Comments on the CASAC Draft Letter to the Administrator on Its Review of US EPA's Integrated Science Assessment for Particulate Matter (External Review Draft – October 2018)

Prepared by
Julie E. Goodman, Ph.D., DABT, FACE, ATS

March 21, 2019

GRADIENT
www.gradientcorp.com
20 University Road
Cambridge, MA 02138
617-395-5000

The United States Environmental Protection Agency (US EPA) Clean Air Scientific Advisory Committee (CASAC) draft letter to the Administrator of US EPA consists of "Consensus Responses to Charge Questions on the Draft ISA" (CASAC, 2019a) and "Individual Comments by CASAC Members on the Draft ISA" (CASAC, 2019b). The draft letter includes very detailed opinions regarding the "Integrated Science Assessment for Particulate Matter (External Review Draft)" (referred to herein as the Draft ISA), which was released in October 2018 (US EPA, 2018a), and provides many useful recommendations for improving the ISA. In general, these opinions are consistent with the comments Gradient provided on the Draft ISA, attached here (Gradient, 2018); however, as discussed below, there are a few areas in which CASAC could provide more detail.

Areas of Expertise

CASAC recommends that it be provided with access to additional expertise (CASAC, 2019a, p. 5). The areas of expertise requested include:

a. Characterization of sampling errors and biases from continuous ambient PM measurements and satellite remote sensing aerosol optical depth (AOD) analysis;
b. Errors and biases in dispersion modeling and photochemical grid modeling;
c. Errors-in-variables methods and effects of exposure (and covariate) estimation errors on epidemiologic study results;
d. Epidemiology of low-dose causal concentration-response functions;
e. Comparative toxicology, dosimetry, and extrapolation of findings in animals to humans;
f. Effects of PM on visibility impairment, climate, and materials. (CASAC, 2019a)

All of these areas are critical to evaluating the effects of PM, and several of these have not been adequately addressed in the Draft ISA (*e.g.*, see Section 3 in our comments on the Draft ISA; Gradient, 2018). However, CASAC does not necessarily need additional experts to evaluate these areas; rather, the ISA should provide more detail so that current CASAC members, with their diverse scientific backgrounds, can evaluate whether these areas have been sufficiently addressed.

Systematic Review Criteria

CASAC states that, "the Draft ISA does not provide a comprehensive or systematic assessment of the available science relevant to understanding the health impacts of exposure to fine particulate matter, nor does it follow widely accepted scientific methods for deriving sound, independently verifiable, scientific conclusions from available data" (CASAC, 2019a). CASAC then goes into more detail in its responses to the charge question in Chapter 1 of the Draft ISA (CASAC, 2019a, p. 12).

While CASAC recommends that the systematic review criteria used in the Draft ISA should be described "explicitly" (CASAC, 2019a), it gives no recommendations on what these criteria should be. CASAC could specifically recommend that the ISA consider the criteria provided in the Toxic Substances Control Act (TSCA) systematic review process (US EPA, 2018b) or those used in other existing evaluation systems, such as the Biomonitoring, Environmental Epidemiology, and Short-lived Chemicals (BEES-C) instrument (LaKind *et al*., 2014), other US EPA systematic review practices (*e.g*., the Integrated Risk Information System [IRIS] Risk of Bias [RoB] evaluation; US EPA, 2013, 2014), the National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) systematic review approach (NTP, 2015), the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (von Elm *et al*., 2007a-e), the Toxicological data Reliability Assessment Tool (ToxRTool; Schneider *et al*., 2009), the Organisation for Economic Co-operation and Development (OECD) guidance document on hazard

assessment (OECD, 2005), and the risk of bias instrument for non-randomized studies of exposures (Morgan *et al*., 2019). Lynch *et al*. (2016) review some of these frameworks, although some have been updated since the publication of this article (*e.g*., LaKind *et al*., 2015; NTP, 2019).

Study Quality and Relevance Criteria

Appendix 1 of the Draft ISA provides study quality criteria in detail (US EPA, 2018a). Issues with these criteria are described briefly in CASAC's draft letter, but more discussion is warranted. The CASAC cover letter and responses to charge questions address study quality in several instances. However, CASAC could be more explicit and specifically indicate that Appendix 1 of the Draft ISA be revised such that:

- **It includes a complete list of study quality criteria.** While the Draft ISA has a list of important study quality aspects for evaluating health effects in Appendix 1 (but no comparable list for welfare studies), it is not a complete list.
- **It indicates that study quality and relevance impact how informative a study is.** Studies of higher quality should be considered more informative, while those with more limitations should be considered less informative. In addition, criteria that must be met for study results to be considered relevant to the US population as a whole, or to "at-risk" populations, should be explicitly stated.
- **It includes quality criteria for *in vitro* and welfare studies.** US EPA cannot determine whether these studies support or call into question a causal association if it has not evaluated study quality.
- **It requires that aspects are tabulated for each individual study.** A systematic review involves reviewing and judging the quality of each individual study in the same manner, and presenting this in a transparent manner. This was not done in the Draft ISA. For practical reasons, quality aspects for individual studies should at least be tabulated for key endpoints that inform the National Ambient Air Quality Standards (NAAQS) (*e.g*., total mortality).

CASAC could also provide good examples of study quality evaluations (*e.g*., see Sections S1.2 in each of the three supplements of Goodman *et al*. [2018] and the Study Quality Evaluation in the Methods and Results and Table 2 in Zu *et al*. [2018]). These systematic reviews explicitly state the metrics used to determine study quality for each discipline and what constitutes higher and lower quality. Quality considerations for each study are tabulated, with studies in rows and study quality aspects in columns. All metrics with lower scores are highlighted, so that the quality of the literature is clear.

Biological Action of PM

CASAC states in its consensus responses to charge questions that, "[t]he 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" (CASAC, 2019a).

The responses to charge questions could further state that the ISA should follow an approach that assesses a complete mechanistic pathway, rather than a "single endpoint" or groups of potentially unrelated endpoints, as recommended for cancer biomarkers for environmental chemicals (*e.g*., Goodson *et al*., 2015), and that the ISA should fully consider the human relevance of the observed endpoints or animal models tested. Many outcomes measured are upstream events, not apical effects. The Draft ISA does not address whether the upstream events were homeostatic changes or reversible effects; the detection of an upstream event alone does not necessarily indicate pathogenesis or disease onset.

Finally, although Table A-1 in Appendix 1 of the Draft ISA (US EPA, 2018a) discusses a number of quality-related considerations for evaluating evidence on $PM_{2.5}$ health effects in various disciplines, including animal toxicology, CASAC could indicate that these considerations are not sufficiently detailed or prescriptive to ensure a consistent evaluation across studies and endpoints, and that the ISA should discuss the quality of toxicity studies it reviews, rather than take the results of individual studies at face value without discussing study- and endpoint-specific methodological limitations. CASAC could also suggest that the ISA consider study quality when integrating evidence across endpoints and studies.

Key Characteristics of Carcinogens

The Draft ISA recommends using the International Agency for Research on Cancer (IARC) 10 key characteristics of carcinogens framework (Smith *et al*., 2016) to identify potential carcinogens. CASAC states that, "[t]he 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)" (CASAC, 2019a, p. 25).

The CASAC letter could mention that the key characteristics framework does not include explicit guidance for evaluating the biological significance of mechanistic endpoints, inter- and intra-individual variability, or study quality and relevance (see Goodman and Lynch, 2017). This framework also does not explicitly address how mechanistic evidence should be integrated with other realms of evidence. Thus, rather than use IARC's framework, the Draft ISA should adapt available frameworks that address study quality and human relevance to evaluate potential carcinogens.

Study Descriptions

CASAC provides comments regarding specific details of individual studies that should be addressed in the ISA. However, CASAC should note that these comments do not constitute a comprehensive or systematic review of individual studies, and that US EPA should not consider these comments as a checklist (in that once they are addressed, the ISA is complete). Rather, CASAC should indicate that all studies be reviewed in a systematic manner, and when doing so, CASAC's comments on individual study details should also be addressed.

Lung Cancer

CASAC indicates in its consensus responses to charge questions that most epidemiology studies of PM and lung cancer evaluated $PM_{2.5}$ exposures a few years before cancer diagnosis or death (CASAC, 2019a, p. 32). In fact, as demonstrated in our comments on the Draft ISA (Gradient, 2018), the exposure periods assessed in most studies included time after the follow-up for cancer started. In some studies, the entire exposure period occurred after the follow-up period ended. This should be added to the CASAC responses to charge questions.

The CASAC letter could also mention that the available epidemiology studies do not adequately account for smoking and family history of lung cancer, and could recommend that the ISA should consider the quality or human relevance of the findings in the experimental studies.

Central Nervous System Effects

In its consensus responses to charge questions, CASAC describes several limitations of the toxicology studies evaluating central nervous system effects. CASAC could also indicate that the Draft ISA does not present any systematic study quality evaluation when it summarizes the available literature on central nervous system effects, nor does it appear to consider study quality when synthesizing the evidence for this group of endpoints. Epidemiology studies of brain volume, cognitive function, and dementia have considerable limitations and uncertainties that undermine the observed associations between long-term $PM_{2.5}$ exposure and neurological endpoints, but these were not considered in the Draft ISA. Findings from animal toxicity studies do not provide evidence for apical central nervous system endpoints and may have limited relevance to humans.

Concentration-Response Functions

In the draft letter, CASAC members appear to be divided regarding the Draft ISA assessment of concentration-response functions. Issues with these functions are described in detail in our comments on the Draft ISA (Gradient, 2018). Briefly, the Draft ISA indicates that recent epidemiology studies of long-term $PM_{2.5}$ exposure and mortality generally support a linear, no-threshold relationship, with confidence in some studies in the range of 5-8 $\mu g/m^3$. The Draft ISA also discusses evidence from cardiovascular endpoints as supportive of $PM_{2.5}$ effects at low concentrations. With regard to short-term $PM_{2.5}$ exposure and mortality, the Draft ISA indicates that epidemiology studies conducted in the US provide evidence for a linear relationship at concentrations as low as 5 $\mu g/m^3$. The Draft ISA does not systematically evaluate the quality of these studies or fully consider potential biases and uncertainties when evaluating the evidence regarding the shape of the concentration-response curves. In addition, the Draft ISA's evaluation is not systematic or consistent across studies or outcomes. Considerable methodological limitations and uncertainties in these epidemiology studies preclude the observed concentration-response data from being used as a basis to revise the level of NAAQS. CASAC members should consider these issues with the evaluation of concentration-response functions in the Draft ISA, and then provide a consensus recommendation.

Populations and Life Stages Potentially at Increased Risk of PM-related Health Effects

Regarding the section of the Draft ISA that "delineates the approach to considering the evidence for at-risk groups and populations," CASAC states that, "[i]t 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" (CASAC, 2019a).

Although US EPA's frameworks for both causal determination and classification of evidence for populations and life stages potentially at increased risk of PM-related health effects are based on consideration of modified Bradford Hill aspects, the former has five categories and the latter four, and the criteria for classification are not the same. It appears that, although defined differently, the four categories in the latter framework are roughly equivalent to *causal relationship*, *suggestive of a causal relationship*, *inadequate to infer a causal relationship*, and *not likely to be a causal relationship* in the Draft ISA's causal framework; the framework does not appear to include an equivalent to the *likely to be a causal relationship* category. The issues with the weight of evidence for causal determination apply to these classifications as well. For example, the Draft ISA defines evidence for a population or life stage as being suggestive of a causal relationship if it "is limited due to some inconsistency within a discipline or, where applicable, a lack of coherence across disciplines" (US EPA, 2018a). If an inconsistency or a lack of coherence is large

enough, it should lead to a conclusion that the weight of evidence is below equipoise. In these circumstances, the evidence is inadequate to make a determination as to whether a population or life stage is at increased risk.

CASAC could indicate that the framework for populations and life stages potentially at increased risk of PM-related health effects and that for causal determination should not differ. Although one is an assessment of direct causation and the other an assessment of factors that can contribute to (or prevent) causation of health effects from PM, in both cases, the goal is to critically, systematically, and transparently review the weight of the scientific evidence. Ideally, the same rules should be applied for both types of analysis; if not, there needs to be justification for using different rules to conduct the same type of analysis. CASAC could recommend that the ISA adopt the Institute of Medicine's (IOM, 2008) recommended categories for the level of evidence required to determine causation, which consider whether the weight of evidence is above or below equipoise.

Disclosure of Interest

During the normal course of her employment, Dr. Julie E. Goodman reviewed the CASAC draft letter to the Administrator, and prepared and submitted these comments to the Designated Federal Officer (DFO) for distribution to CASAC. The American Petroleum Institute provided Gradient with financial support for this work.

References

Clean Air Scientific Advisory Committee (CASAC). 2019a. Letter Report to G. McCarthy (US EPA) re: CASAC Review of the EPA's Integrated Science Assessment for Particulate Matter (External Review Draft - October 2018) (Draft)." 240p., March 7.

Clean Air Scientific Advisory Committee (CASAC). 2019b. "Appendix A: Individual Comments by CASAC Members on the EPA's Integrated Science Assessment for Particulate Matter (External Review Draft - October 2018) (Draft)." 185p., March 7.

Goodman, J; Lynch, H. 2017. "Improving the International Agency for Research on Cancer's consideration of mechanistic evidence." *Toxicol. Appl. Pharmacol.* 319:39-46. doi: 10.1016/j.taap.2017.01.020.

Goodman, JE; Zu, K; Loftus, CT; Lynch, HN; Prueitt, RL; Mohar, I; Shubin, SP; Sax, SN. 2018. "Short-term ozone exposure and asthma severity: Weight-of-evidence analysis." *Environ. Res.* 160:391-397. doi: 10.1016/j.envres.2017.10.018.

Goodson, WH 3rd; Lowe, L; Carpenter, DO; *et al*. 2015. "Assessing the carcinogenic potential of low-dose exposures to chemical mixtures in the environment: The challenge ahead." *Carcinogenesis* 36(Suppl. 1):S254-S296. doi: 10.1093/carcin/bgv039.

Gradient. 2018. "Comments from J. Goodman on EPA's Integrated Science Assessment for Particulate Matter (External Review Draft)" Report to American Petroleum Institute. Submitted to US EPA Docket. EPA-HQ-ORD-2014-0859. 75p., December 10.

Institute of Medicine (IOM). 2008. "Improving the Presumptive Disability Decision-Making Process for Veterans." Committee on Evaluation of the Presumptive Disability Decision-Making Process for Veterans, Board on Military and Veterans Health. National Academies Press, Washington, DC. 781p. Accessed at http://books.nap.edu/openbook.php?record_id=11908&.

LaKind, JS; Sobus, JR; Goodman, M; Barr, DB; Furst, P; Albertini, RJ; Arbuckle, TE; Schoeters, G; Tan, YM; Teeguarden, J; Tornero-Velez, R; Weisel, CP. 2014. "A proposal for assessing study quality: Biomonitoring, Environmental Epidemiology, and Short-lived Chemicals (BEES-C) instrument." *Environ. Int.* 73:195-207. doi: 10.1016/j.envint.2014.07.011.

LaKind, JS; Goodman, M; Barr, DB; Weisel, CP; Schoeters, G. 2015. "Lessons learned from the application of BEES-C: Systematic assessment of study quality of epidemiologic research on BPA, neurodevelopment, and respiratory health." *Environ. Int.* 80:41-71. doi: 10.1016/j.envint.2015.03.015.

Lynch, HN; Goodman, JE; Tabony, JA; Rhomberg, LR. 2016. "Systematic comparison of study quality criteria." *Regul. Toxicol. Pharmacol.* 76:187-198. doi: 10.1016/j.yrtph.2015.12.017.

Morgan, RL; Thayer, KA; Santesso, N; Holloway, AC; Blain, R; Eftim, SE; Goldstone, AE; Ross, P; Ansari, M; Akl, E; Filippini, T; Hansell, A; Meerpohl, JJ; Mustafa, RA; Verbeek, J; Vinceti, M; Whaley, P; Schunemann, HJ; GRADE Working Group. 2019. "A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE." *Environ. Int*. 122:168-184. doi: 10.1016/j.envint.2018.11.004.

National Toxicology Program (NTP). 2015. "OHAT Risk of Bias Rating Tool for Human and Animal Studies." Office of Health Assessment and Translation (OHAT). 37p., January. Accessed at http://ntp.niehs.nih.gov/pubhealth/hat/noms/index-2.html.

National Toxicology Program (NTP). 2019. "Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration." Office of Health Assessment and Translation (OHAT). 101p., March 4. Accessed at https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf.

Organisation for Economic Co-operation and Development (OECD). 2005. "Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment." ENV/JM/MONO(2005)14; OECD Series on Testing and Assessment No. 34. 96p., August 18.

Schneider, K; Schwarz, M; Burkholder, I; Kopp-Schneider, A; Edler, L; Kinsner-Ovaskainen, A; Hartung, T; Hoffmann, S. 2009. "'ToxRTool', a new tool to assess the reliability of toxicological data." *Toxicol. Lett.* 189(2):138-144.

Smith, MT; Guyton, KZ; Gibbons, CF; Fritz, JM; Portier, CJ; Rusyn, I; DeMarini, DM; Caldwell, JC; Kavlock, RJ; Lambert, P; Hecht, SS; Bucher, JR; Stewart, BW; Baan, R; Cogliano, VJ; Straif, K. 2016. "Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis." *Environ. Health Perspect.* 124(6):713-721. doi: 10.1289/ehp.1509912.

US EPA. 2013. "Materials Submitted to the National Research Council Part I: Status of Implementation of Recommendations." Integrated Risk Information System Program. Submitted to National Research Council. 142p., January 30.

US EPA. 2014. "Draft Development Materials for the Integrated Risk Information System (IRIS) Toxicological Review of Inorganic Arsenic [CASRN 7440-38-2]." National Center for Environmental Assessment (NCEA). EPA/630/R-14/101. 731p., April. Accessed at http://www.epa.gov/iris/public meeting/iris_bimonthly-jun2014.

US EPA. 2018a. "Integrated Science Assessment for Particulate Matter (External Review Draft)." National Center for Environmental Assessment (NCEA) - RTP Division. EPA/600/R-18/179. 1881p., October. Accessed at https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=341593.

US EPA. 2018b. "Application of Systematic Review in TSCA Risk Evaluations (Final)." Office of Chemical Safety and Pollution Prevention; US EPA, Office of Pollution Prevention and Toxics. EPA Document # 740-P1-8001. 248p., May. Accessed at https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

von Elm, E; Altman, DG; Egger, M; Pocock, SJ; Gotzsche, PC; Vandenbroucke, JP. 2007a. "Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies." *BMJ* 335(7624):806-808.

von Elm, E; Altman, DG; Egger, M; Pocock, SJ; Gotzsche, PC; Vandenbroucke, JP. 2007b. "STROBE Statement - Checklist of items that should be included in reports of case-control studies." 2p.

von Elm, E; Altman, DG; Egger, M; Pocock, SJ; Gotzsche, PC; Vandenbroucke, JP. 2007c. "STROBE Statement - Items to be included when reporting observational studies in a conference abstract." 1p.

von Elm, E; Altman, DG; Egger, M; Pocock, SJ; Gotzsche, PC; Vandenbroucke, JP. 2007d. "STROBE Statement - Checklist of items that should be included in reports of cross-sectional studies." 2p.

von Elm, E; Altman, DG; Egger, M; Pocock, SJ; Gotzsche, PC; Vandenbroucke, JP. 2007e. "STROBE Statement - Checklist of items that should be included in reports of observational studies." 2p.

Zu, K; Shi, L; Prueitt, RL; Liu, X; Goodman, JE. 2018. "Critical review of long-term ozone exposure and asthma development." *Inhal. Toxicol.* 30(3):99-113. doi: 10.1080/08958378.2018.1455772.

Attachment

Gradient Comments on US EPA's Draft Integrated Science Assessment for Particular Matter

Comments on EPA's Integrated Science Assessment for Particulate Matter (External Review Draft)

Docket ID: EPA-HQ-ORD-2014-0859

Prepared for
American Petroleum Institute
1220 L Street, NW
Washington, DC 20005-4070

December 10, 2018

GRADIENT
www.gradientcorp.com
20 University Road
Cambridge, MA 02138
617-395-5000

Table of Contents

Page

Executive Summary ES-1

1 Introduction 1

2 Overarching Issues in the Draft ISA Evaluation 2
- 2.1 The Draft ISA Lacks a Sufficiently Detailed Systematic Review Protocol 2
- 2.2 Study Quality Is Not Sufficiently Addressed 2
  - 2.2.1 There Are No Quality Criteria for *In Vitro* or Welfare Studies 4
  - 2.2.2 Study Quality Features Impact How Informative a Study Is 5
  - 2.2.3 Study Quality Impacts the Strengths and Limitations of the Collective Body of Evidence 5
  - 2.2.4 Table A-1 Should Contain a Complete List of Study Quality Aspects 6
  - 2.2.5 Quality Aspects Should be Tabulated for Each Individual Study 6
  - 2.2.6 High-quality Studies Should be Considered Key Studies 6
  - 2.2.7 Quality of All Studies Should be Considered 7
- 2.3 Study Relevance Criteria Should Be Explicit 7
- 2.4 Causal Framework Should Be Updated 7
- 2.5 Implications for Causal Determinations of Health and Welfare Effects 20

3 Concentration-response Relationships 21
- 3.1 Epidemiology Studies Do Not Establish a Linear, No-threshold Relationship Between Long-term $PM_{2.5}$ Exposure and Total Mortality or Cardiovascular Mortality/Morbidity 21
  - 3.1.1 Bias and Uncertainty Undermine Mortality Epidemiology Study Results 21
  - 3.1.2 No Evidence for a Linear Relationship with Total Mortality Down to 5-8 $\mu g/m^3$ $PM_{2.5}$ 22
  - 3.1.3 Bias and Uncertainty Undermine Observed Concentration-response Relationships for Cardiovascular Effects 30
  - 3.1.4 The Draft ISA's Evaluation of the Concentration-response Relationships Is Flawed 30
- 3.2 Epidemiology Studies Do Not Establish a Linear, No-threshold Relationship Between Short-term $PM_{2.5}$ Exposure and Total Mortality 39

4 Long-term $PM_{2.5}$ Exposure and Neurological Effects 43
- 4.1 Epidemiology Studies Do Not Indicate Long-term $PM_{2.5}$ Exposure Affects Brain Volume Changes 43
- 4.2 Bias and Uncertainty Undermine Epidemiology Evidence of Long-term $PM_{2.5}$ Exposure and Cognitive Decline/Dementia 44

4.3 The Draft ISA Does Not Consider the Quality and Relevance of the Toxicological Data 45

5 Long-term $PM_{2.5}$ Exposure and Cancer 46
5.1 The Draft ISA Does Not Consider Key Methodological Limitations and Uncertainties in Lung Cancer Epidemiology Studies 46
5.1.1 Epidemiology Studies Did Not Evaluate the Relevant Exposure Window 46
5.1.2 Residual Confounding Likely Undermined the Epidemiology Study Results 48
5.2 The Draft ISA Does Not Fully Consider the Quality and Relevance of the Experimental Studies 52

6 Long-term UFP Exposure and Neurological Effects 53

7 PM and Welfare Effects 55
7.1 Overarching Issues 55
7.2 Visibility Impairment 55
7.3 Effects on Materials 56

8 Recommendations for EPA 57

References 58

List of Tables

Table 2.1 Scientific Considerations for Evaluating the Strength of Inference from Studies on the Health Effects of Particulate Matter

Table 2.2 Institute of Medicine's Recommended Categories for the Level of Evidence for Causation

Table 2.3 EPA's Weight of Evidence for Causal Determination

Table 2.4 EPA's Classification of Evidence for Potential At-risk Factors

Table 3.1 Key Characteristics and Results of Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Total Mortality

Table 3.2 Sources of Bias and Uncertainty in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Total Mortality

Table 3.3 Key Characteristics and Results of Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Morbidity

Table 3.4 Key Characteristics and Results of Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Mortality

Table 3.5 Sources of Bias and Uncertainty in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Morbidity

Table 3.6 Sources of Bias and Uncertainty in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Mortality

Table 3.7 Key Characteristics and Results of Epidemiology Studies of Short-term $PM_{2.5}$ Exposure and Total Mortality

Table 3.8 Sources of Bias and Uncertainty in Epidemiology Studies of Short-term $PM_{2.5}$ Exposure and Total Mortality

Table 5.1 Relative Risks of Lung Cancer Mortality in Male Current *vs*. Never Smokers

Table 5.2 Relative Risks of Lung Cancer Associated with Family History of Lung Cancer

Table 5.3 Adjustment for Selected Confounders in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Lung Cancer

List of Figures

Figure 5.1 Exposure and Follow-up Periods in Studies of Long-term $PM_{2.5}$ Exposure and Lung Cancer Mortality and Incidence

Abbreviations

AOD	Aerosol Optical Depth
BALF	Bronchoalveolar Lavage Fluid
CASAC	Clean Air Scientific Advisory Committee
CI	Confidence Interval
CV	Cardiovascular
EPA	United States Environmental Protection Agency
FOS	Framingham Offspring Study
HSC	Harvard Six City
IARC	International Agency for Research on Cancer
IMPROVE	Interagency Monitoring of Protected Visual Environments
IOM	Institute of Medicine
IRP	Integrated Review Plan
ISA	Integrated Science Assessment
MRI	Magnetic Resonance Imaging
NAAQS	National Ambient Air Quality Standards
NTP	National Toxicology Program
PM	Particulate Matter
REGARDS	REasons for Geographic and Racial Differences in Stroke
RR	Relative Risk
SES	Socioeconomic Status
UFP	Ultrafine Particle
WHIMS	Women's Health Initiatives Memory Study
WoE	Weight of Evidence

Executive Summary

The United States Environmental Protection Agency (EPA) released the Integrated Science Assessment for Particulate Matter (External Review Draft) (herein referred to as the "draft ISA") in October 2018 (US EPA, 2018). The draft ISA indicates that recent epidemiology studies of long-term $PM_{2.5}$ exposure and mortality generally support a linear, no-threshold relationship, with confidence in some studies in the range of 5-8 $\mu g/m^3$. It also concludes that long-term $PM_{2.5}$ exposure is likely causally associated with both nervous system effects and cancer.

In these comments, we discuss a number of issues that we identified regarding the evaluation and analyses in the draft ISA, many of which are consistent with the preliminary comments of Clean Air Scientific Advisory Committee members, Dr. Tony Cox and Dr. Sabine Lange.

Three overarching issues in the draft ISA evaluation of health and welfare effects that undermine its conclusions relate to the systematic review protocol, study quality and relevance, and the causality framework:

Systematic Review Protocol

- **The draft ISA lacks a sufficiently detailed systematic review protocol**. The lack of a sufficiently detailed protocol has led to an evaluation that was not conducted in a systematic, unbiased, or transparent manner. The protocol should include well-developed methods for the literature search strategy; study inclusion and exclusion criteria; a process for data extraction and quality control; specific, prescriptive criteria for evaluating study quality; methods for data analyses; and PM-specific methods for evidence integration and causality determinations.

Study Quality and Relevance

- **Study quality is not sufficiently considered.** While the draft ISA has a list of important study quality aspects for evaluating health effects in Appendix 1 (but no comparable list for welfare studies), it is not complete or sufficiently detailed to allow for a consistent evaluation of individual study quality. Also, only high-quality studies should be considered key studies (*i.e.*, given the most weight in analyses) and the quality of all studies, including new and previously evaluated studies, should be considered for causal determinations.
- **Study quality and relevance impact how informative a study is.** Studies of higher quality should be considered more informative, while those with more limitations should be considered less informative. In addition, criteria that must be met for study results to be considered relevant to the US population as a whole, or to "at-risk" populations, should be explicitly stated.
- **There should be quality criteria for *in vitro* and welfare studies.** EPA cannot determine whether these studies support or call into question a causal association if it has not evaluated study quality.
- **Quality aspects should be tabulated for each individual study.** A systematic review involves reviewing and judging the quality of each individual study in the same manner. This is best accomplished with tables, and this was not done in the draft ISA. For practical reasons, quality

aspects for individual studies should at least be tabulated for key endpoints that inform the National Ambient Air Quality Standards (NAAQS) (*e.g.*, total mortality).

The Causal Framework

- **The causal framework is structured in such a way that biases towards a causal conclusion.** It should be revised to be more balanced.

Because of these overarching issues, the available evidence is not reviewed and integrated in a consistent, systematic way, and consequently, the causal conclusions for health and welfare effects are not warranted based on the weight of scientific evidence.

This is exemplified in the draft ISA's causal determination regarding neurological effects. The draft ISA does not present any systematic study quality evaluation when it summarizes the available literature, nor does it appear to consider study quality when synthesizing the evidence. Epidemiology studies of brain volume, cognitive function, and dementia have considerable limitations and uncertainties that undermined the observed associations between long-term fine particulate matter ($PM_{2.5}$) exposure and neurological endpoints, but this was not considered in the ISA. Findings from animal toxicity studies do not provide evidence for apical endpoints and may have limited relevance to humans.

Similarly, the draft ISA concludes that there is a likely causal relationship between long-term $PM_{2.5}$ exposure and cancer, primarily based on epidemiology studies of lung cancer incidence and mortality, as well as experimental studies that the draft ISA considers to provide evidence for biological plausibility. However, the available epidemiology studies are undermined by considerable methodological limitations; most critically, they do not, or do not adequately, account for latency, smoking, and family history of lung cancer. Also, the draft ISA does not consider the quality or human relevance of the experimental findings. Collectively, the available evidence does not support a likely causal relationship between long-term $PM_{2.5}$ exposure and cancer.

The draft ISA also concludes that there is a likely causal relationship between long-term UFP exposure and neurological effects. Because the draft ISA's evaluation does not take into consideration the quality and human relevance of the animal toxicity studies, the conclusion of a likely causal relationship is not warranted.

With regard to welfare effects, the draft ISA does not acknowledge the uncertainties pertaining to the PM size fractions, which preclude visibility impairment and effects on materials from being used in quantitative risk assessments.

Finally, the draft ISA indicates that recent epidemiology studies of long-term $PM_{2.5}$ exposure and mortality generally support a linear, no-threshold relationship with confidence in some studies in the range of 5-8 $\mu g/m^3$. The draft ISA also discusses evidence from cardiovascular (CV) endpoints as supportive for $PM_{2.5}$ effects at low concentrations. With regard to short-term $PM_{2.5}$ exposure and mortality, the draft ISA indicates that epidemiology studies conducted in the US provide evidence for a linear relationship at concentrations as low as 5 $\mu g/m^3$. The draft ISA does not systematically evaluate the quality of these studies or fully consider potential biases and uncertainties when evaluating the evidence regarding the shape of concentration-response curves. In addition, the draft ISA's evaluation is not systematic or consistent across studies or outcomes. We demonstrate that considerable methodological limitations and uncertainties in these epidemiology studies preclude the observed concentration-response data from being used as a basis to revise the level of NAAQS.

Overall, the draft ISA does not evaluate and integrate the evidence in a transparent, systematic, and unbiased manner. As a result, the causal determinations for health effects are biased towards causation, and undue confidence is placed in observational concentration-response data that contain substantial uncertainties.

1 Introduction

In its last review of particulate matter (PM), the United States Environmental Protection Agency (EPA) concluded that exposure to ambient $PM_{2.5}$ caused or was associated with a wide variety of health effects, and that no threshold had been identified below which these health effects do not occur (US EPA, 2009). EPA released the Integrated Science Assessment for Particulate Matter (External Review Draft) (herein referred to as the "draft ISA") in October 2018 (US EPA, 2018). The draft ISA indicates that recent epidemiology studies of long-term $PM_{2.5}$ exposure and mortality generally support a linear, no-threshold relationship, with confidence in some studies in the range of 5-8 $\mu g/m^3$. It also concludes that long-term $PM_{2.5}$ exposure is likely causally associated with both nervous system effects and cancer, and that long-term exposure to ultrafine particles (UFPs) are likely causally associated with nervous system effects. These three causal determinations are the only ones that changed since the 2009 PM ISA.

As discussed below in Section 2, there are several overarching issues in the draft ISA evaluation that undermine its conclusions, including the lack of a detailed protocol for the entire assessment, the limited evaluation of study quality and relevance, and limitations with the causal framework; all of these issues resulted in individual studies not being reviewed and integrated in a consistent, systematic way, and causal conclusions that are not warranted based on the weight of scientific evidence. Section 3 discusses how concentration-response relationships between $PM_{2.5}$ and mortality/morbidity outcomes observed in epidemiology studies were likely impacted by many biases and uncertainties, both overall and at concentrations in the range of 5-8 $\mu g/m^3$. Sections 4 and 5 discuss the epidemiology and toxicology/mechanistic evidence regarding long-term $PM_{2.5}$ exposure and neurological effects and cancer, respectively. In both cases, the available epidemiology studies are undermined by considerable methodological limitations, and the quality and human relevance of the experimental findings are not considered. Section 6 discusses the toxicology/mechanistic evidence regarding long-term UFP exposure and nervous system effects, and how the quality and human relevance of the experimental findings are not considered in the draft ISA. Section 7 discusses the issues with the evaluation of welfare effects in the draft ISA. Finally, Section 8 provides recommendations for EPA to consider.

Overall, the draft ISA does not evaluate and integrate the evidence in a transparent, systematic, and unbiased manner. As a result, the causal determinations for health effects are biased towards causation, and undue confidence is placed in observational concentration-response data that contain substantial uncertainties.

These comments were prepared with funding from the American Petroleum Institute, but the conclusions and recommendations are based on Gradient's independent review and evaluation.

2 Overarching Issues in the Draft ISA Evaluation

2.1 The Draft ISA Lacks a Sufficiently Detailed Systematic Review Protocol

The draft ISA states, "The U.S. EPA uses a structured and transparent process for evaluating scientific information and determining the causal nature of relationships between air pollution exposures and health effects [details provided in the Preamble to the Integrated Science Assessments (U.S. EPA, 2015)]" (US EPA, 2018, p. P-11). However, the process is not transparent, in that both the draft ISA and Preamble primarily discuss overarching principles. In addition, neither have sufficient detail to ensure that studies are identified and reviewed in a systematic and consistent manner, or integrated in a way that considers study quality and the coherence of results across studies within and across disciplines.

The draft ISA should have included a protocol that includes well-developed methods for the literature search strategy (including keywords and databases to be searched); study inclusion and exclusion criteria; a process for data extraction and quality control; specific, prescriptive criteria for evaluating study quality; methods for data analyses; and PM-specific methods for evidence integration and causality determinations (including plans for assessing data gaps, limitations, and uncertainties in the evidence and the overall systematic review). A detailed protocol would have limited potential biases in the draft ISA and helped ensure that its analyses and results could be reproduced by others. The lack of a sufficiently detailed protocol has led to an evaluation that was not conducted in a systematic, unbiased, or fully transparent manner.

2.2 Study Quality Is Not Sufficiently Addressed

In its comments on the draft Integrated Review Plan (IRP) for the National Ambient Air Quality Standards for Particulate Matter, the Clean Air Scientific Advisory Committee (CASAC, 2016) stated:

> The evaluation of study quality was found to be somewhat vague, and the document would benefit from additional detail and clarification. The IRP describes a "uniform approach" to study quality, but this is not well supported in the text. It is important to be transparent about the process and criteria used in the study quality assessment, and how the quality ratings will be used. For example, it is not clear whether every study will be given some kind of quality rating, who will do the quality assessments, or whether poor quality studies will be rejected from consideration. The studies that will be reviewed for the ISA cross scientific disciplines and include a wide variety of approaches and outcomes. This limits the ability to establish standard quality ratings, as is done in some systematic reviews and meta-analyses. We recommend that the IRP include specific information about the quality assessment process and criteria to be used, acknowledging the limitations and difficulties involved.

As indicated by Dr. Tony Cox and Dr. Sabine Lange in their preliminary comments on the draft ISA (CASAC, 2018), this recommendation was not followed. For example, Dr. Cox stated:

> "Evidence" consisting of published results and conclusions from 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. Conclusions with unknown internal and external validity should not be cited as facts. The Draft ISA does not independently and critically assess the validity of most of the conclusions that it cites or comment on the soundness of the methods that produce them. It leaves unclear exactly how studies were selected, why some apparently valuable ones were not, what makes evidence "sufficient to conclude" something, and how conflicting evidence should be presented and integrated. The final ISA should address each of these points and should be thorough in critically assessing the internal and external validity of the conclusions that it presents and synthesizes.

Dr. Sabine Lange further stated:

> 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 draft ISA has a table with aspects of study quality that should be considered when evaluating scientific evidence on health effects (there is no comparable table for welfare effects). This table is in Appendix 1 of the draft ISA (and reproduced here as Table 2.1). The text immediately preceding the table says (US EPA, 2018, p A-1):

> Table A-1 describes aspects considered in evaluating study quality of controlled human exposure, animal toxicological, and epidemiologic studies. The aspects found in Table A-1 are consistent with current best practices for reporting or evaluating health science data. Additionally, the aspects are compatible with published U.S. EPA guidelines related to cancer, neurotoxicity, reproductive toxicity, and developmental toxicity (U.S. EPA, 2005, 1998, 1996, 1991).
>
> These aspects were not used as a checklist, and judgments were made without considering the results of a study. The presence or absence of particular features in a study did not necessarily lead to the conclusion that a study was less informative or to exclude it from consideration in the ISA. Further, these aspects were not used as criteria for determining causality in the five-level hierarchy. As described in the Preamble, causality determinations were based on judgments of the overall strengths and limitations of the collective body of available studies and the coherence of evidence across scientific disciplines and related outcomes. Table A-1 is not intended to be a complete list of aspects that define a study's ability to inform the relationship between PM and health effects, but it describes the major aspects considered in this ISA to evaluate studies. Where possible, study elements, such as exposure assessment and confounding (*i.e.*, bias due to a relationship with the outcome and correlation with exposures to PM), are considered specifically for PM. Thus, judgments on the ability of a study to inform the relationship between an air pollutant and health can vary depending on the specific pollutant being assessed.

The table is fairly detailed, and the study quality aspects discussed are generally consistent with those considered best practices by several other agencies and organizations. We agree that it is important that these aspects not be used as a checklist, and that they not be used to exclude studies from consideration in

the draft ISA (exclusion should be based solely on relevance). We also agree that study results should not be considered when evaluating study quality.

However, there are some gaps in the quality evaluation system, and the application of the quality criteria has not been performed in a consistent, systematic way. These shortcomings, discussed in more detail below, have resulted in causal conclusions for both health and welfare effects that are not warranted based on the weight of scientific evidence.

2.2.1 There Are No Quality Criteria for *In Vitro* or Welfare Studies

In the Preface, the draft ISA states, "Whereas the ISA tends not to focus the evaluation of the health effects evidence on in vitro studies, for the purposes of examining the mutagenicity of PM *in vitro* systems are discussed because they inform the biological pathways underlying cancer" (US EPA, 2018, p. P-16).

It is not clear how EPA can determine whether an *in vitro* study supports or calls into question a causal association if it has not evaluated study quality. This is particularly true for genotoxicity studies because not all *in vitro* assays of DNA damage predict carcinogenesis. Types of genetic damage associated with cancer involve permanent changes in gene expression, including mutations and structural and numerical chromosome aberrations, but cytotoxic DNA damage will not be sustained in future cell generations (Dearfield *et al*., 2002). Indicator tests are those that evaluate whether a substance can interact with DNA without necessarily causing permanent changes in gene expression, while mutagenicity tests specifically evaluate whether a substance can cause gene mutations or permanent alterations in the structure or number of chromosomes (Eastmond *et al*., 2009; WHO, 2007).

Indicator tests include those that evaluate sister chromatid exchange, DNA strand breaks (such as the comet assay), DNA adducts, and unscheduled DNA synthesis (Eastmond *et al*., 2009; WHO, 2007). Positive results from indicator tests provide suggestive, but not definitive, evidence that a substance is mutagenic. A positive result from a mutagenicity test provides clearer evidence that a substance can cause DNA damage that could potentially lead to cancer. Mutagenicity tests include those that specifically test for mutations (such as the *Salmonella typhimurium* bacterial assay and the mouse lymphoma assay) and those that evaluate effects on chromosomes (such as chromosome aberration and micronucleus assays).

Perhaps more importantly, genotoxicity studies vary considerably in their rates of false positive results. In general, there is a known high rate of false positives with many common substances that do not pose a carcinogenic risk under human exposure conditions (Dearfield and Moore, 2005; Pottenger *et al.*, 2007).

Despite these critical issues, the draft ISA takes all results of genotoxicity assays at face value, without considering the reliability of the available studies. The draft ISA should have considered study quality of all *in vitro* studies, with a particular focus on methods that have been found to lower false positive results, including the use of p53-competent human cells, measures of cytotoxicity based on cell proliferation, quality checks on the source and characterization of the cells used, and tests at reduced maximum concentration (Corvi and Madia, 2017). Several existing study quality evaluation systems are available from which EPA could draw criteria; one of the more well-developed tools is the SciRap tool, which includes a set of criteria for both reporting and methodological quality (in addition to four parameters to evaluate relevance) (Beronius *et al*., 2018).

With regard to welfare effects, there are no specific quality criteria discussed in the draft ISA. The Preamble briefly discusses the importance of using well-established measurement and modeling techniques (US EPA, 2018), but does not provide a comprehensive and detailed set of criteria to fully assess individual study

quality. Detailed quality criteria should be developed for studies of ecological and other welfare effects to allow for a consistent and transparent evaluation of individual study quality.

2.2.2 Study Quality Features Impact How Informative a Study Is

Appendix 1 of the draft ISA (US EPA, 2018) states, "The presence or absence of particular features in a study did not necessarily lead to the conclusion that a study was less informative." Quality and relevance are the only factors that determine whether and to what degree a study is informative. More robust studies should be considered more informative, while those with serious limitations should be considered less informative. For example, if the draft ISA considered a particular statistical model a limitation for one study, it should have considered that study less informative than a study that used a more appropriate model. Similarly, it should have concluded that all studies that used this model were less informative unless there was a reason to conclude otherwise. Also, one particular study strength or limitation could "outweigh" all the others in terms of its impact on the interpretation of results. This critical feature may vary across different endpoints or study designs; however, study quality criteria can be tailored to account for this. EPA should have determined critical features for each type of evidence and outcome *a priori* and applied study quality criteria consistently.

In addition, Section 3 of the draft ISA discusses at length how human exposure to ambient PM can be estimated and how exposure measurement error impacts the interpretation of epidemiology study results. However, there appears to be a disconnect between this section and the health effect evaluations, as exposure measurement techniques and their associated error have not been systematically considered in the evaluation of individual epidemiology studies. Furthermore, exposure assessment is not the only methodological aspect that impacts the validity of epidemiology study results. EPA should systematically evaluate published literature on air pollution epidemiology methodology, regarding not only exposure assessment, but also other methodological aspects such as study design and statistical approaches (several example studies are provided by Dr. Tony Cox in his preliminary comments on the draft ISA; CASAC, 2018). This should then be considered when evaluating study quality and integrating study results.

2.2.3 Study Quality Impacts the Strengths and Limitations of the Collective Body of Evidence

Appendix 1 of the draft ISA (US EPA, 2018) states that "these [quality] aspects were not used as criteria for determining causality in the five-level hierarchy. As described in the Preamble, causality determinations were based on judgments of the overall strengths and limitations of the collective body of available studies and the coherence of evidence across scientific disciplines and related outcomes." The collective body of studies is made up of individual studies. The only way to determine the strengths and limitations of the body of studies is to determine the strengths and limitations of each individual study, and studies that are of higher quality should be weighed more in the causality determination, regardless of results. Furthermore, all bodies of evidence have some level of inconsistency in results across studies, and without considering individual study quality, it is nearly impossible to determine which studies are most likely to reflect the true exposure-response relationship (or lack thereof); as such, those studies should be given the most weight when making conclusions.

In addition, the draft ISA does not sufficiently address study quality when evaluating exposure-response data. While the lack of a thorough, systematic study quality evaluation is an issue for determining causation, it is even more problematic in the context of concentration-response relationships. For causal determinations, studies need to establish the presence of an effect following an exposure, but for concentration-response relationships, studies need to not only establish the presence of an effect, but also the magnitude of an effect in relation to the level of the exposure.

2.2.4 Table A-1 Should Contain a Complete List of Study Quality Aspects

Appendix 1 of the draft ISA (US EPA, 2018) states that "Table A-1 is not intended to be a complete list of aspects that define a study's ability to inform the relationship between PM and health effects." However, the draft ISA *should* have included a full list of every aspect of study quality that EPA used to evaluate how informative each study was for causality determinations. This list should have been developed before the evaluation began, and updated as needed, with the caveat that all updates must be justified and documented. Without a fully comprehensive list, individual studies could be evaluated in a biased and inconsistent manner (*e.g.*, two studies with the same strengths or limitations could carry different weights in the causal analysis).

In addition, the level of detail varies throughout the table. The table should have included enough information so that someone could evaluate the quality of a study in the exact manner as the draft ISA. As it stands, one would not be able to determine how decisions were made for certain studies.

For example, the table should have discussed all of the ways in which PM exposure can be measured, the strengths and limitations of each method, the potential for exposure measurement error, and which methods carry the most weight. There should have also been a discussion of statistical methods used among all studies evaluated and which specific methods are more robust and why for each study design (*e.g.*, whether multiple comparisons have been addressed or whether assumptions in Cox proportional hazard model are appropriate). The table should have addressed specific confounders (*e.g.*, copollutants, socioeconomic status [SES], age, weather) in terms of how they are handled in different studies and their likely impact on results. While some confounders, like age, are universal, others will be specific to the study type, exposure, metric, or outcome (*e.g.*, confounders for neurotoxicity studies will be different than those for cancer studies). For each study type, all known potential confounders should have been listed in the table.

Finally, other factors the draft ISA should have specified in more detail include measurement bias, measurement precision, replicability of observations, data reliability, outliers, and selective outcome reporting.

2.2.5 Quality Aspects Should be Tabulated for Each Individual Study

A systematic review involves reviewing the quality of each individual study in the same manner, and judging the quality of each study in a consistent manner. The best way to do this is by using study quality tables.

Several good examples can be found in Goodman *et al*. (2018) (in Sections S1.2 in each of the three supplements) and Zu *et al*. (2018) (Study Quality Evaluation in Methods and Results and Table 2). These systematic reviews explicitly state the metrics used to determine study quality for each discipline and what constitutes higher and lower quality. Quality considerations for each study are tabulated, with studies in rows and study quality aspects in columns. All metrics with lower scores are highlighted, so that the quality of the literature is clear.

2.2.6 High-quality Studies Should be Considered Key Studies

There is no explicit rationale in the draft ISA regarding the study quality of key studies, or why certain studies are considered key evidence, while others of similar quality are not. Only high-quality studies should be considered key studies, and all studies of similar high quality must be considered and weighed equally. For example, a study with positive results should not be weighed more than a study with null

results if they are both of the same quality. The draft ISA should have included a thorough description of the reasons why specific studies were selected as key evidence and how they relate to other studies that were well conducted but considered as supporting evidence.

2.2.7 Quality of All Studies Should be Considered

The quality of all studies that contribute to the weight of evidence (WoE) needs to be evaluated. The 2009 ISA did not conduct a formal study quality evaluation; thus, all studies included in the 2009 ISA should have been evaluated in the same way as the new studies considered for the current draft ISA. All evidence should have then been re-integrated to determine the causal conclusions, considering the quality of each of the available individual studies, new and old. The draft ISA should have also assessed the quality of studies that do not address the PM-health outcome association specifically (*e.g.*, studies that evaluate "at-risk" factors), because these studies are still fundamental to EPA's decision making.

2.3 Study Relevance Criteria Should Be Explicit

Relevance can be an issue for epidemiology studies (*e.g.*, generalizability or relevance to the US population), but it is always an important consideration for toxicity and mechanistic studies. As discussed by Dr. Tony Cox in his preliminary comments (CASAC, 2018), the draft ISA should explicitly state criteria that must be met for study results to be considered relevant to the US population as a whole, or to "at-risk" populations.

For welfare studies, the Preamble defines ecological effects considered in the ISAs and discusses briefly that studies evaluating effects at or near ambient concentrations and conducted in the US and Canada are considered more relevant (US EPA, 2015a). The draft ISA should include more explicit and complete relevance criteria for welfare studies.

2.4 Causal Framework Should Be Updated

The EPA causal framework for evaluating health effects draws its language from sources across the federal government and scientific community, and particularly relies on an Institute of Medicine (IOM) report titled *Improving the Presumptive Disability Decision-making Process for Veterans* (IOM, 2008). Whereas IOM recommended four categories for the level of evidence for causation (Table 2.2), EPA has five categories for causal relationships (Table 2.3). Based on these categories, the draft ISA determines which health effects will be evaluated in quantitative risk assessments. Notably, the draft ISA uses a different framework (Table 2.4) for classifying effect modifiers (which it calls "at-risk factors") that is much more similar to the IOM framework, although the draft ISA indicates that this framework is based on EPA's causal framework (as shown in Table 2.3).

EPA's causal framework is also ostensibly based on modified Bradford Hill aspects. Both the original and modified Bradford Hill aspects (*i.e.*, strength of association, consistency and coherence, biological plausibility, biological gradient or exposure-response, specificity, temporality of effect, and adversity) are useful tools for evaluating causation; it may be difficult to ascribe observations to causation if these aspects are not met, whereas it may be difficult to ascribe observations to anything other than causation if they are met. In its current form, however, EPA's causal framework is not congruent with the judgments based on the original or modified Bradford Hill aspects. For example, the framework claims to rely heavily on the aspect of consistency across studies in its categorization scheme, but, in practice, EPA does not always fully evaluate consistency or consider other aspects such as coherence, biological plausibility, biological

gradient, and strength of association. In many cases, the draft ISA assumes association indicates causation even when causal modeling may indicate otherwise.

The draft ISA states that evidence is sufficient to conclude a *causal* relationship if "chance, confounding, and other biases [can] be ruled out with reasonable confidence" (US EPA, 2018), yet there is no guidance on what constitutes "reasonable confidence." Based on the current framework, the draft ISA cannot reliably make that determination, because it does not fully explore chance, confounding, and other biases in a consistent manner. The draft ISA suggests that "controlled human exposure studies that demonstrate consistent effects" constitute evidence for a causal relationship (US EPA, 2018), but it should indicate that this is only true if the exposures are at concentrations relevant to ambient exposure and the results are coherent with other lines of evidence. The draft ISA also indicates that "observational studies that cannot be explained by plausible alternatives" constitute evidence for a causal relationship (US EPA, 2018). Yet, the draft ISA does not fully explore alternative explanations for study results. Currently, the draft ISA sets forth a hypothesis (*i.e.*, a criteria pollutant causes a particular health effect) and determines whether the evidence supports that hypothesis. The draft ISA does not, but should have, fully explore whether and to what degree the evidence supports *other* hypotheses (*e.g.*, a confounder, rather than the criteria pollutant, causes a particular health effect). It is only in this manner that alternative hypotheses can truly be ruled out.

The draft ISA states that evidence is sufficient to conclude a *likely causal* relationship if "copollutant exposures are difficult to address and/or other lines of evidence (controlled human exposure, animal, or mode of action information) are limited or inconsistent" or if "animal toxicological evidence from multiple studies from different laboratories demonstrate effects, but limited or no human data are available" (US EPA, 2018). The draft ISA concludes that evidence is *suggestive of a causal relationship* if "at least one high-quality epidemiologic study shows an association with a given health outcome but the results of other studies are inconsistent" or if "a well-conducted toxicological study, such as those conducted in the National Toxicology Program (NTP), shows effects in animal species" (US EPA, 2018).

For making determinations regarding causality, it is important to evaluate all available evidence (positive, null, and negative) in what is referred to as a WoE evaluation. Any WoE evaluation, by definition, involves a consideration of all lines of evidence in a consistent and coherent manner. It is not about resolving all uncertainty; rather, the goal of a WoE evaluation is to determine whether the evidence as a whole supports causation more than it supports a lack of effect. If copollutants cannot be addressed or studies are inconsistent, the WoE may indicate a lack of causality or inadequate evidence to assess causation. If positive effects in high-dose animal studies cannot be related to humans, this does not constitute suggestive evidence; instead, these effects are essentially uninformative regarding causation in humans. Not every study evaluating criteria pollutants is informative for evaluating human health risk, and the draft ISA should not place undue weight on these studies.

It is notable that the EPA causal framework requires only one high-quality study for evidence of a causal relationship to be deemed *suggestive.* Under this definition, high-quality studies that are inconsistent with evidence of an association may exist, but as long as one high-quality study demonstrates an effect, there would still be enough evidence to constitute a suggestive relationship. Instead, all studies should be reviewed using the same criteria, and one should conclude a suggestive causal association only if the WoE indicates that a causal association is more likely than not, based on all the evidence combined. In situations where there are multiple, but inconsistent, high-quality studies, the appropriate conclusion is that the evidence is below equipoise (IOM, 2008).

Finally, evaluating the evidence as a whole means that one should evaluate not only how much evidence can be adduced to support (or to counter) the hypothesized causal effect, but also how separate lines of evidence support (or contradict) one another. It is critical to determine the most likely explanation for

discrepancies across studies by evaluating all of the evidence and not selectively considering evidence that supports or counters a given hypothesis.

Although the frameworks differ slightly, many of the issues noted above also apply to the causal framework for evaluating ecological and other welfare effects. The issues for both health and welfare effects could generally be resolved by updating the draft ISA's categories for causal determination to be more consistent with the IOM framework (on which it was based originally), outlined in Table 2.2. The draft ISA should have evaluated all the evidence in a consistent manner, using well-specified criteria, and determined whether, as a whole, it constitutes evidence for causation or is more likely to be supportive of an alternative hypothesis. EPA should proceed with a risk assessment on a particular health or welfare effect only if the evidence is clearly supportive of causation (*i.e.*, equipoise and above in the IOM framework).

Table 2.1 Scientific Considerations for Evaluating the Strength of Inference from Studies on the Health Effects of Particulate Matter

Study Design	
Controlled Human Exposure	Studies should clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested. Study subjects should be randomly exposed without knowledge of the exposure condition. Preference is given to balanced crossover (repeated measures) or parallel design studies which include control exposures (*e.g.*, to clean filtered air). In crossover studies, a sufficient and specified time between exposure days should be provided to avoid carry over effects from prior exposure days. In parallel design studies, all arms should be matched for individual characteristics such as age, sex, race, anthropometric properties, and health status. In studies evaluating effects of disease, appropriately matched healthy controls are desired for interpretative purposes.
Animal Toxicology	Studies should clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested. Studies should include appropriately matched control exposures (*e.g.*, to clean filtered air, time matched). Studies should use methods to limit differences in baseline characteristics of control and exposure groups. Studies should randomize assignment to exposure groups and where possible conceal allocation to research personnel. Groups should be subjected to identical experimental procedures and conditions; animal care including housing, husbandry, etc. should be identical between groups. Blinding of research personnel to study group may not be possible due to animal welfare and experimental considerations; however, differences in the monitoring or handling of animals in all groups by research personnel should be minimized.
Epidemiology	Inference is stronger for studies that clearly describe the primary and any secondary aims of the study, or specific hypotheses being tested. For short-term exposure, time-series, case crossover, and panel studies are emphasized over cross-sectional studies because they examine temporal correlations and are less prone to confounding by factors that differ between individuals (*e.g.*, SES, age). Panel studies with scripted exposures, in particular, can contribute to inference because they have consistent, well-defined exposure durations across subjects, measure personal ambient pollutant exposures, and measure outcomes at consistent, well-defined lags after exposures. Studies with large sample sizes and conducted over multiple years are considered to produce more reliable results. Additionally, multi-city studies are preferred over single-city studies because they examine associations large diverse geographic areas using a consistent statistical methodology, avoiding the publication bias often associated with single-city studies.[a] If other quality parameters are equal, multicity studies carry more weight than single-city studies because they tend to have larger sample sizes and lower potential for publication bias. For long-term exposure, inference is considered to be stronger for prospective cohort studies and case-control studies nested within a cohort (*e.g.*, for rare diseases) than cross-sectional, other case-control, or ecologic studies. Cohort studies can better inform the temporality of exposure and effect. Other designs can have uncertainty related to the appropriateness of the control group or validity of inference about individuals from group-level data. Study design limitations can bias health effect associations in either direction.

Study Population/Test Model	
Controlled Human Exposure	In general, the subjects recruited into study groups should be similarly matched for age, sex, race, anthropometric properties, and health status. In studies evaluating effects of specific subject characteristics (*e.g.*, disease, genetic polymorphism, etc.), appropriately matched healthy controls are preferred. Relevant characteristics and health status should be reported for each experimental group. Criteria for including and excluding subjects should be clearly indicated. For the examination of populations with an underlying health condition (*e.g.*, asthma), independent, clinical assessment of the health condition is ideal, but self-report of physician diagnosis generally is considered to be reliable for respiratory and cardiovascular disease outcomes.[b] The loss or withdrawal of recruited subjects during the course of a study should be reported. Specific rationale for excluding subject(s) from any portion of a protocol should be explained.
Animal Toxicology	Ideally, studies should report species, strain, substrain, genetic background, age, sex, and weight. Unless data indicate otherwise, all animal species and strains are considered appropriate for evaluating effects of PM exposure. It is preferred that the authors test for effects in both sexes and multiple lifestages, and report the result for each group separately. All animals used in a study should be accounted for, and rationale for exclusion of animals or data should be specified.
Epidemiology	There is greater confidence in results for study populations that are recruited from and representative of the target population. Studies with high participation and low drop-out over time that is not dependent on exposure or health status are considered to have low potential for selection bias. Clearly specified criteria for including and excluding subjects can aid assessment of selection bias. For populations with an underlying health condition, independent, clinical assessment of the health condition is valuable, but self-report of physician diagnosis generally is considered to be reliable for respiratory and cardiovascular diseases.[b] Comparisons of groups with and without an underlying health condition are more informative if groups are from the same source population. Selection bias can influence results in either direction or may not affect the validity of results but rather reduce the generalizability of findings to the target population.
Pollutant	
Controlled Human Exposure	Studies should: (1) include a composite measure of PM (*i.e.*, $PM_{2.5}$, $PM_{10-2.5}$, or ultrafine particles [UFP][c]) or (2) apply some approach (*e.g.*, particle trap or filter) to assess the effects of PM in a complex air pollution mixture (*i.e.*, diesel exhaust, gasoline exhaust, wood smoke).
Animal Toxicology	Studies should: (1) include a composite measure of PM (*i.e.*, $PM_{2.5}$, $PM_{10-2.5}$, or ultrafine particles [UFP][c]) or (2) apply some approach (*e.g.*, particle trap or filter) to assess the effects of PM in a complex air pollution mixture (*i.e.*, diesel exhaust, gasoline exhaust, wood smoke).

Epidemiology	Health effects are evaluated primarily using a composite measure of PM (*i.e.*, $PM_{2.5}$, $PM_{10-2.5}$, or ultrafine particles [UFP][c]) from studies using ambient measurements, model predictions, or a combination of measured and modeled data. Studies of PM components must also include a composite measure of PM. Studies of source-related indicators are also evaluated where the indicator is derived using ambient PM concentrations.
Exposure Assessment or Assignment	
Controlled Human Exposure	For this assessment, the focus is on studies that utilize PM concentrations <2 mg/m^3. Studies that use higher exposure concentrations may provide information relevant to biological plausibility, dosimetry, or inter-species variation. Studies should have well-characterized pollutant concentration, temperature, and relative humidity and/or have measures in place to adequately control the exposure conditions. Preference is given to balanced crossover or parallel design studies which include control exposures (*e.g.*, to clean filtered air). Study subjects should be randomly exposed without knowledge of the exposure condition. Method of exposure (*e.g.*, chamber, facemask, etc.) should be specified and activity level of subjects during exposures should be well characterized.
Animal Toxicology	For this assessment, the focus is on studies that utilize PM concentrations <2 mg/m^3. Studies that use higher exposure concentrations may provide information relevant to biological plausibility, dosimetry, or inter-species variation. Studies should characterize pollutant concentration, temperature, and relative humidity and/or have measures in place to adequately control the exposure conditions. The focus is on inhalation exposure. Non-inhalation exposure experiments (*i.e.*, intratracheal instillation [IT]) are informative for size fractions (*e.g.*, $PM_{10-2.5}$) that cannot penetrate the airway of a study animal and may provide information relevant to biological plausibility and dosimetry. In vitro studies may be included if they provide mechanistic insight or examine similar effects as in vivo studies, but are generally not included. All studies should include exposure control groups (*e.g.*, clean filtered air).

Epidemiology	Of primary relevance are relationships of health effects with the ambient component of PM exposure. However, information about ambient exposure rarely is available for individual subjects; most often, inference is based on ambient concentrations. Studies that compare exposure assessment methods are considered to be particularly informative. Inference is stronger when the duration or lag of the exposure metric corresponds with the time course for physiological changes in the outcome (*e.g.*, up to a few days for symptoms) or latency of disease (*e.g.*, several years for cancer). Given that the spatial variability of PM composite measures varies among size fractions, with more homogeneity for $PM_{2.5}$ than either $PM_{10-2.5}$ or UFP, the need for capturing spatial contrasts is stronger for $PM_{10-2.5}$ or UFP compared with $PM_{2.5}$. Validated measurements, whether averaged across multiple monitors or assigned from the nearest or single available monitor, adequately capture temporal or spatial variation in exposure to $PM_{2.5}$ due to the high correlation between personal exposure and ambient concentration. However, for more spatially heterogeneous $PM_{10-2.5}$ and UFP, the spatial correlation between personal exposure and ambient concentrations is lower. Similarly, PM components show increased spatial variability relative to $PM_{2.5}$. In this case, validated methods that capture the extent of variability for the particular study design (temporal vs. spatial contrasts) and location carry greater weight. Inference based on central site measurements can be adequate if correlated with personal exposures, closely located to study subjects, highly correlated across monitors within a location, used in locations with well-distributed sources, or combined with time-activity information. In studies of short-term exposure, temporal variability of the exposure metric is of primary interest. For all PM size fractions, studies that incorporate time-activity data with personal or microenvironmental monitoring or modeling data may carry greater weight because residential, in-vehicle, and workplace PM exposures may differ in their temporal variability. Results for total personal and indoor PM exposure are other lines of evidence that may inform judgments about causality of PM because inference is based on an individual's microenvironmental exposures and the potential for copollutant confounding may be reduced compared to ambient exposures. Results for total personal exposure can inform understanding of the effects of ambient exposure when well correlated with ambient concentrations. For long-term exposures, methods that well represent within-community spatial variation in individual exposure may be given more weight for spatially-variable ambient $PM_{10-2.5}$ or ultrafine particles. For $PM_{2.5}$, within-community variation in exposure is less important given that $PM_{2.5}$ tends to be more homogeneous. Exposure measurement error often attenuates health effect estimates or increases the imprecision of the association (*i.e.*, width of 95% CIs), particularly associations based on temporal variation in short-term exposure. However, exposure measurement error can bias estimates away from the null in some epidemiologic studies of long-term exposures where the PM size fraction is more spatially heterogeneous (*i.e.*, $PM_{10-2.5}$ or UFP), depending on the locations of the monitor and sources with respect to the study population. To streamline the health effects discussion on studies that are most policy-relevant, for those health categories where the 2009 PM ISA concluded a "causal relationship" the focus is on studies with mean $PM_{2.5}$ concentrations <20 µg/m^3. However, studies that examine a previously identified uncertainty or limitation in the evidence are evaluated even if mean $PM_{2.5}$ concentrations are >20 µg/m^3.

Outcome Assessment/Evaluation	
Controlled Human Exposure	Endpoints should be assessed in the same manner for control and exposure groups (*e.g.*, time after exposure, methods, endpoint evaluator) using valid, reliable methods. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints (*e.g.*, histopathology). For each experiment and each experimental group, including controls, precise details of all procedures carried out should be provided including how, when, and where. Time of the endpoint evaluations is a key consideration that will vary depending on endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions.
Animal Toxicology	Endpoints should be assessed in the same manner for control and exposure groups (*e.g.*, time after exposure, methods, endpoint evaluator) using valid, reliable methods. Blinding of endpoint evaluators is ideal, especially for qualitative endpoints (*e.g.*, histopathology). For each experiment and each experimental group, including controls, precise details of all procedures carried out should be provided including how, when, and where. Time of the endpoint evaluations is a key consideration that will vary depending on endpoint evaluated. Endpoints should be assessed at time points that are appropriate for the research questions.
Epidemiology	Inference is stronger when outcomes are assessed or reported without knowledge of exposure status. Knowledge of exposure status could produce artefactual associations. Confidence is greater when outcomes assessed by interview, self-report, clinical examination, or analysis of biological indicators are defined by consistent criteria and collected by validated, reliable methods. Independent, clinical assessment is valuable for outcomes such as lung function or incidence of disease, but report of physician diagnosis has shown good reliability.[b] When examining short-term exposures, evaluation of the evidence focuses on specific lags based on the evidence presented in individual studies. Specifically, the following hierarchy is used in the process of selecting results from individual studies to assess in the context of results across all studies for a specific health effect or outcome: ▪ Distributed lag models; ▪ Average of multiple days (*e.g.*, 0-2); ▪ If a priori lag days were used by the study authors these are the effect estimates presented; or ▪ If a study focuses on only a series of individual lag days, expert judgment is applied to select the appropriate result to focus on considering the time course for physiologic changes for the health effect or outcome being evaluated. When health effects of long-term exposure are assessed by acute events such as symptoms or hospital admissions, inference is strengthened when results are adjusted for short-term exposure. Validated questionnaires for subjective outcomes such as symptoms are regarded to be reliable,[c] particularly when collected frequently and not subject to long recall. For biological samples, the stability of the compound of interest and the sensitivity and precision of the analytical method is considered. If not based on knowledge of exposure status, errors in outcome assessment tend to bias results toward the null.

Potential Copollutant Confounding	
Controlled Human Exposure	Exposure should be well characterized to evaluate independent effects of PM of various size fractions. Studies should apply some approach (*e.g.*, particle trap or filter) to assess the effects of PM when examining exposures to complex air pollution mixtures (*i.e.*, diesel exhaust, gasoline exhaust, wood smoke).
Animal Toxicology	Exposure should be well characterized to evaluate independent effects of PM of various size fractions. Studies should apply some approach (*e.g.*, particle trap or filter) to assess the effects of PM when examining exposures to complex air pollution mixtures (*i.e.*, diesel exhaust, gasoline exhaust, wood smoke).
Epidemiology	Not accounting for potential copollutant confounding can produce artefactual associations; thus, studies that examine copollutant confounding carry greater weight. The predominant method is copollutant modeling (*i.e.*, two-pollutant models), which is especially informative when correlations are not high. However, when correlations are high ($r > 0.7$), such as those often encountered for UFP and other traffic-related copollutants, copollutant modeling is less informative. Although the use of single-pollutant models to examine the association between PM and a health effect or outcome are informative, ideally studies should also include copollutant analyses. Copollutant confounding is evaluated on an individual study basis considering the extent of correlations observed between the copollutant and PM, and relationships observed with PM and health effects in copollutant models.
Other Potential Confounding Factors[d]	
Controlled Human Exposure	Preference is given to studies utilizing experimental and control groups that are matched for individual level characteristics (*e.g.*, race/ethnicity, sex, body weight, smoking history, age) and time varying factors (*e.g.*, seasonal and diurnal patterns).
Animal Toxicology	Preference is given to studies utilizing experimental and control groups that are matched for individual level characteristics (*e.g.*, strain, sex, body weight, litter size, food and water consumption) and time varying factors (*e.g.*, seasonal and diurnal patterns).
Epidemiology	Factors are considered to be potential confounders if demonstrated in the scientific literature to be related to health effects and correlated with PM. Not accounting for confounders can produce artefactual associations; thus, studies that statistically adjust for multiple factors or control for them in the study design are emphasized. Less weight is placed on studies that adjust for factors that mediate the relationship between PM and health effects, which can bias results toward the null. Confounders vary according to study design, exposure duration, and health effect and may include, but are not limited to the following: Short-term exposure studies: Meteorology, day of week, season, medication use, allergen exposure, and long-term temporal trends. Long-term exposure studies: Socioeconomic status, race, age, medication use, smoking status, stress, noise, and occupational exposures.

Statistical Methodology	
Controlled Human Exposure	Statistical methods should be clearly described and appropriate for the study design and research question (*e.g.*, correction for multiple comparisons). Generally, statistical significance is used to evaluate the findings of controlled human exposure studies. However, consistent trends are also informative. Detection of statistical significance is influenced by a variety of factors including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion; ideally, the sample size should provide adequate power to detect hypothesized effects (*e.g.*, sample sizes less than 3 are considered less informative). Because statistical tests have limitations, consideration is given to both trends in data and reproducibility of results.
Animal Toxicology	Statistical methods should be clearly described and appropriate for the study design and research question (*e.g.*, correction for multiple comparisons). Generally, statistical significance is used to evaluate the findings of animal toxicology studies. However, consistent trends are also informative. Detection of statistical significance is influenced by a variety of factors including, but not limited to, the size of the study, exposure and outcome measurement error, and statistical model specifications. Sample size is not a criterion for exclusion; ideally, the sample size should provide adequate power to detect hypothesized effects (*e.g.*, sample sizes less than 3 are considered less informative). Because statistical tests have limitations, consideration is given to both trends in data and reproducibility of results.
Epidemiology	Multivariable regression models that include potential confounding factors are emphasized. However, multipollutant models (more than two pollutants) are considered to produce too much uncertainty due to copollutant collinearity to be informative. Models with interaction terms aid in the evaluation of potential confounding as well as effect modification. Sensitivity analyses with alternate specifications for potential confounding inform the stability of findings and aid in judgments of the strength of inference from results. In the case of multiple comparisons, consistency in the pattern of association can increase confidence that associations were not found by chance alone. Statistical methods that are appropriate for the power of the study carry greater weight. For example, categorical analyses with small sample sizes can be prone to bias results toward or away from the null. Statistical tests such as *t*-tests and Chi-squared tests are not considered sensitive enough for adequate inferences regarding PM-health effect associations. For all methods, the effect estimate and precision of the estimate (*i.e.*, width of 95% CI) are important considerations rather than statistical significance.

Notes:
CI = Confidence Interval; ISA = Integrated Science Assessment; PM = Particulate Matter; SES = Socioeconomic Status; UFP = Ultrafine Particle.
(a) US EPA (2008, as cited in US EPA, 2018).
(b) Murgia *et al*. (2014); Weakley *et al*. (2013); Yang *et al*. (2011); Heckbert *et al*. (2004); Barr *et al*. (2002); Muhajarine *et al*. (1997); Toren *et al*. (1993); Burney *et al*. (1989), all as cited in US EPA (2018).
(c) UFPs are defined as particles <100 nm in size, but studies often include size fractions larger than 100 nm in the assessment of the relationship between UFP exposure and health effects.
(d) Many factors evaluated as potential confounders can be effect measure modifiers (*e.g.*, season, comorbid health condition) or mediators of health effects related to PM (comorbid health condition). The relationship between an air pollutant and health can vary depending on the specific pollutant being assessed.
Source: Adapted from US EPA (2018, Table A-1).

Table 2.2 Institute of Medicine's Recommended Categories for the Level of Evidence for Causation

Causal Determination	Evidence
Sufficient	The evidence is sufficient to conclude that a causal relationship exists. For example: a) replicated and consistent evidence of an association from several high-quality epidemiologic studies that cannot be explained by plausible noncausal alternatives (*e.g.*, chance, bias, or confounding); or b) evidence of causation from animal studies and mechanistic knowledge; or c) compelling evidence from animal studies and strong mechanistic evidence from studies in exposed humans, consistent with (*i.e.*, not contradicted by) the epidemiologic evidence.
Equipoise and above	The evidence is sufficient to conclude that a causal relationship is at least as likely as not, but not sufficient to conclude that a causal relationship exists. For example: a) evidence of an association from the preponderance of several high-quality epidemiologic studies that cannot be explained by plausible noncausal alternatives (*e.g.*, chance, bias, or confounding) as well as animal evidence and biological knowledge consistent with a causal relationship; or b) strong evidence from animal studies or mechanistic evidence that is not contradicted by human or other evidence.
Below equipoise	The evidence is not sufficient to conclude that a causal relationship is at least as likely as not, or is not sufficient to make a scientifically informed judgment. For example: a) consistent human evidence of an association that is limited by the inability to rule out chance, bias, or confounding with confidence, and weak animal or mechanistic evidence; or b) animal evidence suggestive of a causal relationship, but weak or inconsistent human and mechanistic evidence; or c) mechanistic evidence suggestive of a causal relationship, but weak or inconsistent animal and human evidence; or d) the evidence base is very thin.
Against	The evidence suggests the lack of a causal relationship. For example: a) consistent human evidence of no causal association from multiple studies covering the full range of exposures encountered by humans; or b) animal or mechanistic evidence supportive of a lack of a causal relationship.

Note:
Source: IOM (2008).

Table 2.3 EPA's Weight of Evidence for Causal Determination

Causal Determination	Health Effects	Ecological and Other Welfare Effects
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (*e.g.*, doses or exposures generally within one to two orders of magnitude of recent concentrations). That is, the pollutant has been shown to result in health effects in studies in which chance, confounding, and other biases could be ruled out with reasonable confidence. For example: (1) controlled human exposure studies that demonstrate consistent effects, or (2) 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.	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. Controlled exposure studies (laboratory or small- to medium-scale field studies) provide the strongest evidence for causality, but the scope of inference may be limited. Generally, the determination is based on multiple studies conducted by multiple research groups, and evidence that is considered sufficient to infer a causal relationship is usually obtained from the joint consideration of many lines of evidence that reinforce each other.
Likely to be a causal relationship	Evidence is sufficient to conclude that a causal relationship is likely to exist with relevant pollutant exposures. That is, the pollutant has been shown to result in health effects in studies where results are not explained by chance, confounding, and other biases, but uncertainties remain in the evidence overall. For example: 1) observational studies show an association, but copollutant exposures are difficult to address and/or other lines of evidence (controlled human exposure, animal, or mode of action information) are limited or inconsistent, or (2) 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.	Evidence is sufficient to conclude that there is a likely causal association with relevant pollutant exposures. That is, an association has been observed between the pollutant and the outcome in studies in which chance, confounding, and other biases are minimized but uncertainties remain. For example, field studies show a relationship, but suspected interacting factors cannot be controlled, and other lines of evidence are limited or inconsistent. Generally, the determination is based on multiple studies by multiple research groups.

Causal Determination	Health Effects	Ecological and Other Welfare Effects
Suggestive of a causal relationship	Evidence is suggestive of a causal relationship with relevant pollutant exposures but is limited, and chance, confounding, and other biases cannot be ruled out. For example: (1) when the body of evidence is relatively small, at least one high-quality epidemiologic study shows an association with a given health outcome and/or at least one high-quality toxicological study shows effects relevant to humans in animal species, or (2) when the body of evidence is relatively large, evidence from studies of varying quality is generally supportive but not entirely consistent, and there may be coherence across lines of evidence (*e.g.*, animal studies or mode of action information) to support the determination.	Evidence is suggestive of a causal relationship with relevant pollutant exposures, but chance, confounding, and other biases cannot be ruled out. For example, at least one high-quality study shows an effect, but the results of other studies are inconsistent.
Inadequate to infer a causal relationship	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.	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect.
Not likely to be a causal relationship	Evidence indicates there is no causal relationship with relevant pollutant exposures. Several adequate studies, covering the full range of levels of exposure that human beings are known to encounter and considering at-risk populations and lifestages, are mutually consistent in not showing an effect at any level of exposure.	Evidence indicates there is no causal relationship with relevant pollutant exposures. Several adequate studies examining relationships with relevant exposures are consistent in failing to show an effect at any level of exposure.

Notes:
EPA = United States Environmental Protection Agency.
Source: US EPA (2015b, Table III).

Table 2.4 EPA's Classification of Evidence for Potential At-risk Factors[a]

Classification	Health Effects
Adequate evidence	There is substantial, consistent evidence within a discipline to conclude that a factor results in a population or lifestage being at increased risk of air pollutant-related health effect(s) relative to some reference population or lifestage. Where applicable, this evidence includes coherence across disciplines. Evidence includes multiple high-quality studies.
Suggestive evidence	The collective evidence suggests that a factor results in a population or lifestage being at increased risk of an air pollutant-related health effect(s) relative to some reference population or lifestage, but the evidence is limited due to some inconsistency within a discipline or, where applicable, a lack of coherence across disciplines.
Inadequate evidence	The collective evidence is inadequate to determine if a factor results in a population or lifestage being at increased risk of an air pollutant-related health effect(s) relative to some reference population or lifestage. The available studies are of insufficient quantity, quality, consistency, and/or statistical power to permit a conclusion to be drawn.
Evidence of no effect	There is substantial, consistent evidence within a discipline to conclude that a factor does not result in a population or lifestage being at increased risk of air pollutant-related health effect(s) relative to some reference population or lifestage. Where applicable, the evidence includes coherence across disciplines. Evidence includes multiple high-quality studies.

Notes:
EPA = United States Environmental Protection Agency.
(a) An "at-risk factor" is best described as an effect modifier, which is a technical term defined in epidemiology as a variable that differentially modifies the observed effect of a risk factor on disease status.
Source: US EPA (2015b, Table III).

2.5 Implications for Causal Determinations of Health and Welfare Effects

In his preliminary comments on the draft ISA, Dr. Tony Cox stated:

> [T]he ISA should [e]xplain more explicitly how individual studies and evidence are selected, evaluated, combined or synthesized, and resolved when they conflict, in reaching the ISA's conclusions…
>
> Throughout the ISA, conclusions from cited studies should not be presented as evidence until their internal and external validity have been carefully, critically, and independently evaluated and documented as part of the ISA process. Unwarranted, unsound, and unvalidated conclusions appear to be prevalent in this literature… Therefore it is important for the ISA not to passively repeat and summarize conclusions taken at face value, but to actively engage in critical evaluation and synthesis.

As discussed above, the review process lacks an *a priori* detailed protocol, and as a result, is not systematic and not consistent across studies, endpoints, or disciplines. Study quality is not sufficiently considered when appraising and integrating evidence. In addition, the causal framework employed by EPA is biased towards causality. These limitations call into question the validity of the causal determinations with regard to the health and welfare effects of PM in the draft ISA. The draft ISA's conclusions regarding the causal and likely causal relationships between PM exposures and various health and welfare effects are not based on a systematic and unbiased evaluation.

3 Concentration-response Relationships

The draft ISA indicates that recent epidemiology studies of long-term $PM_{2.5}$ exposure and mortality generally "support a linear, no-threshold relationship, especially at lower ambient $PM_{2.5}$ concentrations, with confidence in some studies in the range of 5-8 μg/m^3" (US EPA, 2018, p. 1-50). The draft ISA also indicates that while most epidemiology studies of long-term $PM_{2.5}$ exposure and cardiovascular (CV) outcomes (*i.e.*, morbidity or mortality) support a linear, no-threshold relationship, some studies suggest a supralinear concentration-response relationship. With regard to short-term $PM_{2.5}$ exposure and mortality, the draft ISA indicates that epidemiology studies conducted in the US provide evidence for a linear relationship at concentrations as low as 5 μg/m^3.

The draft ISA does not systematically evaluate the quality of these studies or fully consider potential biases and uncertainties when evaluating the evidence regarding the shape of concentration-response curves. In addition, the draft ISA's evaluation of concentration-response relationships is not systematic or consistent across outcomes. Below, we discuss in more detail how the observed concentration-response relationships between $PM_{2.5}$ and mortality/morbidity outcomes observed in epidemiology studies were likely impacted by many biases and uncertainties, and thus should not be a basis to set the level of the National Ambient Air Quality Standards (NAAQS). We also point out several issues in the draft ISA's evaluation of this topic; some of these and other issues were discussed by Dr. Tony Cox and Dr. Sabine Lange in their preliminary comments (CASAC, 2018).

3.1 Epidemiology Studies Do Not Establish a Linear, No-threshold Relationship Between Long-term $PM_{2.5}$ Exposure and Total Mortality or Cardiovascular Mortality/Morbidity

3.1.1 Bias and Uncertainty Undermine Mortality Epidemiology Study Results

The draft ISA discusses a number of epidemiology studies that evaluated the concentration-response relationship between long-term $PM_{2.5}$ exposure and total mortality (Section 11.2.4, Table 11-7 in the draft ISA). Here, we present key characteristics and main results of all of these studies in Table 3.1, and sources of bias and uncertainty in Table 3.2.

Based on the study quality considerations outlined in the Preamble for the ISAs (US EPA, 2015a) and Appendix 1 in the draft ISA (US EPA, 2018), we considered several broad categories where biases and uncertainties could arise, including exposure assessment, adjustment for individual-level covariates and ecological covariates, evaluation of copollutants, and statistical analyses. Within each category, we considered various methodological issues that impact study quality and the potential for bias. For example, with regard to exposure assessment, we considered whether a study only used central site monitoring data with low spatial resolutions, whether $PM_{2.5}$ exposure estimates were validated, whether the temporal variation or residential mobility was accounted for, whether $PM_{2.5}$ exposures in multiple time periods were evaluated to identify the most relevant exposure windows, and whether the exposure period appropriately matched the follow-up period for mortality.

As shown in Table 3.2, the studies of long-term $PM_{2.5}$ exposure and total mortality had many methodological limitations that likely led to substantial biases and/or uncertainties in the results. While

most studies used and validated $PM_{2.5}$ exposure estimates at a relatively high spatial resolution, the potential for exposure measurement error was likely high for several other aspects of the exposure assessment. A striking limitation of most of these studies is a mismatch between the $PM_{2.5}$ exposure period and the follow-up period for mortality. For at least some of the participants, the $PM_{2.5}$ exposure periods included time after death, which violates the temporality rule in causality (*i.e.*, the cause has to occur before the effect). In addition, several studies did not account for temporal variation in the $PM_{2.5}$ exposure, using a time-invariant exposure estimate in the analyses. Also, more than half of the studies did not account for residential mobility, likely resulting in considerable exposure measurement error. Most studies did not assess $PM_{2.5}$ in multiple time periods to identify the most relevant exposure window.

Confounding is another major issue in these studies. Although most studies considered a number of individual-level and community-level covariates, residential and unmeasured confounding were likely present. For example, recent studies have shown that both individual and community SES have a considerable impact on mortality (Stringhini *et al*., 2017; Steel *et al*., 2018). Although most epidemiology studies adjusted for some socioeconomic factors at individual and/or community level when evaluating the concentration-response relationship between long-term $PM_{2.5}$ exposure and mortality, these socioeconomic factors were measured crudely and likely did not entirely account for the effects of individual and community SES on mortality, thus residual confounding by SES is likely. In addition, few studies accounted for individual smoking, diet, and exercise, or community-level confounders such as access to and quality of health care and violence. The lack of robust adjustment for these factors significantly increased the uncertainty in the study results.

In addition, most studies did not assess or adjust for copollutants; thus the observed concentration-response relationships in these studies may not reflect the independent effects of $PM_{2.5}$.

With regard to statistical analyses, none of the studies accounted for multiple comparisons. Most studies did not test the assumptions of the statistical models used, statistically test nonlinearity, or specifically assess the presence of a potential threshold. When studies used natural splines to examine the shape of the concentration-response curves, the curves were sensitive to the degree of freedom chosen, indicating the results were not robust.

Collectively, the epidemiology studies of long-term $PM_{2.5}$ and total mortality suffered from considerable methodological limitations which likely had substantial impact on the validity of the study results. These studies are not sufficiently robust to establish a linear, no-threshold concentration-response relationship.

3.1.2 No Evidence for a Linear Relationship with Total Mortality Down to 5-8 $\mu g/m^3$ $PM_{2.5}$

The draft ISA indicates that Lepeule *et al*. (2012), Shi *et al*. (2016), and Di *et al*. (2017a) "observed linear, no-threshold concentration-response relationships for total (nonaccidental) mortality, with confidence in the relationship down to a concentration of 8, 5, and 6 $\mu g/m^3$, respectively." However, each of these studies, as presented in Tables 3.1 and 3.2, suffered from several key methodological limitations that considerably increased the uncertainty in the study results and likely undermined the validity of the observed concentration-response relationships.

Lepeule *et al*. (2012) conducted an updated analysis of the Harvard Six City (HSC) cohort with mortality follow-up from 1974 to 2009. To estimate individual $PM_{2.5}$ exposures, Lepeule *et al*. (2012) relied on one fixed-site monitor in each of the six cities from 1979 to 1986-1988, then estimated $PM_{2.5}$ concentrations from monitored PM_{10} data and visibility data between 1986-1988 and 1998, and finally used direct measurements of $PM_{2.5}$ by EPA monitors. This process considerably increased the uncertainty in the $PM_{2.5}$ exposure estimates.

Another critical limitation of Lepeule *et al*. (2012) is confounding. Only several individual-level covariates (including smoking status and pack-years) were adjusted for in the statistical analyses, with education level being the only socioeconomic-related measure. Residual confounding by socioeconomic factors and unmeasured confounding (*e.g*., diet and physical activity) were likely present. In addition, Lepeule *et al*. (2012) did not adjust for any community-level covariates or copollutants.

Exposure measurement error and confounding severely undermine the observed concentration-response relationship; thus Lepeule *et al*. (2012) did not establish a linear, no-threshold concentration-response relationship for total mortality, with confidence in the relationship down to a concentration of 8 $\mu g/m^3$.

The analysis by Shi *et al*. (2016) was conducted among Medicare enrollees in the New England area in the US from 2003 to 2008. While the authors used validated models to estimate the 12-month average $PM_{2.5}$ concentrations prior to death or censoring, the validity of the $PM_{2.5}$ estimates was limited by the quality of the input variables such as the Aerosol Optical Depth (AOD) data, as satellite-based AOD measurements can be biased by unresolved cloud, water vapor, and smoke. Because Medicare records do not provide information on address changes, the authors had to assume that subjects remained at the same address for the duration of the study period. Also, considering the potential mechanisms underlying the $PM_{2.5}$ effect on mortality, the 12-month period prior to death likely was not the relevant exposure window. In addition, Shi *et al*. (2016) did not exclude deaths from unnatural causes, which likely biased the results. Finally, no individual-level confounders were adjusted for in the analyses, which severely undermined the validity of the observed concentration-response relationship.

Di *et al*. (2017a) evaluated the relationship between long-term $PM_{2.5}$ exposure and total mortality in Medicare enrollees in the continental US from 2000 to 2012. They used a different model than Shi *et al*. (2016) to estimate $PM_{2.5}$ concentrations. Although this model was validated and more flexible regarding complex nonlinear relationships, it still depended on the same input variables as the exposure model used by Shi *et al*. (2016). Thus, the validity of $PM_{2.5}$ estimates was still impacted by the issues discussed above with these input data. In addition, because Medicare records were used, residential mobility was not accounted for and deaths from unnatural causes were not excluded, resulting in errors in exposure and outcome assessments. The annual $PM_{2.5}$ concentration in the year prior to death or censoring was evaluated in the concentration-response analysis, which likely was not the relevant exposure window, as discussed above. Regarding the adjustment for confounders, while Di *et al*. (2017a) included several individual-level covariates, important confounders such as smoking and BMI were not available for the Medicare cohort.

In light of these methodological limitations, the concentration-response relationships reported by Shi *et al*. (2016) and Di *et al*. (2017a) are not sufficiently robust in general, and do not establish a linear, no-threshold relationship for total mortality down to $PM_{2.5}$ concentrations of 5-6 $\mu g/m^3$.

Table 3.1 Key Characteristics and Results of Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Total Mortality

Characteristics and Results		Crouse *et al.* (2012)		Crouse *et al.* (2015)		Villeneuve *et al* . (2015)	
Location		Canada		Canada		Canada	
Study Type		Cohort		Cohort		Cohort	
Source Population		CanCHEC		CanCHEC		CNBSS	
Sample Size		2.1 million		2.5 million		89,248	
Study Period		1991-2001 (11-year)		1991-2006 (16-year)		1980-2005 (26-year)	
Exposure	Metric	Annual mean $PM_{2.5}$		Median annual $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	2001-2006		1998-2006		1998-2006	
	Exposure windows	Single		Single		Single	
	Spatial scale	10 km grid		10 km grid		10 km grid	
	Temporal variation	Time-invariant		Time-invariant		Time-invariant	
	Residential mobility	Not considered		Considered		Not considered	
	Source	Satellite-based estimates		Satellite-based estimates		Satellite-based estimates, GEOS-chem models	
	Level (μg/m^3)	Mean ± SD	8.7 ± 3.9	IQR	5.8	Median (range)	9.1 (1.3, 17.6)
Outcome	Endpoint	Non-accidental mortality		Non-accidental mortality		Non-accidental mortality	
	Source	Canadian Mortality Database		Canadian Mortality Database		Canadian Mortality Database	
Individual Covariates	Variables	Age, sex, aboriginal ancestry, visible minority, marital status, highest level of education, employment status, occupational classification		Age, sex, aboriginal ancestry, visible minority status, education, marital status, immigrant status, employment status, occupational classification, quintiles of household income		Age at entry, occupation, marital status, education, BMI, cigarette pack-years	
	Source	Questionnaire at enrollment		Questionnaire at enrollment		Questionnaire at enrollment	
Ecological Covariates	Variables	Unemployed adults (%), adults without high school diplomas (%), subjects in the lowest income quintile (%), population size of home		Immigrants (%), adults without high school diplomas (%), subjects in the lowest income quintile (%)		Mean income, with high school education (%), low income households (%), unemployment rate	
	Source	Canada Census (1991)		Canada Census (1991, 1996, 2001, 2006)		Canada Census (1991)	
Copollutants	Variable	None		Warm-season	Annual mean NO_2	None	
	Time period			2002-2009	2006		
	Spatial scale			21 km grid	Surface		
	Temporal scale			Time-invariant	Time-invariant		
	Source			Canadian air quality forecast models and monitors	LUR		
Linear Association Analyses	Exposure contrast	Per 10 μg/m^3		Per mean-5th percentile increment (5 μg/m^3)		Per 10 μg/m^3	
	Statistical models	Standard Cox models	Random-effects Cox models	Single-pollutant Cox models	Multi-pollutant Cox models	Standard Cox models	
	All-cause/Non-accidental	1.15 (1.13, 1.16)	1.10 (1.05, 1.15)	1.035 (1.029, 1.041)	1.011 (1.003, 1.020)	1.12 (1.04, 1.19)	
C-R Analyses	Outcome	Non-accidental mortality		Non-accidental mortality		Non-accidental mortality	
	Splines (*df*)	Natural splines (≤4 *df*)		Natural splines (2 *df*)		Natural splines (3 *df*)	
	Nonlinearity statistically tested	Tested, p-value not reported		$p < 0.0001$		Not tested	
	Observed/reported shape of the C-R curve	Linear		Supralinear (*i.e.* , larger changes in risk for low concentrations)		Nonlinear V-shaped	
Threshold Analyses	Threshold model	Not conducted		Not conducted		Hockey-stick linear splines (assuming no association below the threshold), starting from 2 to 14 μg/m^3 at 1-μg/m^3 increments	
	Threshold identification					Minimizing the -2*log-likelihood	
	Outcome					Non-accidental mortality	
	Threshold estimate					11 μg/m^3 ($p = 0.004$)	

Characteristics and Results		Chen *et al.* (2016)		Pinault *et al.* (2016)		Wong *et al.* (2015)	
Location		Canada		Canada		China	
Study Type		Cohort		Cohort		Cohort	
Source Population		EFFECT		CCHS		Elderly Health Centre Study	
Sample Size		8,873		299,500		66,820	
Study Period		1999-2011 (13-year)		2000-2011 (12-year)		1998-2011 (13-year)	
Exposure	Metric	Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	2001-2010		1998-2010		2000-2011	
	Exposure windows	Single		Single		Single	
	Spatial scale	10 km grid		1 km grid		1 km grid	
	Temporal variation	Time-invariant		3-year moving average		Time-invariant	
	Residential mobility	Considered		Not considered		Not considered	
	Source	Satellite-based estimates, GEOS-chem models		Satellite-based estimates, GEOS-chem models		Satellite-based estimates	
	Level (μg/m^3)	Mean (range)	10.7 (2.2-16.5)	Mean ± SD	6.3 ± 2.5	Median (range)	35.3 (26.4-44.6)[c]
Outcome	Endpoint	Non-accidental mortality		Non-accidental mortality		Non-accidental mortality	
	Source	Ontario Registrar General's Death Database		Canadian Mortality Database		Death Registry	
Individual Covariates	Variables	Age, sex, region, marital status, employment, major cardiac risk factors (including smoking status, family history of coronary artery disease, diabetes, hyperlipidemia, hypertension, stroke, previous AMI, and previous percutaneous coronary intervention), AMI type (STEMI/Non-STEMI), GRACE risk score, acute pulmonary edema, in-hospital care, medications, select comorbidities		Age, sex, immigrant status, visible minority status, aboriginal status, marital status, income adequacy quintile, educational attainment, employment status, smoking status, alcohol consumption, fruit and vegetable consumption, BMI		Age, sex, BMI, smoking status, physical exercise, education, monthly expenses	
	Source	Medical records		CCHS survey		In-person interviews	
Ecological Covariates	Variables	Subjects aged 15+ without high school diplomas (%), unemployment rate (%), mean household income		Recent immigrants (%), completed high school (%), low income household (%)		TPU-level: subjects aged 65+ (%), subjects with > secondary education (%), average monthly income; District-level: smokers aged 15+ years from 1998 to 2011 (%)	
	Source	Canada Census (2001)		Canada Census (2001 or 2006)		Hong Kong Census (2001)	
Copollutants	Variable	None		None		None	
	Time period						
	Spatial scale						
	Temporal scale						
	Source						
Linear Association Analyses	Exposure contrast	Per 10 μg/m^3		Per 10 μg/m^3		Per 10 μg/m^3	
	Statistical models	Random-effects Cox models		Standard Cox models		Standard Cox models	
	All-cause/Non-accidental	1.22 (1.03, 1.45)		1.261 (1.190, 1.336)		1.14 (1.07, 1.22)	
C-R Analyses	Outcome	Non-accidental mortality		Non-accidental mortality		Non-accidental mortality	
	Splines (*df*)	Natural splines (≤ 4 *df*)		Splines		Natural splines (*df* NR)	
	Nonlinearity statistically tested	Tested, p-value not reported		Not tested		p-value: 0.772	
	Observed/reported shape of the C-R curve	Linear		Not specified; appears supralinear		Linear	
Threshold Analyses	Threshold model	Not conducted		Hockey-stick linear splines (assuming no association below the threshold), starting from 1 to 10 μg/m^3 at 1-μg/m^3 increments		Not conducted	
	Threshold identification			Minimizing the -2*log-likelihood			
	Outcome			Non-accidental mortality			
	Threshold estimate			0 μg/m^3 (+95% CI: 4.5 μg/m^3)			

Characteristics and Results		Beelen *et al.* (2014)			Cesaroni *et al.* (2013)	
Location		Europe (13 countries)			Italy	
Study Type		Meta analysis of 22 Cohorts			Cohort	
Source Population		ESCAPE			RoLS	
Sample Size		367,251			1.2 million	
Study Period		1985-2010			2001-2010 (9-year)	
Exposure	Metric	Annual mean $PM_{2.5}$			Annual mean $PM_{2.5}$	
	Exposure period	2008-2011			2005	
	Exposure windows	Single			Single	
	Spatial scale	Surface			1 km grid	
	Temporal variation	Time-invariant			Time-invariant	
	Residential mobility	Not considered			Considered	
	Source	LUR			Eulerian dispersion model	
	Level (μg/m^3)	Range of means	6.6-31		Mean ± SD	23 ± 4.4
Outcome	Endpoint	Non-accidental mortality			Non-accidental mortality	
	Source	Mortality registries			Rome Municipal Register	
Individual Covariates	Variables	Age, sex, smoking status, smoking intensity, smoking duration, environmental tobacco smoke, fruit intake, vegetables intake, alcohol consumption, BMI, educational level, occupational class, employment status, and marital status			Age, sex, marital status, place of birth, education, occupation	
	Source	Questionnaire at enrollment			Rome Municipal Register	
Ecological Covariates	Variables	SES variables (mostly mean income of the neighborhood or municipality)			Census-block socioeconomic position index (derived based on a factor analysis including education, occupation, house ownership, family composition, crowding, and immigrant status)	
	Source	NR			Rome Census (2001) ascertained from Rome Municipal Register	
Copollutants	Variable	Annual mean NO_2	Annual mean PM_{coarse}		Annual mean NO_2	
	Time period	2008-2011	2008-2011		2007	
	Spatial scale	Surface	Surface		Surface	
	Temporal scale	Time-invariant	Time-invariant		Time-invariant	
	Source	LUR	LUR		LUR	
Linear Association Analyses	Exposure contrast	Per 5 μg/m^3			Per 10 μg/m^3	
	Statistical models	Cox models and meta analysis			Single-pollutant Cox models	Multi-pollutant Cox models
		Single-pollutant	Two-pollutant (adjust for NO_2)	Two-pollutant (adjust for PM_{coarse})		
	All-cause/Non-accidental	1.07 (1.02, 1.13)	1.06 (0.98, 1.15)	1.07 (1.01, 1.14)	1.04 (1.03, 1.05)	1.01 (0.99, 1.02)
C-R Analyses	Outcome	Non-accidental mortality			Non-accidental mortality	
	Splines (*df*)	Natural splines (2 *df*)			Natural splines (≤ 4 *df*)	
	Nonlinearity statistically tested	p-value range: 0.03-0.95[f]			Tested, p-value not reported	
	Observed/reported shape of the C-R curve	Linear			Linear	
Threshold Analyses	Threshold model	Restricting analyses to subjects with exposure levels below pre-specified thresholds (e.g., 25, 20, 15, 10 μg/m^3)			Not conducted	
	Threshold identification	Slope became insignificant				
	Outcome	Non-accidental mortality				
	Threshold estimate	Between 15 and 20 μg/m^3				

Characteristics and Results		Thurston *et al.* (2016)		Di *et al.* (2017a)		
Location		US		US		
Study Type		Cohort		Cohort		
Source Population		NIH-AARP		Medicare population		
Sample Size		517,041		61 million		
Study Period		2000-2009 (10-year)		2000-2012 (13-year)		
Exposure	Metric	Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$		
	Exposure period	1999-2008		2000-2012		
	Exposure windows	Single		Single		
	Spatial scale	Surface		1 km grid		
	Temporal variation	Time-invariant (time-independent); Yearly (1-year lag) (time-dependent)		Yearly		
	Residential mobility	Considered		Considered		
	Source	LUR, BME		Satellite-based estimates, neural network, CTM		
	Level (μg/m^3)	Range	2.9-28	Mean	11	
Outcome	Endpoint	All-cause mortality		All-cause mortality		
	Source	Social Security Administration Death Master File		Centers for Medicare and Medicaid services		
Individual Covariates	Variables	Age, sex, race, education, marital status, BMI, alcohol consumption, smoking history, region indicator		Age at study entry, sex, race, eligible for Medicaid, region indicator		
	Source	Questionnaire at enrollment		Centers for Medicare and Medicaid services		
Ecological Covariates	Variables	Median income, ≤ high school education (%)		Hispanic (%), Black (%), Median household income, median value of housing, below poverty level (%), did not complete high school (%), owner-occupied housing (%), population density, low-density lipoprotein level measured (%), Glycated hemoglobin level measured (%), ≥1 Ambulatory visits (%), county-level BMI, ever smoked		
	Source	US Census (2000)		US Census (2000, 2010), American Community Survey (2005-2012), CDC BRFSS		
Copollutants	Variable	Annual mean O_3		Warm-season mean O_3		
	Time period	2000		2000-2012		
	Spatial scale	PMSA-level		1 km grid		
	Temporal scale	Time-invariant		Yearly		
	Source	Fixed-site monitors		Satellite-based estimates, neural network, CTM		
Linear Association Analyses	Exposure contrast	Per 10 μg/m^3		Per 10 μg/m^3		
	Statistical models	Standard Cox models		Single pollutant Cox models	Two-pollutant Cox models	
		Time-dependent exposure	Time-independent exposure		Full cohort	Low-exposure
	All-cause/Non-accidental	1.03 (0.99, 1.05)	1.03 (1.00, 1.05)	1.084 (1.081, 1.086)	1.073 (1.071, 1.075)	1.136 (1.131, 1.141)
C-R Analyses	Outcome	All-cause mortality		All-cause mortality		
	Splines (*df*)	Natural splines (≤ 4 *df*)		Penalized splines		
	Nonlinearity statistically tested	Not tested		Not tested		
	Observed/reported shape of the C-R curve	Authors specify monotonically increasing; appears non-linear		Linear with no signal of a threshold down to 5 μg/m^3		
Threshold Analyses	Threshold model	Not conducted		Not conducted		
	Threshold identification					
	Outcome					
	Threshold estimate					

Characteristics and Results		Lepeule *et al.* (2012)		Hart *et al.* (2015)		Shi *et al.* (2016)	
Location		US		US		US	
Study Type		Cohort		Cohort		Ecologic	
Source Population		HSC		NHS		Medicare population	
Sample Size		8,096		108,767		551,024	
Study Period		1974-2009 (36-year)		2000-2006 (6-year)		2003-2008 (6-year)	
Exposure	Metric	Annual mean $PM_{2.5}$		Monthly mean $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	1974-2009		2000-2006		2003-2008	
	Exposure windows	Multiple		Single		Single	
	Spatial scale	City-level		Surface		1 km grid	
	Temporal variation	N-year moving average[a]		12-month moving average		1-year moving average	
	Residential mobility	Not considered		Considered		Not considered	
	Source	Fixed-site monitors[b]		Spatio-temporal models		Satellite-based estimates	
	Level (μg/m^3)	Mean	15.9	Mean ± SD	12 ± 2.8	Mean ± SD	8.12 ± 2.28
Outcome	Endpoint	All-cause mortality		Non-accidental mortality		All-cause mortality	
	Source	National Death Index		State vital statistics records, National Death Index		Centers for Medicare and Medicaid services	
Individual Covariates	Variables	Age, sex, BMI, education, smoking status, cumulative smoking (pack-years)		Age, race, physical activity, BMI, hypercholesterolemia, family history of MI, smoking pack-years, current smoking status, diet, education, parents' occupation, marital status, husband's education		None	
	Source	Questionnaire at enrollment		Questionnaires		NA	
Ecological Covariates	Variables	None		Median income, median house value		Zip-code level: race, education, median household income; county-level: every day smoking prevalence	
	Source	NA		US Census (2000)		US Census (2000), CDC BRFSS	
Copollutants	Variable	None		None		None	
	Time period						
	Spatial scale						
	Temporal scale						
	Source						
Linear Association Analyses	Exposure contrast	Per 10 μg/m^3		Per 10 μg/m^3		Per 10 μg/m^3	
	Statistical models	Standard Cox models		Standard Cox models		Time series analyzed	
				Spatio-temporal models	Nearest monitors	Full cohort	Low-exposure[e]
	All-cause/Non-accidental	1.14 (1.07, 1.22)		1.18 (1.02, 1.36)	1.22 (1.02, 1.45)	7.52 ± 5.73[d]	9.28 ± 8.88[d]
C-R Analyses	Outcome	All-cause mortality		Non-accidental mortality		All-cause mortality	
	Splines (*df*)	Penalized splines		Restricted cubic splines		Penalized splines	
	Nonlinearity statistically tested	p-value range: 0.24-0.43		Not tested		Not tested	
	Observed/reported shape of the C-R curve	Linear		Not specified; appears linear		Linear down to 6 μg/m^3, and attenuated below 6 μg/m^3	
Threshold Analyses	Threshold model	Not conducted		Not conducted		Not conducted	
	Threshold identification						
	Outcome						
	Threshold estimate						

Notes:

AMI = Acute Myocardial Infarction; BME = Bayesian Maximum Entropy; BMI = Body Mass Index; CI = Confidence Interval; COPD = Chronic Obstructive Pulmonary Disease; CPS-II = Cancer Prevention Study II; C-R = Concentration-response; CTM = Chemical Transport Model; *df* = Degrees of Freedom; IQR = Interquartile Range; LUR = Land-use Regression; MI = Myocardial Infarction; NA = Not Applicable; NO_2 = Nitrogen Dioxide; NR = Not Reported; O_3 = Ozone; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter; PMSA = Primary Metropolitan Statistical Area; ppb = Parts Per Billion; SD = Standard Deviation; SES = Socioeconomic Status.

(a) $PM_{2.5}$ moving average was 1 year before death or censor for all-cause deaths, 1-3 years for cardiovascular and lung cancer deaths, and 1-5 years for COPD deaths.

(b) $PM_{2.5}$ estimates were obtained from centrally located monitors from 1974 to 1986-1988, depending on the city; EPA PM_{10} monitors from the end of monitoring until 1998; and EPA $PM_{2.5}$ monitors from 1999 to 2009.

(c) Mean ± SD of $PM_{2.5}$ quartiles: Q1: 32.6 ± 1.03; Q2: 34.6 ± 0.43; Q3: 36.2 ± 0.53; Q4: 38.8 ± 1.34.

(d) Results presented as percent increase in mortality.

(e) The analysis was restricted only to person time with chronic $PM_{2.5}$ < 10 μg/m^3.

(f) The p-value for the Norway cohort was significant (0.03); all other cohort p-values were > 0.05.

Table 3.2 Sources of Bias and Uncertainty in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Total Mortality This table summarizes several broad methodological categories where biases and uncertainties could arise in estimated concentration-response relationships between long-term $PM_{2.5}$ exposure and total mortality, including exposure assessment, individual-level covariates, ecological covariates, evaluation of copollutants, and statistical analyses. Red shading indicates the potential for bias and/or the presence of uncertainty with regard to a specific methodological characteristics, but does not reflect the magnitude of such a bias/uncertainty on study results. Unshaded cells indicate there are no apparent biases/uncertainties. For example, Crouse *et al.* (2012) did not account for temporal variation when assessing $PM_{2.5}$ exposure, thus the red shading with an "X" reflects the potential for bias and the presence of uncertainty in this aspect. In addition, Crouse *et al.* (2012) did not report how information on covariates was collected, therefore, red shading with an "NR" indicates the potential for bias and the presence of uncertainty with regard to information bias.

Sources of Bias and Uncertainty		Crouse *et al.* (2012)	Crouse *et al.* (2015)	Villeneuve *et al.* (2015)	Chen *et al.* (2016)	Pinault *et al.* (2016)	Wong *et al.* (2015)	Beelen *et al.* (2014)	Cesaroni *et al.* (2013)	Thurston *et al.* (2016)	Di *et al.* (2017a)	Lepeule *et al.* (2012)	Hart *et al.* (2015)	Shi *et al.* (2016)
$PM_{2.5}$ Exposure Assessment	Central site monitoring (low spatial resolution)											X		
	No validation for $PM_{2.5}$ data						X			X				
	Temporal variation not accounted for	X	X	X	X			X	X					
	Residential mobility not accounted for	X		X		X	X	X				X		X
	No evaluation on multiple exposure windows	X	X	X	X	X	X	X	X	X	X		X	X
	Personal activities not accounted for (*e.g.,* time spent indoors)	X	X	X	X	X	X	X	X	X	X	X	X	X
	Mismatch of $PM_{2.5}$ exposure window and mortality	X	X	X	X		X	X	X		X			
Individual Covariates	No adjustment of individual covariates													X
	Information bias (*e.g.,* self-reported covariates)	X	X	X		X	X	X		X		X	X	
	Temporal variation not accounted for	X	X	X	X	X	X	X	X	X		X		
	Unmeasured confounding (*e.g.,* pre-existing conditions)	X	X	X	X	X	X	X	X	X	X	X	X	X
Ecological Covariates	No adjustment of ecological covariates											X		
	Temporal variation not accounted for	X		X	X	X	X	X	X	X			X	X
	Residential mobility not accounted for	X		X		X	X	X						X
	Unmeasured confounding (*e.g.,* access to health care, violence)	X	X	X	X	X	X	X	X	X	X	X	X	X
Evaluation of Copollutants	No adjustment of copollutants	X		X	X	X	X					X	X	X
	Central site monitoring (low spatial resolution)									X				
	No validation for copollutants data									X				
	Temporal variation not accounted for		X					X	X	X				
	Residential mobility not accounted for							X		X				
	Personal activities not accounted for (*e.g.,* time spent indoors)		X					X	X	X	X			
	Collinearity/nonlinear relationship with $PM_{2.5}$ not addressed/accounted for									X				
	Mismatch of copollutants window and mortality		X					X	X	X	X			
Statistical Analyses	Model assumptions not tested/relaxed		X	X		X	X	X	X	X	X		X	X
	C-R curves sensitive to *df* (natural splines)	X	X	X	X	NR	X	X	X	X				
	Nonlinearity not assessed statistically			X		X				X	X		X	X
	Threshold not assessed	X	X		X		X		X	X	X	X	X	X

Notes:
C-R = Concentration-response; df = Degrees of Freedom; NR = Not Reported; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter.

3.1.3 Bias and Uncertainty Undermine Observed Concentration-response Relationships for Cardiovascular Effects

The draft ISA discusses a number of epidemiology studies of CV morbidity (Section 6.2.16, Table 6-51 of the draft ISA) and mortality (Section 6.2.16, Table 6-52 of the draft ISA) that evaluated the concentration-response relationships with long-term $PM_{2.5}$ exposure. Similar to the total mortality studies, we tabulated key characteristics and main results of these studies in Tables 3.3 (morbidity) and 3.4 (mortality), and sources of bias and uncertainty in Tables 3.5 (morbidity) and 3.6 (mortality).

As demonstrated in Tables 3.5 and 3.6, studies of CV morbidity and mortality had similar methodological limitations as the total mortality studies, with the most bias and uncertainty in the exposure assessments and confounding adjustment. Most studies had mismatched exposure and follow-up periods, did not account for time variation in PM exposures, or did not try to identify the most relevant exposure windows. Exposure measurement error in these studies were likely substantial. Confounding at the individual level (*e.g.*, physical activity, SES) and/or the community level (*e.g.*, access to and quality of health care, violence) was also a major issue in these studies. In addition, the statistical analyses were generally insufficient to establish the shape of the concentration-response curves because, as indicated in the draft ISA, most studies did not conduct a thorough evaluation of alternatives to linearity.

In her comments on the draft ISA (CASAC, 2018), Dr. Sabine Lange stated:

> This combination of data makes it difficult to draw any conclusions about the shape of the C-R function [for CV morbidity]. 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)." She also stated, "This section completely skates over the negative evidence and EPA's own 'inconclusive' determinations for different endpoints.

Overall, in light of study limitations, epidemiology studies do not provide strong evidence for any specific concentration-response relationship between long-term $PM_{2.5}$ exposure and CV effects, particularly in the low $PM_{2.5}$ concentrations.

3.1.4 The Draft ISA's Evaluation of the Concentration-response Relationships Is Flawed

There are several issues with the draft ISA's evaluation of concentration-response data. As discussed in Section 2, the draft ISA does not sufficiently address study quality. While the lack of a thorough, systematic study quality evaluation is an issue for determining causation, it is even more problematic in the context of concentration-response relationships. For causal determinations, studies need to establish the presence of an effect following an exposure, but for concentration-response relationships, studies need to not only establish the presence of an effect, but also the magnitude of an effect in relation to the level of the exposure.

The draft ISA does not present any study quality evaluations for the epidemiology studies on which it relies for concentration-response relationships. For example, the draft ISA indicates that these epidemiology studies used a variety of statistical methods but does not discuss the strengths and limitations of these statistical methods or consider whether these methods were appropriately used in the studies. The draft ISA also does not consider sensitivity analyses in studies where the observed concentration-response relationships were sensitive to the degrees of freedom chosen for the natural spline, indicating that the results were not robust.

A major source of bias and uncertainty in epidemiology studies is exposure assessment. Analyses by Rhomberg *et al*. (2011) and Cox (2018) demonstrates that exposure measurement error tends to linearize the estimated concentration-response relationship and mask any true threshold. Despite acknowledging that exposure measurement error can lead to bias in either direction regarding estimation of health effects, the draft ISA does not consider this issue when evaluating concentration-response data.

Setting aside the issues of study quality, the draft ISA does not fully consider the consistency of the results across studies. For total mortality, as acknowledged in Section 11.2.4 of the draft ISA, while several studies observed a linear relationship, some studies suggested a supralinear relationship or the presence of a threshold. The draft ISA concludes a linear, no-threshold concentration-response relationship without giving any rationale for disregarding studies that support nonlinear or threshold relationships.

In contrast, for CV effects, the draft ISA states that the interpretation of the concentration-response data is complicated by "both the lack of thorough empirical evaluations of alternatives to linearity as well as the results from cut-point analyses that provide some potential indication for nonlinearity in the relationship between long-term $PM_{2.5}$ exposure and cardiovascular disease" (Section 6.2.16, P 6-203). These issues are also applicable to studies of total mortality, but the draft ISA does not address them when evaluating the concentration-response relationship for total mortality. This indicates that the draft ISA does not take a consistent approach to evaluate concentration-response data across endpoints.

Table 3.3 Key Characteristics and Results of Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Morbidity

Characteristics and Results		Chen *et al.* (2014)		Cesaroni *et al.* (2014)		Miller *et al.* (2007)	
Location		Canada		Europe (5 countries)		US	
Study Type		Cohort		Meta analysis of 11 cohorts		Cohort	
Source Population		NPHS and CCHS respondents		ESCAPE		WHI	
Sample Size		35,303		100,166		65,893	
Study Period		1996-2010 (15-year)		Start 1997, end not reported[c]		1994-2003 (10-year)	
Exposure	Metric	Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	2001-2006		2008-2011		2000	
	Exposure windows	Multiple		Single		Single	
	Spatial scale	10 km grid		Surface		Community-level	
	Temporal scale	Time-invariant		Time-invariant		Time-invariant	
	Residential mobility	Considered		Considered		Not considered	
	Source	Satellite-based estimates		LUR		Fixed-site monitors	
	Level (μg/m^3)	Mean (range)	10.7 (2.9, 19.2)	Range of means	7-31	Mean (range)	13.5 (3.4, 28.3)
Outcome	Endpoint	Hypertension		Acute coronary disease (ACD)		Cardiovascular disease (CVD)	
	Source	Ontario Hypertension Database		Hospital discharge and mortality		Medical records, Annual	
Individual Covariates	Variables	Age, sex, marital status, education, household income adequacy, race, BMI, smoking, physical activity, drinking, diet, urban residency, preexisting diabetes mellitus or COPD, region indicator		Age, sex, year of enrolment, marital status, education, occupation, smoking status, smoking duration, smoking intensity		Age, race, education, household income, smoking status, systolic blood pressure, BMI, presence/absence of diabetes, hypertension, or hypercholesterolemia	
	Source	Survey at baseline		NR		Questionnaire at enrollment	
Ecological Covariates	Variables	Unemployment rate, less than high school education (%), mean household income		Socioeconomic status indicator		None	
	Source	Canada Census (1996, 2001, 2006)		NR		NA	
Copollutants	Variable	None		None		PM_{coarse}, CO, SO_2, NO_2, O_3	
	Time period					NR	
	Spatial scale					Community-level	
	Temporal scale					Time-invariant	
	Source					Fixed-site monitors	
Linear Association Analyses	Exposure contrast	Per 10 μg/m^3		Per 5 μg/m^3		Per 10 μg/m^3	
	Statistical models	Random-effects Cox models		Cox models and meta analysis		Standard Cox models	
						Single-pollutant	Multi-pollutant
	Cardiovascular event	NR		NR		1.24 (1.09, 1.41)	1.53 (1.21, 1.94)[a]
	Coronary heart disease	NR		NR		1.21 (1.04, 1.42)	NR
	Acute coronary events	NR		1.13 (0.98, 1.30)		NR	NR
	Cerebrovascular	NR		NR		1.35 (1.08, 1.68)	NR
	Myocardial infarction	NR		NR		1.06 (0.85, 1.34)	NR
	Coronary	NR		NR		1.20 (1.00, 1.43)	NR
	Hypertension	1.13 (1.05, 1.22)		NR		NR	NR
	Stroke	NR		NR		1.28 (1.02, 1.61)	NR
C-R Analyses	Outcome	Hypertension		Not conducted		CVD	
	Splines	Natural splines (2, 3 *df*)				Splines	
	Nonlinearity statistically tested	Tested, p-value not reported				Not tested	
	Observed/reported shape of the C-R curve	Linear				Authors did not specify; appears supralinear	
Threshold Analyses	Threshold model	Not conducted		Restricting analyses to subjects with		Not conducted	
	Threshold identification			Slope changed from significant to not			
	Outcome			ACD			
	Threshold estimate			No threshold			

Characteristics and Results		Dorans *et al.* (2016)		Kaufman *et al.* (2016)			
Location		US		US			
Study Type		Cohort		Cohort			
Source Population		FHS Offspring and Third Generation cohorts		MESA Air			
Sample Size		3,399		6,795			
Study Period		2002-2005, 2008-2011		2000-2012 (13-year)			
Exposure	Metric	Annual mean $PM_{2.5}$		Long-term average $PM_{2.5}$ between baseline/follow-up exams			
	Exposure period	2003		1999-2012			
	Exposure windows	Single		Single			
	Spatial scale	1 km grid		Surface			
	Temporal scale	Time-invariant		Time-invariant			
	Residential mobility	Not considered		Considered			
	Source	Spatiotemporal model		Spatiotemporal model			
	Level (μg/m^3)	Median (IQR)	10.7 (1.4)[a]	Mean (SD)		14.3 (2.5)	
Outcome	Endpoint	Coronary artery calcium (CAC)		Coronary artery calcium (CAC) progression			
	Source	MDCT scans		CT scans			
Individual Covariates	Variables	Age at scan, sex, BMI, smoking status, pack-years, education, cohort, scan, date of scan, number of days between scan and examination		Baseline age, sex, ethnicity, city, income, CT scanner type, BMI, physical activity, smoking status and second-hand smoke exposure, employment, adiposity, cholesterol, statin use			
	Source	Physician interview and physical examination		Questionnaire and physical examinations			
Ecological Covariates	Variables	Median value of owner-occupied		Socioeconomic index, level of education, income			
	Source	US Census (2000)		NR			
Copollutants	Variable	None		NO_x, NO_2			BC
	Time period			1999-2012			2006-2008
	Spatial scale			Surface			
	Temporal scale			Time-invariant			
	Source			Spatiotemporal model			
Linear Association Analyses	Exposure contrast	Per IQR (1.4 μg/m^3)		Per 5 μg/m^3			
	Statistical models	Logistic regression		Mixed effects model			
				Single-pollutant	NO_x adjusted	NO_2 adjusted	BC adjusted
	Coronary artery calcium	1.02 (0.91, 1.14)		4.1 (1.4, 6.8)[b]	3.1 (-1.3, 7.6)[b]	4.1 (0.2, 8.1)[b]	4.5 (1.8, 7.3)[b]
C-R Analyses	Outcome	CAC progression (log-transformed)		CAC progression			
	Splines	Restricted cubic spline (5 knots)		Thin plate regression spline (5 *df*)			
	Nonlinearity statistically tested	Not tested		Not tested			
	Observed/reported shape of the C-R curve	Nonlinear (suggested positive association at lower PM levels, suggested negative association at higher levels)		Authors did not specify; appears supralinear (attenuation at higher levels)			
Threshold Analyses	Threshold model	Not conducted		Not conducted			
	Threshold identification						
	Outcome						
	Threshold estimate						

Notes:
BC = Black Carbon; BMI = Body Mass Index; CO = Carbon Monoxide; COPD = Chronic Obstructive Pulmonary Disease; CPS-II = Cancer Prevention Study II; CT = Computed Tomography; C-R = Concentration-response; *df* = Degrees of Freedom; IQR = Interquartile Range; LUR = Land-use Regression; MDCT = Multiple Detector Computed Tomography; NA = Not Applicable; NO_2 = Nitrogen Dioxide; NR = Not Reported; O_3 = Ozone; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter; ppb = Parts Per Billion; SO_2 = Sulfur Dioxide.
(a) For the year 2003.
(b) Results presented in Agatston units/year.
(c) End of follow-up period was not reported; mean duration of follow-up was 11.5 years.

Table 3.4 Key Characteristics and Results of Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Mortality

Characteristics and Results		Gan *et al.* (2011)		Crouse *et al.* (2012)		Villeneuve *et al.* (2015)	
Location		Canada		Canada		Canada	
Study Type		Cohort		Cohort		Cohort	
Source Population		Metropolitan Vancouver residents		CanCHEC		CNBSS	
Sample Size		452,735		2.1 million		89,248	
Study Period		1999-2002 (4-year)		1991-2001 (11-year)		1980-2005 (26-year)	
Exposure	Metric	Monthly mean $PM_{2.5}$		Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	1994-1998		2001-2006		1998-2006	
	Exposure windows	Single		Single		Single	
	Spatial scale	10 m grid		10 km grid		10 km grid	
	Temporal scale	Time-invariant		Time-invariant		Time-invariant	
	Residential mobility	Considered		Not considered		Not considered	
	Source	LUR		Satellite-based estimates		Satellite-based estimates, GEOS-chem	
	Level ($\mu g/m^3$)	Mean ± SD	4.08 ± 1.63	Mean ± SD	8.7 ± 3.9	Median (range)	9.1 (1.3, 17.6)
Outcome	Endpoint	Mortality: Coronary heart disease (CHD)		Mortality: Cardiovascular disease (CVD), cerebrovascular disease (CBVD), Ischemic heart disease (IHD)		Mortality: Cardiovascular disease (CVD), Ischemic heart disease (IHD)	
	Source	Provincial death registration database		Canadian Mortality Database		Canadian Mortality Database	
Individual Covariates	Variables	Age, sex, preexisting comorbidities (diabetes, COPD, hypertensive heart disease)		Age, sex, aboriginal ancestry, visible minority, marital status, highest level of education, employment status, occupational classification		Age at entry, occupation, marital status, education, BMI, cigarette pack-years	
	Source	Provincial hospitalization records and death registration records		Questionnaire at enrollment		Questionnaire at enrollment	
Ecological Covariates	Variables	Neighborhood SES		Unemployed adults (%), adults without high school diplomas (%), subjects in the lowest income quintile (%), population size of home		Mean income, with high school education (%), low income households (%), unemployment rate	
	Source	Canada Census (2001)		Canada Census (1991)		Canada Census (1991)	
Copollutants	Variable	Monthly mean	Monthly mean BC	None		None	
	Time period	1994-1998					
	Spatial scale	10 m grid					
	Temporal scale	Time-invariant					
	Source	LUR					
Linear Association Analyses	Exposure contrast	Per IQR (1.58 $\mu g/m^3$)		Per 10 $\mu g/m^3$		Per 10 $\mu g/m^3$	
	Statistical models	Single-pollutant Cox models	Multi-pollutant Cox models	Standard Cox models	Random-effects Cox models	Standard Cox models	
	Cardiovascular disease	NR	NR	1.16 (1.13, 1.19)	1.15 (1.07, 1.24)	1.32 (1.14, 1.53)	
	Coronary heart disease	1.01 (0.98, 1.05)	1.00 (0.96, 1.03)	NR	NR	NR	
	Ischemic heart disease	NR	NR	1.31 (1.27, 1.35)	1.30 (1.18, 1.43)	1.34 (1.09, 1.66)	
	Cerebrovascular disease	NR	NR	1.04 (0.99, 1.10)	1.04 (0.93, 1.16)	NR	
	Circulatory disease	NR	NR	1.16 (1.13, 1.18)	1.14 (1.06, 1.22)	1.32 (1.14, 1.52)	
C-R Analyses	Outcome	CHD mortality		CVD/CBVD	IHD mortality	CVD/IHD mortality	
	Splines/linear trend test	Linear trend test across quintiles		Natural splines (4 *df*)		Natural splines (3 *df*)	
	Nonlinearity statistically tested	$p_{trend} = 0.813$		Tested, p-value not reported		Not tested	
	Observed/reported shape of the C-R curve	Nonlinear		Linear	Nonlinear	Linear	
Threshold Analyses	Threshold model	Not conducted		Not conducted		Hockey-stick linear splines (assuming no association below the threshold), starting from 2 to 14 $\mu g/m^3$ at 1-$\mu g/m^3$ increments	
	Threshold identification					Minimizing the -2*log-likelihood	
	Outcome					CVD/IHD mortality	
	Threshold estimate					No threshold	

Characteristics and Results		Cesaroni *et al.* (2013)		Miller *et al.* (2007)		Lepeule *et al.* (2012)	
Location		Italy		US		US	
Study Type		Cohort		Cohort		Cohort	
Source Population		RoLS		WHI		HSC	
Sample Size		1.2 million		65,893		8,096	
Study Period		2001-2010 (9-year)		1994-2003 (10-year)		1974-2009 (36-year)	
Exposure	Metric	Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	2005		2000		1974-2009	
	Exposure windows	Single		Single		Multiple	
	Spatial scale	1 km grid		Community-level		City-level	
	Temporal scale	Time-invariant		Time-invariant		N-year moving average[a]	
	Residential mobility	Considered		Not considered		Not considered	
	Source	Eulerian dispersion model		Fixed-site monitors		Fixed-site monitors[b]	
	Level (μg/m^3)	Mean ± SD	23 ± 4.4	Mean (range)	13.5 (3.4, 28.3)	Mean	15.9
Outcome	Endpoint	Mortality: Cardiovascular disease (CVD), Ischemic heart disease (IHD)		Mortality: Cardiovascular disease (CVD)		Mortality: Cardiovascular disease (CVD)	
	Source	Rome Municipal Register		Proxy reports, National Death Index		National Death Index	
Individual Covariates	Variables	Age, sex, marital status, place of birth, education, occupation		Age, race, education, household income, smoking status, cigarettes per day, smoking duration, systolic blood pressure, BMI, presence/absence of diabetes, hypertension, or		Age, sex, BMI, education, smoking history	
	Source	Rome Municipal Register		Questionnaire at enrollment		Questionnaire at enrollment	
Ecological Covariates	Variables	Census-block socioeconomic position index (derived based on a factor analysis including education, occupation, house ownership, family composition, crowding, and immigrant status)		None		None	
	Source	Rome Census (2001) ascertained from Rome Municipal Register		NA		NA	
Copollutants	Variable	Annual mean NO_2		None		None	
	Time period	2007					
	Spatial scale	Surface					
	Temporal scale	Time-invariant					
	Source	LUR					
Linear Association Analyses	Exposure contrast	Per 10 μg/m^3		Per 10 μg/m^3		Per 10 μg/m^3	
	Statistical models	Single-pollutant Cox models	Multi-pollutant Cox models	Standard Cox models		Standard Cox models	
	Cardiovascular disease	1.06 (1.04, 1.08)	NR	1.76 (1.25, 2.47)		1.26 (1.14, 1.40)	
	Coronary heart disease	NR	NR	2.21 (1.17, 4.16)[e]		NR	
	Ischemic heart disease	1.10 (1.06, 1.13)	NR	NR		NR	
	Cerebrovascular disease	1.08 (1.04, 1.13)	NR	1.83 (1.11, 3.00)		NR	
C-R Analyses	Outcome	CVD/IHD mortality		CVD mortality		CVD mortality	
	Splines/linear trend test	Natural splines (≤ 4 *df*)		Splines		Penalized splines	
	Nonlinearity statistically tested	Tested, p-value not reported		Not tested		p-value range: 0.24-0.43	
	Observed/reported shape of the C-R curve	Linear		Generally linear		Linear	
Threshold Analyses	Threshold model	Not conducted		Not conducted		Not conducted	
	Threshold identification						
	Outcome						
	Threshold estimate						

Characteristics and Results		Weichenthal *et al.* (2014)		Jerrett *et al.* (2017)		Thurston *et al.* (2016)	
Location		US		US		US	
Study Type		Cohort		Cohort		Cohort	
Source Population		AHS		CPS-II		NIH-AARP	
Sample Size		83,378		668,629		517,041	
Study Period		1993-2009 (17-year)		1982-2004 (23-year)		2000-2009 (10-year)	
Exposure	Metric	Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$		Annual mean $PM_{2.5}$	
	Exposure period	2001-2006		2002-2004[c]		1999-2008	
	Exposure windows	Single		Single		Single	
	Spatial scale	10 km grid		36 km, 10 km, 1 km, 30 m grid[d]		Surface	
	Temporal scale	Time-invariant		Time-invariant		Time-invariant (time-independent); Yearly (1-year lag) (time-dependent)	
	Residential mobility	Not considered		Not considered		Considered	
	Source	Satellite-based estimates, GEOS-chem models		Satellite-based estimates, GEOS-chem models, LUR, BME		LUR, BME	
	Level ($\mu g/m^3$)	Mean (range)	9.44 (5.7, 19.2)	Range	0.7-26.6	Range	2.9-28
Outcome	Endpoint	Mortality: Cardiovascular disease		Mortality: Circulatory disease		Mortality: Cardiovascular disease	
	Source	State Death registries, National Death Index		National Death Index		Social Security Administration Death Master File	
Individual Covariates	Variables	Age, sex, state of enrollment, birth year, smoking pack-years, BMI, marital status, education, alcohol consumption, vegetable intake		Age, sex, race, smoking history, secondhand smoke, occupational PM2.5 exposure, occupational dust/fumes exposure, marital status, BMI, alcohol use, diet, education		Age, sex, race, education, marital status, BMI, alcohol consumption, smoking history, region indicator	
	Source	Questionnaire at enrollment		Questionnaire at enrollment		Questionnaire at enrollment	
Ecological Covariates	Variables	None		Median household income, income disparity, unemployment, educational level, population who were Black or Hispanic (%)		Median income, high school or less (%)	
	Source	NA		US Census (1990)		US Census (2000)	
Copollutants	Variable	None		None		Annual mean O_3	
	Time period					2000	
	Spatial scale					PMSA-level	
	Temporal scale					Time-invariant	
	Source					Fixed-site monitors	
Linear Association Analyses	Exposure contrast	Per 10 $\mu g/m^3$		Per 10 $\mu g/m^3$		Per 10 $\mu g/m^3$	
	Statistical models	Standard Cox models		Standard Cox models		Standard Cox models	
						Time-dependent exposure	Time-independent exposure
	Cardiovascular disease	1.15 (0.76, 1.72)		NR		1.11 (1.06, 1.16)	1.10 (1.05, 1.15)
	Ischemic heart disease	NR		1.15 (1.11, 1.19)		NR	NR
	Circulatory disease	NR		1.14 (1.11, 1.17)		NR	NR
C-R Analyses	Outcome	CVD mortality		Circulatory mortality		CVD mortality	
	Splines/linear trend test	Natural splines (2 *df*)		Natural splines (2 *df*)		Natural splines (≤ 4 *df*)	
	Nonlinearity statistically tested	Not tested		Not tested		Not tested	
	Observed/reported shape of the C-R curve	Authors specify linear; appears non-linear		Supralinear		Authors specify monotonically increasing; appears linear	
Threshold Analyses	Threshold model	Not conducted		Not conducted		Not conducted	
	Threshold identification						
	Outcome						
	Threshold estimate						

Notes:

Maximum Entropy; BMI = Body Mass Index; CHD = Coronary Heart Disease; COPD = Chronic Obstructive Pulmonary Disease; CPS-II = Cancer Prevention Study II; C-R = Concentration-response; EPA = United States Environmental Protection Agency; df = Degrees of Freedom; HR = Hazard Ratio; LUR = Land-use Regression; NO_2 = Nitrogen Dioxide; NR = Not Reported; O_3 = Ozone; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter; PMSA = Primary Metropolitan Statistical Area; ppb = Parts Per Billion; SD = Standard Deviation; SES = Socioeconomic Status.

(a) $PM_{2.5}$ moving average was 1 year before death or censor for all-cause deaths, 1-3 years for cardiovascular and lung cancer deaths, and 1-5 years for COPD deaths.

(b) Centrally located monitors: from 1974 to 1986-1988, depending on the city; EPA PM_{10} monitors: from the end of monitoring until 1998; EPA $PM_{2.5}$ monitors: from 1999 to 2009.

(c) Most models averaged data for 2002-2004, except one model ($PM_{2.5}$ RS 01-06) that averaged data for 2001-2006.

(d) Jerrett *et al*. (2016) used six exposure models described in other published studies plus one model developed for this study: $PM_{2.5}$ HBMCMAQ 02-04 yielded estimates at 36 km spatial resolution; $PM_{2.5}$ RS 01-06 and $PM_{2.5}$ BME 02-04, at 10 km; $PM_{2.5}$ No GWR RS 02-04 and $PM_{2.5}$ GWR RS 02-04, at 1 km; $PM_{2.5}$ BMELUR 02-04 and $PM_{2.5}$ BMELURRS 02-04, at 30 m.

(e) HR (95% CI) was 2.21 (1.17, 4.16) for definite diagnosis of CHD, 1.26 (0.62, 2.56) for possible diagnosis of CHD.

Table 3.5 Sources of Bias and Uncertainty in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Morbidity This table summarizes several broad methodological categories where biases and uncertainties could arise in estimated concentration-response relationships between long-term $PM_{2.5}$ exposure and cardiovascular morbidity, including exposure assessment, individual-level covariates, ecological covariates, evaluation of copollutants, and statistical analyses. Red shading indicates the potential for bias and/or the presence of uncertainty with regard to a specific methodological characteristics, but does not reflect the magnitude of such a bias/uncertainty on study results. Unshaded cells indicate there are no apparent biases/uncertainties. For example, Cesaroni *et al.* (2014) did not account for temporal variation when assessing $PM_{2.5}$ exposure, thus the red shading with an "X" reflects the potential for bias and the presence of uncertainty in this aspect. Also, Cesaroni *et al.* (2014) did not report how information on covariates was collected, therefore, red shading with an "NR" indicates the potential for bias and the presence of uncertainty with regard to information bias. In addition, Cesaroni *et al.* (2014) did not use natural splines to estimate the concentration-response curve, so the unshaded cell with an "NA" indicates that there is no apparent bias or uncertainty in this aspect.

Sources of Bias and Uncertainty		**Chen *et al.* (2014)**	**Cesaroni *et al.* (2014)**	**Miller *et al.* (2007)**	**Dorans *et al.* (2016)**	**Kaufman *et al.* (2016)**
$PM_{2.5}$ Exposure Assessment	Central site monitoring (low spatial resolution)		X	X		
	No validation for $PM_{2.5}$ data			X		
	Temporal variation not accounted for	X	X	X	X	X
	Residential mobility not accounted for			X	X	
	No evaluation on multiple exposure windows		X	X	X	X
	Personal activities not accounted for (*e.g.,* time spent indoors)	X	X	X	X	
	Mismatch of $PM_{2.5}$ exposure window and outcome	X	X	X	X	X
Individual Covariates	No adjustment of individual covariates					
	Information bias (*e.g.,* self-reported covariates)	X	NR	X	NR	X
	Temporal variation not accounted for	X	X	X	NR	X
	Unmeasured confounding (*e.g.,* pre-existing conditions)	X	X	X	X	X
Ecological Covariates	No adjustment of ecological covariates			X		
	Temporal variation not accounted for	X	X		X	X
	Residential mobility not accounted for	X	X		X	X
	Unmeasured confounding (*e.g.,* access to health care, violence)	X	X	X	X	X
Evaluation of Copollutants	No adjustment of copollutants	X	X		X	
	Central site monitoring (low spatial resolution)			X		
	No validation for copollutants data			X		
	Temporal variation not accounted for			NR		X
	Residential mobility not accounted for			X		
	Personal activities not accounted for (*e.g.,* time spent indoors)			X		X
	Collinearity/nonlinear relationship with $PM_{2.5}$ not addressed/accounted for			X		
	Mismatch of copollutants window and outcome			NR		X
Statistical Analyses	Model assumptions not tested/relaxed		X	X		NR
	C-R curves sensitive to *df* (natural splines)	X	NA	NR		X
	Nonlinearity not assessed statistically		X	X	X	X
	Threshold not assessed	X		X	X	X

Notes:
C-R = Concentration-response; df = Degrees of Freedom; NA = Not Applicable; NR = Not Reported; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter.

Table 3.6 Sources of Bias and Uncertainty in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Cardiovascular Mortality This table summarizes several broad methodological categories where biases and uncertainties could arise in estimated concentration-response relationships between long-term $PM_{2.5}$ exposure and cardiovascular mortality, including exposure assessment, individual-level covariates, ecological covariates, evaluation of copollutants, and statistical analyses. Red shading indicates the potential for bias and/or the presence of uncertainty with regard to a specific methodological characteristics, but does not reflect the magnitude of such a bias/uncertainty on study results. Unshaded cells indicate there are no apparent biases/uncertainties. For example, Miller *et al.* (2007) did not account for temporal variation when assessing $PM_{2.5}$ exposure, thus the red shading with an "X" reflects the potential for bias and the presence of uncertainty in this aspect. Also, Miller *et al.* (2007) did not report what spline function was used when assessing the concentration-response relationship, therefore, red shading with an "NR" indicates the potential for bias and the presence of uncertainty. In contrast, Gan *et al.* (2011) did not use natural splines to estimate the concentration-response curve, so the unshaded cell with an "NA" indicates that there is no apparent bias or uncertainty in this aspect.

Sources of Bias and Uncertainty		**Gan *et al*. (2011)**	**Crouse *et al*. (2012)**	**Villeneuve *et al.* (2015)**	**Cesaroni *et al.* (2013)**	**Miller *et al*. (2007)**	**Lepeule *et al.* (2012)**	**Weichenthal *et al.* (2014)**	**Jerrett *et al.* (2017)**	**Thurston *et al.* (2016)**
$PM_{2.5}$ Exposure Assessment	Central site monitoring (low spatial resolution)					X	X			
	No validation for $PM_{2.5}$ data					X				X
	Temporal variation not accounted for	X	X	X	X	X		X	X	
	Residential mobility not accounted for	X	X	X		X	X	X	X	
	No evaluation on multiple exposure windows	X	X	X	X	X		X	X	X
	Personal activities not accounted for (*e.g.,* time spent indoors)	X	X	X	X	X	X	X	X	X
	Mismatch of $PM_{2.5}$ exposure window and mortality	X	X	X	X	X		X	X	
Individual Covariates	No adjustment of individual covariates									
	Information bias (*e.g.,* self-reported covariates)		X	X		X	X	X	X	X
	Temporal variation not accounted for	X	X	X	X	X	X	X	X	X
	Unmeasured confounding (*e.g.,* pre-existing conditions)	X	X	X	X	X	X	X	X	X
Ecological Covariates	No adjustment of ecological covariates									
	Temporal variation not accounted for	X	X	X	X	X	X	X	X	X
	Residential mobility not accounted for	X	X	X					X	
	Unmeasured confounding (*e.g.,* access to health care, violence)	X	X	X	X	X	X	X	X	X
Evaluation of Copollutants	No adjustment of copollutants									
	Central site monitoring (low spatial resolution)									X
	No validation for copollutants data									X
	Temporal variation not accounted for	X	X	X	X	X	X	X	X	X
	Residential mobility not accounted for	X								X
	Personal activities not accounted for (*e.g.,* time spent indoors)	X			X					X
	Collinearity/nonlinear relationship with $PM_{2.5}$ not addressed/accounted for									X
	Mismatch of copollutants window and mortality	X			X					X
Statistical Analyses	Model assumptions not tested/relaxed	X		X	X	X		X	X	X
	C-R curves sensitive to *df* (natural splines)	NA	X	X	X	NR		X	X	X
	Nonlinearity not assessed statistically			X		X		X	X	X
	Threshold not assessed	X	X		X	X	X	X	X	X

Notes:
C-R = Concentration-response; df = Degrees of Freedom; NA = Not Applicable; NR = Not Reported; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter.

3.2 Epidemiology Studies Do Not Establish a Linear, No-threshold Relationship Between Short-term $PM_{2.5}$ Exposure and Total Mortality

The draft ISA discusses recent studies conducted in the US (Lee *et al*., 2015; Shi *et al*., 2016; Di *et al*., 2017b) and Europe (Samoli *et al*., 2013) that evaluated the concentration-response relationship between short-term $PM_{2.5}$ exposure and total mortality (Section 11.1.10 of the draft ISA). Here, we present key characteristics and main results of these studies in Table 3.7, and sources of bias and uncertainty in Table 3.8. For studies of short-term $PM_{2.5}$ exposures and total mortality, the bias and uncertainty categories we considered include exposure assessment, adjustment for individual-level, ecological, meteorological, and temporal covariates, evaluation of copollutants, and statistical analysis.

Similar to mortality studies of long-term $PM_{2.5}$ exposure, mortality studies of short-term $PM_{2.5}$ exposure have considerable uncertainties in exposure assessments. Samoli *et al*. (2013) only relied on centrally located monitors to estimate $PM_{2.5}$ concentrations at low spatial resolution. While the US studies (Lee *et al*., 2015; Shi *et al*., 2016; Di *et al*., 2017b) used validated models to estimate $PM_{2.5}$ concentrations, they were still limited by the quality of the input AOD data, as discussed in Section 3.1.2. None of the studies considered individual activity patterns, such as time spent indoors, outside residential areas, or commuting.

Confounding is likely an issue in all of these studies Samoli *et al*. (2013) and Shi *et al*. (2016) conducted time-series analyses, which are ecological in nature. Unmeasured confounding, such as fluctuations in society stress levels, is likely. Lee *et al*. (2016) and Di *et al*. (2017b) used a case-crossover design; while time-invariant factors were automatically controlled for, time-variant confounders such as physical exertion, stress, or influenza epidemics were not accounted for. For all four studies, adjustment for meteorological and temporal factors was not sufficient, which likely had considerable impact on the study results. This is because most of the studies only adjusted for temperature, but not humidity, and these studies did not conduct sufficient sensitivity analyses with alternative forms and lag times for the meteorological covariates. In addition, except for Samoli *et al*. (2013), these studies did not adjust for holidays, thus residual confounding by temporal factors is likely.

In addition, only one study conducted specific threshold analyses, and none statistically tested for nonlinearity of the concentration-response relationship.

The draft ISA does not consider these study quality issues when discussing the short-term $PM_{2.5}$ concentration-response data. Notably, the draft ISA acknowledges that recent studies have not addressed the difficulties in assessing the $PM_{2.5}$-mortality concentration-response relationships (*e.g*., sparse data at low and high $PM_{2.5}$ concentrations, influence of exposure measurement error), as identified in the 2009 PM ISA (US EPA, 2009), and that they have not conducted systematic evaluations of alternatives to linearity. Yet the draft ISA concludes that these studies provide evidence of a no-threshold linear relationship with confidence down to 5 $\mu g/m^3$ $PM_{2.5}$. This conclusion is not supported by the available evidence and contradicts the draft ISA's statement of uncertainty.

Given their limitations and uncertainties, the epidemiology studies do not establish a linear, no-threshold relationship between short-term $PM_{2.5}$ exposure and total mortality in general, let alone with low $PM_{2.5}$ concentrations.

Table 3.7 Key Characteristics and Results of Epidemiology Studies of Short-term $PM_{2.5}$ Exposure and Total Mortality

Characteristics and Results		Samoli *et al.* (2013)					Shi *et al.* (2015)	
Location		Europe (4 countries)					US	
Study Type		Time series analysis					Time series analysis	
Source Population		MED-PARTICLES					Medicare population	
Sample Size		15 million					551,024	
Study Period		2001-2010					2003-2008 (6-year)	
Exposure	Metric	$PM_{2.5}$ lag01					$PM_{2.5}$ lag01	
	Exposure period	2001-2010 (same with study period for each city)					2003-2008	
	Exposure windows	Multiple					Multiple	
	Spatial scale	City-level					1 km grid	
	Temporal scale	Daily					Daily	
	Residential mobility	Not considered					Not considered	
	Source	Fixed site monitors					Satellite-based estimates	
	Level (μg/m^3)	Range of medians			13.6-27.7		Mean ± SD	8.21 ± 5.10
Outcome	Endpoint	All-cause mortality					All-cause mortality	
	Source	NR					Centers for Medicare and Medicaid services	
Individual Covariates	Variables	None					None	
	Source	NA					NA	
Ecological Covariates	Variables	Influenza, summer population decrease					Zip-code level: race, education, median household income; county-level: every day smoking prevalence	
	Source	Hospital admission records for influenza epidemics					US Census (2000), CDC BRFSS	
Meteorological Covariates	Variables	Temperature					Temperature	
	Forms	Spline (3 *df*)					Spline (3 *df*)	
	Source	NR					NCDC monitors	
Temporal	Variables	Seasonality, day of the week, holiday					Time/season trend, day of the week	
Copollutants	Variable	PM_{coarse}		SO_2	NO_2	O_3	None	
	Time period	2001-2010 (same as study period for each city)						
	Spatial scale	City-level						
	Temporal scale	Daily						
	Source	Fixed site monitors						
Linear Association Analyses[a]	Exposure contrast	Per 10 μg/m^3					Per 10 μg/m^3	
	Statistical models	Time series analyses					Time series analyses	
		Single-pollutant	Adjust for PM_{coarse}	Adjust for SO_2	Adjust for NO_2	Adjust for O_3	Full cohort	Low daily exposure[b]
	All-cause/non-accidental	0.55 (0.27, 0.84)	0.59 (0.00, 1.18)	0.33 (-0.37, 1.03)	0.28 (-0.12, 0.68)	0.46 (0.16, 0.76)	2.14 ± 0.75	2.14 ± 0.81
C-R Analyses	Outcome	Not conducted					All-cause mortality	
	Splines/linear trend test						Penalized splines	
	Nonlinearity statistically tested						Not tested	
	Observed/reported shape of the C-R curve						Linear	
Threshold Analyses	Threshold model	Hockey-stick linear splines (assuming no association below the threshold), starting from 0 to 35 μg/m^3 at 5-μg/m^3 increments					Not conducted	
	Threshold identification	Minimizing the mean deviance						
	Outcome	All-cause mortality						
	Threshold estimate	No threshold						

Characteristics and Results		Lee *et al.* (2016)			Di *et al.* (2017b)		
Location		US			US		
Study Type		Case-crossover			Case-crossover		
Source Population		North Carolina, South Carolina, and Georgia residents			Medicare population		
Sample Size		848,270			22 million		
Study Period		2007-2011 (5-year)			2000-2012 (13-year)		
Exposure	Metric	PM2.5 lag01			PM2.5 lag01		
	Exposure period	2007-2011			2000-2012		
	Exposure windows	Multiple			Multiple		
	Spatial scale	1 km grid			1 km grid		
	Temporal scale	Daily			Daily		
	Residential mobility	Not considered			Not considered		
	Source	Satellite-based estimates			Satellite-based estimates, neural network, CTM		
	Level (μg/m^3)	Mean ± SD		11.1 ± 4.4	Mean		11.6
Outcome	Endpoint	Non-accidental mortality			All-cause mortality		
	Source	State departments of public health			Centers for Medicare and Medicaid services		
Individual Covariates	Variables	None			None		
	Source	NA			NA		
Ecological Covariates	Variables	None			None		
	Source	NA			NA		
Meteorological Covariates	Variables	Temperature			Temperature, dew point temperature		
	Forms	Continuous, spline (3 df)			Spline (3,6,9 df)		
	Source	NCDC monitors			NARR		
Temporal Covariates	Variables	Day of the week, season			Day of the week, time trend, seasonal and subseasonal patterns		
Copollutants	Variable	None			O_3		
	Time period				2000-2012 (warm-season)		
	Spatial scale				1 km grid		
	Temporal scale				Daily (8-hour maximum)		
	Source				Satellite-based estimates, neural network, CTM		
Linear Association Analyses[a]	Exposure contrast	Per 10 μg/m^3			Per 10 μg/m^3		
	Statistical models	Conditional logistic regression model			Conditional logistic regression model		
		Full cohort	Low annual exposure[c]	Low daily and annual exposure[d]	Single-pollutant	Two-pollutant (full cohort)	Two-pollutant (low exposure)[e]
	All-cause/non-accidental	1.56 (1.19, 1.94)	2.06 (1.97, 2.15)	2.08 (1.99, 2.17)	1.18 (1.09, 1.28)	1.05 (0.95, 1.15)	1.61 (1.48, 1.74)
C-R Analyses	Outcome	Not conducted			All-cause mortality		
	Splines/linear trend test				Penalized splines		
	Nonlinearity statistically tested				Not tested		
	Observed/reported shape of the C-R curve				Supralinear with less steep slope > 15 μg/m^3		
Threshold Analyses	Threshold model	Not conducted			Not conducted		
	Threshold identification						
	Outcome						
	Threshold estimate						

Notes:

BMI = Body Mass Index; C-R = Concentration-response; CTM = Chemical Transport Model; *df* = Degrees of Freedom; NA = Not Applicable; NARR = North American Regional Reanalysis; NCDC = National Climatic Data Center; NO_2 = Nitrogen Dioxide; NR = Not Reported; O_3 = Ozone; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter; $PM_{2.5}$ lag01 = the average of $PM_{2.5}$ levels on the same day and the previous day; ppb = Parts Per Billion; SD = Standard Deviation; SO_2 = Sulfur Dioxide.

(a) Results presented as percent increase in mortality with 95% CI.

(b) The analysis was restricted only to person time with daily $PM_{2.5}$ < 30 μg/m^3.

(c) The analysis was restricted only to person time with annual $PM_{2.5}$ < 12 μg/m^3.

(d) The analysis was restricted only to person time with annual $PM_{2.5}$ < 12 μg/m^3 and daily $PM_{2.5}$ < 35 μg/m^3.

(e) The analysis was restricted only to person time with daily $PM_{2.5}$ < 25 μg/m^3.

Table 3.8 Sources of Bias and Uncertainty in Epidemiology Studies of Short-term $PM_{2.5}$ Exposure and Total Mortality This table summarizes several broad methodological categories where biases and uncertainties could arise in estimated concentration-response relationships between short-term $PM_{2.5}$ exposure and total mortality, including exposure assessment, individual-level covariates, ecological covariates, meteorological covariates, temporal covariates, evaluation of copollutants, and statistical analyses. Red shading indicates the potential for bias and/or the presence of uncertainty with regard to a specific methodological characteristics, but does not reflect the magnitude of such a bias/uncertainty on study results. Unshaded cells indicate there are no apparent biases/uncertainties. For example, Samoli *et al.* (2013) only relied on central site monitoring when assessing $PM_{2.5}$ exposure, thus the red shading with an "X" reflects the potential for bias and the presence of uncertainty in this aspect. Also, because Samoli *et al.* (2013) conducted a time-series analysis, an ecological study design, unshaded cells with an "NA" indicate there is no apparent bias or uncertainty with regard to adjustment for individual-level covariates. In addition, Samoli *et al.* (2013) did not use natural splines to estimate the concentration-response curve, so the unshaded cell with an "NA" indicates that there is no apparent bias or uncertainty in this aspect. In contrast, Lee *et al.* (2015) did not conduct any spline analysis when assessing the concentration-response relationship, therefore, red shading with an "NA" indicates there is increased potential for bias and uncertainty in this aspect.

Sources of Bias and Uncertainty		**Samoli *et al.* (2013)**	**Shi *et al.* (2015)**	**Lee *et al.* (2016)**	**Di *et al.* (2017b)**
$PM_{2.5}$ Exposure Assessment	Central site monitoring (low spatial resolution)	X			
	No validation for $PM_{2.5}$ data	X			
	No evaluation on multiple lags				
	Personal activities not accounted for (*e.g.,* time spent indoors,	X	X	X	X
Individual Covariates	No adjustment of individual covariates statistically or by study design	NA	NA		
	Information bias (*e.g.,* self-reported covariates)				
	Temporal variation not accounted for			X	X
	Unmeasured/residual confounding			X	X
Ecological Covariates	No adjustment of ecological covariates			X	X
	Unmeasured/residual confounding	X	X		
Meteorological Covariates	No adjustment of meteorological covariates				
	No evaluation on multiple forms (e.g., varying *df*, lags)	X	X		
	Unmeasured/Residual confounding	X	X	X	X
Temporal Covariates	No adjustment of temporal covariates				
	Unmeasured/residual confounding	X	X	X	X
Evaluation of Copollutants	No adjustment of copollutants		X	X	
	Central site monitoring (low spatial resolution)	X			
	No validation for copollutants data	X			
	Personal activities not accounted for (*e.g.,* time spent indoors,	X			X
	Collinearity/nonlinear relationship with $PM_{2.5}$ not addressed/accounted				X
Statistical Analyses	C-R curves sensitive to *df* (natural splines)	NA		NA	
	Nonlinearity not assessed statistically	X	X	X	X
	Threshold not assessed		X	X	X

Notes:
C-R = Concentration-response; *df* = Degrees of Freedom; NA = Not Applicable; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter.

4 Long-term $PM_{2.5}$ Exposure and Neurological Effects

The draft ISA concludes that there is a likely causal relationship between long-term $PM_{2.5}$ exposure and nervous system effects, primarily based on animal toxicity studies of inflammation, oxidative stress, morphological changes, and neurodegeneration in the brain and epidemiology studies of brain volume, cognitive function, and neurodegenerative diseases.

The draft ISA does not present a systematic study quality evaluation in its summary of the available literature, nor does it appear to consider study quality when synthesizing the evidence. As discussed below, epidemiology studies of brain volume, cognitive function, and dementia have considerable limitations and uncertainties that undermine the observed associations between long-term $PM_{2.5}$ exposure and neurological endpoints. In addition, findings from animal toxicity studies do not necessarily provide evidence for apical endpoints, and may have limited relevance to humans.

As stated by Dr. Sabine Lange in her preliminary comments on the draft ISA (CASAC, 2018):

> 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.

4.1 Epidemiology Studies Do Not Indicate Long-term $PM_{2.5}$ Exposure Affects Brain Volume Changes

The draft ISA concludes that epidemiology studies based on the Women's Health Initiatives Memory Study (WHIMS) cohort (Chen *et al*., 2015; Casanova *et al*., 2016) and the Framingham Offspring Study (FOS) cohort (Wilker *et al*., 2015) provide key evidence that long-term $PM_{2.5}$ exposure is associated with reductions in brain volume. The draft ISA's conclusion is not warranted because all three studies evaluated brain volume cross-sectionally and thus are not informative regarding changes in brain volume over time. In addition, the draft ISA does not present any quality evaluation of these studies, nor does it appear to consider study quality when evaluating the results. This is disconcerting because all three studies share common critical methodological limitations that preclude their utility in causal inference.

Chen *et al*. (2015), Casanova *et al*. (2016), and Wilker *et al*. (2015) used magnetic resonance imaging (MRI) to assess brain volumes in study participants. Each participant only underwent one MRI scan; therefore, only cross-sectional measurements of brain volume, instead of changes over time, were available for these participants. These studies compared the inter-individual differences in brain volumes across $PM_{2.5}$ concentrations but not how $PM_{2.5}$ concentrations relate to within-individual changes in brain volume. The cross-sectional nature of the outcome assessment makes these studies hypothesis-generating at best, and inappropriate for causal inference.

In addition, there is a high degree of inter- and intra-individual variability in brain volume. Total brain volume can vary by nearly two-fold among typically developing humans of the same age, and brain size variation is coupled with brain shape diversity (Reardon *et al*., 2018). Even within an individual, brain volume changes because of various physiological and/or pathological processes and factors such as

hydration levels and neurodegenerative diseases (Duning *et al*., 2005; Nakamura *et al*., 2014; Maclaren *et al*., 2014). Moreover, brain volume measurements by MRI vary considerably due to factors such as the scanner, imaging protocol, and software used for data processing (Maclaren *et al*., 2014). Volumetric measurements from multiple MRI scans of the same individual can fluctuate by as much as 9% for various brain regions (Maclaren *et al*., 2014). Even for multicenter studies with harmonized protocols, which is the case for the studies by Chen *et al*. (2015) and Casanova *et al*. (2016), site differences could lead to severe biases in volumetric analyses (Shinohara *et al*., 2017). Taken together, these factors indicate that outcome measurement error was likely substantial when brain volume was estimated from a single MRI scan in the studies by Chen *et al*. (2015), Casanova *et al*. (2016), and Wilker *et al*. (2015).

The draft ISA does not consider these critical limitations in the cross-sectional study design and outcome measurements, and thus its conclusion that long-term $PM_{2.5}$ exposure is associated with reductions in brain volume is not appropriate.

4.2 Bias and Uncertainty Undermine Epidemiology Evidence of Long-term $PM_{2.5}$ Exposure and Cognitive Decline/Dementia

The draft ISA discusses a number of epidemiology studies that evaluated the associations between long-term $PM_{2.5}$ exposure and cognitive function, measured dichotomously (Section 8.2.5.2, Figure 8-3 in the draft ISA) or continuously (Section 8.2.5.2, Figure 8-4 in the draft ISA), and neurodegenerative diseases (Section 8.2.6, Figure 8-6 in the draft ISA). The draft ISA concludes that epidemiology studies report consistent associations with cognitive decrements and with all-cause dementia, and that these provide evidence for a likely causal relationship between long-term $PM_{2.5}$ exposure and nervous system effects.

As discussed in Section 2, the draft ISA does not sufficiently address study quality when evaluating evidence for causal determinations. This is the case with epidemiology studies of cognitive function and neurodegenerative diseases. For example, the draft ISA does not discuss the variability and reliability of various cognitive function tests used in epidemiology studies or consider how these issues could limit the utility of study results, particularly those from cross-sectional analyses, in causal inference. The draft ISA also does not appear to consider the potential for exposure measurement error when epidemiology studies relied on residential addresses to estimate $PM_{2.5}$ exposure without considering residential mobility or activity patterns (*e.g*., time spent indoors). Although most studies adjusted for a number of potential confounders in their analyses, residual confounding was still likely an issue because the information on covariates was usually assessed at baseline only (*i.e*., not accounting for changes over time) and by self-report.

Setting aside study quality issues, the draft ISA does not appear to appraise individual studies or endpoints in a consistent manner. For example, Loop *et al*. (2013) conducted a longitudinal analysis in the REasons for Geographic and Racial Differences in Stroke (REGARDS) Cohort to determine whether long-term $PM_{2.5}$ exposure was associated with incident cognitive impairment among participants who were cognitively intact at baseline. Loop *et al*. (2013) reported a null association after robust adjustment for potential confounders. Despite discussing and acknowledging the null results reported by Loop *et al*. (2013), the draft ISA does not consider this study as key evidence in its causal determination (Table 8-20) or address the inconsistency between results reported by Loop *et al*. (2013) and others (Weuve *et al*., 2012; Cacciottolo *et al*., 2017).

In addition, Tonne *et al*. (2014) conducted a longitudinal analysis of cognitive decline in a large cohort in the UK. While the point estimates of Z-scores for various cognitive test were negative (*i.e*., in the direction of adversity), they were small in magnitude, ranging from -0.03 to -0.003, and none were statistically

significant. Yet the draft ISA inappropriately considers these null results as supporting a negative impact of $PM_{2.5}$ exposure on cognitive function.

Moreover, the draft ISA groups the studies of neurodegenerative diseases, including Parkinson's disease, Alzheimer's disease, and dementia, and concludes that evidence is inconsistent. It states that "[h]igh quality studies relying on neurologist confirmed PD [Parkinson's disease] provided no evidence of an association" and that there is an "[a]ssociation with all-cause dementia determined by physician adjudication observed in WHIMS but not in registry based follow-up study of Alzheimer's disease in China" (Table 8-20, P 8-63). However, the draft ISA only discusses results from select studies of all-cause dementia as being coherent with animal toxicity data, without providing any rationale for not considering the inconsistent results from studies of Parkinson's disease and Alzheimer's disease.

Overall, epidemiology evidence, undermined by methodological limitations and inconsistent results, does not establish an association between long-term $PM_{2.5}$ exposure and cognitive decline/dementia.

4.3 The Draft ISA Does Not Consider the Quality and Relevance of the Toxicological Data

The draft ISA indicates that "[t]he strongest evidence of an effect of long-term exposure to $PM_{2.5}$ on the nervous system is provided by animal toxicological studies that show inflammation, oxidative stress, morphologic changes, and neurodegeneration in multiple brain regions following long-term exposure to $PM_{2.5}$ CAPs [concentrated ambient particles]" (Section 8.2.9, P 8-61). However, the draft ISA does not consider the quality and relevance of the toxicity evidence.

Table A-1 in Appendix 1 of the draft ISA discusses a number of quality-related considerations for evaluating evidence on $PM_{2.5}$ health effects in various disciplines, including animal toxicology. These considerations include study design, test model, pollutant, exposure assignment, outcome assessment/evaluation, potential copollutant confounding, other confounding factors, and statistical methodology. As discussed in Section 2, these criteria are not sufficiently detailed or prescriptive to ensure a consistent evaluation across studies and endpoints.

In fact, the draft ISA does not discuss the quality of toxicity studies, apparently taking the results of individual studies at face value without discussing study- and endpoint-specific methodological limitations. The draft ISA also does not consider study quality when integrating evidence across endpoints and studies.

In addition, the draft ISA does not sufficiently consider the relevance of exposure doses in experimental studies. For example, Table 8-20 in the draft ISA presents the $PM_{2.5}$ concentrations associated with brain inflammation and oxidative stress in toxicological studies. These concentrations are generally an order of magnitude higher than the current NAAQS of 12 $\mu g/m^3$. It is unclear whether similar molecular and cellular events occur in humans exposed to ambient $PM_{2.5}$ concentrations.

Finally, the draft ISA does not fully consider the human relevance of the observed neurological endpoints or animal models tested. Many outcomes measured are upstream events, not apical effects. The draft ISA did not address whether they were homeostatic changes or reversible effects; the detection of an upstream event alone does not necessarily indicate pathogenesis or disease onset.

Overall, the draft ISA should have assessed these issues when evaluating toxicity evidence for causal determination on nervous system effects.

5 Long-term $PM_{2.5}$ Exposure and Cancer

The draft ISA concludes that there is a likely to be causal relationship between long-term $PM_{2.5}$ exposure and cancer, primarily based on epidemiology studies of lung cancer incidence and mortality, as well as experimental studies that the draft ISA concludes provide evidence for biological plausibility. However, as discussed by Dr. Sabine Lange in her preliminary comments on the draft ISA (CASAC, 2018) and below, the available epidemiology studies are undermined by considerable methodological limitations; most critically, they do not, or do not adequately, account for latency, smoking, and family history of lung cancer. Also, the draft ISA does not consider the quality and human relevance of the experimental findings. As stated by Dr. Lange, "[The] lack of coherence does not suggest a likely causal relationship" (CASAC, 2018).

5.1 The Draft ISA Does Not Consider Key Methodological Limitations and Uncertainties in Lung Cancer Epidemiology Studies

Although the draft ISA indicates that multiple epidemiology studies evaluating long-term $PM_{2.5}$ and cancer are of high quality, it does not present any study quality evaluation. As discussed in Section 2.2, there are many aspects of study quality that can impact the interpretation of results (*e.g.*, confounding). Below we discuss two major limitations in the epidemiology studies of long-term $PM_{2.5}$ exposure and lung cancer, and how they significantly increased the uncertainty in the observed associations.

5.1.1 Epidemiology Studies Did Not Evaluate the Relevant Exposure Window

Lung cancer is a chronic disease with a long latency period. For example, an analysis of over 350,000 lung cancer cases indicated that the time between cancer initiation and diagnosis was approximately 13.6 years (Nadler and Zurbenko, 2014). Therefore, when evaluating potential lung carcinogens, epidemiology studies should consider exposure windows at least a decade prior to cancer diagnosis.

This was not done in the epidemiology studies that evaluated associations between long-term $PM_{2.5}$ exposure and lung cancer mortality and incidence. As shown in Figure 5.1, none of the studies reviewed in the draft ISA included sufficient lag time between the exposure period and follow-up period to account for latency. Moreover, in most studies, exposure periods included time after the follow-up for cancer ended, resulting in considerable exposure measurement error. In several studies, the exposure periods occurred entirely <u>after</u> the cancer follow-up periods, thus violating the temporality rule of causation (*i.e.*, the cause has to occur before the effect).

Because the relevant exposure window for lung cancer was not evaluated and the potential for substantial exposure measurement error was high, observed associations between $PM_{2.5}$ and lung cancer in these studies are not reliable. This was not addressed in the draft ISA.

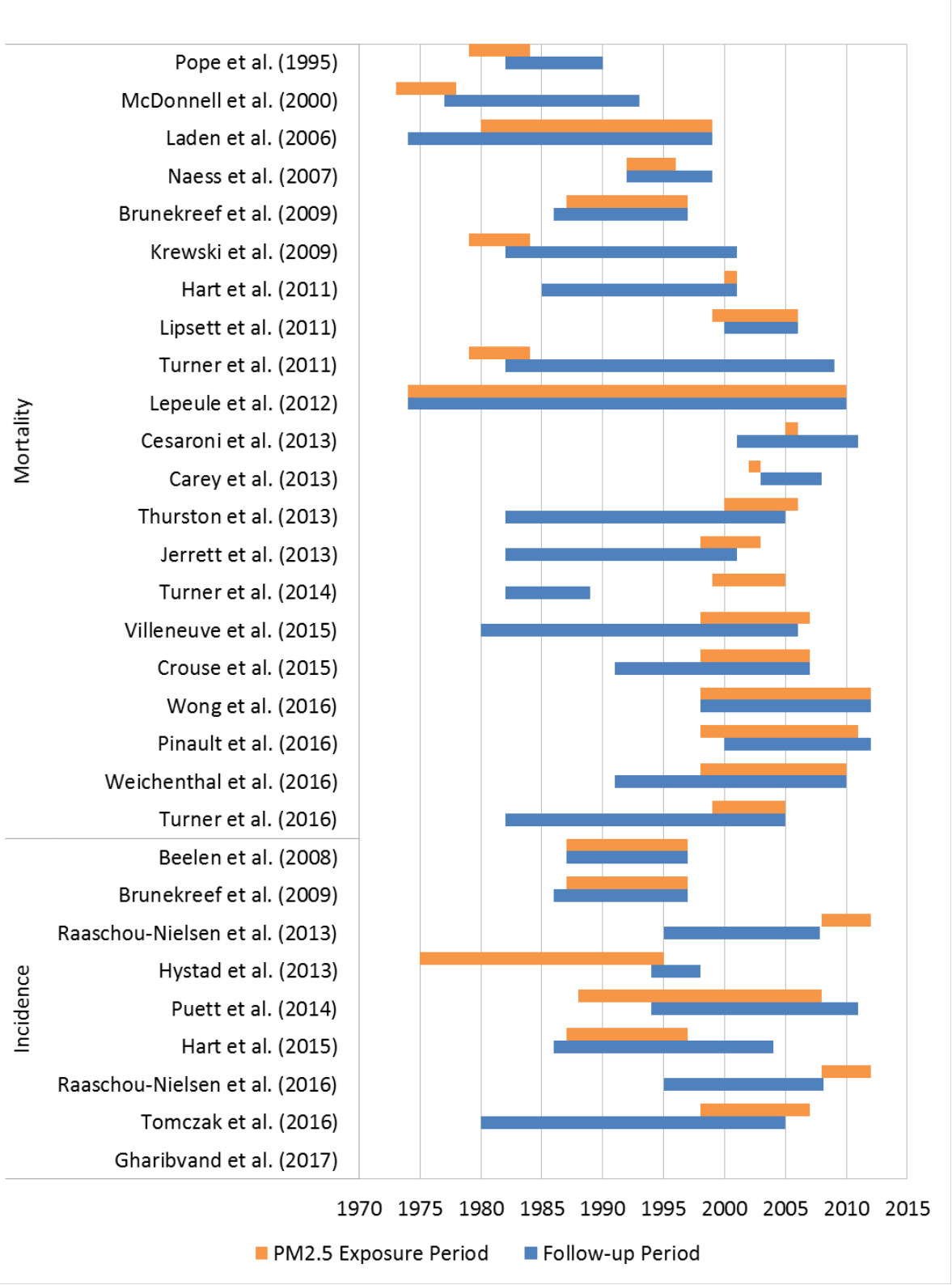

Figure 5.1 Exposure and Follow-up Periods in Studies of Long-term $PM_{2.5}$ Exposure and Lung Cancer Mortality and Incidence

5.1.2 Residual Confounding Likely Undermined the Epidemiology Study Results

Lung cancer has several known risk factors, including smoking, exposure to second-hand smoke, having a family history of lung cancer, and exposure to radon gas, asbestos, arsenic, chromium, and nickel (ACS, 2016a,b).

Smoking accounts for approximately 90% of all lung cancer cases (CDC, 2013). A recent analysis by Thun *et al*. (2013), based on close to one million people pooled from five large contemporary cohorts in the US, showed that compared to men who never smoked, men who are current smokers are nearly 25 times more likely to die from lung cancer (relative risk [RR] = 24.97, 95% confidence interval [CI]: 24.30-30.70). In addition, the risk of lung cancer mortality increased with increasing number of cigarettes smoked per day (intensity), longer duration of smoking, younger age at initiating smoking, and older age at quitting smoking, and decreased with the number of years since quitting smoking. The RRs of lung cancer mortality according to number of cigarettes smoked per day and duration of smoking for male current smokers are presented in Table 5.1.

Table 5.1 Relative Risks of Lung Cancer Mortality in Male Current *vs*. Never Smokers[a]

Cigarettes/Day	< 10	10-19	20-39	40+	P for Trend
RR	15.83	23.61	32.42	41.72	< 0.0001
95% CI	13.45-18.65	20.58-27.09	28.33-37.11	33.18-52.46	
Duration (Years)	**< 30**	**30-39**	**40-49**	**50+**	**P for Trend**
RR	3.17	9.58	20.06	29.40	< 0.0001
95% CI	1.01-9.99	4.66-19.69	14.40-27.94	23.29-37.12	

Notes:
CI = Confidence Interval; RR = Relative Risk.
(a) Recreated from Table S3 in Thun *et al*. (2013).

In addition to active smoking, environmental (passive or secondhand) tobacco smoke is also considered to be a cause of lung cancer, based on findings from epidemiology studies in nonsmokers (NTP, 2011; US Public Health Service, 2006). Exposure to spousal smoking is associated with an increase of 20%-30% in lung cancer risk, while exposure to secondhand smoke at workplace is associated with an increase of 24% in lung cancer risk in the US.

People with a family history of lung cancer are also at increased risk for developing the disease (CDC, 2013). Having a family history of lung cancer is associated with an increase of over 80% in lung cancer risk, as reported in two meta-analyses (Matakidou *et al*., 2005; Gu *et al*., 2010). Lung cancer risks associated with a family history of lung cancer by categories of proband[1] are presented in Table 5.2.

[1] A proband is a person serving as the starting point for the genetic study of a family.

Table 5.2 Relative Risks of Lung Cancer Associated with Family History of Lung Cancer[a]

Categories		Relative Risk	95% CI	P
Relative Affected	Father	1.62	1.43-1.82	< 0.001
	Mother	1.96	1.60-2.41	< 0.001
	Siblings	1.92	1.68-2.19	< 0.001
Smoking Status	Smoking	1.73	1.54-1.94	< 0.001
	Non-smoking	1.42	1.06-1.91	0.02

Notes:
CI = Confidence Interval.
(a) Adapted from Table 3 in Gu *et al*. (2010).

When evaluating a potential lung carcinogen in humans, it is important to adjust for these risk factors, as they are likely confounders. This is especially true for smoking. Because smoking is such a strong risk factor and multiple smoking metrics are independently associated with lung cancer risk, it is critical to measure and account for different metrics such as smoking intensity, duration, age at starting smoking, and years since quitting.

As shown in Table 5.3, this is not the case for the epidemiology studies of long-term $PM_{2.5}$ exposure and lung cancer. None of the studies accounted for family history of lung cancer or all of the relevant smoking metrics. Several studies did not adjust for smoking status at all, which makes their findings completely unreliable. Exposure to secondhand smoking, if adjusted for, was assessed with different metrics across studies. The impact of residual confounding by active and passive smoking and potential confounding by family history was not given any consideration in the draft ISA and likely biased the observed associations between long-term $PM_{2.5}$ exposure and lung cancer.

Table 5.3 Adjustment for Selected Confounders in Epidemiology Studies of Long-term $PM_{2.5}$ Exposure and Lung Cancer

Study	Location	Smoking					Exposure to Secondhand Smoking	Family History
		Status	Intensity	Duration	Pack-years	Other		
Mortality								
Pope *et al.* (1995)	US	Y	Y	Y	N	N	Hours/day exposed to smoking	N
McDonnell *et al.* (2000)	US (CA)	Y	N	N	Y	N	N	N
Laden *et al.* (2006)	US	Y	N	N	Y	N	N	N
Naess *et al.* (2007)	Norway	N	N	N	N	N	N	N
Brunekreef *et al.* (2009)	Netherlands	Y	Y	Y	N	N	Partner's smoking status	N
Krewski *et al.* (2009)	US	Y	Y	Y	N	Age starting smoking	Hours/day exposed to smoking	N
Hart *et al.* (2011)	US	N	N	N	N	N	N	N
Lipsett *et al.* (2011)	US (CA)	Y	N	N	Y	N	Exposure to secondhand smoking at home (yes/no)	N
Turner *et al.* (2011)	US	Y	N	N	N	N	Passive smoking (hours)	N
Lepeule *et al.* (2012)	US	Y	N	N	Y	N	N	N
Cesaroni *et al.* (2013)	Italy	N	N	N	N	N	N	N
Carey *et al.* (2013)	UK	Y	Y	N	N	N	N	N
Jerrett *et al.* (2013)	US (CA)	Y	Y	Y	N	Age starting smoking	Hours/day exposed to smoking	N
Thurston *et al.* (2013)	US	Y	Y	Y	N	Age starting smoking < 18	Hours/day exposed to smoking	N
Turner *et al.* (2014)	US	Y	N	N	N	N	Hours/day exposed to smoking	N
Villeneuve *et al.* (2015)	Canada	Y	N	N	Y	N	N	N
Crouse *et al.* (2015)	Canada	N	N	N	N	N	N	N
Wong *et al.* (2016)	Hong Kong	Y	N	N	N	N	Percentage of smokers in district	N
Pinault *et al.* (2016)	Canada	Y	N	N	N	N	N	N
Weichenthal *et al.* (2016)	Canada	N	N	N	N	N	N	N
Turner *et al.* (2016)	US	Y	Y	Y	N	Age starting smoking < 18	Hours passive smoking	N

Study	Location	Smoking					Exposure to Secondhand Smoking	Family History
		Status	Intensity	Duration	Pack-years	Other		
Incidence								
Beelen *et al.* (2008)	Netherlands	Y	Y	Y	N	N	Partner's smoking status	N
Brunekreef *et al.* (2009)	Netherlands	Y	Y	Y	N	N	Partner's smoking status	N
Raaschou-Nielsen *et al.* (2013)	Europe	Y	Y	Y	N	Time since quitting	Exposure (yes/no)	N
Hystad *et al.* (2013)	Canada	Y	N	N	Y	Time since quitting	Person-years of residential and occupational exposure	N
Puett *et al.* (2014)	US	Y	N	N	Y	Time since quitting	Exposure at home, at work, and during childhood	N
Hart *et al.* (2015)	Netherlands	Y	Y	Y	N	N	Partner's smoking status	N
Raaschou-Nielsen *et al.* (2016)	Europe	Y	Y	Y	N	Time since quitting	Exposure (yes/no)	N
Tomczak *et al.* (2016)	Canada	Y	N	N	Y	N	N	N
Gharibvand *et al.* (2017)	US	Y	Y	N	N	Time since quitting	N	N

Note:
Red shading indicates that the results were in the direction of increased bias or uncertainty.

5.2 The Draft ISA Does Not Fully Consider the Quality and Relevance of the Experimental Studies

The draft ISA's evaluation of mechanistic studies does not appear to follow EPA's framework for evaluating carcinogenic modes of action (US EPA, 2005). Specifically, EPA (2005) requires that:

> All pertinent studies are reviewed in analyzing a mode of action, and an overall weighing of evidence is performed, laying out the strengths, weaknesses, and uncertainties of the case as well as potential alternative positions and rationales.

This was not done in the ISA. Instead, the draft ISA indicates that "extensive" experimental evidence provides support for biological plausibility because toxicity studies show that $PM_{2.5}$ exhibits several of the 10 key characteristics of carcinogens defined in Smith *et al*. (2016) and currently being used by the International Agency for Research on Cancer (IARC) in its evaluations of mechanistic evidence. However, the ISA does not fully consider the quality, external validity, and relevance of the evidence regarding these characteristics.

As discussed in Section 2, while the draft ISA includes some overarching principles regarding evaluating the quality of experimental animal studies (but notably not *in vitro* studies), it does not provide detailed quality criteria or appear to consider study quality when integrating evidence. As outlined in Section 4.3, multiple aspects in the design, implementation, and analysis of animal toxicity studies can impact study quality, and the draft ISA should have considered these factors and their potential influence on the interpretation of results.

Even setting this aside, the draft ISA does not consider that many of these key characteristics of carcinogens are also common to substances that do not cause cancer, so their presence does not necessarily support a causal association (Goodman and Lynch, 2017). In fact, the ability to predict cancer hazard by these key characteristics, either alone or in combination, has been demonstrated to be no better than chance (Becker *et al*., 2017). Yet, the draft ISA relies on experimental studies as evidence supporting upstream events (*e.g*., oxidative stress), and does not acknowledge that these events are not necessarily indicative of carcinogenesis. That is to say, the draft ISA does not consider the biological relevance of specific endpoints to humans. It can be difficult to determine whether observed biological perturbations represent homeostatic changes or molecular initiating events that may lead to cancer (Miller *et al*., 2016). This is particularly true for genotoxicity, where different assays differ in their predictive ability with regard to cancer (as discussed in Section 2.2.1).

Furthermore, as discussed by Dr. Sabine Lange (CASAC, 2018):

> The notes that $PM_{2.5}$ components can be mutagenic in vitro, and that one study showed enhanced urethane-induced tumors, does not complete a pathway of biological plausibility of $PM_{2.5}$ causing cancer. And it doesn't provide a plausibility pathway for mortality associated with relatively short-term $PM_{2.5}$ exposure – months to several years. Those are two different things and should be separated out, because some genotoxicity in the years before cancer death aren't contributing to that cancer formation.

Finally, as presented in Table 10-8 of the draft ISA, the exposure doses tested in the animal toxicity studies were generally at least an order of magnitude higher than what is usually experienced by the general populations in the US – a critical issue that the draft ISA does not consider. The draft ISA also does not discuss the possibility that the observed effects are only manifest at high exposure doses when the cellular defensive mechanisms are overwhelmed.

6 Long-term UFP Exposure and Neurological Effects

The draft ISA concludes that there is a likely causal relationship between long-term UFP exposure and nervous system effects, primarily based on animal toxicity studies of inflammation, oxidative stress, morphologic changes in the brain, cognitive and behavioral effects, and neurodevelopmental effects. The draft ISA does not consider study quality when integrating the evidence, however. In addition, the findings from animal toxicity studies do not necessarily provide evidence for apical endpoints and may have limited relevance to humans. In light of these issues, the evidence is inadequate to infer a causal relationship between long-term exposure to UFP and neurological effects.

Dr. Sabine Lange came to a similar conclusion in her comments on the draft ISA (CASAC, 2018). She stated:

> Likely to be causal 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. That type of information would make a far stronger case for this causality determination that is supposed to be applicable to humans at ambient concentrations… [T]he neurodevelopmental data [aren't] 'extensive' – it is just two studies but subdivided into half a dozen papers.

Regarding long-term exposure to UFP, the draft ISA indicates that "[t]he strongest evidence is provided by animal toxicological studies showing inflammation, oxidative stress, and neurodegeneration in adult mice and Alzheimer's disease pathology in a susceptible animal model. In addition, pre- and early postnatal exposure to UFP results in behavioral effects, inflammation, and persistent morphologic changes" (Section 8.6.7, P. 8-104). However, as with the toxicity evidence for the effects of long-term exposure to $PM_{2.5}$ discussed above in Section 4.3, the draft ISA does not consider the quality and relevance of the toxicity evidence for effects of long-term exposure to UFP.

The draft ISA does not discuss the quality of toxicity studies; rather, the results of individual studies are apparently taken at face value without any discussion of study- and endpoint-specific methodological limitations. The draft ISA also does not consider study quality when integrating evidence across endpoints and studies. Table 8-38 in the draft ISA lists key evidence from UFP toxicity studies, with a footnote indicating that such evidence can be supporting or contradicting, yet only positive results are included in this table. There is no discussion of the basis for considering evidence from a particular study to be key evidence, such as study quality or other considerations.

Although Table A-1 in Appendix 1 of the draft ISA discusses a number of quality-related considerations for evaluating animal toxicology evidence on UFP health effects (including study design, test model, pollutant, exposure assignment, outcome assessment/evaluation, potential copollutant confounding, other confounding factors, and statistical methodology), as discussed above in Section 2, these criteria are not sufficiently detailed or prescriptive to ensure a consistent evaluation across studies and endpoints. It is unclear whether any of these considerations were incorporated in the evaluation of individual UFP studies, as they are not discussed for any of these studies in the draft ISA.

In addition, the draft ISA does not sufficiently consider the relevance of exposure doses in experimental studies. For example, Table 8-38 in the draft ISA presents the UFP concentrations associated with the

various neurological effects reported in toxicological studies. All of these concentrations are generally an order of magnitude higher than the current $PM_{2.5}$ NAAQS of 12 μg/m^3, of which there is a variable fraction of UFPs. It is unclear whether similar molecular and cellular events occur in humans exposed to the UFP fraction at ambient $PM_{2.5}$ concentrations.

Further, the draft ISA does not fully consider the human relevance of the observed neurological endpoints or animal models tested. Many outcomes measured are upstream events, not apical effects. As discussed above in Section 4.3, the draft ISA does not address whether they were homeostatic changes or reversible effects; the detection of an upstream event alone does not necessarily indicate pathogenesis or disease onset.

For example, the draft ISA discusses several studies that evaluated the expression of genes related to inflammation in various sections of the brain, but these studies did not provide an indication as to whether there was confirmatory evidence for inflammation in the brain. Tyler *et al*. (2016) reported increased expression of several genes related to inflammation in mice exposed to UFPs, but stated that there were only minimal inflammatory effects observed in bronchoalveolar lavage fluid (BALF). This does not fit with the biological pathway for nervous system effects of long-term UFP exposure proposed in the draft ISA (*i.e*., pulmonary inflammation leads to systemic inflammation and neuroinflammation) and calls into question the reliability of the gene expression data for predicting apical effects.

Finally, the draft ISA does not fully discuss the consistency of the various neurological endpoints reported in the UFP toxicity studies. Most of the effects were evaluated in only one study and need to be confirmed in other studies before firm conclusions on causality can be made. The draft ISA also does not explicitly state how the different endpoints are related to each other and whether the results across endpoints are consistent with a particular outcome. For example, there is no discussion of whether the particular cognitive and behavioral effects observed in mice would be expected from the reported morphological changes in the brain. Rather, the integration of the evidence across studies consists of a summary of positive results with no clear indication of their relevance to each other.

Overall, the draft ISA should have assessed these issues when evaluating the toxicity evidence for a causal determination on nervous system effects from long-term exposure to UFP. In light of these issues, the evidence is inadequate to infer a causal relationship between long-term exposure to UFP and neurological effects.

7 PM and Welfare Effects

7.1 Overarching Issues

As discussed in Section 2, the draft ISA lacks a detailed systematic review protocol. No information is given regarding literature search strategy; study inclusion and exclusion criteria; a process for data extraction and quality control; specific, prescriptive criteria for evaluating study quality; methods for data analyses; or PM-specific methods for evidence integration and causality determinations.

In addition, the draft ISA does not consider study quality when evaluating evidence on welfare effects of PM. Specifically:

- While the Preamble to the ISAs discusses several quality considerations for evaluating studies of welfare effects, these considerations do not include all of the methodological aspects that may impact the interpretation of the study results. In addition, these considerations are not sufficiently detailed to allow a systematic and transparent evaluation of individual study quality.
- In the draft ISA, there are no specific quality criteria for studies of welfare effects, like those for studies of health effects in Appendix 1. Evaluation of individual studies of welfare effects is highly descriptive without much, if any, discussion on study quality.
- Individual study quality is not appropriately evaluated in the draft ISA; consequently, study quality is not sufficiently considered when integrating the evidence across studies and endpoints. There is no indication that the draft ISA gives higher-quality studies more weight or considers the overall WoE in the causal determination.

The draft ISA does not explicitly specify relevance criteria for studies of welfare effects. There is little discussion regarding whether study findings from various PM concentrations, experimental approaches, and measured outcomes are applicable to welfare effects of ambient PM in the US.

Similar to the causal framework for health effects, the causal framework for welfare effects is biased towards a causal relationship. EPA should update the causal framework for welfare effects to the IOM framework, as well.

Unlike the evaluation of health effects, the draft ISA's evaluation of welfare effects generally discusses PM as a whole without considering different size fractions. The draft ISA should conduct separate analyses for different size fractions (*i.e.*, PM_{10}, $PM_{2.5\text{-}10}$, $PM_{2.5}$, UFP) and various welfare effects (*i.e.*, visibility, climate, and effects on materials).

7.2 Visibility Impairment

As in the 2009 PM ISA, the draft ISA concludes that there is a causal relationship between PM exposure and visibility impairment. The draft ISA states that visibility impairment by atmospheric PM, with strongest effects in the size range of 0.1 to 1.0 μm, is supported by historical data, as well as more recent studies that are based on measurements of $PM_{2.5}$ and light extinction. However, many aspects of visibility are dependent upon weather, which also introduces uncertainty.

The Interagency Monitoring of Protected Visual Environments (IMPROVE) algorithm models PM effects on light extinction and has changed since the last PM review cycle (*i.e.*, addition of a sea salt term, calculation that relates particulate organic matter concentration from organic carbon concentration, elevation and mean temperature variable for gas scattering). However, it is not clear whether modeled visibility impairment takes into account the variability between species, region, season, and whether a location is urban or rural. Because many processes that influence PM are strongly affected by the weather, a focused effort to include meteorological processes into the algorithm is necessary to interpret model outputs.

In addition, light extinction efficiencies can be highly variable between species (up to a factor of 10 has been reported, as shown in Figure 13-1 in the draft ISA). PM species vary by region and season and by whether a location is urban or rural, and this also can impact light extinction. It is unclear if these large variabilities were taken into account in the assessment of PM's effects on visibility.

It is clear that there is a generic causal relationship between PM exposure and visibility impairment. However, the exposure levels as a function of size fractions ($PM_{2.5}$, $PM_{2.5\text{-}10}$, and PM_{10}) are not well characterized. Because it is not known which specific PM size fractions cause visibility impairment, the draft ISA should acknowledge this uncertainty and the fact that this endpoint cannot be used as the basis of a quantitative risk assessment.

7.3 Effects on Materials

The 2009 and current draft ISA conclude that there is a causal relationship between PM exposure and effects on materials. The 2009 ISA focused on examining PM impacts on stone used for historic monuments and buildings. The current draft ISA presents new information for glass and metals, including modeling of glass soiling and identifying which pollutants are most influential in metal corrosion in a multipollutant environment, and how that varies between metals. The draft ISA indicates that new research supports a causal relationship for the deposition of PM on metals, building materials, and glass.

There are several aspects of the assessment that remain unclear, including exposure-response relationships, damage functions, and interaction of copollutants. In addition, some uncertainties remain, such as quantitative relationships between particle concentration and frequency of repair, deposition rates of airborne PM to surfaces, and the interaction of copollutants with regard to materials damage effects.

Thus, while the evidence supports a causal association between PM and effects on materials, it is not clear which size fractions cause the effects or at what exposure level this occurs. Thus, similar to visibility impairment, the draft ISA should acknowledge the uncertainty pertaining to size fraction and that effects on materials should not be used in a quantitative risk assessment.

8 Recommendations for EPA

EPA should address three overarching issues in the draft ISA that undermine its evaluations of health and welfare effects. These relate to the systematic review protocol, study quality and relevance, and the causality framework. Specifically, EPA should:

- Include and follow a sufficiently detailed systematic review protocol;
- Sufficiently address study quality by providing detailed study quality criteria, tabulating study quality characteristics for individual studies, and specifying how individual study quality impacts evidence integration;
- Explicitly state study relevance criteria; and
- Update the causal framework in such a way that does not inherently bias towards a causal conclusion.

EPA should re-evaluate causality once these overarching issues with the evaluation process are addressed. While re-evaluating all endpoints may not be feasible, EPA should at least re-evaluate the associations that form the basis of the NAAQS and for which causal conclusions in the current draft ISA differ from those in the 2009 ISA. These include long-term $PM_{2.5}$ exposure and total mortality, nervous system effects, and cancer, and long-term UFP exposure and nervous system effects.

Furthermore, the current lack of a thorough, systematic study quality evaluation was noted by CASAC in its review of the PM IRP (CASAC, 2016) and the draft ISA (CASAC, 2018), and is a serious issue for determining causation, and it is even more problematic in the context of concentration-response relationships. This is because for causal determinations, studies need to establish the presence of an effect; however, for concentration-response relationships, studies also need to calculate the magnitude of an effect in relation to the level of exposure. EPA should conduct and include in the ISA a thorough, systematic quality evaluation of studies of concentration-response relationships between PM exposures and mortality, and fully consider the impact of potential biases and uncertainties on the study results.

Finally, EPA should consider and discuss in the ISA the uncertainties associated with PM size fractions, which preclude visibility impairment and effects on materials from being used in a quantitative risk assessment.

These recommendations will allow EPA to evaluate and integrate the evidence in a transparent, systematic, and unbiased manner. As a result, the causal determinations for health and welfare effects will not be inherently biased towards causation, and undue confidence will not be placed in observational concentration-response data that have substantial uncertainties.

References

American Cancer Society (ACS). 2016a. "Non-small cell lung cancer risk factors." May 16. Accessed at http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-risk-factors.

American Cancer Society (ACS). 2016b. "Small cell lung cancer risk factors." February 26. Accessed at http://www.cancer.org/cancer/lungcancer-smallcell/detailedguide/small-cell-lung-cancer-risk-factors.

Becker, RA; Dreier, DA; Manibusan, MK; Cox, LAT; Simon, TW; Bus, JS. 2017. "How well can carcinogenicity be predicted by high throughput 'characteristics of carcinogens' mechanistic data?" *Regul. Toxicol. Pharmacol.* 90:185-196. doi: 10.1016/j.yrtph.2017.08.021.

Beelen, R; Hoek, G; van den Brandt, PA; Goldbohm, RA; Fischer, P; Schouten, LJ; Armstrong, B; Brunekreef, B. 2008. "Long-term exposure to traffic-related air pollution and lung cancer risk." *Epidemiology* 19(5):702-710. doi: 10.1097/EDE.0b013e318181b3ca.

Beelen, R; Raaschou-Nielsen, O; Stafoggia, M; *et al*. 2014. "Effects of long-term exposure to air pollution on natural-cause mortality: An analysis of 22 European cohorts within the multicentre ESCAPE project." *Lancet* 383(9919):785-795. doi: 10.1016/S0140-6736(13)62158-3.

Beronius, A; Molander, L; Zilliacus, J; Ruden, C; Hanberg, A. 2018. "Testing and refining the Science in Risk Assessment and Policy (SciRAP) web-based platform for evaluating the reliability and relevance of in vivo toxicity studies." *J. Appl. Toxicol.* doi: 10.1002/jat.3648.

Brunekreef, B; Beelen, R; Hoek, G; Schouten, L; Bausch-Goldbohm, S; Fischer, P; Armstrong, B; Hughes, E; Jerrett, M; van den Brandt, P. March 2009. "Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study." HEI Research Report 139. 106p.

Cacciottolo, M; Wang, X; Driscoll, I; Woodward, N; Saffari, A; Reyes, J; Serre, ML; Vizuete, W; Sioutas, C; Morgan, TE; Gatz, M; Chui, HC; Shumaker, SA; Resnick, SM; Espeland, MA; Finch, CE; Chen, JC. 2017. "Particulate air pollutants, APOE alleles and their contributions to cognitive impairment in older women and to amyloidogenesis in experimental models." *Transl. Psychiatry* 7(1):e1022. doi: 10.1038/tp.2016.280.

Carey, IM; Atkinson, RW; Kent, AJ; van Staa, T; Cook, DG; Anderson, HR. 2013. "Mortality associations with long-term exposure to outdoor air pollution in a national English cohort." *Am. J. Respir. Crit. Care Med.* 187(11):1226-1233.

Casanova, R; Wang, X; Reyes, J; Akita, Y; Serre, ML; Vizuete, W; Chui, HC; Driscoll, I; Resnick, SM; Espeland, MA; Chen, JC. 2016. "A voxel-based morphometry study reveals local brain structural alterations associated with ambient fine particles in older women." *Front. Hum. Neurosci.* 10:495. doi: 10.3389/fnhum.2016.00495.

Centers for Disease Control and Prevention (CDC). 2013. "Lung cancer - What are the risk factors?" Accessed on July 29, 2014 at http://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm.

Cesaroni, G; Badaloni, C; Gariazzo, C; Stafoggia, M; Sozzi, R; Davoli, M; Forastiere, F. 2013. "Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome." *Environ. Health Perspect.* 121(3):324-331. doi: 10.1289/ehp.1205862.

Cesaroni, G; Forastiere, F; Stafoggia, M; et al. 2014. "Long term exposure to ambient air pollution and incidence of acute coronary events: Prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project." *BMJ* 348:f7412. doi: 10.1136/bmj.f7412.

Chen, H; Burnett, RT; Copes, R; Kwong, JC; Villeneuve, PJ; Goldberg, MS; Brook, RD; van Donkelaar, A; Jerrett, M; Martin, RV; Brook, JR; Kopp, A; Tu, JV. 2016. "Ambient fine particulate matter and mortality among survivors of myocardial infarction: Population-based cohort study." *Environ. Health Perspect.* 124(9):1421-1428. doi: 10.1289/EHP185.

Chen, H; Burnett, RT; Kwong, JC; Villeneuve, PJ; Goldberg, MS; Brook, RD; Van Donkelaar, A; Jerrett, M; Martin, RV; Kopp, A; Brook, JR; Copes, R. 2014. "Spatial association between ambient fine particulate matter and incident hypertension." *Circulation* 129(5):562-569. doi: 10.1161/CIRCULATIONAHA.113.003532.

Chen, JC; Wang, X; Wellenius, GA; Serre, ML; Driscoll, I; Casanova, R; McArdle, JJ; Manson, JE; Chui, HC; Espeland, MA. 2015. "Ambient air pollution and neurotoxicity on brain structure: Evidence from Women's Health Initiative Memory Study." *Ann. Neurol.* 78(3):466-476. doi: 10.1002/ana.24460.

Clean Air Scientific Advisory Committee (CASAC). 2016. Letter to G. McCarthy (US EPA) re: CASAC Review of the EPA's Integrated Review Plan for the National Ambient Air Quality Standards for Particulate Matter (External Review Draft – April 2016). EPA-CASAC-16-003. 68p., August 31. Accessed at https://yosemite.epa.gov/sab/sabproduct.nsf/ea5d9a9b55cc319285256cbd005a472e/34ccb8a1e98e12e885257fee00485971%21OpenDocument.

Clean Air Scientific Advisory Committee (CASAC). 2018. "Preliminary Comments from Members of the CASAC on EPA's Integrated Science Assessment for Particulate Matter (External Review Draft – October 2018) Received as of 11-27-18." 114p., December 10. Accessed on December 10, 2018 at https://yosemite.epa.gov/sab/sabproduct.nsf/E9B02B3DC858DC258525835F005FAC0A/$File/Preliminary+CASAC+PM+ISA+Comments-121018.pdf.

Corvi, R; Madia, F. 2017. "In vitro genotoxicity testing - Can the performance be enhanced?" *Food Chem. Toxicol.* 106(Part B):600-608. doi: 10.1016/j.fct.2016.08.024.

Cox, LAT. 2018. "Effects of exposure estimation errors on estimated exposure-response relations for PM2.5." *Environ. Res.* 164:636-646. doi: 10.1016/j.envres.2018.03.038.

Crouse, DL; Peters, PA; Hystad, P; Brook, JR; van Donkelaar, A; Martin, RV; Villeneuve, PJ; Jerrett, M; Goldberg, MS; Pope, CA III; Brauer, M; Brook, RD; Robichaud, A; Menard, R; Burnett, RT. 2015. "Ambient PM2.5, O3, and NO2 exposures and associations with mortality over 16 years of follow-up in the Canadian Census Health and Environment Cohort (CanCHEC)." *Environ. Health Perspect.* 123(11):1180-1186. doi: 10.1289/ehp.1409276.

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. 2012. "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.* 120(5):708-714. doi: 10.1289/ehp.1104049.

Dearfield, KL; Cimino, MC; McCarroll, NE; Mauer, I; Valcovic, LR. 2002. "Genotoxicity risk assessment: a proposed classification strategy." *Mutat. Res.* 521(1-2):121-135.

Dearfield, KL; Moore, MM. 2005. "Use of genetic toxicology information for risk assessment." *Environ. Mol. Mutagen.* 46(4):362-390.

Di, Q; Dai, L; Wang, Y; Zanobetti, A; Choirat, C; Schwartz, JD; Dominici, F. 2017b. "Association of short-term exposure to air pollution with mortality in older adults." *JAMA* 318(24):2446-2456. doi: 10.1001/jama.2017.17923.

Di, Q; Wang, Y; Zanobetti, A; Wang, Y; Koutrakis, P; Choirat, C; Dominici, F; Schwartz, JD. 2017a. "Air pollution and mortality in the Medicare population." *N. Engl. J. Med.* 376(26):2513-2522. doi: 10.1056/NEJMoa1702747.

Dorans, KS; Wilker, EH; Li, W; Rice, MB; Ljungman, PL; Schwartz, J; Coull, BA; Kloog, I; Koutrakis, P; D'Agostino, RB Sr.; Massaro, JM; Hoffmann, U; O'Donnell, CJ; Mittleman, MA. 2016. "Residential proximity to major roads, exposure to fine particulate matter, and coronary artery calcium: The Framingham Heart Study." *Arterioscler. Thromb. Vasc. Biol.* 36(8):1679-1685. doi: 10.1161/ATVBAHA.116.307141.

Duning, T; Kloska, S; Steinstrater, O; Kugel, H; Heindel, W; Knecht, S. 2005. "Dehydration confounds the assessment of brain atrophy." *Neurology* 64(3):548-550. doi: 10.1212/01.WNL.0000150542.16969.CC.

Eastmond, DA; Hartwig, A; Anderson, D; Anwar, WA; Cimino, MC; Dobrev, I; Douglas, GR; Nohmi, T; Phillips, DH; Vickers, C. 2009. "Mutagenicity testing for chemical risk assessment: Update of the WHO/IPCS Harmonized Scheme." *Mutagenesis* 24(4):341-349.

Gan, WQ; Koehoorn, M; Davies, HW; Demers, PA; Tamburic, L; Brauer, M. 2011. "Long-term exposure to traffic-related air pollution and the risk of coronary heart disease hospitalization and mortality." *Environ. Health Perspect.* 119(4):501-507.

Gharibvand, L; Shavlik, D; Ghamsary, M; Beeson, WL; Soret, S; Knutsen, R; Knutsen, SF. 2017. "The association between ambient fine particulate air pollution and lung cancer incidence: Results from the AHSMOG-2 Study." *Environ. Health Perspect.* 125(3):378-384. doi: 10.1289/EHP124.

Goodman, J; Lynch, H. 2017. "Improving the International Agency for Research on Cancer's consideration of mechanistic evidence." *Toxicol. Appl. Pharmacol.* 319:39-46. doi: 10.1016/j.taap.2017.01.020.

Goodman, JE; Zu, K; Loftus, CT; Lynch, HN; Prueitt, RL; Mohar, I; Shubin, SP; Sax, SN. 2018. "Short-term ozone exposure and asthma severity: Weight-of-evidence analysis." *Environ. Res.* 160:391-397. doi: 10.1016/j.envres.2017.10.018.

Gu, J; Hua, F; Zhong, D; Chen, J; Liu, H; Zhou, Q. 2010. "Systemic review of the relationship between family history of lung cancer and lung cancer risk." *Chin. J. Lung Cancer* 13(3):224-229.

Hart, JE; Garshick, E; Dockery, DW; Smith, TJ; Ryan, L; Laden, F. 2011. "Long-term ambient multipollutant exposures and mortality." *Am. J. Respir. Crit. Care Med.* 183(1):73-78. doi: 10.1164/rccm.200912-1903OC.

Hart, JE; Spiegelman, D; Beelen, R; Hoek, G; Brunekreef, B; Schouten, LJ; van den Brandt, P. 2015. "Long-term ambient residential traffic–related exposures and measurement error–adjusted risk of incident lung cancer in the Netherlands Cohort Study on diet and cancer." *Environ. Health Perspect.* 123:860-866. doi: 10.1289/ehp.1408762.

Hystad, P; Demers, PA; Johnson, KC; Carpiano, RM; Brauer, M. 2013. "Long-term residential exposure to air pollution and lung cancer risk." *Epidemiology* 24(5):762-772. doi: 10.1097/EDE.0b013e3182949ae7.

Institute of Medicine (IOM). 2008. "Improving the Presumptive Disability Decision-Making Process for Veterans." Committee on Evaluation of the Presumptive Disability Decision-Making Process for Veterans, Board on Military and Veterans Health, National Academies Press (Washington, DC), 781p. Accessed at http://books.nap.edu/openbook.php?record_id=11908&.

Jerrett, M; Burnett, RT; Beckerman, BS; Turner, MC; Krewski, D; Thurston, G; Martin, RV; van Donkelaar, A; Hughes, E; Shi, Y; Gapstur, SM; Thun, MJ; Pope, CA III. 2013. "Spatial analysis of air pollution and mortality in California." *Am. J. Respir. Crit. Care Med.* 188(5):593-599. doi: 10.1164/rccm.201303-0609OC.

Jerrett, M; Turner, M; Beckerman, BS; Pope, CA III; van Donkelaar, A; Martin, RV; Serre, M; Crouse, D; Gapstur, SM; Krewski, D; Diver, WR; Coogan, PF; Thurston, GD; Burnett, RT. 2017. "Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates." *Environ. Health Perspect.* 125(4):552-559. doi: 10.1289/EHP575.

Kaufman, JD; Adar, SD; Barr, RG; Budoff, M; Burke, GL; Curl, CL; *et al*. 2016. "Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): A longitudinal cohort study." *Lancet* 388(10045):696-704. doi: 10.1016/S0140-6736(16)00378-0.

Krewski, D; Jerrett, M; Burnett, RT; Ma, R; Hughes, E; Shi, Y; Turner, MC; Pope, A; Thurston, G; Calle, EE; Thun, MJ. May 2009. "Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." HEI Research Report 140. 154p.

Laden, F; Schwartz, J; Speizer, FE; Dockery, DW. 2006. "Reduction in fine particulate air pollution and mortality: Extended follow-up to the Harvard Six Cities Study." *Am. J. Respir. Crit. Care Med.* 173(6):667-672.

Lee, M; Koutrakis, P; Coull, B; Kloog, I; Schwartz, J. 2016. "Acute effect of fine particulate matter on mortality in three Southeastern states from 2007-2011." *J. Expo. Anal. Environ. Epidemiol.* 26(2):173-179. doi: 10.1038/jes.2015.47.

Lepeule, J; Laden, F; Dockery, D; Schwartz, J. 2012. "Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities study from 1974 to 2009." *Environ. Health Perspect.* 120(7):965-970. doi: 10.1289/ehp.1104660.

Lipsett, MJ; Ostro, BD; Reynolds, P; Goldberg, D; Hertz, A; Jerrett, M; Smith, DF; Garcia, C; Chang, ET; Bernstein, L. 2011. "Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort." *Am. J. Respir. Crit. Care Med.* 184(7):828-835.

Loop, MS; Kent, ST; Al-Hamdan, MZ; Crosson, WL; Estes, SM; Estes, MG Jr.; Quattrochi, DA; Hemmings, SN; Wadley, VG; McClure, LA. 2013. "Fine particulate matter and incident cognitive impairment in the REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort." *PLoS ONE* 8(9):e75001. doi: 10.1371/journal.pone.0075001.

Maclaren, J; Han, Z; Vos, SB; Fischbein, N; Bammer, R. 2014. "Reliability of brain volume measurements: A test-retest dataset." *Sci. Data* 1:140037. doi: 10.1038/sdata.2014.37.

Matakidou, A; Eisen, T; Houlston, RS. 2005. "Systematic review of the relationship between family history and lung cancer risk." *Br. J. Cancer* 93(7):825-833.

McDonnell, WF; Nishino-Ishikawa, N; Petersen, FF; Chen, LH; Abbey, DE. 2000. "Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers." *J. Expo. Anal. Environ. Epidemiol.* 10(5):427-436.

Miller, KA; Siscovick, DS; Sheppard, L; Shepherd, K; Sullivan, JH; Anderson, GL; Kaufman, JD. 2007. "Long-term exposure to air pollution and incidence of cardiovascular events in women." *N. Engl. J. Med.* 356(5):447-458.

Miller, MF; Goodson, WH III; Manjili, MH; Kleinstreuer, N; Bisson, WH; Lowe, L. 2016. "Low-Dose Mixture Hypothesis of Carcinogenesis Workshop: Scientific underpinnings and research recommendations." *Environ. Health Perspect.* doi: 10.1289/EHP411.

Nadler, DL; Zurbenko, IG. 2014. "Estimating cancer latency times using a Weibull model." *Adv. Epidemiol.* 2014:746769. doi: 10.1155/2014/746769.

Naess, O; Nafstad, P; Aamodt, G; Claussen, B; Rosland, P. 2007. "Relation between concentration of air pollution and cause-specific mortality: Four-year exposures to nitrogen dioxide and particulate matter pollutants in 470 neighborhoods in Oslo, Norway." *Am. J. Epidemiol.* 165(4):435-443.

Nakamura, K; Brown, RA; Araujo, D; Narayanan, S; Arnold, DL. 2014. "Correlation between brain volume change and T2 relaxation time induced by dehydration and rehydration: Implications for monitoring atrophy in clinical studies." *Neuroimage Clin.* 6:166-170. doi: 10.1016/j.nicl.2014.08.014.

National Toxicology Program (NTP). 2011. *Report on Carcinogens (Twelfth Edition)*. 499p. Accessed at http://ntp.niehs.nih.gov/go/roc12.

Pinault, L; Tjepkema, M; Crouse, DL; Weichenthal, S; van Donkelaar, A; Martin, RV; Brauer, M; Chen, H; Burnett, RT. 2016. "Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort." *Environ. Health* 15:18. doi: 10.1186/s12940-016-0111-6.

Pope, CA; Thun, MJ; Namboodiri, MM; Dockery, DW; Evans, JS; Speizer, FE; Heath, CW. 1995. "Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults." *Am. J. Respir. Crit. Care Med.* 151(3 pt. 1):669-674. doi: 10.1164/ajrccm/151.3_Pt_1.669.

Pottenger, LH; Bus, JS; Gollapudi, BB. 2007. "Genetic toxicity assessment: Employing the best science for human safety evaluation Part VI: When salt and sugar and vegetables are positive, how can genotoxicity data serve to inform risk assessment?" *Toxicol. Sci.* 98(2):327-331.

Puett, RC; Hart, JE; Yanosky, JD; Spiegelman, D; Wang, M; Fisher, JA; Hong, B; Laden, F. 2014. "Particulate matter air pollution exposure, distance to road, and incident lung cancer in the Nurses' Health Study cohort." *Environ. Health Perspect.* 122(9):926-932. doi: 10.1289/ehp.1307490.

Raaschou-Nielsen, O; Andersen, ZJ; Beelen, R; Samoli, E; Stafoggia, M; Weinmayr, G; *et al*. 2013. "Air pollution and lung cancer incidence in 17 European cohorts: Prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)." *Lancet Oncol.* 14(9):813-822. doi: 10.1016/S1470-2045(13)70279-1.

Raaschou-Nielsen, O; Beelen, R; Wang, M; *et al*. 2016. "Particulate matter air pollution components and risk for lung cancer." *Environ. Int.* 87:66-73. doi: 10.1016/j.envint.2015.11.007.

Reardon, PK; Seidlitz, J; Vandekar, S; Liu, S; Patel, R; Park, MTM; Alexander-Bloch, A; Clasen, LS; Blumenthal, JD; Lalonde, FM; Giedd, JN; Gur, RC; Gur, RE; Lerch, JP; Chakravarty, MM; Satterthwaite, TD; Shinohara, RT; Raznahan, A. 2018. "Normative brain size variation and brain shape diversity in humans." *Science* 360(6394):1222-1227. doi: 10.1126/science.aar2578.

Rhomberg, LR; Chandalia, JK; Long, CM; Goodman, JE. 2011. "Measurement error in environmental epidemiology and the shape of exposure-response curves." *Crit. Rev. Toxicol.* 41(8):651-671.

Samoli, E; Stafoggia, M; Rodopoulou, S; Ostro, B; Declercq, C; Alessandrini, E; Diaz, J; Karanasiou, A; Kelessis, AG; Le Tertre, A; Pandolfi, P; Randi, G; Scarinzi, C; Zauli-Sajani, S; Katsouyanni, K; Forastiere, F; MED-PARTICLES Study Group. 2013. "Associations between fine and coarse particles and mortality in Mediterranean cities: Results from the MED-PARTICLES project." *Environ. Health Perspect.* 121(8):932-938. doi: 10.1289/ehp.1206124.

Shi, L; Zanobetti, A; Kloog, I; Coull, BA; Koutrakis, P; Melly, SJ; Schwartz, JD. 2016. "Low-concentration PM2.5 and mortality: Estimating acute and chronic effects in a population-based study." *Environ. Health Perspect.* 124(1):46-52. doi: 10.1289/ehp.1409111.

Shinohara, RT; Oh, J; Nair, G; Calabresi, PA; Davatzikos, C; Doshi, J; Henry, RG; Kim, G; Linn, KA; Papinutto, N; Pelletier, D; Pham, DL; Reich, DS; Rooney, W; Roy, S; Stern, W; Tummala, S; Yousuf, F; Zhu, A; Sicotte, NL; Bakshi, R; NAIMS Cooperative. 2017. "Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis." *AJNR Am. J. Neuroradiol.* 38(8):1501-1509. doi: 10.3174/ajnr.A5254.

Smith, MT; Guyton, KZ; Gibbons, CF; Fritz, JM; Portier, CJ; Rusyn, I; DeMarini, DM; Caldwell, JC; Kavlock, RJ; Lambert, P; Hecht, SS; Bucher, JR; Stewart, BW; Baan, R; Cogliano, VJ; Straif, K. 2016. "Key characteristics of carcinogens as a basis for organizing data on mechanisms of carcinogenesis." *Environ. Health Perspect.* 124(6):713-721. doi: 10.1289/ehp.1509912.

Steel, N; Ford, JA; Newton, JN; *et al*. 2018. "Changes in health in the countries of the UK and 150 English Local Authority areas 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016." Lancet 392 (10158) : 1647-1661. doi: 10.1016/S0140-6736(18)32207-4.

Stringhini, S; Carmeli, C; Jokela, M; *et al*. 2017. "Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: A multicohort study and meta-analysis of 1·7 million men and women." Lancet 389 (10075) : 1229-1237. doi: 10.1016/S0140-6736(16)32380-7.

Thun, MJ; Carter, BD; Feskanich, D; Freedman, ND; Prentice, R; Lopez, AD; Hartge, P; Gapstur, SM. 2013. "50-year trends in smoking-related mortality in the United States." *N. Engl. J. Med.* 368(4):351-364. doi: 10.1056/NEJMsa1211127.

Thurston, GD; Ahn, J; Cromar, KR; Shao, Y; Reynolds, HR; Jerrett, M; Lim, CC; Shanley, R; Park, Y; Hayes, RB. 2016. "Ambient particulate matter air pollution exposure and mortality in the NIH-AARP Diet and Health Cohort." *Environ. Health Perspect.* 124(4):484-490. doi: 10.1289/ehp.1509676.

Thurston, GD; Ito, K; Lall, R; Burnett, RT; Turner, MC; Krewski, D; Shi, Y; Jerrett, M; Gapstur, SM; Diver, WR; Pope, CA III. 2013. "NPACT Study 4. Mortality and long-term exposure to PM2.5 and its components in the American Cancer Society's Cancer Prevention Study II Cohort." HEI Research Report 177 Part 4. In National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiologic and Toxicologic Studies of the Health Effects of Particulate Matter Components. Health Effects Institute, Boston, MA. p127-166. Accessed at https://www.healtheffects.org/publication/national-particle-component-toxicity-npact-initiative-integrated-epidemiologic-and.

Tomczak, A; Miller, AB; Weichenthal, SA; To, T; Wall, C; van Donkelaar, A; Martin, RV; Crouse, DL; Villeneuve, PJ. 2016. "Long-term exposure to fine particulate matter air pollution and the risk of lung cancer among participants of the Canadian National Breast Screening Study." *Int. J. Cancer* 139(9):1958-1966. doi: 10.1002/ijc.30255.

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. doi: 10.1097/EDE.0000000000000144.

Turner, MC; Cohen, A; Jerrett, M; Gapstur, SM; Diver, WR; Pope, CA III; Krewski, D; Beckerman, BS; Samet, JM. 2014. "Interactions between cigarette smoking and fine particulate matter in the Risk of Lung Cancer Mortality in Cancer Prevention Study II." *Am. J. Epidemiol.* 180(12):1145-1149. doi: 10.1093/aje/kwu275.

Turner, MC; Jerrett, M; Pope, CA III; Krewski, D; Gapstur, SM; Diver, WR; Beckerman, BS; Marshall, JD; Su, J; Crouse, DL; Burnett, RT. 2016. "Long-term ozone exposure and mortality in a large prospective study." *Am. J. Respir. Crit. Care Med.* 193(10):1134-1142. doi: 10.1164/rccm.201508-1633OC.

Turner, MC; Krewski, D; Pope, CA III; Chen, Y; Gapstur, SM; Thun, MJ. 2011. "Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers." *Am. J. Respir. Crit. Care Med.* 184(12):1374-1381. doi: 10.1164/rccm.201106-1011OC.

Tyler, CR; Zychowski, KE; Sanchez, BN; Rivero, V; Lucas, S; Herbert, G; Liu, J; Irshad, H; McDonald, JD; Bleske, BE; Campen, MJ. 2016. "Surface area-dependence of gas-particle interactions influences pulmonary and neuroinflammatory outcomes." *Part. Fibre Toxicol.* 13(1):64. doi: 10.1186/s12989-016-0177-x.

US EPA. 2005. "Guidelines for Carcinogen Risk Assessment." Risk Assessment Forum, EPA/630/P-03/001F, 166p., March.

US EPA. 2009. "Integrated Science Assessment for Particulate Matter (Final)." Office of Research and Development, National Center for Environmental Assessment (NCEA) - RTP Division, EPA/600/R-08/139F. 2228p., December. Accessed at http://www.epa.gov/ncea/pdfs/partmatt/Dec2009/PM_ISA_full.pdf.

US EPA. 2015a. "Preamble to the Integrated Science Assessments." EPA/600/R-15/067. November. Accessed on December 1, 2015 at http://www.epa.gov/isa.

US EPA. 2015b. "Integrated Science Assessment for Oxides of Nitrogen–Health Criteria (Second External Review Draft)." National Center for Environmental Assessment (NCEA), EPA/600/R-14/006. 1135p., January. Accessed on February 01, 2015 at http://www.epa.gov/ncea/isa.

US EPA. 2018. "Integrated Science Assessment for Particulate Matter (External Review Draft)." National Center for Environmental Assessment (NCEA) - RTP Division, EPA/600/R-18/179. 1881p., October. Accessed on October 24, 2018 at https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=341593.

US Public Health Service. 2006. "The Health Consequences of Involuntary Exposure to Tobacco Smoke - A Report of the Surgeon General." Office of the Surgeon General. 727p.

Villeneuve, PJ; Weichenthal, SA; Crouse, D; Miller, AB; To, T; Martin, RV; van Donkelaar, A; Wall, C; Burnett, RT. 2015. "Long-term exposure to fine particulate matter air pollution and mortality among Canadian women." *Epidemiology* 26(4):536-545. doi: 10.1097/EDE.0000000000000294.

Weichenthal, S; Crouse, DL; Pinault, L; Godri-Pollitt, K; Lavigne, E; Evans, G; van Donkelaar, A; Martin, RV; Burnett, RT. 2016. "Oxidative burden of fine particulate air pollution and risk of cause-specific mortality in the Canadian Census Health and Environment Cohort (CanCHEC)." *Environ. Res.* 146:92-99. doi: 10.1016/j.envres.2015.12.013.

Weichenthal, S; Villeneuve, PJ; Burnett, RT; Van Donkelaar A; Martin, RV; Jones, RR; DellaValle, CT; Sandler, DP; Ward, MH; Hoppin, JA. 2014. "Long-term exposure to fine particulate matter: Association with nonaccidental and cardiovascular mortality in the agricultural health study cohort." *Environ. Health Perspect.* 122(6):609-615. doi: 10.1289/ehp.1307277.

Weuve, J; Puett, RC; Schwartz, J; Yanosky, JD; Laden, F; Grodstein, F. 2012. "Exposure to particulate air pollution and cognitive decline in older women." *Arch. Intern. Med.* 172(3):219-227. doi: 10.1001/archinternmed.2011.683.

Wilker, EH; Preis, SR; Beiser, AS; Wolf, PA; Au, R; Kloog, I; Li, W; Schwartz, J; Koutrakis, P; DeCarli, C; Seshadri, S; Mittleman, MA. 2015. "Long-term exposure to fine particulate matter, residential proximity to major roads and measures of brain structure." *Stroke* 46(5):1161-1166. doi: 10.1161/STROKEAHA.114.008348.

Wong, CM; Tsang, H; Lai, HK; Thomas, GN; Lam, KB; Chan, KP; Zheng, Q; Ayres, JG; Lee, SY; Lam, TH; Thach, TQ. 2016. "Cancer mortality risks from long-term exposure to ambient fine particle." *Cancer Epidemiol. Biomarkers Prev.* 25(5):839-845. doi: 10.1158/1055-9965.EPI-15-0626.

World Health Organization (WHO). 2007. "Mutagenicity Testing for Chemical Risk Assessment (Draft)." 31p., September.

Zu, K; Shi, L; Prueitt, RL; Liu, X; Goodman, JE. 2018. "Critical review of long-term ozone exposure and asthma development." *Inhal. Toxicol.* 30(3):99-113. doi: 10.1080/08958378.2018.1455772.