattern seen by Dai 2014 (lower effects of PM at high and low temperatures) is not consistent with the (imprecise) results of Pascal 2014, who showed increased PM2.5 effects at high temperatures. Does EPA or the study authors have an explanation for this discrepancy? Similarly, Sun 2015, who shows more effects at lower temperatures is inconsistent with both Dai 2014 and Pascal 2014. And Li 2015 showed higher effects at both lower AND higher temperatures. These four studies have literally run the gamut of options and demonstrate unexplained heterogeneity in study estimates.
- City and Regional Heterogeneity – Two US studies conducted in similar ways did not show the same potential components as being responsible for the regional heterogeneity in PM2.5-mortality effect estimates (Lippmann 2013a, Dai 2014). And in Boston Zanobetti 2014a reported that the strongest association was found on winter days with higher primary PM, even though the EPA reports that the warm season has the strongest associations (in the seasonal section). Davis 2011, when looking for city-clustering by PM2.5 components, showed a North-South division, whereas the PM2.5-effects have typically shown an East-West division. Baxter 2013 also could not identify sources or components to explain the observed regional heterogeneity. All together it seems that despite multiple attempts, this heterogeneity is not explained by sources or PM components. It would be helpful if the EPA could report the amount of PM2.5 effect estimate attributable to the various components reported from Lippmann 2013a and Baxter 2014 (I.e. the R2), as is done with Baxter 2017. This section demonstrates that despite numerous attempts by researchers, there is still substantial unexplained heterogeneity in the effect estimates between PM2.5 and mortality that need to be considered when assessing the causal connection between the two.

Exposure Assessment Techniques

- Monitor representativeness wouldn't just provide information about regional heterogeneity (maybe), it would provide information about the biases and potential validity of epidemiology study results. EPA states that Di 2017 found a smaller association using the nearest monitor versus the modeled estimates, and that this was consistent with Kloog 2013. But Kloog 2013 found a larger association with the nearest monitor method. Either there is a discrepancy here, or the EPA needs to better explain what it is comparing. There is also a substantial difference in effect estimates (0.8 % vs 4.5%). When discussing the differences between monitored and modeled data, it would be helpful if the EPA reported how well the study's model predicted monitored concentrations, as a measure of the accuracy of the model.

Timing of Effects and Exposure Metrics:

- A figure of the various lagged effects would be useful, particularly because many of the effect estimates are not presented elsewhere in the ISA. This would allow the readers to more easily understand the lagged patterns.

Concentration Response and Threshold Analyses:

- "2004 AQCD and 2009 PM ISA stated that conducting C-R and threshold analyses is challenging due to the "(1) limited range of available concentration levels (i.e., sparse data at the low and high end); (2) heterogeneity of [at-risk] populations [between cities]; and (3) influence of measurement error" (U.S. EPA, 2004). Even with these inherent limitations, studies have continued to examine the PM-mortality C-R relationship and whether a threshold exists." These considerations don't just make threshold analysis challenging, they render results of an analysis

with those limitations inaccurate or uninterpretable. Just because study authors did these analyses anyway, doesn't mean that the EPA must take the results at face value, knowing that these problems exist. Rhomberg 2011, Yoshimura 1990, and Cox 2018 all demonstrated that you can't detect a threshold or nonlinear response with this much variability in the data, and a reviewer at the ozone workshop held in Oct-Nov 2018 said the same thing. This concern is discussed more thoroughly in the general section of my comments.

- For the Shi 2015 study (Figure 11-11), the C-R function doesn't look linear. Do the authors ever use the default as non-linear, and then test if linear fits any better than non-linear? (This has to do with your perspective and default assumptions). And if you have less confidence in the curve less at concentrations less than 5 because of the width of the confidence intervals, then you should also have less confidence in the curve at 10-15 ug/m3 – because the confidence interval widths are the same, down to about 2.5 ug/m3, and it certainly looks like there is a threshold somewhere between 5 and 10 ug/m3.
- It looks like the Di 2017a analysis forced the curve through the origin – if it did, how could it identify a threshold of no effect? Upon reading the text of that manuscript, it is not clear if they did force it through the origin or not – this should be clarified by EPA. What should also be clarified by the EPA is the risk metric used by the study authors. Di 2017a presents the risk metric as a percent increase in relative risk (RR) per 10 ug/m3 PM2.5, which is a non-standard metric. Presumably this can be interpreted as a % increase in risk of total mortality per 10 ug/m3 PM2.5, but EPA should clarify this, so readers know that the results can be compared to other study results. Interestingly, the authors also present the absolute risk difference, per 1 million persons at risk. This is a very helpful metric that shows in the main analysis that there are 1.4 people with increased risk per 1 million at risk. This puts the risk in the context of other assessments, such as carcinogen assessments, and would allow the EPA to use similar frameworks for acceptable and unacceptable risk.

PM2.5 Components and Sources:

- In Figure 11-13, why is there no lag information in the orange boxes (null or non-statistically significant negative associations)? Also, why are some of the lag numbers in the boxes (primarily in the PM2.5 total row) bolder? This should be included in the legend.
- A forest plot would be useful in this section, to demonstrate whether or not the individual components had larger or smaller effect estimates compared to total PM2.5.
- In Figure 11-14 of Lippmann 2013a, most of the effect estimates weren't statistically significant, and many were negative. In addition, soil was more consistently positive than combustion products.

Summary and Causal Determination:

- 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, there are studies showing completely inconsistent temperature relationships, 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. At what point do you go back to your underlying assumptions (i.e that short-term exposure to PM2.5 causes mortality) and ask whether they are valid?

- "Collectively, recent studies indicate that the heterogeneity in PM2.5-mortality risk estimates cannot be attributed to one factor, but instead a combination of factors including, but not limited to, compositional and source differences as well as exposure differences." This statement is misleading – there is no data presented that looks at all the possible components together to show that combined they impact the observed heterogeneity. What we do have is several studies that show almost no impact of composition, source, and exposure differences.
- "However, to date, studies have not conducted extensive analyses exploring alternatives to linearity when examining the shape of the PM2.5-mortality C-R relationship." It is hard to be confident that the shape of the C-R function is linear, when you haven't examined alternatives to linearity.

Long-Term PM2.5 Exposure and Total Mortality

- Again, the EPA should clarify why their focus is on 20 ug/m3 for long-term PM2.5 concentrations. This is above the current (and former) standard and shouldn't be equated to the 24-hour standard (which has a different form).
- "The evidence in this section will focus on epidemiologic studies because experimental studies of long-term exposure and mortality are generally not conducted." But long-term exposure studies in animals have been conducted, typically at higher-than-ambient concentrations – these studies could be investigated to explore whether animals have experienced increased mortality with exposure to PM2.5.

Associations between LT-PM2.5 and Mortality:

- Harvard 6 Cities and ACS Cohorts – The EPA notes that there are discrepancies between the findings of Pope 1995 for the early ACS (found a positive statistically significant effect) and Enstrom 2017 (did not find a positive statistically significant effect), but that there was a difference between the two in that Enstrom used 85 counties and Pope used 50 MSAs. However, the Enstrom study had a finer resolution, so one would guess that it had less exposure error and therefore possibly a greater effect estimate, or one with narrower confidence intervals, compared to Pope. As the EPA notes, the many re-analyses of the ACS cohort present the opportunity to explore the effects of different types of exposure estimates. Instead of just saying that the results are generally consistent in magnitude and direction, EPA should take this opportunity to look for more patterns in the data, based on the assumptions that are made. For example, the EPA could ask whether a better exposure estimate moves the effect away from the null or narrows the confidence intervals. From Figure 11-17, it doesn't seem that any of the investigated subtypes of mortality particularly show an increase in effect estimate, or a decrease in width of confidence intervals, compared to total mortality. The regional estimate of mean PM2.5 concentration from Turner 2016 is shown as 0.5 ug/m3 in Figure 11-17. Is this right?
- Other North American Analyses – As with the H6S and ACS cohorts, Figure 11-19 shows that there is no increase in the association, or particular decrease in the effect estimate confidence intervals, of CVD or respiratory mortality compared to total mortality. The Medicare cohort also offers the opportunity to compare results from different exposure models – the Shi 2015 study in New England used a sophisticated model, and Kioumourtzoglou et al. (2016) used fixed site

monitors, and the effect estimate for Shi 2015 was smaller with about the same precision, so this does not support the hypotheses that improved exposure measurement techniques show greater, more precise effect estimates. The model accuracy of Thurston 2015 should be mentioned, because if it only uses a LUR model, these are known to be inaccurate in locations where they weren't specifically designed. The Crouse 2012 and 2015 results were somewhat inconsistent (IHD mortality in 2012 was 1.31, and in 2015 was 1.09) with the difference in studies being 5 more years of follow up. The EPA should offer some explanation or discussion of this result, if one was presented by the authors. Again, the same data was used in Lipsett 2011 and Ostro 2015, but Ostro had a better exposure estimate method, but the effect estimates were lower (not biased away from the null). European studies are discussed in this section, so I would recommend that the section be retitled "Other North American and European Analyses". Also, results from the European analyses should be included in the Forest plot for Figure 11-19, or on their own forest plot.

- Causal Analysis – Did the studies cited adequately discuss whether the SUTVA assumptions were met in their studies? This section should also include a discussion about Tony Cox's causality papers (e.g. (Cox and Popken, 2015)), as well as the Greven et al., 2011; and Pun et al., 2017 papers that did a difference-in-difference analysis to determine the effects of local changes in PM compared to national trends and found that there was no association between PM2.5 and mortality.
- Life Expectancy – what is a doubly-robust additive hazards model, and what kind of information does it provide that allows the authors to estimate 5400 fewer deaths from a decrease of 1 ug/m3 in annual average PM2.5? The authors (Wang 2017a) call it a causal analysis, but it is not clear why this "double robust" method is causal. Also, the effect estimate from Wang 2017b (upon which Wang 2017a is supposedly based) is pretty high – 1.021 per 1 ug/m3, whereas Di 2017c is 1.04 per 5 ug/m3. What kind of controls for confounding did Bacarelli 2016 use? Were the results comparable to the other life expectancy studies?

Potential Copollutant Confounding:

- What is the reference for the meta-analysis that only looked at copollutant models with $r < 0.7$?

Shape of the C-R Curve:

- EPA reports that many studies have shown evidence for LNT, some have shown supralinear shapes, and other studies have shown thresholds. As above, these likely aren't the best kind of studies to use to determine the shape of the C-R function, given the variability and errors in the estimates (Rhomberg 2011).

Factors that May Influence PM2.5 Associations:

- It is interesting that Lee 2011 showed that monitors are more accurate within a 98 km distance, but AOD is more accurate outside of that. That is a pretty bad track record, given the supposed resolution of satellite measurements.
- When discussing the Jerret et al. 2016 results, EPA should note that the effect estimate doesn't tell you how accurate the model is at estimating exposures.
- Results from the Hart 2015 study show that enhanced exposure estimate accuracy doesn't change the estimate, which is not what would be predicted. They themselves show that a bias correction increases the effect estimate. Therefore, either the more sophisticated exposure estimate methods

aren't actually more accurate, or there is some other reason for the observed association that is not impacted by the method used to estimate exposure, possibly the complexity of the variables in the model (Corrothers & Evan, 2000; Fewell 2007).

- When talking about exposure windows, and specifically the Wong 2015 study, EPA noted that risks decrease (presumably from PM) in ages over 70 or 75. Is that just in the Wong 2015 study, or is it a general result? If it is a general result, that is not what would be expected for a population dying from CVD or respiratory illnesses and is not consistent with the hypothesis of a vulnerable population having enhanced mortality from PM2.5 exposure.

PM2.5 Sources and Components:

- The Wolf 2015 results likely aren't useful, because of such poor LUR model performance.

Short-Term PM10-2.5 Exposure and Total Mortality

Associations with All-Year Mortality:

- While all the associations presented are positive, few are statistically significant, and many have the added exposure error of estimating PM10-2.5 by subtracting PM10 from PM2.5 county-average measurements.

C-R Relationships and Thresholds:

- Both the studies cited that looked at different concentration cut-points and the association between PM10-2.5 and mortality found that the highest concentrations and/or extreme events like dust-storms, had the lowest associations with PM10-2.5. Does this make sense? The EPA should address the lack of concentration response here.

Summary and Causal Conclusions:

- EPA states that "recent studies provide initial evidence that informs additional uncertainties and limitations identified in the studies evaluated in the 2009 PM ISA, specifically potential copollutant confounding; effect modification (e.g., temperature, season); and the shape of the C-R relationship and whether a threshold exists." However, most of these sections did not provide information that further informed the uncertainties in these areas. The studies looked at these potential limitation areas, but no real conclusions could be drawn about the results.
- Suggestive of causality doesn't seem supported here, because there were almost no statistically-significant effects, and many unaddressed uncertainties (bias, chance, confounding).

Long-Term PM10-2.5 Exposure and Total Mortality

- The EPA's final causal conclusion in this section is suggestive, but based on what? There is limited biological plausibility data, a number of epidemiology studies that show no effect, and those that do show effects that are often attenuated when PM2.5 is included in the model; the exposure estimates are very uncertain (all used the subtraction model for estimating PM10-2.5); and there is no other information about model specification, temperature, etc. There is one French study that shows positive effects, but is this enough (in the face of a lot of negative evidence) to call the endpoint suggestive?

Long-Term UFP Exposure and Total Mortality

- What did the experts in Hoek et al. 2009 base their recommendations on, if there was no data for the endpoint? What is the point of asking experts for their advice on something without any information? I don't think that the Hoek study should be included in this review, because it is apparently based on opinion and not data.

Chapter 12: Populations and Lifestages Potentially at Increased Risk from PM Exposure

- The summary of this chapter should describe the conclusions for the chapter (I.e. the life stages or populations considered to be at increased risk), as is done with the other chapter summaries.
- "similar to the characterization of epidemiologic evidence in Chapters 5-11, statistical significance is not the sole criterion by which effect modification and evidence of increased risk is determined; emphasis is placed on patterns or trends in results across these epidemiologic studies." I agree that statistical significance should not be the sole criterion for judging the validity of evidence. However, chance does need to be considered as one of the aspects that could potentially lead to an association between two variables (along with bias and confounding). Therefore, it is important that statistical significance is discussed in this and other chapters of this ISA. Unfortunately, this does not appear to be this case, in this chapter or elsewhere.
- The figures in the supplement do include notations for determination of whether there is a statistically significant difference between the two groups, and this should be incorporated into the main text.
- In the supplement the figure legends need to be changed so that they are endpoint-specific, and there needs to be an explanation of the up and down arrows in the colored boxes.

Pre-Existing Diseases/Conditions:

- CVD – The EPA concludes that there is only suggestive evidence of pre-existing CVD increasing the risk for PM2.5-mediated effects, but this isn't consistent with the hypothesis that those who are already vulnerable are the ones who would have more extreme effects from PM exposure (like mortality), or that PM2.5 targets the CV system. Does it make sense that there is a causal association between PM2.5 and CV effects, but only a suggestive association between PM2.5 and people with CVD? Is the EPA suggesting that PM2.5 is causing these effects de novo (i.e. in healthy people without pre-existing disease)?
- Diabetes – I agree with the EPA that the evidence is insufficient to determine if diabetics have an increased risk from PM2.5 exposure. There was inconsistent evidence and issues with the studies, and this is the appropriate conclusion to derive from that combination of data.
- Obesity – The text is not clear as to why obesity gets a suggestive designation, whereas diabetes got an insufficient designation. Both have a fair number of studies showing mostly inconsistent associations. Similar with elevated cholesterol – why is that insufficient, and obesity is suggestive?
- Elevated cholesterol – similar to diabetes, I agree that there is insufficient data.
- Respiratory Diseases – it is unclear in this section whether data from Chapter 5 about asthma and COPD are being used to draw the suggestive conclusion here, or if it is the limited data presented

where comparisons are made between people with and without the disease. In the introduction to this chapter the EPA specifies that the at-risk category is particularly in comparison to people without the potential risk factor, but the conclusions from this section seem to be mostly from the Chapter 5 conclusions.

Genetic Factors:

- The conclusion "the evidence is suggestive that individuals with variants in the glutathione pathway are at increased risk for PM2.5-related health effects compared to those without a variant genotype." should be more specific to which variant type – for example, does this refer to glutathione variants that reduce the ability for the glutathione pathway to reduce oxidant stress? There is also not much evidence presented for the glutathione pathway, and no information about the active or inactive or less active forms of the gene, and if those are related to PM2.5 health effects.
- There should be a description of the functional effect of the NFE2L2 rs1364725 allele. Does this allele decrease the function of the NFE2L2 protein? Were the people heterozygous for the allele (and therefore any effect would have to be dominant), or homozygous? Also, the referenced paper (Hampel 2010) used 16 sec ECG recordings, but my understanding is that there needs to be a longer duration of ECG recordings to get reliable variability data. In addition, that study did not see associations with the GSTM1 deletion allele. Similarly, for the cited SNPs in GSTP1, TNF, and TLR4 – did those reduce the activity of the gene products? Just because it is a minor allele, doesn't make it less functional. The same with the micro-RNA processing gene GEMIN4 – if you don't know the effect of the SNP, then what does it tell you if there is an association?

Sociodemographic Factors:

- Children – To test the hypothesis that increased oral breathing in boys increases their risk of PM2.5-mediated effects, are there any studies that address the effects of PM2.5 in boys vs girls? This section also states that children tend to spend more time outdoors, but the exposure chapter section that discusses the CHAD database states that according to that data, children spend less time outdoors than adults (pg 3-65). So, either the CHAD database is flawed, or this statement is. This section states that "there has been little evidence from stratified analyses to demonstrate children being at increased risk of the health effects associated with PM2.5 exposure compared to adults. That is, positive effect estimates are often observed in stratified analyses of children, but these effect estimates are similar in magnitude to those observed for adults (Supplemental Table S12-7) (U.S. EPA, 2018)." This doesn't seem consistent with the conclusion that there is adequate evidence that children are at increased risk from PM2.5 exposure. The conclusion seems to be based on the data showing positive associations in children for things that were only measured in children (e.g. lung function development), but that seems to be inconsistent with what EPA said is the point of this section, which is comparison to a reference group. These aspects should be divided or specifically addressed.
- Older Adults – this document notes that there is no consistent evidence that older adults have greater health effects associated with exposure to PM2.5 than younger adults. This seems inconsistent with the general hypothesis that people with greater risk of the health effect are more susceptible to something that contributes to that health effect. In addition, the EPA provides plenty of evidence (not insufficient evidence) that there is actually no increased risk amongst older adults.

- Race – in the summary section, the EPA concludes that there is adequate evidence demonstrating an increased risk of PM2.5-related effects in non-whites, in part due to disparities in exposure. However, there is almost no discussion in this section of whether the epidemiology studies tested if the disparity was due to increased exposure, or due to other factors. This should be explicitly discussed in this chapter. If the conclusion is that the increased risk is due to some other non-exposure factor, there should be discussion of what that factor is.

Residential Location:

- Urban v Rural – There is a several page list of study results in this section, describing largely inconsistent results comparing urban v rural locations. However, there needs to be synthesis of all these results – are there patterns? Why might you expect there to be higher associations in urban compared to rural locations, or vice versa? What would other data suggest? Did the authors control for the higher PM in urban locations? Just listing study results does not help the reader synthesize the conclusions.
- Proximity to roadway – again, why would you expect certain vulnerabilities near roadways, especially if there is no evidence showing that that the PM concentrations are higher near roadways? Is it noise, or other SES factors, or other pollutants? Were these controlled for in the studies (proximity studies being notoriously problematic for drawing conclusions)? Both of the cited animal toxicology studies have interpretation issues – there may be other, non-PM related reasons why more effects were seen with animals closer to a major road (e.g. stress, noise) in Kleinman 2005, and the other cited study (Farraj 2006a, b) showed a concentration response, not a response to proximity to roadways (which the EPA says earlier does not increase PM2.5 concentrations).

Behavioral and Other Factors:

- Smoking – Lung cancer should be included as an endpoint in this section, because it is probably the best studied for differences between smokers and non-smokers, and EPA concluded that there was a likely causal association. From the summaries in Chapter 10, it seemed like there was more associations between PM2.5 and lung cancer in never smokers than in ever or current smokers. In general, one would predict that PM2.5 would increase effects in smokers, because it would generally increase their dose, and because they already have a lot of the health effect precursors that are being attributed to PM (inflammation, CV changes, respiratory effects). Conversely, perhaps there is an adaptive process that protects the lungs of smokers from additional exposure. Either way, these are important issues to discuss in addition to just listing the study results.

References

Avery, C.L., Mills, K.T., Williams, R., McGraw, K.A., Poole, C., Smith, R.L., Whitsel, E.A., 2010a. 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

Avery, C.L., Mills, K.T., Williams, R., McGraw, K.A., Poole, C., Smith, R.L., Whitsel, E.A., 2010b. 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

Bates, J.H.T., Irvin, C.G., 2003. Measuring lung function in mice: the phenotyping uncertainty principle. J. Appl. Physiol. 94, 1297–1306. https://doi.org/10.1152/japplphysiol.00706.2002

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.

Bräuner, E.V., Forchhammer, L., Möller, P., Simonsen, J., Glasius, M., Wåhlin, P., Raaschou-Nielsen, O., Loft, S., 2007. Exposure to ultrafine particles from ambient air and oxidative stress-induced DNA damage. Environ. Health Perspect. 115, 1177–1182. https://doi.org/10.1289/ehp.9984

Brook, R.D., Brook, J.R., Urch, B., Vincent, R., Rajagopalan, S., Silverman, F., 2002. Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. Circulation 105, 1534–1536. https://doi.org/10.1161/01.CIR.0000013838.94747.64

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.

Cefalu, M., Dominici, F., 2014. Does exposure prediction bias health-effect estimation?: The relationship between confounding adjustment and exposure prediction. Epidemiology 25, 583–90. https://doi.org/10.1097/EDE.0000000000000099

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

Cox, L.A.T., Popken, D.A., 2015. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? Ann. Epidemiol. 25, 162–173. https://doi.org/10.1016/j.annepidem.2014.11.006

Fakhri, A.A., Ilic, L.M., Wellenius, G.A., Urch, B., Silverman, F., Gold, D.R., Mittleman, M.A., 2009. Autonomic effects of controlled fine particulate exposure in young healthy adults: Effect modification by ozone. Environ. Health Perspect. 117, 1287–1292. https://doi.org/10.1289/ehp.0900541

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

Frampton, M.W., Stewart, J.C., Oberdörster, G., Morrow, P.E., Chalupa, D., Pietropaoli, A.P., Frasier, L.M., Speers, D.M., Cox, C., Huang, L.S., Utell, M.J., 2006. Inhalation of ultrafine particles alters blood leukocyte expression of adhesion molecules in humans. Environ. Health Perspect. 114, 51–58. https://doi.org/10.1289/ehp.7962

Frampton, M.W., Utell, M.J., Zareba, W., Oberdorster, G., Cox, C., Huang, L.S., Morrow, P.E., Lee, F.E., Chalupa, D., Frasier, L.M., Speers, D.M., Stewart, J., 2004. Effects of exposure to ultrafine carbon particles in healthy subjects and subjects with asthma. Res. Rep. - Heal. Eff. Inst. 1–63.

Fuller, W.A., 1987. Measurement Error Models. John Wiley & Sons, 1987.

Ghio, A.J., Hall, A., Bassett, M.A., Cascio, W.E., Devlin, R.B., 2003. Exposure to concentrated ambient air particles alters hematologic indices in humans. Inhal. Toxicol. 15, 1465–1478. https://doi.org/10.1080/08958370390249111

Goldman, G.T., Mulholland, J.A., Russell, A.G., Strickland, M.J., Klein, M., Waller, L.A., Tolbert, P.E., 2011. Impact of exposure measurement error in air pollution epidemiology: Effect of error type in time-series studies. Environ. Heal. A Glob. Access Sci. Source 10, 61. https://doi.org/10.1186/1476-069X-10-61

Graff, D.W., Cascio, W.E., Rappold, A., Zhou, H., Huang, Y.C.T., Devlin, R.B., 2009. Exposure to concentrated coarse air pollution particles causes mild cardiopulmonary effects in healthy young adults. Environ. Health Perspect. 117, 1089–1094. https://doi.org/10.1289/ehp.0900558

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

Harder, S.D., Soukup, J.M., Ghio, A.J., Devlin, R.B., Becker, S., 2001. Inhalation of PM2.5 does not modulate host defense or immune parameters in blood or lung of normal human subjects. Environ. Health Perspect. 109 Suppl, 599–604. https://doi.org/10.1289/ehp.01109s4599

Hernán, M.A., 2013. The Hazards of Hazard Ratios. Epidemiology 21, 13–15. https://doi.org/10.1097/EDE.0b013e3181c1ea43.The

Holgate, S.T., Sandström, T., Frew, A.J., Stenfors, N., Nördenhall, C., Salvi, S., Blomberg, A., Helleday, R., Södenberg, M., Devlin, R.B., Wilson, S.J., 2003. Health Effects of Acute Exposure to Air Pollution Exposed to Diesel Exhaust Concentrated Ambient Particles. Heal. Eff. Inst. Res. Rep.

Jurek, A.M., Greenland, S., Maldonado, G., 2008. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? Int. J. Epidemiol. 37, 382–5. https://doi.org/10.1093/ije/dym291

Jurek, A.M., Greenland, S., Maldonado, G., Church, T.R., 2005. Proper interpretation of non-differential misclassification effects: expectations vs observations. Int. J. Epidemiol. 34, 680–7. https://doi.org/10.1093/ije/dyi060

Langrish, J.P., Watts, S.J., Hunter, A.J., Shah, A.S.V., Bosson, J.A., Unosson, J., Barath, S., Lundbäck, M., Cassee, F.R., Donaldson, K., Sandström, T., Blomberg, A., Newby, D.E., Mills, N.L., 2014. Controlled exposures to air pollutants and risk of cardiac arrhythmia. Environ. Health Perspect. 122, 747–753. https://doi.org/10.1289/ehp.1307337

Lay, J.C., Zeman, K.L., Ghio, A.J., Bennett, W.D., 2001. Effects of inhaled iron oxide particles on alveolar epithelial permeability in normal subjects. Inhal. Toxicol. 13, 1065–1078. https://doi.org/10.1080/08958370152647618

Lioy, P., Waldman, J., Buckley, T., Butler, J., Pietarinen, C., 1990. The personal, indoor and outdoor concentrations of PM-10 measured in an industrial community during the winter. Atmos. Environ. 24, 57–66.

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.

McGuinn, L.A., Ward-Caviness, C., Neas, L.M., Schneider, A., Di, Q., Chudnovsky, A., Schwartz, J., Koutrakis, P., Russell, A.G., Garcia, V., Kraus, W.E., Hauser, E.R., Cascio, W., Diaz-Sanchez, D., Devlin, R.B., 2017. Fine particulate matter and cardiovascular disease: Comparison of assessment methods for long-term exposure. Environ. Res. 159, 16–23. https://doi.org/10.1016/j.envres.2017.07.041

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

Stare, J., Maucort-Boulch, D., 2016. Odds ratio, hazard ratio and relative risk. Metod. Zv. 13, 59–67.

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

US EPA, 2018. Application of Systematic Review in TSCA Risk Evaluations.

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.

Dr. Timothy E. Lewis

Comments on Chapter 13

Please comment on the identification, evaluation and characterization of the available scientific evidence from studies of PM on non-ecological welfare effects of visibility impairment, climate, and materials and the application of information from these studies, as presented in Chapter 13, to inform causality determinations and uncertainty characterizations for these welfare outcomes.

The information presented in Chapter 13 supports a causal relationship between PM and visibility impairment, climate effects, and effects on materials.

The evaluation of welfare effects often lumps PM together as a whole without considering different size fractions. It is recommended that EPA perform more analyses for different size fractions to determine whether various effects on visibility, climate, and materials are observed.

Specific quality criteria targets for inclusion or exclusion of welfare effects studies should be articulated up front in each section. The studies presented in this chapter are mostly descriptive with little reference to quality. There is little discussion of how study findings that consist of different PM concentrations, different mixtures, different experimental design questions, and different ambient conditions apply directly to non-ecological welfare effects in the U.S. A "Research Needs" section should be added to the final ISA. In addition, line numbers should be added for pages 13-1 through 13-56.

For visibility effects, a thorough discussion of the instrumentation used for measuring visibility is provided. It would be very useful if the instruments were shown in a table with the figures of merit associated with each, and how well each instrument provides the most policy relevant measurements. The distinction between anthropogenic PM impairment versus natural PM impairment needs to be more clearly separated and explained. How this distinction can or will be used for setting a secondary standard needs to be included in the document.

The document does a good job more firmly establishing a causal relationship between PM and visibility. However, it is challenging to tease out the complex nature of PM across the country and how the variation in PM composition affects visibility differently. Setting a secondary standard given such variability will be very difficult. A discussion on the direct effect of PM or other pollutants (e.g., photochemical oxidant) on visual acuity should be included. Instruments would not be responsive to these eye irritants. Also, comparing perceived visibility impairment of urban versus more "bucolic" settings may have inherent biases. Some viewers of these scenes may not find urban viewscapes to be very appealing no matter how clear the image may be. Moreover, regional differences in perceived visibility may be due to societal differences. Westerners may have greater expectations of clear mountain vistas than Easterners.

The color maps, bar charts, and other graphical data presentations are very helpful. The uncertainty associated with the size fraction and visibility impairment needs to be stated clearly.

For climate effects, uncertainty in the effects of complex aerosol composition on climate needs to be better resolved. If there is new evidence that increased atmospheric turbidity is increasing cloud-to-ground lightning strikes and hence increased forest fires, that information should be added to the document.

For effects on materials, it was difficult to determine from the literature review presented in the ISA at what level damage to materials was unacceptable and how that relates back to PM concentration, size, and mixture. It is laudable that data from other countries were included in the assessment. The document should discuss if there is sufficient meta data available to fully characterize the data quality attributes associated with these data.

It would be helpful for the PM CASAC seven-member panel to have access to a much larger review panel that would allow for additional input into the non-ecological welfare effects and better inform our recommendations on the appropriate level for a secondary standard.

Dr. Corey Masuca

2.2 Atmospheric Size Distributions

Atmospheric particle formation (secondary) nucleation, accumulation, and coarse modes

2.3 Primary Sources and Atmospheric Formation

Primary PM – source-derived

Secondary PM – gas-phase chemical compounds

2.3.1 – Primary PM2.5 Emissions
2.3.1.1 – National Scale Emissions

Uncertainties in emission estimates

Dust and fire – significant portion of PM2.5 2014 NEI

Dust includes agricultural and road dust, predominately

2.3.1.2. – Urban Scale Emissions

Great variability from city to city in PM2.5 primary emissions

Mobile sources are a major source of primary PM at urban scales

2.3.2 Secondary PM2.5 Formation

Secondary emissions account for a substantial fraction of PM2.5 mass with both natural and anthropogenic sources, forming by way of atmospheric photochemical oxidation reactions of both organic and inorganic gas-phase precursors.

2.3.2.1 Precursor Emissions

Ammonia plays important role in the formation of sulfate and nitrate PM. Oxidation of VOCs may also yield semi- and nonvolatile compounds that contribute to PM and the formation of new particles.

Sulfur dioxides emissions are mainly from electricity generation units (EGUs). Nitrogen oxides are emitted by a range of combustion sources, including various mobile sources

2.3.2.2 Secondary Inorganic Aerosols (SIA)

Particulate sulfate, nitrate, and ammonium formation processes help to form oxides of sulfur and nitrogen. Together, these PM2.5 components produced by secondary formation often account for the majority of PM2.5 mass.

Both H2SO4 and HNO3 react with atmospheric ammonia. Atmospheric particulate NH4NO3 is in equilibrium with gas phase NH3 and HNO3. Lower temperature and higher relative humidity shifts the equilibrium towards particulate NH4NO3 because of the large sensitivity of the equilibrium constant to temperature.

2.3.2.3 Secondary Organic Aerosols (SOA)

In the presence of high NOx concentrations, the oxidation of biogenic hydrocarbons is observed to produce larger quantities of SOA. High ambient NOx concentrations in the atmosphere are typically due to anthropogenic emissions. Mixtures of both organic and anthropogenic precursors produce greater SOA yields than mixtures dominated by just one class of precursors.

2.3.3 Primary PM10-2.5 Emissions

Crustal materials dominate the PM10-2.5 fraction throughout the US and fugitive dust has been identified as the largest sources of measured PM10 in many locations in the western US. Mineral dust, organic debris, and sea spray have also been identified as mainly in the coarse fraction. Road and construction dust represent a mechanism for suspension of crustal material on paved and unpaved roads.

Any potential for secondary coarse PM formation?

2.3.4 Ultrafine Particles

Ambient UFP originate from two distinct processes: primary emissions and new particle formation (NPF). Primary UFP originate from a large variety of sources such as transportation (road, traffic, ships, and aircraft), power plants, municipal waste incineration, construction and demolition, vegetation fires, domestic biomass burning, cooking, and cigarette smoke.

2.3.4.1 Primary Sources

Motor vehicles are a major, if not the most important, source of UFP in urban environments.
Most of the particles emitted by marine and aircraft engines are in the ultrafine size range. Emissions of UFPs appears to be a strong function of fuel sulfur content, with reduced emissions from lower sulfur fuels.

Biomass burning is also a major source of UFP.

PM2.5 and PM10 have various degrees of inorganic metals such as chromium, cadmium, manganese, arsenic, etc.

No discussion of natural background concentrations or anthropogenic transport between cities/states/regions, etc. of primary PM

2.4 Measurement, Monitoring, and Modeling
2.4.1 PM2.5 and PM10

FRMs and FEMs

In practice, a large fraction of the FEM monitors in operation form PM are automated and designed to provide hourly data, while FRMs for PM2.5 PM10, and PM10-2.5 require sampling for 24-hours and provide a daily average PM2.5 concentration, including pre-and post sampling gravimetric laboratory analysis.

Section 2.4.1 also discusses the difference between FRM and FEM monitors and describes the three most widely used FEMs. FRMs typically measure 24-hour integrated samples every third day. Short time resolution automated FEMs can measure hourly samples every day. In the past, FEMs typically measured higher $PM_{2.5}$ concentrations than FRMs; therefore, some states were reluctant to switch to FEMs. However, the new Teledyne optical spectrometer FEMs are much more accurate and many states are now converting their FRMs to FEMs. In July of 2017, Georgia EPD ran two regulatory FEMs. Currently, Georgia EPD runs nine regulatory FEMs and will be running twelve regulatory FEMs by June of 2019. A similar trend is occurring across many parts of the country which will produce significantly more $PM_{2.5}$ data at hourly resolution.

41% reduction from 2000 through 2017 for PM2.5 – annual average

2.4.2 PM10-2.5

Although the PM10-2.5 FRM and FEMs were already discussed in the 2009 ISA, the state of technology for PM10-2.5 measured is reviewed here because the large data set of nationwide PM10-2.5 network measurements is reported for the here, in this document, for the first time. PM10-2.5 FRM and FEMs new used for routine network monitoring are considerably improved compared to methods (i.e., subtraction methods) used in the previous key analyses of PM10-2.5 sampling issues. New results reveal changing trends in PM2.5/PM10 ratios.

2.4.4 Chemical Compounds

Measurement of PM components is potentially useful for providing insight into what sources contribute to PM mass as well as for discerning differential toxicity. Sulfate, nitrate, ammonium, organic carbon and elemental carbon as well as a suite of elements are measured in national speciation monitoring networks and intensive field studies mainly by collection on filters.

2.4.5 Satellite Remote Sensing

Satellite instruments measure radiance (electromagnetic energy flux), that can then be used to provide information on the aerosol column amount, or the aerosol optical depth (AOD). Because PM2.5 is not directly measured, computational algorithms involving a range of assumptions must be applied to obtain

estimates of PM2.5 concentrations. These inferred measurements involve potential errors that are not encountered with the FRM or other ground-based PM2.5 measurements.

Data cannot be collected when clouds and snow are present or from excessive amounts of smoke being mistaken for clouds.

The many factors that impact the relationship between AOD and PM2.5 concentrations lead to widely varying and sometimes relatively low, correlations when linear relationships are developed.

Limitations in accurately measuring PM concentrations

Any studies conducted to compare concentrations from satellite remote sensing with FRM/FEM monitoring?

2.4.6 Monitoring Networks

Extensive new PM monitoring efforts now complement long-standing networks by providing additional data supporting multiple objectives, including for PM research. These new monitoring efforts including Near-Road Monitoring for PM2.5 and the National Core (NCore) network for multipollutant measurements that are associated with special projects or are complementary to other networks, including particle number, black carbon, and continuous component monitoring.

Limitations of three to six day sample collection using FRM; FEM while continuous, not primary method

2.4.7 Chemistry-Transport Models (CTMs)

Key observations were that the largest errors in photochemical modeling were still thought to arise from the meteorological and emissions inputs to the model and that additional uncertainty was introduced by the parameterization of meteorological and chemical processes.

With respect to the concentrations derived from CTMs, the modeled concentrations are significantly higher than the observed concentrations at the speciation monitors. The reason for the overprediction is that there is no adjustment for near-source removal due to small sub-grid scale turbulence and impaction on building and vegetative surfaces (Pouliot G., *et al*., Assessing the Anthropogenic Fugitive Dust Emission Inventory and Temporal Allocation Using an Updated Speciation of Particulate Matter, January 2012, DOI: 10.1007/978-94-007-1359-8_97). It is estimated that local source removal typically accounts for 75% of total removal of fine particulate matter nationally (T.G. Pace, "Methodology to Estimate the Transportable Fraction (TF) of Fugitive Dust Emissions for Regional and Urban Scale Air Quality Analyses", U.S. EPA, Research Triangle Park NC, August 2005, https://www.nrc.gov/docs/ML1321/ML13213A386.pdf). This removal factor is defined as a "capture fraction" and varies by location. The amount that is not removed is defined as the "transportable fraction." A discussion of capture fraction and transportable fraction should be included in this chapter to help place the importance of dust emissions into proper perspective.

No additional discussion of limitations and/or uncertainties of CTM. Irrespective, Section 2.4.7 does a good job of documenting the scientific advances in CTMs.

2.5 Ambient Concentrations
2.5.1.1 Variability Across the US
2.5.1.1.1 PM2.5

Both annual average and 98th percentile concentrations are generally lower than what was observed in the 2005-2007 period, continuing the downward trend.

The mean of annual average concentrations based on 24-hour samples across all sites during the 3-year period (2013-2015) was 8.6 micrograms/cubic meter. This compares to a mean of annual average concentrations of 12 micrograms/cubic meter from 2005 to 2007.

There were a few notes discrepancies in the noted figures for the State of Georgia. Particularly, Figure 2-14 shows the 98th percentile 24-hour $PM_{2.5}$ concentrations for 2013-2015. The red monitor in southern Georgia appears to be Albany (13-095-0007). However, according to certified AQS data, the 24-hour 2013-2015 design value for Albany is 23 $\mu g/m^3$ (should be a blue dot, not red dot).

2.5.1.1.2 PM10

During the period from 2013-2015, the national average concentration was 21.1 micrograms/cubic meter, which is 15% lower than the average for 2005-2007

Summer concentrations appear to be typically higher than other seasons, with the highest average concentration as well as the highest concentrations at all percentiles up to the 95th percentile for summer. Winter concentrations are lower at all percentiles with average concentrations of 6 micrograms/cubic meter lower in winter than in summer.

There were a few notes discrepancies in the noted figures for the State of Georgia. Particularly, Figure 2-15 shows the 98th percentile PM_{10} concentrations for 2013-2015. There are no measurements shown in Georgia although Georgia has three PM_{10} monitors (13-089-0002, 13-121-0039, and 13-245-0091) with certified data in AQS from 2013-2015. The 98th percentile PM_{10} concentrations for all three PM_{10} monitors in Georgia are well below 75 $\mu g/m^3$ (blue dots).

2.5.1.1.6 PM2.5 Components

A major change in PM2.5 composition compared to the 2009 PM ISA is the reduction in sulfate concentrations, resulting in smaller sulfate contribution to PM2.5 mass in 2013-2015 compared to 2005-2007, especially in the Eastern US. As a result, at many locations sulfate has been replaced as the greatest single contributor to PM2.5 mass by organic material of nitrate.

2.5.2 Temporal Variability
2.5.2.1 Region Trends
2.5.2.1.5 Chemical Compounds

In the 2009 PM ISA, sulfate is described as the most abundant component of PM2.5 on a national average, with nitrate, particulate organic matter and sometimes crustal material also contributing substantially to PM2.5 mass.

Decreases in sulfate concentrations have led to decreases in PM2.5 concentrations since sulfates accounted for a large fraction of PM2.5 mass

2.5.2.2 Seasonal Variations
2.5.2.2.1 PM2.5

Averaged over all locations and years from 2001 – 2006, seasonal average PM2.5 concentrations were approximately 12 micrograms/cubic meter in summer and winter, but declined to approximately 9 micrograms/cubic meter in the spring and fall.
Observations that the highest seasonal average concentrations occurred in the Eastern US and in winter in the Western US.

The observed reduction in summer PM2.5 concentrations in the East to the extent that summer is no longer the season with the highest national average PM2.5 concentrations is a major development, and is a predictable consequence of successful reduction of SO2 emissions.

2.5.2.2.4 PM Components

Sulfate and OC together accounted for the majority of PM2.5 mass in many metropolitan areas in the summer, while higher nitrate concentrations were observed in the winter.

2.5.2.3 Hourly and Weekday-Weekend Variability

A two-peaked diel pattern was observed in diverse urban locations and attributed to rush-hour traffic for the morning peak and a combination of rush hour traffic, decreasing atmospheric dilution, and nucleation for the afternoon/evening peak.

2.5.3 Common Patterns of Particulate Matter Characteristics in the US

Historically, PM2.5 has been highest in the summer and has been largely accounted for by sulfate over a large area that that encompasses most of the Eastern US, extending into the Great Plains.

At all of the locations reported sulfate was the most abundant component measured for the period 2003-2005, accounting for close to half of the overall average PM2.5 mass.

Ammonium nitrate and organic PM from diverse combustion sources are the main contributors to PM2.5 under winter conditions.

A common characteristic of PM in both California and the dryer areas of the Western US that contrasts with the Eastern US is higher fraction of PM10 accounted for by PM10-2.5, with PM10-2.5 accounting for the most PM10 mass in the West, but PM2.5 accounting for most PM10 in the East.

PM2.5 concentrations averaged over the 11-year period from 1998-2008 over the entire contiguous US were reported to be 2.6 micrograms/cubic meter higher on days under stagnant conditions than for non-stagnant days. When all US data over a multiyear period are considered, temperature is positively correlated with PM2.5.

Importance of confounding for temperature and/or relative humidity in either controlled human studies and/or epidemiological studies.

2.5.4 Background Particulate Matter

Missing discussion on regional (i.e., state-to-state) transport for both PM2.5 and PM10.

Background PM concentrations can be best characterized with chemical transport modeling simulations via source apportionment modeling or estimating what the residual PM concentrations would be were the US anthropogenic emissions entirely removed (i.e., "zero-out" modeling).

2.5.4.2 Intercontinental Transport

Transport at midlatitudes is dominated by westerly winds, which transport East Asian emission across the North Pacific Ocean to North America.

Observed trends in PM are usually more closely related to local emission trends than to long-range transport, and at monitoring sites throughout the US intercontinental influences are small.

Missing discussion on regional (i.e., state-to-state) transport for both PM2.5 and PM10.

Dr. Steven Packham

A second draft of the Draft ISA is needed to better integrate evidence from toxicology and human studies. Biological pathways should be outlined showing pulmonary inflammation as a key causal link between inhaled PM and adverse health effects as suggested below in Figure 1. There are numerous studies published on the respiratory system's responses to PM and the systemic mediation of acute pulmonary inflammation. Many are not referenced in the Draft ISA. The time has come to properly and fully integrate risk assessment with biological theory and evidence.

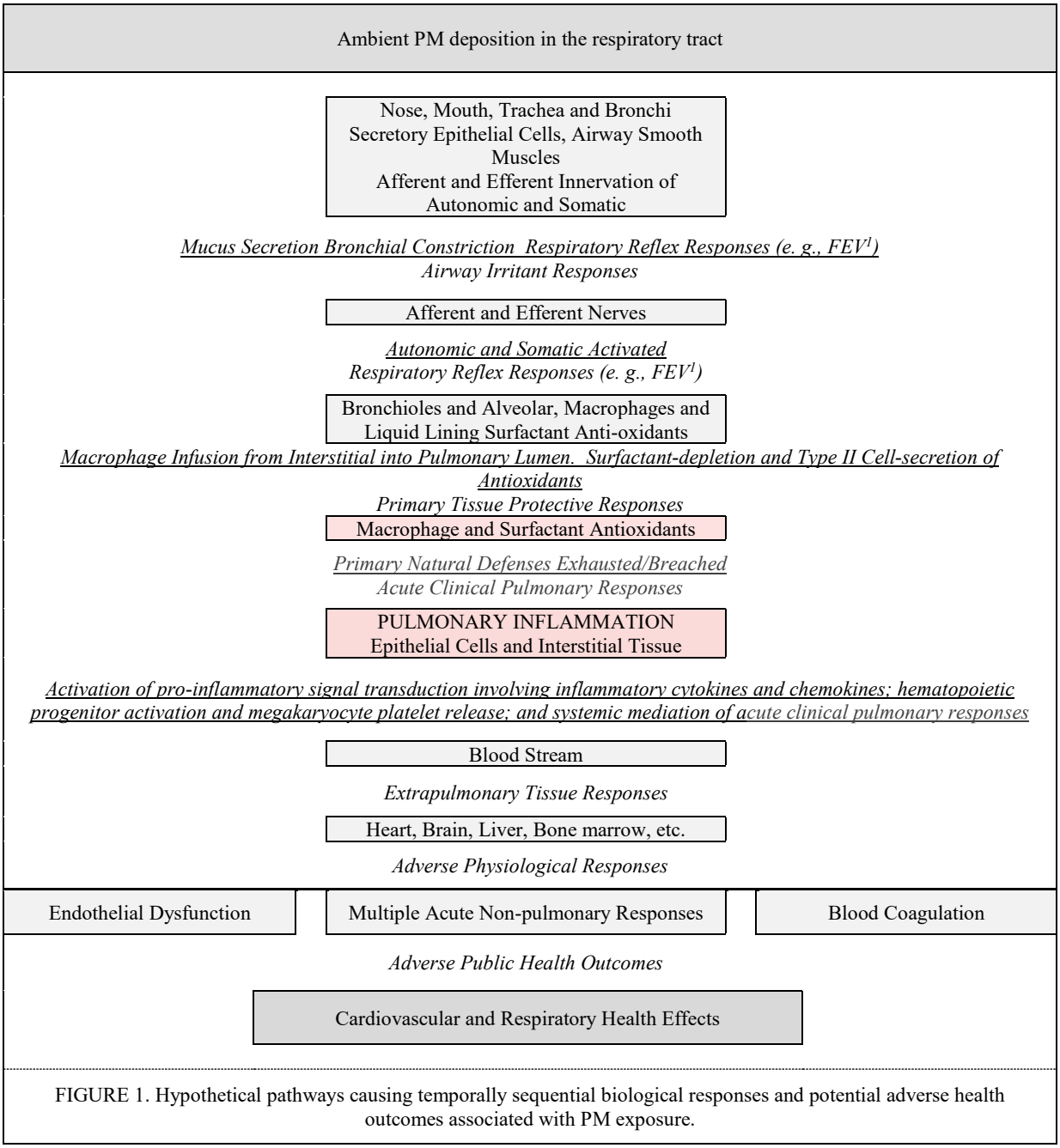


FIGURE 1. Hypothetical pathways causing temporally sequential biological responses and potential adverse health outcomes associated with PM exposure.

1. The final PM ISA should use a more comprehensive review of relevant human studies and exposure-dose-response evidence to estimate inflammatory response thresholds for ambient PM exposures and to hypothesize PM exposure concentration levels that would be protective of normal and sensitive individuals. [1, 2]
2. The final PM ISA should reinstate the integrative approach schematically characterized initially in Figures 10.1 and 10.2 (U. S. 1996 AQCD) and methods described on pages 10-1 through 10-5 of that document. Similar methods were outlined again only fifteen years ago (U. S. 2004 AQCD, pages 6-1, 6-2, 6-74, 7-1 and 7-2).

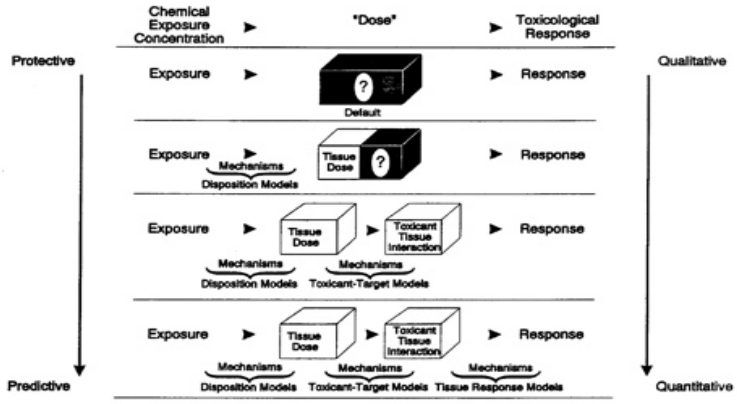


Figure 10-1. Schematic characterization of comprehensive exposure-dose-response continuum and the evolution of protective to predictive dose-response estimates.

Source: Adapted from Conolly (1990) and Andersen et al. (1992).

Characterization of the exposure-dose-response continuum for PM requires the elucidation and understanding of the mechanistic determinants of inhaled particle dose, toxicant-target interactions, and tissue responses. The exposure-dose-response continuum can be represented as events in the progression from exposure to a specific adverse public health outcome as illustrated in Figure 10-2.

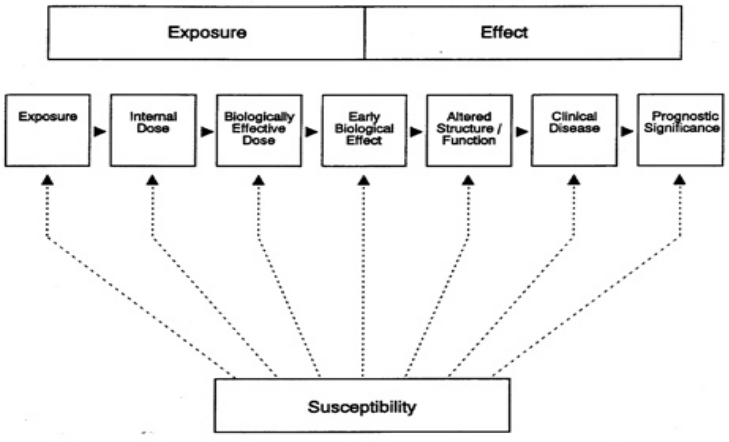


Figure 10-2. Biological marker components in sequential progression between exposure and disease.

3. The final PM ISA should use this integrative philosophy and approach to test the protective efficacy of estimated pulmonary inflammation thresholds and *tissue response models* to validate predicted causal determinations.
4. The final PM ISA should include a thorough analysis and review of the fundamental differences between continuous graded dose-response relationships and all-or-none quantal concentration-risk (C-R) relationships.
5. The final PM ISA should cite references associated with Figure 14.[3]

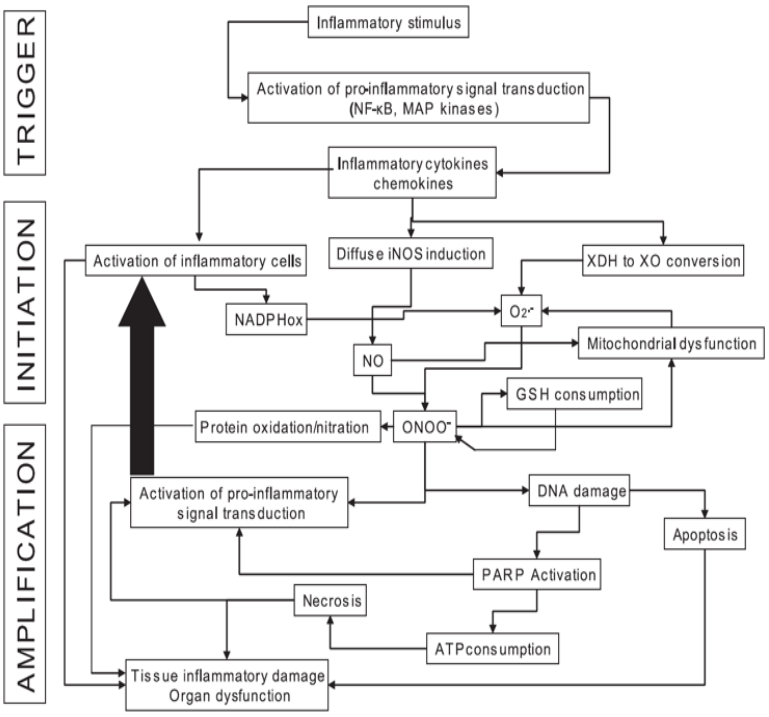


FIG. 14. Mechanisms of amplification of inflammation by peroxynitrite. Inflammation is triggered by the activation of multiple signaling cascades culminating in the upregulated production of an array of proinflammatory cytokines and chemokines. Those initiate a more complex inflammatory reaction characterized by the activation of inflammatory cells and the stimulated activity of enzymes, including inducible NO synthase (iNOS), which produces high amounts of NO, and the superoxide ($O_2^{\bullet-}$)-producing enzymes NADPH oxidase (NADPHox) and xanthine oxidase (XO). The simultaneous production of NO and $O_2^{\bullet-}$ results in the generation of peroxynitrite ($ONOO^-$), which in turn damages target molecules including proteins, glutathione (GSH), mitochondria, and DNA. DNA damage can initiate apoptotic cell death and is also the obligatory trigger for the activation of poly(ADP-ribose) polymerase (PARP), which may induce cell necrosis by ATP depletion. Both $ONOO^-$ and PARP further participate to the upregulation of proinflammatory signal transduction pathways, thereby producing a self-amplifying cycle of inflammatory cell injury, as indicated by the black arrow.

Physiol Rev • VOL 87 • JANUARY 2007 • www.prv.org

Figure 14

6. The final PM ISA should also cite selected references from Pacher, P. el. al. (2007) used in the following Sub-sections:
   A. Nitric oxide and peroxynitrite in cardiac diseases 343-51
   B. Nitric oxide and peroxynitrite in vascular diseases 352-54
   C. Nitric oxide and peroxynitrite in circulatory shock 355-61
   E. Nitric oxide and peroxynitrite in cancer 366
   F. Nitric oxide and peroxynitrite in stroke and other forms of reperfusion injury 367-375
   G. Nitric oxide and peroxynitrite in neurodegenerative disorders 376-79, and
   H. Nitric oxide and peroxynitrite in diabetes and diabetic complications 380-384

A General Comment

It would be helpful for the CASAC to have ready access to experts in toxicology and human physiology comparable in number and prominence to those in epidemiology and risk assessment. Experts with knowledge of human biology, particularly respiratory physiology and natural pulmonary defense mechanisms against most air contaminants, would allow for a better understanding of plausibility and coherence, the determination of causal mechanisms, the interpretation of graded dose-response relationships, the identification of dose-dependent thresholds, and the establishment of requisite margins of safety for clinical adverse health effects.

Specific Charge Questions (Director Vandenberg)

Comments on ISA, Chapter 4

To what extent does the discussion clearly and accurately convey the dosimetry of inhaled PM and the processes of deposition, clearance, retention, and translocation?

Section 4.2 is reasonably accurate with respect to dosimetry in terms of patterns of deposition in various anatomical regions of the respiratory system. Particle deposition density per cm^2 of surface area in various anatomical regions is an important factor. Particularly in evaluating studies of pulmonary defense mechanisms. A table of particle deposition densities should be added.

Sub-section 4.1.2.3: Epithelial Lining Fluid should include a reference to the study by Kendall, et al[4] in which PM2.5 was collected directly into normal lung lining liquid (surfactant). The particles aggregated into larger (>5 um) dense structures compared with samples collected in air or into saline. The control showed that the agglomeration effects were not due to drying *per se* but were specifically associated with the protein-rich surfactant solution. Studies of surface chemistry for urban and smoking PM2.5 show significant modification by BALF. Findings of increased attractive and adhesive forces in BALF suggest that aggregation is enhanced by the surfactant lining the surface of the lungs and respiratory airways."

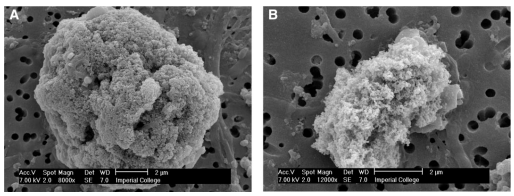


Fig. 2. Densely agglomerated 35-nm particle conglomerates (>5 um) found in particle samples collected by sampling $PM_{2.5}$ directly into lung lining liquid. Samples were subsequently filtered onto 0.4 um Nucleopore filters for SEM analysis. Kendall, et al. Page L112

This same study should be referenced in sections and sub-sections of Chapter 4 on the subjects of *deposition, clearance, retention, and translocation*; that are specifically mentioned in the charge question. The apparent aggregation of smaller particles into larger ones has huge implications on theories, assumptions, and frank speculations on plausibility determinations as well as theories of causal

biological mechanisms: All plausibility determinations and theories, assumptions and speculations relating to causation in the ISA need to be modified and corrected in the ISA narrative.

Please comment on the identification, evaluation and characterization of the available scientific evidence from epidemiologic, controlled human exposure, toxicological and associated human exposure and atmospheric sciences studies and the application of information from these studies to inform causality determinations and uncertainty characterizations for human health outcomes.

The identification, evaluation and characterization of scientific evidence from controlled human exposure, toxicological and associated human exposures have not been properly or adequately applied to the determination of causality and the issue of uncertainty in this or in the other Chapters of the ISA.

Comments to Specific Charge Questions for Chapters 5-11

Please comment on the characterization of the evidence within these chapters.

Please comment on the portrayal and discussion of the biological plausibility evidence presented at the outset of Chapters 5 – 11 and the extent to which: (1) the organization adequately captures the current state of the science with respect to potential pathways by which PM could impart health effects, and (2) as currently constructed, inform causality determinations.

For comments that otherwise would be made here, please see comments to Specific Charge Questions for Chapters 4 and 12, and to the Overarching Charge Questions.

Comments to Specific Charge Questions for Chapter 12

Please comment on the extent to which the available scientific evidence from epidemiologic, controlled human exposure and toxicological studies [has] been integrated to inform conclusions on populations and/or lifestages potentially at increased risk of a PM-related health effect.

Chapter 12 has neither identified nor properly evaluated scientific evidence, knowledge, or concepts from the scientific disciplines of toxicology and human physiology with sufficient clarity to assess whether air quality standards protect 'at risk' populations and vulnerable lifestages with an adequate margin of safety. At the heart of the problem are the issues of the two types of dose-response relationships and population vs individual thresholds.

<u>Toxicology</u>

Toxicology is the science[5] of adverse effects of substances on living organisms.[6,7]. The dose response relationship between the degree of response of the biological organism and the amount of substance causing a biological response is the most fundamental and pervasive concept in toxicology.[8]

From a practical perspective, there are two types of dose-response relationships: 1) the individual dose-response relationship, which describes the response of an individual organism to varying doses, often

referred to as a 'graded' response because the measured effect is continuous over a range of doses, and 2) a 'quantal' dose-response relationship, which characterizes the distribution of responses to different doses in a population of individual organisms. This distribution can vary depending on the specific substance and dosage range being studied. It's not atypical for it to assume a normal, or bell-shaped curve, as shown in Figure 1:[9]

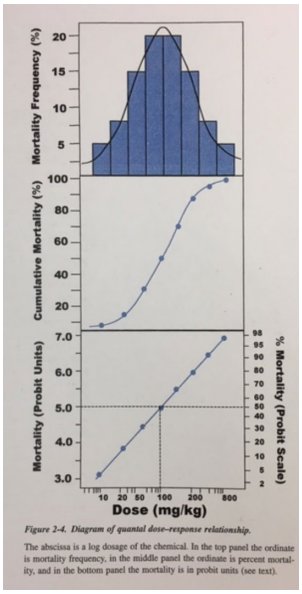


Figure 1[10]

Thresholds and Dose-response Relationships:

In contrast to the 'graded' or continuous-scale dose-response in individuals the dose response relationship in a population is 'quantal' in nature. Meaning that at any given dose an individual is classified as either being a "responder" or a "nonresponder."

In a probabilistic transformation of a population frequency distribution, one approaches a response of 0 percent as the dose decreases and 100 percent as the dose is increased. Theoretically, the *probability* of the response never reaches 0 or 100 percent. However, the smallest effective dose of any chemical or substance that causes and evokes a stated all-or-none response is called a *threshold dose* even though theoretically it can't be determined experimentally. The most important assumption that must be considered before a dose-response relationship can be used appropriately is that the substance or pollutant is *causing* the response.[11]

Integration of Toxicology and Human Physiology

Chapter 12 does not present the fundamental principles of graded and quantal dose-response relationships and thresholds. Nor does it define and discuss the critical concepts of *threshold* and *margin of safety* in context of graded vs quantal dose-response relationships. These omissions are also found in the *Preamble to the Integrated Science Assessments* (hereafter *Preamble*) rendering it inadequate for

outlining a general process for integrating evidence from epidemiological studies with those of controlled human and animal exposures.

If chapters and sections on the fundamental principles of toxicology and human physiology could be added to the *Preamble*, they could significantly enhance the CASAC's accelerated review of draft IRP and ISA documents.

Standard-setting Options and Alternatives

A range of options and alternatives for setting standards in terms of indicators, averaging times, form, and levels can be developed and scientifically supported with empirically defined and quantified thresholds and margins of safety. Please consider the alternative NAAQS concept presented in Figures 2a and 2b using underlying concepts of graded and quantal dose-response relationships.

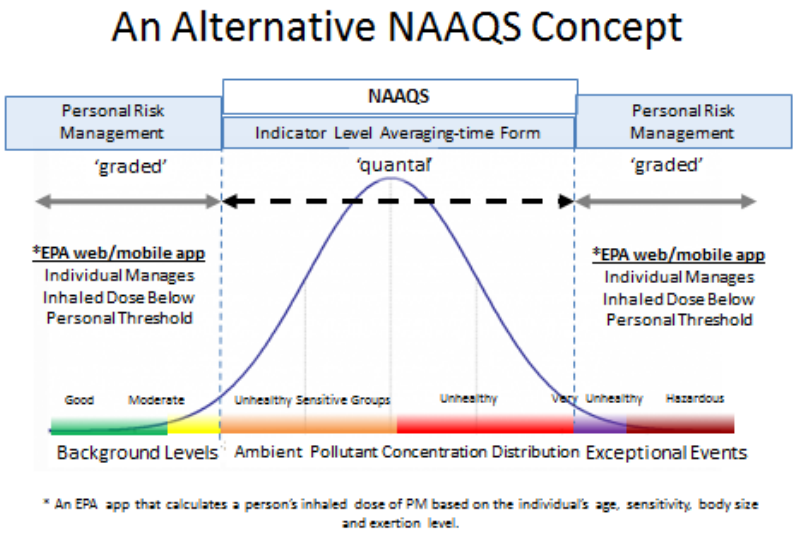


Figure 2a

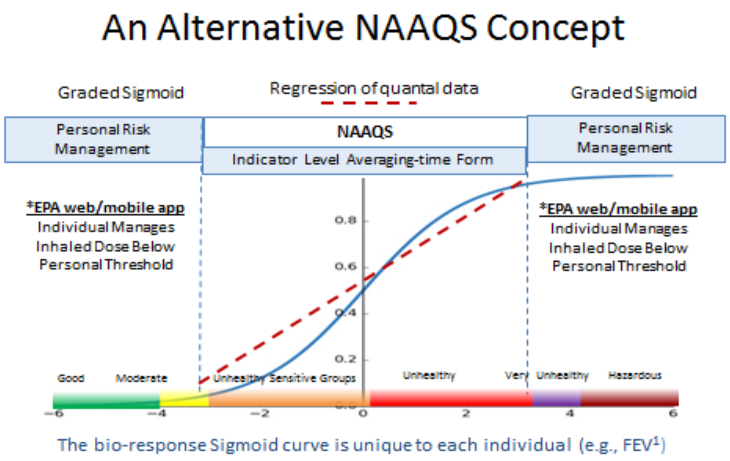


Figure 2b

Overarching Charge Questions

In the context of Principle 2: What scientific evidence has been developed since the last review to indicate if the current primary and/or secondary NAAQS need to be revised or if an alternative level or form of these standards is needed to protect public health and/or public welfare? Please recommend to the Administrator any new NAAQS or revisions of existing criteria and standards as may be appropriate. In providing advice, please consider a range of options for standard setting, in terms of indicators, averaging times, form, and levels for any alternative standards, along with a description of the alternative underlying interpretations of the scientific evidence and risk/exposure information that might support such alternative standards

Critical concepts relating to human physiology and animal toxicology are not well represented in the ISA. The CASAC's ability to provide advice on standard setting options is consequently limited.

For instance, the fundamental difference between a) *graded* dose-response data collect from individual organisms and b) *quantal* data collected from population databases is not included in the *Preamble* or the ISA.

An example of an alternative strategy for setting NAAQS and modifying the AQI using compatible strengths of graded and quantal dose response studies is provided in comments to Specific Charge Questions for Chapter 12.

It is my opinion that a chapter on principles of toxicology and basic human respiratory physiology should be added in this review cycle to the *Preamble* to be used in drafting this and all future ISAs and IRPs.

Are there areas in which additional knowledge is required

Yes, fundamental principles of toxicology and human physiology should be added to the *Preamble* and current knowledge and evidence relating to biologic causal mechanisms, dose response thresholds, and margins of safety need to be presented independently from risk assessments and then integrated in the ISA

Do key studies, analyses, and assessments which may inform the Administrator's decision to revise the NAAQS properly address or characterize uncertainty and causality? Are there appropriate criteria to ensure transparency in the evaluation, assessment, and characterization of key scientific evidence for this review?

Key studies from all relevant scientific disciplines, particularly those from the sciences of human physiology and toxicology, need to be properly integrated into review plans and ISA air quality criteria.

Whether the ISA *properly* addresses uncertainty is a matter of one's epistemology and philosophy of science, and one's definition of science. [12] My philosophy of science favors Karl Popper's principle of empirical falsification: Meaning, a theory in the empirical sciences can never be proven, but it can be falsified, and should be scrutinized by decisive experiments. Metaphysically, I consider individual human organisms as ontological antecedents to communities, and cohorts, i.e., to groups and cohorts of individuals whether aggregated in study designs or grouped by nature, e.g., DNA profiles, families, etc.; or by physical living conditions and locations, e.g., altitude, climate, etc.; or political boundaries such as cities, counties, states, and nations.

Lastly, I accept the definition of science to be the intellectual and practical activity encompassing the systematic study of the structure and behavior of the *physical* and *natural* world through observation and experiment, and I would like to see a more thorough treatment of scientific knowledge of human biology, particularly respiratory physiology and pulmonary defense mechanisms by which the body protects itself from environmental stressors encounter on a daily basis.

---

[1] Andersen, M. E., Krishnan. K. Conolly, R. B.; McClellan, R. O. (1992) Mechanistic toxicology research and biologically-based modeling: partners for improving quantitative risk assessments. CIIT Activities 12(1): 1-7.
[2] Conolly, R. B. (1990) Biologically-based models for toxic effects: tools for hypothesis testing and improving health risk assessments. CIIT Activities 10: 1-8.
[3] Pacher, P. et. al. (2007) *Nitric Oxide and Peroxynitrite in Health and Disease* (Sub-section D. *Nitric Oxide and Peroxynitrite in Local Inflammation*) Physiological Reviews, 87(1), 315–424. https://doi.org/10.1152/physrev.00029.2006
[4] Kendall M, Tetley T D, Wigzell E, Hutton B, Nieuwenhuijsen M, And Luckham P. Lung ling liquid modifies $PM_{2.5}$ in favor of particle aggregation: a protective mechanism. *Am J Physiol Lung Cell Mol Physiol* ,282: L109–L114, 2002.

[5] Science is the intellectual and practical activity encompassing the systematic study if the structure and behavour of the physical world through observation and experiment. Oxford Dictionary of English
[6] Casarett and Doull's Toxicology: The Basic Science of Poisons. 2nd Edition. Chapter 2. Evaluation of Safety: Toxicologic Evaluation Curtis D. Klaassen and John Doull. 1980
[7] Casarett and Doull's Toxicology: The Basic Science of Poisons. 6th Edition. Chapter 2, Principles of Toxicology David L. Eaton and Curtis D. Kaassen. 2001
[8] Ibid. pp 17-18.
[9] Ibid. pp 18-19.
[10] Ibid. p 19.
[11] Cause is determined by noting a relationship between actions or events such that one or more are the result of the other or others. In legal rulings of causation this may be stated as, But for Event A, Event B would not have resulted.
[12] Oxford English Dictionary.