NCASI

IMPACT. SCIENCE. SOLUTIONS.

Proposed Systematic Review Methodology for the Particulate Matter Integrated Science Assessment

EXECUTIVE SUMMARY

In October, 2018, the US EPA published the External Review Draft of the Integrated Science Assessment for Particulate Matter (PM ISA) in order to fulfill the requirement of the Clean Air Act under Section 109(d)(1) that mandates the US EPA administrator complete a thorough review of the criteria published under Section 108 and the national ambient air quality standards every 5 years, and to make such revisions in criteria and standards and promulgate such new standards as may be appropriate. In this effort, the PM ISA reviews thousands of potentially relevant scientific articles related to particulate matter exposure and the potential for human health impact.

While the PM ISA successfully summarizes a substantial amount of literature related to the potential human health impact of particulate matter exposure, the approach used to critically evaluate and integrate lines of evidence produced in the literature search lacks several critical features of modern systematic review practices. Areas of the PM ISA that demonstrate the potential for improvement include the selection of policy relevant research questions, the evaluation of risk of bias, and prescribed approaches to the integration of evidence. Use of modern systematic review approaches in these areas will improve both the transparency of how scientific conclusions are arrived at in the PM ISA process, as well as the reliability of the conclusions drawn in the PM ISA that will ultimately be relied on for policy decision making.

In an effort to further the dialog with US EPA staff scientists on efforts to continually improve upon the ISA process and ensure that best science is used to inform public health policy, NCASI staff in collaboration with subject matter experts, have developed an example of a systematic review protocol that demonstrates how modern systematic review approaches would improve and potentially alter the conclusions drawn by a contemporary analysis of the particulate matter human health literature.

Please find attached a proposed systematic review protocol to evaluate the impact of PM2.5 exposure on the outcomes of mortality and ischemic heart disease. In the appendices, two scientific articles highlighted in the External Review Draft PM ISA as being salient to conclusions drawn regarding the relationship between PM2.5 exposure these outcomes are treated with the approach outlined in the protocol as a demonstration of the impact of these modern systematic review methodologies on the interpretation of scientific evidence.

NATIONAL COUNCIL FOR AIR AND STREAM IMPROVEMENT, INC.

402 SW 140th Terrace
Newberry, FL 32669-3000

(352) 331-1745
(352) 331-1766
ncasi.org

Submitted respectfully,

Giffe Johnson, PhD
Principal Scientist
Toxicology, Epidemiology, and Risk Assessment
National Council for Air and Stream Improvement (NCASI)

Protocol for a Systematic Review to Evaluate the Potential Association Between Respiratory and Cardiovascular Related Mortality or Ischemic Heart Disease and Exposure to Policy Relevant Concentrations of PM2.5.

Table of Contents

1.0 INTRODUCTION
1.1 Background and rationale
1.2 Objectives (specific aims)

2.0 METHODS
2.1 Eligibility criteria (types of studies)
- PECOS statement
- Exclusion criteria and prioritization (duplicates, not primary data, no relevant exposure, no relevant outcome)

2.2 Search strategy
- Databases
- Search terms and/or search strings/query translations from PubMed
- Screening studies for eligibility
- Study Selection

2.3 Data extraction
- See example templates for human and animal studies

2.4 Assess Quality of Evidence / Risk of Bias
- OHAT Risk of Bias Tool (see example template; modify with explicit criteria)
- Discuss responses to OHAT risk of bias tool (e.g., high risk of bias, probably high risk, unclear, etc.)

2.5 Structure for the Body of Evidence
- Summary of findings table (organized by the primary outcomes

2.8 Evidence Integration and Conclusion Development
- Risk of bias across studies
- Unexplained inconsistency
- Indirectness
- Imprecision
- Publication bias

3.0 REFERENCES

APPENDICES

1.0 INTRODUCTION

1.1 Background and rationale

In developing the Integrated Science Assessment (ISA) for Particulate Matter (PM), the U.S. EPA reviews and summarizes the evidence from studies on atmospheric sciences, dosimetry, human exposure, animal toxicology, mode of action, controlled human exposure, epidemiology, biogeochemistry, and/or terrestrial and aquatic ecology and other welfare effects in order to inform risk management and policy decisions under the National Ambient Air Quality Standards (NAAQS). Two principal documents, the NAAQS Integrated Review Plan (US EPA, 2016) and the NAAQS Preamble (US EPA, 2015), discuss overarching principles and provide guidance for conducting reviews under the ISA process and suggest the incorporation of some elements of systematic review, but fall short of prescribing the use of a fully featured systematic review framework to address key policy questions. In addition, the Clean Air Act places strict deadlines on EPA to complete the ISA process, making it an impracticable endeavor to evaluate the numerous lines of scientific evidence that inform the health effects of PM exposures within a robust systematic review framework, while adhering to modern practices of literature selection, study quality evaluation, and evidence integration.

However, in the absence of such a framework, the process becomes less robust and reliable conclusions cannot be reached regarding the relationship between exposure and health outcomes. This is largely a result of the inherent limitations of epidemiological studies in describing well defined exposure-response relationships at low levels of exposure amid the substantial methodological issues of controlling uncertainty under those research conditions. A body of evidence of this nature requires detailed and well-defined criteria of evaluation in order to reach scientifically defensible conclusions. EPA's recent evaluation of potential health impacts from exposures to decreasing concentrations of particulate matter ≤2.5 microns in diameter (PM2.5) within the framework of the ISA by using traditional methods has this drawback.

Formal systematic review protocols are one way to balance the need for efficient, yet detailed and transparent, literature reviews within ISAs, while conforming to modern approaches for the literature selection, study quality assessment, and evidence integration. These protocols should address key policy relevant research questions within the broader context of scientific literature as a whole. The purpose of the systematic review protocol presented here is to provide a framework for evaluating the relationship between policy relevant exposure concentrations of PM2.5 and three specific outcomes: respiratory and cardiovascular-related mortality and ischemic heart disease (IHD) to serve as a model that EPA can apply under the NAAQS ISA framework. The goal is to provide EPA with the necessary tools for the review of research questions that are targeted and narrower in scope, but still require robust conclusions with a high degree of certainty to inform policy decisions.

1.2 Objectives (specific aims)

The primary objectives of this document are 1) to provide a protocol for addressing two policy relevant research questions using contemporary approaches to systematic review; and 2) to demonstrate applicability of this protocol to address a specific area in the NAAQS PM ISA. This is consistent with stated goals of the Integrated Review Plan (IRP) (3.1 Scope of the PM ISA):

In order to provide a more focused evaluation of the scientific evidence for health and non-ecological welfare related effects, the PM ISA will discuss the most important topics that address policy-relevant questions. Therefore, the PM ISA will more fully evaluate those health and non-ecological welfare effects for which the evidence in the 2009 PM ISA was less certain (i.e., effects where the causal determination was "likely to be causal", "suggestive", or "inadequate" as detailed below in section 3.4.3) and where there is now a larger body of evidence (e.g., diabetes, nervous system effects, etc.). For those health and non-ecological welfare effects where the 2009 PM ISA concluded that the evidence was sufficient to infer a causal relationship (i.e., health: short- and long-term PM2.5 exposures and cardiovascular effects; short- and long-term PM2.5 exposures and mortality; and welfare: PM exposures and effects on visibility, climate, and materials), the PM ISA will focus more specifically on policy-relevant considerations, such as the level at which effects are

observed, and on characterizing the extent to which new studies address key uncertainties and limitations identified in the previous review or provide insight on new issues [p. 3-2]
The specific aims of the study protocol are to provide a systematic review and evidence integration framework to address the following research questions:

1) Among the adult population, are respiratory and cardiovascular mortality effects, observed in studies accounting for confounding and other biases, related to exposures to annual average PM2.5 < 12 μg/m^3 (equivalent to the current PM2.5 NAAQS)?

and;

2) Among the adult population, are ischemic heart disease-related emergency department visit effects, observed in studies accounting for confounding and other biases, related to exposures to annual average PM2.5 < 12 μg/m^3 (equivalent to the current PM2.5 NAAQS)?

While previous ISAs have described the potential causal link between PM2.5 exposure and the outcomes of mortality and ischemic heart disease, these causal links were not focused on exposure concentrations relevant to current PM2.5 NAAQS standards. Therefore, while EPA has heretofore classified these outcomes as being causally associated with PM2.5 exposures, the approach has ignored the exposure concentration. The protocol proposed here specifically addresses this issue, seeking to study whether effects are observed at levels below the current NAAQS, in studies in which confounding, bias, and chance are not likely to explain the observed association.

2.0 METHODS

2.1 Eligibility criteria (types of studies)

"Participants," "Exposure," "Comparator," "Outcomes," and "Study Design" (PECOS) statements were developed for each health outcome included in the framework. Studies will be selected for inclusion according to the criteria specified in the PECOS statements below for the outcome of cardiovascular and respiratory mortality (Table 1) and the outcome of ischemic heart disease (Table 2). Note that only human studies will be included for the mortality outcomes, whereas animal and mechanistic studies will be included for the IHD outcome.

Table 1: PECOS Statement for the Outcome of Cardiovascular and Respiratory Mortality

Population	Adult population (18 years or older)
Exposure	Particulate matter up to 2.5 µm in diameter (PM2.5) through inhalation route at or below exposure concentrations at 12 ug/m^3 or below; studies that rely on surrogate or indirect exposure measurements (e.g. distance from exposure source) will be excluded
Comparator	Population of the same demographics as target population with PM2.5 exposures that include discrete concentration ranges below the current NAAQS standards to be compared; for time series studies, other days exposed will be the comparator
Outcome	Cardiovascular- and respiratory-related mortality as indicated by hospital death certificates
Study Design	Studies that compare discrete categories of exposure at and below the current NAAQS standard in a prospective cohort, retrospective cohort, case cross-over, or time series designs. Only studies with estimates of relative risk in relation to quantitative estimates of exposure will be included.

Table 2: PECOS Statement for the Outcome of Ischemic Heart Disease (IHD)

	Human	Animal	In Vitro
Population	Adult population (18 years and above)	Rodents or primates of adult age	Human or animal cardiac tissue or cells in full state of maturation (e.g. no partially differentiated tissue)
Exposure	Particulate matter up to 2.5 µm in diameter (PM2.5) through inhalation route at a concentration of 12 ug/m^3 or below	Particulate matter up to 2.5 µm in diameter (PM2.5) through inhalation route at a human extrapolation equivalence concentration of 12 ug/m^3 or below	Particulate matter up to 2.5 µm in diameter (PM2.5) through airborne exposure or aqueous solution
Comparator	Population of the same demographics as target population with PM2.5 exposures that include discrete concentration ranges below 12 ug/m^3; for time series studies the comparator metric is days of exposure	Animal group of the same species, sex, and age, exposed to filtered air controls	Unexposed human or animal cardiac tissue or cells in full state of maturation (e.g. no partially differentiated tissue)
Outcome	Cardiovascular-related emergency department visits and hospital admissions IHD	Pathophysiological determination of IHD based on chronic narrowing of a coronary artery by deposition of atheroma plaques or occlusion by thrombosis or relevant biomarkers of IHD (e.g. troponin and AST)	Elevated biomarkers of IHD e.g. troponin and AST
Study Design	Studies that compare discrete categories of exposure at and below the current NAAQS standard in a prospective cohort, retrospective cohort, or time series design	Studies of adequate statistical power that use appropriate pathophysiologic determination of outcome and a negative control group	Studies of adequate statistical power that use appropriate biochemical determination of outcome and a negative control group

Publications to be excluded include: reviews and secondary research, editorials, and studies that analyze surrogates of exposure to PM2.5 (e.g., distance between residence and roadways, traffic counts), irrelevant exposures or outcomes, or that are absent stated requirements in the PECOS tables.

Additional studies that will be used to inform the analysis but are separate from the formal systematic review are studies that inform on the potential for uncertainty, confounding, or publication bias among the selected studies. These studies will be incorporated into the evidence integration process to assist in evaluating the overall degree of uncertainty present in the body of evidence.

2.2 Search strategy

The search for relevant articles will be conducted using the EPA HERO database, as well as PubMED, SCOPUS, and BASE and the following keywords 'PM2.5', 'particulate matter', 'mortality', 'ischemic heart disease', 'heart disease', 'cardiovascular disease', 'respiratory mortality' and 'cardiovascular mortality' with duplicates removed.
The flow of record search to final inclusion will proceed as shown in Figure 1A found in the appendices.

2.3 Data extraction

Data extraction will be conducted using the appropriate template by study type as shown in the appendices.

2.4 Assess Quality of Evidence / Risk of Bias

The risk of bias for each included study is assessed using a modified instrument developed based on the OHAT Risk of Bias tool for the generally accepted risk of bias domains: selection bias, confounding, performance bias, attrition bias, detection bias and reporting bias (NTP 2019). The tool is modified to address methods for assessing PM exposure to include considerations of modeling, monitoring, and other methodologies.

For each domain, OHAT criteria, modified to address PM ecological studies and specific issues related to PM health effects evaluations, is used. Table 2 lists the specific domains, domain questions and judgement criteria. The risk of bias is assessed for each study question using a rating system with four categories as follows: low risk of bias, probably low risk of bias, probably high risk of bias, and high risk of bias. Although general criteria are provided for each category, it is expected that each reviewer will provide supporting rationale for each rating given to each study.

Two reviewers will independently rate each study for each of the domains and the reviewers will discuss and resolve any discrepancies.

The templates for Risk of Bias profiles are presented separately for human (Table S1) and animal/mechanistic studies (Table 1A) in the appendices.

Risk of Bias Profile and Judgements for each Study

Not all risk of bias domains should be considered of equal importance in the overall evaluation of study quality. Studies can be placed into tiers that reflect the two most important domains: the quality of the exposure assessment and the consideration of confounding and effect modification.

For example, to be considered a higher quality study (Tier 1), a human study must be rated as "definitely low" or "probably low" for both of the following risk of bias domains:

- Exposure assessment (can we be confident in the exposure characterization?)
- Confounding (does the study design or analysis account for important confounding and modifying variables?)

Similarly, an animal study will be considered a higher quality study (Tier 1) when the following risk of bias domains are judged "definitely low" or "probably low" for both:

- Exposure relevance to humans (are exposures relevant to exposures humans would experience?)
- Controls (are animal controls adequate?)

In contrast, a human study will be considered a lower quality study (Tier 3) when the risk of bias domains are rated as “definitely high” or “probably high” for both:

- Exposure assessment
- Confounding

Similarly, an animal study will be considered a lower quality study (Tier 3) when the following risk of bias domains are judged “definitely high” or “probably high” for:

- Controls are inadequate

Other studies will be considered to be Tier 2 studies when either one of the above critical domains is “probably high” or “definitely high”, or other domain considerations help inform the reviewer that additional sources of bias exist in a study.

With regards to the exposure characterization, health impacts must be presented in relation to quantitative metrics of PM2.5 and higher quality studies will rely upon better measures of exposure, for example, personal exposure monitoring is preferred; exposure estimated using personal activity records (for example) to modify measurements from centrally located monitors or modeling are also acceptable. Studies in which exposure is measured ecologically (that is, estimated using measurements from one or more central monitors) will be considered to be of the lowest tier (Tier 3).

With regards to confounding/effect modification, Tier 1 studies will include studies where confounding was considered and effect modification was evaluated, for example including age, gender, socioeconomic status, meteorological parameters, respiratory or influenza patterns, and co-pollutant exposures. Reviewers will also evaluate and note evidence of unmeasured confounding or uncertainty relevant to a particular study from the broader literature search in order to bring this finding into the evidence integration process.

2.5 Structure for the Body of Evidence

The structure of an evaluation of overall body of evidence will include a heat map based on the risk of bias profiles for each study (see appendices). In particular, the risk of bias considerations will determine which studies fall into each study quality Tier. Only studies that fall into Tier 1 and 2 will be considered for evaluation in the evidence integration (see below) process to evaluate the relationship between policy relevant exposures to PM2.5 and target health endpoints. Table 3 below illustrates the organization of evidence obtained from literature selection and risk of bias analysis. This organization depicts relevant domains of study quality and relevance in order to draw conclusions regarding the confidence of the available evidence for addressing the proposed research questions. A brief rationale is provided in the table, we more detailed discussion in the text. A similar table would be developed for IHD and include the evidence from epidemiological, animal and mechanistic studies.

Table 3. Synthesis of Evidence for Epidemiological Studies

Outcome	Number of epidemiology studies	Overall risk of bias	Inconsistency of the evidence	Indirectness of evidence	Publication bias	Overall conclusion	Confidence in the conclusion
Cardiovascular OR Respiratory mortality from long-term exposure	X studies List type of study design	Severe OR Not severe Include rationale based on the heat map, other external evidence of confounding or other bias(discordant findings	Severe OR Not severe Include heterogeneity of results	Severe OR not severe Include rationale for rating, e.g., outcome data not assessed individually; and/or exposure not assessed	Detected OR Not detected OR Not assessed Include rationale, for example, meta-analyses that report evidence of bias OR no evaluations conducted.	Increases in cardiovascular mortality below 12 μg/m^3 … Include strength of the effect?	Slight/Moderate/Robust/Indeterminate/Compelling evidence of no effect Overall confidence in conclusion based on risk of bias , consistency, strength, publication bias.

2.6 Evidence Integration and Conclusions

Once the results of individual studies have been assessed in the context of methodological strengths and limitations, they should be integrated both within and across evidence realms. For mortality outcomes, only epidemiological studies will be assessed and thus evidence integration across realms will not be conducted. For IHD, Although the vast majority of studies on criteria air pollutants are expected to be epidemiology studies, but various types of experimental studies, including animal studies including controlled human exposure studies, toxicity studies, and mechanistic/mode-of-action (MoA) studies, should will be considered for certain health outcomes (e.g., for IHD in this case). Table 4 shows how these latter studies should could be evaluated to determine whether they support causation and human relevance for IHD; Table 5 shows how this should be applied to determine biological plausibility in humans.

Table 4. Criteria for Evaluating Experimental Studies

Criterion	Considerations
Causation	Consistency: Repeatability of key events and effects across species/study designs Magnitude: Large, considering type of effect, background prevalence, species and dose range, exposure pattern Essentiality: Reversibility of effects if exposure is stopped or a key event prevented Specificity: Apical effect is likely to occur following key event Temporality: Observation of key events in a hypothesized order, before toxicity is apparent Exposure-response: Key events observed at exposures below or similar to those associated with the adverse effect Biological Concordance: Proposed mode of action is consistent with current biological knowledge of the toxicological outcome Analogy: Proposed mode of action is consistent with what is known for other related chemicals with a well-defined mode of action
Human Relevance	Relevant groups and life stages Comparative developmental processes and their relative timing Differences in ontogeny that affect dose metrics (*e.g.*, placental or lactational transfer, key metabolic enzymes) Consequences of interaction of chemical with cells, tissues, and organs Magnitude of exposure differences for observation of key events or apical outcome

Notes:
(a) Adapted from Boobis *et al.* (2008), Meek *et al.* (2014), and NTP (2019).

Table 5. Confidence in Biological Plausibility

	Human Relevance	Inadequate Evidence for Human Relevance	No Human Relevance
Supports effects at low levels	High	Moderate	Inadequate
Inadequate Evidence for effects at low levels	Inadequate	Inadequate	Inadequate
Supports Lack of effects at low levels	High (not plausible)	Moderate (not plausible)	Inadequate

Table 4 is based on both the International Programme on Chemical Safety (IPCS) Mode-of-Action/Human Relevance framework (Boobis *et al*., 2008; Meek *et al*., 2014) and the Office of Health Assessment and Translation (OHAT) framework for evaluating the confidence in the toxicology body of literature (NTP, 2019). The IPCS framework is intended to evaluate key events from mechanistic and MoA studies; however, we have modified the framework so that it can also be used to evaluate toxicity studies that assess apical effects and whether reported effects in these studies (or a lack thereof) are consistent and coherent with key events identified in mechanistic and MoA studies. We note this framework is quite similar to the criteria OHAT uses to evaluate confidence in the evidence (*i.e*., risk of bias, unexplained inconsistency, indirectness, imprecision, magnitude, dose-response, consistency across models/species/related outcomes). All of these criteria are incorporated into Table 4 except for risk of bias; this is because risk of bias should be determined at the individual study level and should be incorporated in the initial evaluation of study results (*i.e*., results of studies with moderate or high risk of bias should not be considered reliable and, as such, should not be considered when evidence is being integrated).

If all aspects of causation in Table 4 are met, one can conclude that the evidence supports causation in these experimental systems. If not all are met, then one must explain whether it is

more likely that the evidence as a whole supports causation (*i.e*., one must provide likely explanations for any aspect not being met), that the evidence supports no causation, or that the evidence is inadequate to determine causation. One should evaluate human relevance in a similar manner. Once this is complete, one can determine whether there is high, moderate, or inadequate confidence in biological plausibility, or whether the evidence indicates high or moderate confidence for a lack of biological plausibility, as shown in Table 5.

This biological plausibility assessment is then incorporated into the evaluation of epidemiology evidence, using modified Bradford Hill aspects. The Integrated Science Assessment (ISA) Preamble notes that the Bradford Hill aspects provide a framework for assessing evidence but should not be considered as fixed rules of evidence (*i.e*., a checklist) for developing causal conclusions (US EPA, 2015). Rather, they provide a framework for systematic evaluation of the weight of the evidence (WoE) for inferring causality. As such, not meeting one or more of the aspects does not necessarily preclude a judgment of causality.

The aspects as listed in the Preamble are shown in Table 6. We have modified the explanations of each aspect to be more succinct. The experimental evidence provides information on the likelihood of biological plausibility, as demonstrated in Table 6, and is informative regarding coherence (*i.e*., whether all the evidence fits together). Similar to the framework for experimental studies in Table 4, if all Bradford Hill aspects are met, one can conclude that the evidence as a whole supports causation, or in the case of the current protocol, supports a conclusion of effect induction at target exposure concentrations. If not all are met, then one must explain whether it is most likely that the evidence as a whole supports causation (*i.e*., one must provide likely explanations for any aspect that is not met), is suggestive of causation, supports no causation, or is inadequate to determine causation.

Table 6. Criteria for Evidence Integration

	Criteria for Evidence Integration
Aspect	**Explanation**
Strength of Association	Large and precise risk estimates are less likely to be due to chance, bias, or other factors
Consistency	Evidence is stronger if consistent effects are observed among studies of different designs, people, places, circumstances, and times
Specificity	Evidence is stronger when disease is specific to an exposure or exposure is specific to disease
Temporality	Exposure must precede the occurrence of disease
Exposure-response	Evidence is stronger when a well-characterized exposure-response relationship exists (*e.g.*, disease risk increases with greater exposure intensity and/or duration) -– at the relevant exposure levels (in this case below the current NAAQS)
Biological Plausibility	Evidence on the biological mechanism of an effect allows a scientifically defensible determination for causation at relevant exposure levels
Coherence	All of the known facts related to the observed association from the various evidence streams fit together in a coherent manner
Experiment	"Natural experiments" can provide strong evidence when an intervention or cessation of exposure results in a change in disease risks
Analogy	Evidence is stronger when a similar substance is an established causal factor for a similar effect

Figure 1 shows how evidence across realms should integrated to evaluate causation. The current US National Ambient Air Quality Standard (NAAQS) causal framework has five categories for causation (causal, likely causal, suggestive, inadequate, and not likely causal). Figure 1 has four categories (it does not include likely causal). The NAAQS 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. However, because all studies should be reviewed using the same criteria, it is more appropriate to 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 inadequate (IOM, 2008). With this definition of suggestive, the likely causal category is not necessary. This four-tiered framework is consistent with other causal frameworks, such as that defined in the Institute of Medicine (IOM) report titled *Improving the Presumptive Disability Decision-Making Process for Veterans* (IOM, 2008) and, notably, the framework in the ISA Preamble for potential at-risk factors.

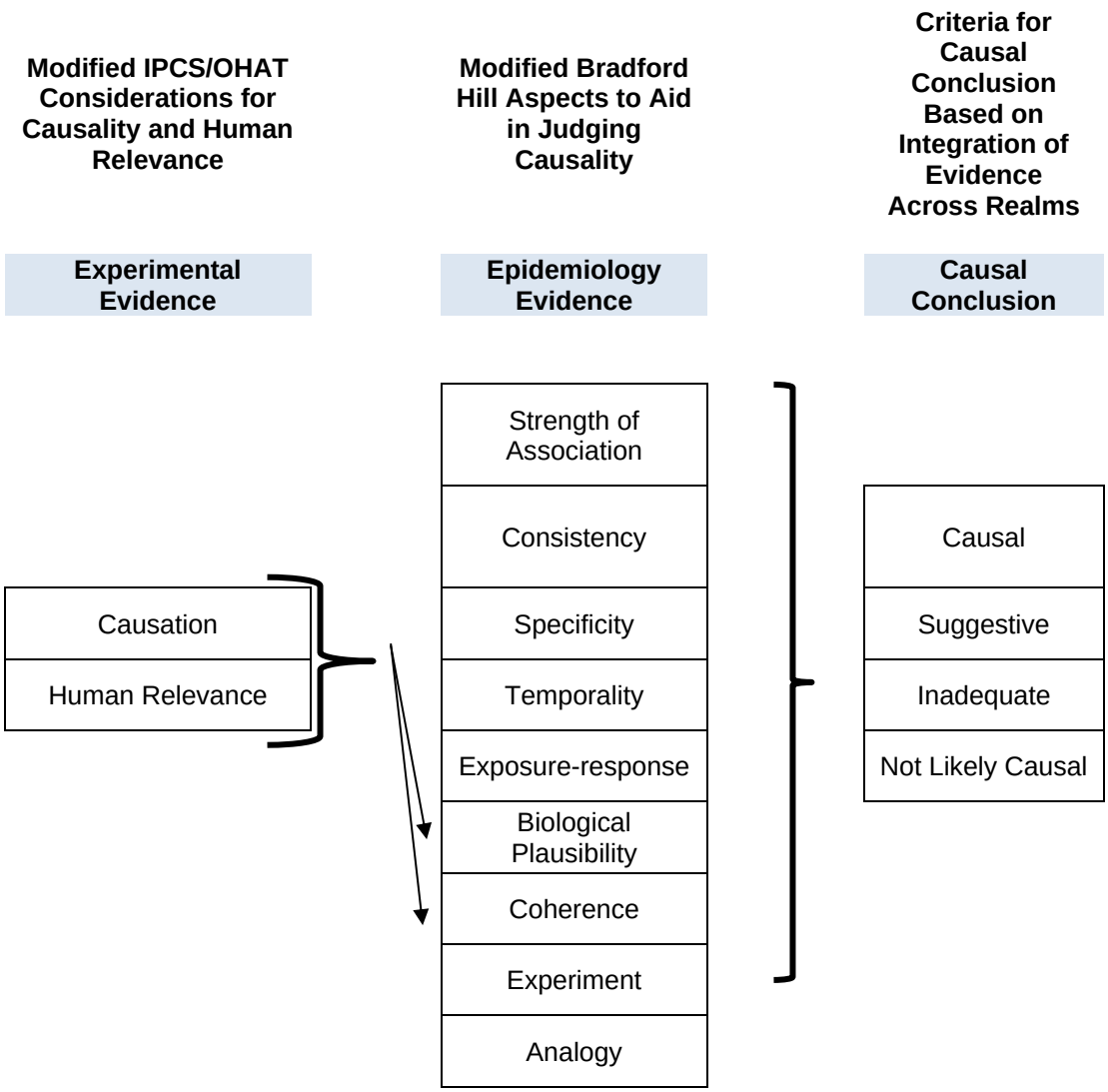


Figure 1. Causal Conclusion for an Effect Based on Experimental and Epidemiology Evidence. IPCS = International Programme on Chemical Safety; OHAT = Office of Health Assessment and Translation.

A *causal* relationship should be concluded when all modified Bradford Hill aspects are met or most are met, and there is a likely explanation for each that is not met. A *suggestive* relationship should be concluded when an assessment of the evidence indicates that a causal relationship is more likely than not, but some of the modified Bradford Hill aspects have inadequate information, and all other aspects are met or there is a likely explanation for each that is not met. The *inadequate* category should be concluded when most or all modified Bradford Hill aspects have inadequate information or are not met but there is no likely explanation for each that is not met. The *not likely causal* category should be concluded when evidence indicates there is no causal relationship, based on the modified Bradford Hill aspects (*e.g.*, there is a consistent lack of association in robust epidemiology studies and the experimental evidence indicates a lack of biological plausibility.

For the purpose of this protocol, which evaluates research questions related to the potential of effects occurring within specific exposure concentrations rather than questions of general causation, the question of ‘causation’ becomes significantly more narrow in scope and requires

refining the application of evidence integration in order to be effectively 'fit for purpose.' In a review that seeks to explore questions of general causation between an exposure and health effects, it is expected that a broad range of study types, and the use of a variety of methodologies, might be available to inform the conclusion of the review. However, for a protocol designed to address a specific aspect of the relationship between exposure and a health outcome within the larger issue of general causation will have more specific requirements in terms of study features and methodological choices in order to reach conclusions with high confidence.

For this protocol, evidence integration domains that provide evidence for magnitude of effect and exposure-response are essential in order to address the stated research questions. As such, methodological choices that impact the interpretation of these domains should be documented in careful detail by reviewers to ensure an appropriate conclusion is drawn for each research question, with an accurate characterization of the confidence associated with those conclusions. Conclusions with high confidence are valuable for informing risk assessment and risk management practices. Conclusions with poor confidence are valuable for informing the research community as to the data gaps that need to be addressed in order to adequately answer the questions being asked.

3.0 REFERENCES

Boobis, AR; Doe, JE; Heinrich-Hirsch, B; Meek, ME; Munn, S; Ruchirawat, M; Schlatter, J; Seed, J; Vickers, C. 2008. "IPCS framework for analyzing the relevance of a noncancer mode of action for humans." *Crit. Rev. Toxicol.* 38(2):87-96.

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

Meek, ME; Palermo, CM; Bachman, AN; North, CM; Lewis, RJ. 2014. "Mode of action human relevance (species concordance) framework: Evolution of the Bradford Hill considerations and comparative analysis of weight of evidence." *J. Appl. Toxicol.* doi: 10.1002/jat.2984.

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., https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf.

US EPA. 2015. "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. 2016. "Integrated Review Plan for the National Ambient Air Quality Standards for Particulate Matter." EPA-452/R-16-005, December. Available at https://www3.epa.gov/ttn/naaqs/standards/pm/data/201612-final-integrated-review-plan.pdf

Appendices

Appendix Supplemental (S): Protocol Tables and Figures
Appendix (A) Data collection Example for Research Question 1: Mortality
Appendix (B) Data collection Example for Research Question 2: Ischemic Heart Disease

Appendix S. Protocol Tables and Figures

Figure 1A: Study selection process.

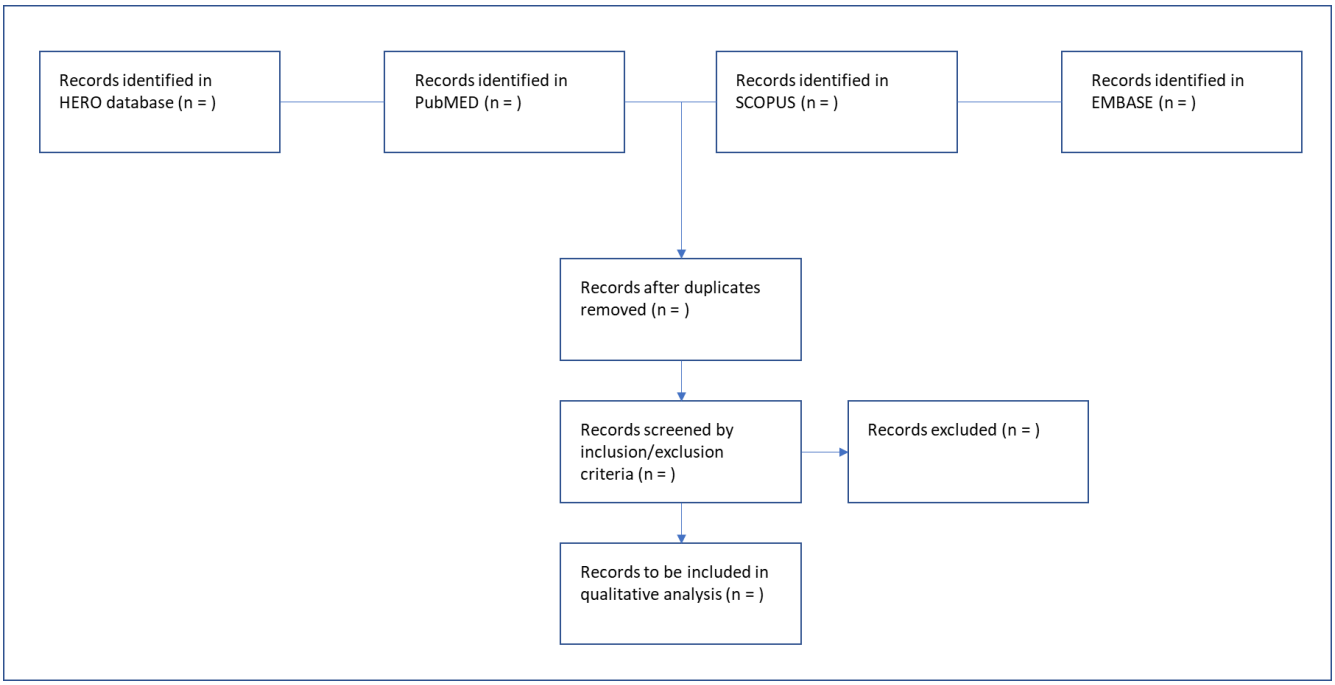


Table 1A. Risk-of-Bias Guidelines

Study type/design e.g., Co (cohort), CaCo (case-control), CrSe (cross-sectional), CaS (case series/case report), Eco (Ecological), EA (experimental animal)

Rating (low, prob low, prob high, high risk, unclear, N/A)

For more details refer to OHAT Risk of Bias Tool (NTP, 2013)
https://ntp.niehs.nih.gov/ntp/ohat/pubs/riskofbiastool_508.pdf

Bias domain **Question**	**Judgment guidelines**
Selection bias	
1. Did selection of study participants in appropriate comparison groups? Applies to: Co, CaCo, CrSe, Eco	Low/prob low – **Cohort, Cross-Sectional**: There is direct/indirect evidence that subjects (both exposed and nonexposed) were similar (e.g., recruited from the same eligible population, recruited with the same method of ascertainment using the same inclusion and exclusion criteria, and were of similar age and health status), recruited within the same time frame, and had the similar participation/response rates. **Case Control**: There is direct/indirect evidence that cases and controls were similar (e.g., recruited from the same eligible population including being of similar age, gender, ethnicity, and eligibility criteria other than outcome of interest as appropriate), recruited within the same time frame, and controls are described as having no history of the outcome. Note: A study will be considered low risk of bias if baseline characteristics of groups differed, but these differences

Bias domain Question	Judgment guidelines
	were considered as potential confounding or stratification variables (see Question 4). **Time-series**: For ecological studies, a table of information or text on potential differences in characteristics that could bias results is provided, and these characteristics are adjusted for as potential confounders. There is direct evidence that subjects (both exposure groups and referent groups) were similar (e.g., of similar geographic region, ethnicity, socioeconomic status, etc.), **OR** baseline characteristics of groups differed but these differences were considered as potential confounding or stratification variables in analyses (see Question 4). **Additional Guidance:** Comparison groups selected adequately. Study provides table of subject characteristics by exposure levels and/or by case status. Cross-sectional studies can be considered low risk of bias if a general table of subject characteristics is provided and analyses are adjusted for confounders. Prob high/high - **Cohort, Cross-Sectional**: There is indirect/direct evidence that subjects (both exposed and nonexposed) were not similar, recruited within very different time frames, or had very different participation/response rates, **OR** there is insufficient information provided about the comparison group including a different rate of nonresponse without an explanation. **Case Control**: There is indirect/direct evidence that controls were drawn from a very dissimilar population than cases or recruited within very different time frames, **OR** there is insufficient information provided about the appropriateness of controls including rate of response reported for cases only. **Time-series**: There is indirect/direct evidence that subjects (both exposure groups and referent groups) were not similar (e.g., of similar geographic region, ethnicity, socioeconomic status), **OR** there is insufficient information provided about the appropriateness of comparison groups. At least one known difference between the groups was not accounted for (e.g., the study authors acknowledged that the groups were different with respect to a variable that is a potential confounder not considered in the analysis), **OR** recruitment methods were very different (e.g., recruitment completed during different time frames, different criteria were used for recruitment).
Confounding	
4. Did the study design or analysis account for important confounding and modifying variables? Applies to: Co, CaCo, CrSe, CaS	Low/prob low – **Cohort, Cross-Sectional, Case Series/Report**: There is direct/indirect evidence that appropriate adjustments or explicit considerations were made for primary covariates and confounders in the final analyses through statistical models to reduce research-specific bias including standardization, case matching, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified. Acceptable consideration of appropriate adjustment factors includes cases when the

Bias domain Question	Judgment guidelines
	factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included. **Case Control**: There is direct/indirect evidence that appropriate adjustments were made for primary covariates and confounders in the final analyses through statistical models to reduce research specific bias including standardization, matching of cases and controls, adjustment in multivariate model, stratification, propensity scoring, or other methods were appropriately justified. **Time-series**: There is direct/indirect evidence that appropriate adjustments or explicit considerations were made for covariates and confounders in the final analyses through statistical models (e.g., standardization, multivariate adjustment). Acceptable consideration of appropriate adjustment factors includes cases when the factor is not included in the final adjustment model because the author conducted analyses that indicated it did not need to be included. **Additional Guidance:** Study adjusted for or addressed important potential confounders. Age, gender, education, and socioeconomic status are potential confounders that need to be addressed and considered in the study design or analyses. In addition, specific important confounders for this assessment depend on the health outcome (e.g., smoking for lung cancer) Other confounders might also be judged important for certain health outcomes. A low risk-of-bias rating was assigned for this question if potential confounders deemed important were adequately addressed (e.g., distribution of variables was compared between groups, and there was no statistically significant difference). Prob high/high - **Cohort, Cross-Sectional, Case Series/Report**: There is indirect/direct evidence that the distribution of primary covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, **OR** there is insufficient information provided about the distribution of known confounders. **Case Control**: There is indirect/direct evidence that the distribution of primary covariates and known confounders differed between cases and controls and was not investigated further, **OR** there is insufficient information provided about the distribution of known confounders in cases and controls. **Time-series**: There is indirect/direct evidence that the distribution of covariates and known confounders differed between the groups and was not appropriately adjusted for in the final analyses, **OR** there is insufficient information provided about the distribution of known confounders. **Additional Guidance:** Design or analysis did not adjust for important potential confounders. Adjustments were made for some potential confounders, but at least one major confounder was not addressed for a particular health outcome (e.g., no adjustment for smoking when evaluating lung cancer).

Bias domain **Question**	**Judgment guidelines**
5. **Did researchers adjust or control for other exposures that are anticipated to bias results?** Applies to: Co, CaCo, CrSe, CaS, Eco	Low/prob low – **Cohort, Case Control, Cross-Sectional, Case Series/Report/Ecological**: There is direct/indirect evidence that other exposures anticipated to bias results were not present or were appropriately adjusted for. For occupational studies or studies of contaminated sites, other chemical exposures known to be associated with those settings were appropriately considered. **Additional Guidance:** Researchers adjusted for other chemicals or accounted for occupational exposures likely to be associated with the outcome (low), **OR** it is deemed that coexposures present would not appreciably bias results (prob low). Note, as discussed above, this includes insufficient information provided on coexposures in general population studies. Prob high/high - **Cohort, Cross-Sectional, Case Series/Report**: There is indirect/direct evidence that there was an unbalanced provision of additional coexposures across the primary study groups, which were not appropriately adjusted for, **OR** there is insufficient information provided about coexposures in occupational studies or other studies where exposures to other air pollutants would have been reasonably anticipated. **Case Control**: There is indirect/direct evidence that there was an unbalanced provision of additional coexposures across cases and controls, which were not appropriately adjusted for, **OR** there is insufficient information provided about coexposures in occupational studies or other studies where exposures to other air pollutants would have been reasonably anticipated. **Ecological and Semi-individual**: There is indirect/direct evidence that there was an unbalanced provision of additional coexposures, which were not appropriately adjusted for, **OR** there is insufficient information provided about coexposures in studies where exposures to other air pollutants would have been reasonably anticipated. **Additional Guidance:** There is evidence that coexposures might not have been addressed. Examples include any study of populations that may be exposed to numerous ambient air pollutants including gases (ozone, nitrogen oxides and sulfur oxides) but coexposures to these additional air pollutants is not addressed; or a study with known coexposures, but the relevance of the coexposure to PM effects is unknown, or it is not clear if other compounds were adjusted for in the analyses. Known differential exposure to other air pollutants also associated with the health outcome of interest occurred with PM, and exposure was not addressed by the study authors.
Attrition/ Exclusion bias	
7. Were outcome data complete without attrition or exclusion from analysis? Applies to: Co, CaCo, CrSe, Eco	Low/prob low – **Cohort**: There is direct/indirect evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study. Acceptable handling of subject attrition includes: very little missing outcome data; reasons

Bias domain Question	Judgment guidelines
	for missing subjects unlikely to be related to outcome (for survival data, censoring unlikely to be introducing bias); missing outcome data balanced in numbers across study groups, with similar reasons for missing data across groups; **OR** missing data have been imputed using appropriate methods, **AND** characteristics of subjects lost to follow up or with unavailable records are described in an identical way and are not significantly different from those of the study participants. **Case Control, Cross-Sectional**: There is direct/indirect evidence that exclusion of subjects from analyses was adequately addressed, and reasons were documented when subjects were removed from the study or excluded from analyses (low) OR were deemed not to bias the results (prob low). **Time-series**: There is direct/indirect evidence that there was no loss of subjects (e.g., due to moving or migration) or data during the study and outcome data were complete, **OR** incomplete outcome data were adequately addressed, **AND** characteristics of subjects lost to follow up or with unavailable records are described in an identical way and are not significantly different from those of the study participants. **Additional Guidance:** There are no reported data lost to attrition, and the numbers in the results tables sum to the total number of subjects, **OR** less than 10% of data are missing, **OR** there are some missing outcome data but study report clearly identifies missing data and how it was handled (e.g., loss to follow-up for a cohort study is determined to be minimal if there are some missing data for either the exposure or outcome for certain subjects at a specific time measured and the authors clearly explain what happened to everyone and which results were used in the analyses). For ecological studies specifically, there are no reported data lost to attrition, **OR** there are some missing data but study report clearly identifies missing data and how they were handled (e.g., migration in and out of study area and residence location within study area were tracked and accounted for or references provided to verify that population migration within or in/out of study area is not a concern for this population), and characteristics of subjects lost to attrition do not differ significantly from those included in study. Prob high - **Cohort**: There is direct evidence that loss of subjects (i.e., incomplete outcome data) was adequately addressed and reasons were documented when human subjects were removed from a study, **OR** it is deemed that the proportion lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow-up or with unavailable records from those of the study participants. Generally, the higher the ratio of participants with missing data to participants with events, the greater potential there is for bias. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable. **Case Control, Cross-Sectional**: There is indirect evidence that exclusion of subjects from analyses was adequately

Bias domain **Question**	**Judgment guidelines**
	addressed, and reasons were documented when subjects were removed from the study or excluded from analyses. **Ecological and Semi-individual**: There is indirect evidence that there was no loss of subjects (e.g., due to migration during the study) and outcome data were complete, **OR** it is deemed that the proportion of subjects lost to follow-up would not appreciably bias results. This would include reports of no statistical differences in characteristics of subjects lost to follow-up or with unavailable records of outcomes. For studies with a long duration of follow-up, some withdrawals for such reasons are inevitable. **Additional Guidance:** No direct evidence of loss to follow-up, attrition, or loss of subjects due to migration/moving provided. The tables of results do not include the number of subjects and it is not stated that there was any loss data missing; **OR** there appear to be no or very few missing data; **OR** in a cohort study, there is no mention of loss to follow-up. High – **Cohort**: There is indirect evidence that loss of subjects (i.e., incomplete outcome data) was unacceptably large and not adequately addressed, **OR** there is insufficient information provided about numbers of subjects lost to follow-up. **Case Control, Cross-Sectional**: There is indirect evidence that exclusion of subjects from analyses was not adequately addressed, **OR** there is insufficient information provided about why subjects were removed from the study or excluded from analyses. **Ecological and Semi-individual**: There is direct/indirect evidence that incomplete outcome data (e.g., due to subject migration or moving) were unacceptably large [greater than 20% in each group] and not adequately addressed, **OR** there is insufficient information provided about missing outcome data. **Additional Guidance:** Missing outcome data with no explanation of why data were missing, and it is unclear from the characteristics table or other information provided in the report why the data might be missing.
Detection bias	
8. Can we be confident in the exposure characterization? Applies to: EA, Co, CaCo, CrSe, CaS, Eco	Low/prob low – **Cohort, Case Control, Cross-Sectional, Case Series/Report**: There is direct evidence that appropriate measurements were taken and that the most reliable methods for sampling are conducted. The best measurements would include personal measurement of PM concentrations. If fixed-site monitors are used, then modeling of PM concentrations to estimate personal exposures would be preferred. The least preferred exposure measurement would be single or multiple fixed-site monitors. If PM2.5 data are not available and must be estimated, this would add to the uncertainty in the exposure measurements. Modeled estimates should include validation of model estimates against measured concentrations. **Ecological and Semi-individual**: This rating is not applicable. Only studies with individual-level exposure characterization can earn this rating. If individual-level exposure data are provided, the study is not an ecological study, and should

Bias domain **Question**	**Judgment guidelines**
	be reclassified and rated according to other study type ROB criteria. Prob high/high - **Cohort, Case Control, Cross-Sectional, Case Series/Report**: There is direct or indirect evidence that data are based on single or a few fixed-site monitor locations that would not adequately describe personal exposures. A surrogate for PM2.5 was used to estimate concentrations (e.g., PM10) **Ecological and Semi-individual**: There is indirect/direct evidence that the chemical in question was not adequately characterized by appropriate measures and methods (e.g., no historical monitoring, isolated or remote-time samples taken to be representative of large areas).
9. Can we be confident in the outcome assessment? Applies to: Co, CaCo, CrSe, CaS, Eco	Low/prob low – **Case Control**: There is direct/indirect evidence that the outcome was assessed in cases using well-established methods (the gold standard) and subjects had been followed for the same length of time in all study groups. **Cross-Sectional, Case Series/Report**: There is direct/indirect evidence that the outcome was assessed using well-established methods (the gold standard). **Ecological and Semi-individual**: There is direct/indirect evidence that the outcome was assessed using well-established methods, the gold standard (e.g., individual-level outcome data were assessed, as in the case of semi-individual ecological studies), and subjects have been followed for the same length of time in all study groups. Acceptable assessment methods will depend on the outcome, but examples of such methods may include: objectively measured with diagnostic methods, measured by trained interviewers, obtained from reliable registries or records. **Additional Guidance:** Cancer cases are histologically confirmed, **OR** data obtained from nationwide registry are accepted as valid and complete, **OR** outcome diagnosed by physician, **OR** outcome obtained from medical record data or validated with such data (if self-reported). Prob high/high - **Case Control**: There is indirect/direct evidence that the outcome was assessed in cases using an insensitive instrument or was not adequately validated, **OR** there is insufficient information provided about how cases were identified. **Cross-Sectional, Case Series/Report**: There is indirect/direct evidence that the outcome assessment method is an insensitive instrument or was not adequately validated, **OR** there is insufficient information provided about validation of outcome assessment method. **Ecological and Semi-individual**: There is indirect/direct evidence that the authors did not validate the methods used, or the length of follow-up differed by study group, **OR** there is insufficient information provided about validation of outcome assessment method. **Additional Guidance:** Outcome is self-reported (e.g., “ever been diagnosed by a physician”) and not verified by medical records or other

Bias domain **Question**	**Judgment guidelines**
	means. There is insufficient information on quality of self-report or validation of answers. Outcome is assessed by nurses and there is no information on assessor agreement.
Selective reporting bias	
10. Were all measured outcomes reported? Applies to: HCT, Co, CaCo, CrSe, CaS	Low/prob low – **Cohort, Case Control, Cross-Sectional, Case Series/Report, Ecological**: There is direct/indirect evidence that all the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction, **OR** analyses that had not been planned at the outset of the study (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such, and it is deemed that the omitted analyses were not appropriate and selective reporting would not appreciably bias results. This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not). **Cohort, Case Control, Cross-Sectional, Case Series/Report, Ecological**: There is indirect evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported, **OR** there is insufficient information provided about selective outcome reporting. **Additional Guidance:** An outcome mentioned in a part of the study report is obviously missing from the results.
Other Bias	
11. Were there no other potential threats to internal validity (e.g., statistical methods were appropriate, and researchers adhered to the study protocol)? MODEL SPECIFICATION	Low/prob low – Cohort, Case Control, Cross-Sectional, Case Series/Report, Ecological: There is direct/indirect evidence that there was no impact of model selection on depicting the linearity or otherwise shape of the concentration response curve Additional Guidance: Taking into consideration that linear models (e.g. Cox Proportional Hazards, etc.) rely on a model assumption of linearity and evaluate model validity over a large spectrum of exposure data that may not be relevant to specific research questions of this protocol, studies that employ linear models at exposure ranges outside the range of interest may mischaracterize the concentration/response function at specific ranges of exposure and introduce bias. Prob high/high - Cohort, Case Control, Cross-Sectional, Case Series/report, Ecological: There is direct/indirect evidence that model validity was assessed over

Bias domain **Question**	**Judgment guidelines**
	exposure ranges outside the range of interest to the research question or model validity was not assessed at all. Low/prob low – **Assessment-Specific Clarification:** Statistical analyses were appropriate and no other threats to internal validity were identified. Study authors might acknowledge limitations, but these are not expected to affect the study's internal validity. There are study limitations likely to bias the results toward or away from the null, but adequate sample size was available in each cell ($n \geq 5$), **OR** sample size is small and acknowledged as a potential limitation by study authors, but significant results were still observed. Prob high/high - There are study limitations likely to bias results towards or away from the null, **OR** analyses were conducted on a small number of subjects ($n < 5$ in any given cell) and no statistically significant results were observed.
Did researchers adhere to study protocol?	Low/prob low – **Cohort, Case Control, Cross-Sectional, Case Series/Report, Ecological**: There is direct/indirect evidence that there were no deviations from the protocol (i.e., authors reported no deviations/did not report any deviations), **OR** deviations from the protocol are described and it is deemed that they would not appreciably bias results. **Additional Guidance:** Taking into consideration typical reporting practices, it seems unlikely that deviations from the protocol will be explicitly reported in most studies. Thus, unless stated otherwise by the authors (i.e., evidence of deviation is reported), or it is clear from the study report that deviations from the planned approach occurred, assume that no deviations occurred. It is anticipated that this approach will result in a rating of "probably low risk of bias" for most studies. If there are deviations, the rating reflects how the deviations changed direction, magnitude, and/or significance of the results. Prob high/high - **Cohort, Case Control, Cross-Sectional, Case Series/report, Ecological**: There is direct/indirect evidence that there were large deviations from the protocol as outlined in the methods or study report. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not prespecified or reporting outcomes not prespecified (unless clear justification for their reporting is provided, such as an unexpected effect).

Risk of Bias Profile (HUMAN)

Table S.1 Study Reference

Study Element	Description
Participants	
Exposure	
Comparator	
Outcome	
Study Design	

Bias domain	Reviewer's judgment	Support for judgment
Source population representation (Did selection of study participants result in appropriate comparison groups?)		
Confounding (Did the study design or analysis account for important confounding and modifying variables?)		
Incomplete outcome data (Were outcome data complete without attrition or exclusion from analysis?)		
Exposure assessment (Can we be confident in the exposure characterization?)		
Outcome assessment (Can we be confident in the outcome assessment?)		
Selective outcome reporting (Were all measured outcomes reported?)		
Other potential threats to internal validity (Were statistical methods appropriate? Did researchers adhere to study protocol?)		

Model Specification (Were statistical models evaluated for validity of underlying assumptions or was model validity assessed with exposure ranges outside the range of interest for the research question?)		
External evidence of bias (does external research indicate there may be unmeasured uncertainty, bias, or confounding present in the study desing or dataset?)		

Risk-of-Bias Profile (Experimental Animal)

Table S.1 Study Reference

Study Element	Description
Design	
Participants	
Exposure	
Comparator	
Outcome	

Bias domain	Reviewer's judgment	Support for judgment
Randomization of dose / exposure level (Was administered dose or exposure level adequately randomized?)		
Inadequate concealment of allocation (Was allocation to study groups adequately concealed?)		
Experimental conditions (Were experimental conditions identical across study groups?)		
Blinding (Were research personnel and human subjects blinded to study group during the study?)		
Incomplete outcome data (Were outcome data complete without attrition or exclusion from analysis?)		
Exposure assessment (Can we be confident in the exposure characterization?)		
Outcome assessment (Can we be confident in the outcome assessment?)		
Selective reporting (Were all measured outcomes reported?)		
Other potential threats to internal validity (Were statistical methods appropriate? Did researchers adhere to study protocol? Did study design or analysis account important confounding or modifying variables (including unintended co-exposures)?		

Table S2. Experimental Studies Integration

Criterion	Considerations	Reviewer's Judgement	Support for Judgement	Rating
Causation	Consistency			
	Magnitude			
	Essentiality			
	Specificity			
	Temporality			
	Exposure-response			
	Biological Concordance			
	Analogy			
Human Relevance	Relevant groups and life stages			
	Comparative developmental processes and their relative timing			
	Differences in ontogeny			
	Consequences of interaction of chemical with cells, tissues, and organs			
	Magnitude of exposure differences for observation of key events or apical outcome			
Confidence in Biological Plausibility				

Table S3. Bradford Hill Criteria for Evidence Integration

Aspect	Reviewer's Judgement	Support for Judgement	Causal Conclusion
Strength of Association			
Consistency			
Specificity			
Temporality			
Exposure-response			
Biological Plausibility			
Coherence			
Experiment			
Analogy			

Appendix A Data Collection Example for Research Question 1: Mortality

Data Collection and Summary for Kloog et al. 2013

Kloog, I; Ridgway, B; Koutrakis, P; Coull, BA; Schwartz, JD. (2013). Long- and short-term exposure to PM2.5 and mortality: Using novel exposure models. Epidemiology 24: 555-561. http://dx.doi.org/10.1097/EDE.0b013e318294beaa

Background

In the Integrated Science Assessment for Particulate Matter External Draft Review 2018, mortality is identified as a health outcome of importance causally related to PM2.5 exposure:

> As in the 2009 PM ISA, the current ISA concludes there is a "causal relationship" between short-term PM2.5 exposure and total (nonaccidental) mortality (Section 11.1). Recent multicity studies conducted in the U.S., Canada, Europe, and Asia in combination with the single- and multicity studies evaluated in the 2009 PM ISA continue to provide evidence of consistent, positive associations between short-term PM2.5 exposure and total mortality. Draft ISA p. ES-13

Of the articles cited in the Draft ISA, Kloog et al. 2013 is cited as demonstrating one of the strongest magnitudes of effect for the impact of PM2.5 on mortality at 2.8% per 10 ug/m^3 increase in concentration. However, the specific research question regarding these effects at concentrations similar to and below current NAAQS policy values has not been evaluated through systematic review. Due to the reliance upon Kloog et al. 2013 in the Draft ISA to infer that PM2.5 may impact mortality rates at policy relevant concentrations, this article was selected to provide an example of data extraction, analysis of risk of bias, and evaluate elements of evidence integration that would result from conducting a systematic review using the provided systematic review protocol. As the systematic review protocol presented here focuses on a specific aspect of causal analysis (e.g. evaluating causation relevant to existing NAAQS policy concentrations of PM2.5), processing this article through this systematic review protocol demonstrates the impact of utilizing a structured systematic review framework on the interpretation of this literature and provides insight on potential alternative conclusions that might be reached by using this framework for addressing specific, policy relevant research questions.

Risk of Bias Profile (HUMAN)

Table A.1 Kloog et al. 2013 PECOS Evaluation

Study Element	Description
Participants	468,570 death records; 46% were men, 94% were white, and 20% had higher than a high school education; average age at death was 75 years
Exposure	10 km x 10 km grid based PM2.5 estimated exposure concentration assignment
Comparator	10 km x 10 km exposure grid compares outcomes among varying estimated concentrations of PM2.5 concentrations
Outcome	Mortality
Study Design	Time-series analysis combined with chronic exposure analysis

Table A.2 Kloog et al. 2013 Risk of Bias Evaluation

Bias domain	Reviewer's judgment	Support for judgment
Source population representation (Did selection of study participants result in appropriate comparison groups?)	Probably low risk of bias	Large population across exposure ranges allows for internal comparison of relatively similar demographics amongst higher and lower exposed subjects; some risk of bias may still be present due to use of a localized population restricted to the state of Massachusetts that may not be generalizable.
Confounding (Did the study design or analysis account for important confounding and modifying variables?)	Probably high risk of bias	Socio-economic confounder covariates are estimated at tract level and not applied to individual subjects resulting in potential misclassification for these covariates. Smoking status is estimated using lung cancer as a surrogate likely resulting in misclassification for this confounder.
Incomplete outcome data (Were outcome data complete without attrition or exclusion from analysis?)	Low risk of bias	Outcome data is based on state of Massachusetts mortality records and is likely to be near complete.
Exposure assessment (Can we be confident in the exposure characterization?)	Probably high risk of bias	Prediction models estimate exposure concentrations in 10 km x 10 km grid. No subject level intraday concentration measurements are available. There is likely subject level exposure misclassification from applying tract level exposure estimations to individual subjects.
Outcome assessment (Can we be confident in the outcome assessment?)	Low risk of bias	Outcome data is based on state of Massachusetts mortality records and is likely to very accurate.
Selective outcome reporting (Were all measured outcomes reported?)	Low risk of bias	Outcome data is based on state of Massachusetts mortality records and is likely to be near complete for the stated time period.
Other potential threats to internal validity (Were statistical methods appropriate? Did researchers adhere to study protocol?)	--	--

Model Specification (Were statistical models evaluated for validity of underlying assumptions or was model validity assessed with exposure ranges outside the range of interest for the research question?)	Probably high risk of bias	Short term association with mortality was evaluated through a linear model based on 10 ug/m^3 changes in estimated PM2.5 concentration at day and with one day lag in effect. Model validity is not independently assessed at policy relevant PM2.5 concentrations and may be impacted by model validity at higher concentrations. The model specification creates a positive bias for determining statistically significant associations at the lower range of exposure estimates due to use of a model that relies on the underlying assumption of linearity.
External evidence of bias (does external research indicate there may be unmeasured uncertainty, bias, or confounding present in the study design or dataset?)	Probably high risk of bias	No attempt is made to account for and adjust for unmeasured residual confounding.

Summary of Risk of Bias Findings

As indicated in the protocol, studies will be classified as Tier 3 when the risk of bias domains are rated as “high” or “probably high” for both exposure assessment or confounding. Given the ranking of exposure assessment as “probably high” due to the indirect application of estimated exposure concentrations to individual exposure assessment (e.g. ecological exposure assessment) and the “probably high” rating for control for confounding due to a lack of subject level data on confounding covariates, Kloog et al. 2013 would be classified as a Tier 3. This designation is further supported by a finding of probably high risk of bias from model specification, which is indicated as potentially introducing bias when attempting to draw inference on the stated research question for this systematic review, which refers specifically to policy relevant concentrations of PM2.5 (occurring at the lower end of the estimated exposure concentration range).

Summary of Evidence Integration

Given that this appendix serves as an example for the treatment of a specific article under the proposed systematic review protocol and not a full implementation of the protocol, a full evidence integration analysis is not possible. However, using examples from the current Draft ISA, it is possible to place Kloog et al. 2013 within the broader context relevant literature to explore its potential impact in forming a conclusion on the charge question of the proposed systematic review. The Draft ISA identified Kloog et al. 2013 as demonstrating the highest magnitude of effect of PM2.5 on mortality (e.g. 2.8% increase per 10ug/m^3 increase in PM2.5) of recent studies, with 18 others from the current review cycle that potentially supported this conclusion (Lippmann et al. 2013; Baxter et al. 2017; Dai et al. 2014; Krall et al. 2013; Lee et al. 2015; Janssen et al 2013; Samoli et al. 2013; Stafoggia et al. 2017; Lanzinger et al. 2016; Pascal et al. 2014; Lee et al. 2015; Di et al. 2017; Zanobetti et al. 2014; Shi et al. 2015; Young et al. 2017; Ueda et al. 2009; Atkinson et al. 2014; Adar et al. 2014) and 8 from the previous review cycle (Burnett and Goldberg 2003; Klemm and Mason 2003; Burnett et al. 2004; Zanobetti and Schwartz 2009; Dominici et al. 2007; Franklin et al. 2007; Franklin et al. 2008; Ostro et al. 2006).

However, as we begin to consider evidence integration criteria (Table S.3), it is notable that low magnitudes of association exist in these studies (e.g. all but one are lower than a 3% increase over a broad range of PM2.5 exposure increase) and several are statistically non-significant. Most studies are characterized by mixed exposures, making the element of specificity difficult to demonstrate, and most studies fail to demonstrate increased hazard ratios as exposure concentration increases in a consistent manner (e.g. dose response).

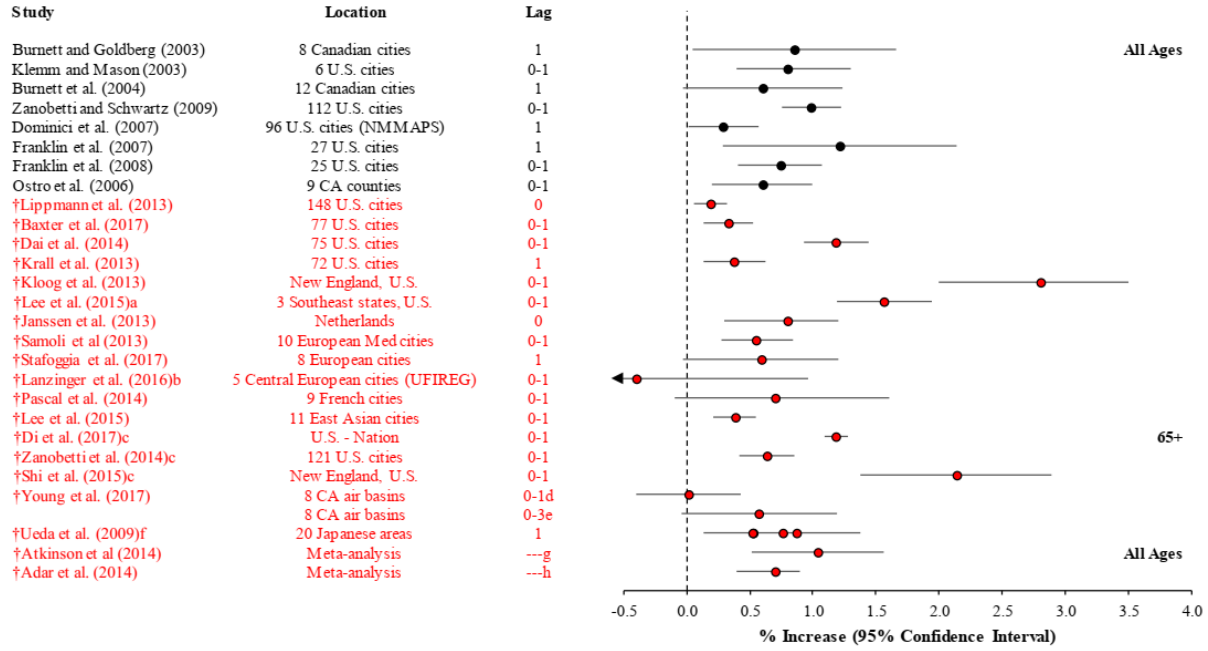
Table A.3 Kloog et al. 2013 Criteria for Evidence Integration

	Criteria for Evidence Integration
Aspect	**Explanation**
Strength of Association	Large and precise risk estimates are less likely to be due to chance, bias, or other factors
Consistency	Evidence is stronger if consistent effects are observed among studies of different designs, people, places, circumstances, and times
Specificity	Evidence is stronger when disease is specific to an exposure or exposure is specific to disease
Temporality	Exposure must precede the occurrence of disease

Exposure-response	Evidence is stronger when a well-characterized exposure-response relationship exists (*e.g.*, disease risk increases with greater exposure intensity and/or duration) -– at the relevant exposure levels (in this case below the current NAAQS)
Biological Plausibility	Evidence on the biological mechanism of an effect allows a scientifically defensible determination for causation at relevant exposure levels
Coherence	All of the known facts related to the observed association from the various evidence streams fit together in a coherent manner
Experiment	"Natural experiments" can provide strong evidence when an intervention or cessation of exposure results in a change in disease risks
Analogy	Evidence is stronger when a similar substance is an established causal factor for a similar effect

Considering the 27 total articles (Fig. A.1) that inform on this issue, none would have superior risk of bias ratings for exposure characterization or model specification than Kloog et al. 2014. This places all human health evidence for the outcome of mortality at policy relevant PM2.5 concentrations in either Tier 2 or Tier 3. This is largely a result of limited exposure characterization and the persistent methodological choice among these studies to use a linearized statistical model over a broad range of exposures (typically a per 10 ug/m3 basis) that prevents assessing the specific, policy relevant research question and substantially lowers the confidence of conclusions drawn from this analysis.

Figure A.1: Summary of Mortality Studies Evaluated in the Draft ISA*



*Adapted from Draft PM ISA p. 11-10

Summary of Findings

Coming to a specific conclusion regarding the indications of systematically integrated evidence analysis of this literature would require a full literature search and analysis as prescribed by the protocol. However, by examining some sentinel features of the literature relied upon in the Draft ISA to draw inferences on the relationship between mortality and PM2.5 exposure, it is clear that a systematic evaluation of this literature could potentially provide for alternate conclusions than those reached in the Draft ISA. Evaluating Kloog et al. 2013 for risk of bias and contextualizing the

findings with evidence integration criteria demonstrates how a systematic approach may provide a more detailed ranking and weighting of information and provide a better characterization of confidence in conclusions than a less systematic approach to literature review.

As well, this approach demonstrates the importance of specificity in the research question when drafting the criteria for a systematic review. The proposed protocol refers to a research question that targets a very specific range of exposure, and subsequently, study quality criteria reflect the need to accurately characterize that range of exposure. As a result, study quality evaluation, the informative value of a particular study, and the confidence evaluation could be substantially different under this protocol than one that asked a less specific research question or one that refers to a different set of exposure metrics. The specific question asked under this protocol provides an evaluation that may be more informative to policy decision makers that are charged with critically evaluating the efficacy of current NAAQS concentrations for protecting public health.

References

Adar, SD; Filigrana, PA; Clements, N; Peel, JL. (2014). Ambient coarse particulate matter and human health: A systematic review and meta-analysis [Review]. Curr Environ Health Rep 1: 258-274. http://dx.doi.org/10.1007/s40572-014-0022-z

Atkinson, RW; Kang, S; Anderson, HR; Mills, IC; Walton, HA. (2014). Epidemiological time series studies of PM2.5 and daily mortality and hospital admissions: a systematic review and meta-analysis [Review]. Thorax 69: 660-665. http://dx.doi.org/10.1136/thoraxjnl-2013-204492

Baxter, LK; Crooks, JL; Sacks, JD. (2017). Influence of exposure differences on city-to-city heterogeneity in PM2.5-mortality associations in US cities. Environ Health 16: 1. http://dx.doi.org/10.1186/s12940-016-0208-y

Burnett, RT; Goldberg, MS. (2003). Size-fractionated particulate mass and daily mortality in eight Canadian cities. Boston, MA: Health Effects Institute.

Burnett, RT; Stieb, D; Brook, JR; Cakmak, S; Dales, R; Raizenne, M; Vincent, R; Dann, T. (2004). Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities. Arch Environ Occup Health 59: 228-236. http://dx.doi.org/10.3200/AEOH.59.5.228-236

Dai, L; Zanobetti, A; Koutrakis, P; Schwartz, JD. (2014). Associations of fine particulate matter species with mortality in the United States: a multicity time-series analysis. Environ Health Perspect 122: 837-842. http://dx.doi.org/10.1289/ehp.1307568

Di, Q; Wang, Y; Zanobetti, A; Wang, Y; Koutrakis, P; Choirat, C; Dominici, F; Schwartz, JD. (2017c). Air pollution and mortality in the Medicare population. N Engl J Med 376: 2513-2522. http://dx.doi.org/10.1056/NEJMoa170274

Dominici, F; Peng, RD; Zeger, SL; White, RH; Samet, JM. (2007). Particulate air pollution and mortality in the United States: did the risks change from 1987 to 2000? Am J Epidemiol 166: 880-888. http://dx.doi.org/10.1093/aje/kwm222

Franklin, M; Koutrakis, P; J, S. (2008). The role of particle composition on the association between PM2.5 and mortality. Epidemiology 19: 680-689. http://dx.doi.org/10.1097/EDE.0b013e3181812bb7

Franklin, M; Zeka, A; Schwartz, J. (2007). Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. J Expo Sci Environ Epidemiol 17: 279-287. http://dx.doi.org/10.1038/sj.jes.7500530

Janssen, NAH; Fischer, P; Marra, M; Ameling, C; Cassee, FR. (2013). Short-term effects of PM2.5, PM10 and PM2.5-10 on daily mortality in the Netherlands. Sci Total Environ 463: 20-26. http://dx.doi.org/10.1016/j.scitotenv.2013.05.062

Klemm, RJ; Mason, R. (2003). Replication of reanalysis of Harvard Six-City mortality study. In HEI Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, Part II. Boston, MA: Health Effects Institute.

Kloog, I; Ridgway, B; Koutrakis, P; Coull, BA; Schwartz, JD. (2013). Long- and short-term exposure to PM2.5 and mortality: Using novel exposure models. Epidemiology 24: 555-561. http://dx.doi.org/10.1097/EDE.0b013e318294beaa

Krall, JR; Anderson, GB; Dominici, F; Bell, ML; Peng, RD. (2013). Short-term exposure to particulate matter constituents and mortality in a national study of U.S. urban communities. Environ Health Perspect 121: 1148-1153. http://dx.doi.org/10.1289/ehp.1206185

Lanzinger, S; Schneider, A; Breitner, S; Stafoggia, M; Erzen, I; Dostal, M; Pastorkova, A; Bastian, S; Cyrys, J; Zscheppang, A; Kolodnitska, T; Peters, A. (2016). Associations between ultrafine and fine particles and mortality in five central European cities - Results from the UFIREG study. Environ Int 88: 44-52. http://dx.doi.org/10.1016/j.envint.2015.12.006

Lee, H; Honda, Y; Hashizume, M; Guo, YL; Wu, CF; Kan, H; Jung, K; Lim, YH; Yi, S; Kim, H. (2015a). Short term exposure to fine and coarse particles and mortality: Amulticity time-series study in East Asia. Environ Pollut 207: 43-51. http://dx.doi.org/10.1016/j.envpol.2015.08.036

Lee, M; Koutrakis, P; Coull, B; Kloog, I; Schwartz, J. (2015c). Acute effect of fine particulate matter on mortality in three Southeastern states from 2007-2011. J Expo Sci Environ Epidemiol 26: 173-179. http://dx.doi.org/10.1038/jes.2015.47

Lippmann, M; Chen, LC; Gordon, T; Ito, K; Thurston, GD. (2013b). National Particle Component Toxicity (NPACT) initiative: Study 3. Time-series analysis of mortality, hospitalizations, and ambient PM2.5 and its components. Appendix G. Supplemental information [HEI]. (177). Boston, MA: Health Effects Institute. https://www.healtheffects.org/system/files/RR177-Lippmann-Study3-AppendixG_0.pdf

Ostro, B; Broadwin, R; Green, S; Feng, WY; Lipsett, M. (2006). Fine particulate air pollution and mortality in nine California counties: Results from CALFINE. Environ Health Perspect 114: 29-33. http://dx.doi.org/10.1289/ehp.8335

Pascal, M; Falq, G; Wagner, V; Chatignoux, E; Corso, M; Blanchard, M; Host, S; Pascal, L; Larrieu, S. (2014). Short-term impacts of particulate matter (PM10, PM10-2.5, PM2.5) on mortality in nine French cities. Atmos Environ 95: 175-184. http://dx.doi.org/10.1016/j.atmosenv.2014.06.030

Samoli, E; Stafoggia, M; Rodopoulou, S; Ostro, B; Declercq, C; Alessandrini, E; Díaz, J; Karanasiou, A; Kelessis, AG; Le Tertre, A; Pandolfi, P; Randi, G; Scarinzi, C; Zauli-Sajani, S; Katsouyanni, K; Forastiere, F; group, tMS. (2013). Associations between Fine and Coarse Particles and Mortality in Mediterranean Cities: Results from the MED-PARTICLES Project. Environ Health Perspect 121: 932-938. http://dx.doi.org/10.1289/ehp.1206124

Shi, L; Zanobetti, A; Kloog, I; Coull, BA; Koutrakis, P; Melly, SJ; Schwartz, JD. (2015). Low-concentration PM2.5 and mortality: estimating acute and chronic effects in a population-based study. Environ Health Perspect 124: 46-52. http://dx.doi.org/10.1289/ehp.1409111

Stafoggia, M; Schneider, A; Cyrys, J; Samoli, E; Andersen, ZJ; Bedada, GB; Bellander, T; Cattani, G; Eleftheriadis, K; Faustini, A; Hoffmann, B; Jacquemin, B; Katsouyanni, K; Massling, A; Pekkanen, J; Perez, N; Peters, A; Quass, U; Yli-Tuomi, T; Forastiere, F. (2017). Association between short-term exposure to ultrafine particles and mortality in eight European urban areas. Epidemiology 28: 172-180. http://dx.doi.org/10.1097/EDE.0000000000000599

Ueda, K; Nitta, H; Ono, M; Takeuchi, A. (2009). Estimating mortality effects of fine particulate matter in Japan: A comparison of time-series and case-crossover analyses. J Air Waste Manag Assoc 59: 1212-1218. http://dx.doi.org/10.3155/1047-3289.59.10.1212

Young, SS; Smith, RL; Lopiano, KK. (2017). Air quality and acute deaths in California, 2000-2012. Regul Toxicol Pharmacol 88: 173-184. http://dx.doi.org/10.1016/j.yrtph.2017.06.003

Zanobetti, A; Schwartz, J. (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. Environ Health Perspect 117: 1-40. http://dx.doi.org/10.1289/ehp.0800108

Zanobetti, A; Dominici, F; Wang, Y, un; Schwartz, JD. (2014b). A national case-crossover analysis of the shortterm effect of PM2.5 on hospitalizations and mortality in subjects with diabetes and neurological disorders. Environ Health 13: 38. http://dx.doi.org/10.1186/1476-069X-13-38

Appendix B Data Collection Example for Research Question 2: Ischemic Heart Disease

Data Collection and Summary for Cesaroni et al. 2014.

Cesaroni, G; Forastiere, F; Stafoggia, M; Andersen, ZJ; Badaloni, C; 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 2014;348:f7412 doi: 10.1136/bmj.f7412

Background

In the Integrated Science Assessment for Particulate Matter External Draft Review 2018, ischemic heart disease is identified as a health outcome of importance potentially related to PM2.5 exposure:

> The strongest evidence comes from epidemiologic studies that reported consistent, positive associations between short-term PM2.5 exposure and cardiovascular-related emergency department visits and hospital admissions particularly for ischemic heart disease (IHD)… ISA p. ES-11

Of the literature reviewed in the Integrated Science Assessment, Cesaroni et al. 2014 is cited as a principal reference for demonstrating the potential causal association between PM2.5 and ischemic heart disease (IHD) authored within the current review cycle:

> In summary, some well-conducted prospective studies indicate an association between long-term 19 exposure to PM2.5 and IHD outcomes in post-menopausal women (Miller et al., 2007) and in a 20 meta-analysis of European cohorts (Cesaroni et al., 2014). ISA p. 6-148

Due to the emphasis placed on Cesaroni et al. 2014 in the Integrated Science Assessment for demonstrating the potential causal relationship between PM2.5 exposure and IHD, this article was selected to provide an example of data extraction, analysis of risk of bias, and evaluate elements of evidence integration that would result from conducting a systematic review using the provided systematic review protocol. As the systematic review protocol presented here focuses on a specific aspect of causal analysis (e.g. evaluating causation relevant to existing NAAQS policy concentrations of PM2.5), processing this article through this systematic review protocol demonstrates the impact of utilizing a structured systematic review framework on the interpretation of this literature and provides insight on potential alternative conclusions that might be reached by using this framework for addressing specific, policy relevant research questions.

Risk of Bias Profile (HUMAN)

Table B.1 Cesaroni et al. 2014 PECOS Evaluation

Study Element	Description
Participants	100,166 people were enrolled from 1997 to 2007 and followed for an average of 11.5 years. Participants were free from previous coronary events at baseline.
Exposure	Modelled concentrations of particulate matter <2.5 µm (PM2.5), 2.5-10 µm (PMcoarse), and <10 µm (PM10) in aerodynamic diameter, soot (PM2.5 absorbance), nitrogen oxides, and traffic exposure at the home address based on measurements of air pollution conducted in 2008-12.
Comparator	Cohort specific hazard ratios for incidence of acute coronary events (myocardial infarction and unstable angina) per fixed increments of the pollutants with adjustment for sociodemographic and lifestyle risk factors, and pooled random effects meta-analytic hazard ratios.
Outcome	Incidence of acute coronary events (myocardial infarction and unstable angina).
Study Design	Cohort/Meta-analysis: included 11 European cohorts from five countries with information about incident cases of acute coronary events.

Table B.2 Cesaroni et al. 2014 Risk of Bias Evaluation

Bias domain	Reviewer's judgment	Support for judgment
Source population representation (Did selection of study participants result in appropriate comparison groups?)	Probably low risk of bias	Large population across exposure ranges allows for internal comparison of relatively similar demographics amongst higher and lower exposed subjects; some risk of bias may still be present in the form of ecological fallacy due to disproportionate sizes of cohorts among disparate cultural/geographic locations.
Confounding (Did the study design or analysis account for important confounding and modifying variables?)	Probably low risk of bias	Major confounders including co-pollutants, SES, estimates of physical activity, BMI, smoking, gender, and education are included where available; some cohorts lack specific confounder data that could not be incorporated into the meta-analysis; much of the confounder data is self-reported, potentially leading to bias; as well, measurements for specific confounders are different among some cohorts.
Incomplete outcome data (Were outcome data complete without attrition or exclusion from analysis?)	Probably low risk of bias	The criteria under which subjects were classified as loss to follow-up (e.g. an IHD event occurred but was not included in the data) is not well described in the methods and likely differed among cohorts that were evaluated; however, there is no indication that this was a prevalent phenomena in the data set.
Exposure assessment (Can we be confident in the exposure characterization?)	Probably high risk of bias	Exposure to PM2.5 is based on a total of 6 weeks of ambient air sampling, which was then input to a land use regression model and applied to specific, non sample site address extrapolated over a period of 3 years. Given that inferences on association are drawn based on a 5 um/m3 change in PM 2.5exposure, even minor

Bias domain	Reviewer's judgment	Support for judgment
		unmeasured error in the exposure assessment has the potential to substantially alter the measures of association produced. Directionality of this potential bias is unknown. Additionally, this study did not specifically address PM2.5 exposures at the 12 ug/m3 threshold (e.g. 15, 20, and 25 ug/m3 were evaluated). Though 12 ug/m3 and lower exposures were included in the dataset, inference drawn on a per 5 ug/m3 may be biased towards results related to thresholds specifically evaluated by authors.
Outcome assessment (Can we be confident in the outcome assessment?)	Low risk of bias	Outcomes are associated with subject ID and medical record and are likely to be accurate in most cases.
Selective outcome reporting (Were all measured outcomes reported?)	Low risk of bias	Authors did not state that any unnecessary censoring was employed.
Other potential threats to internal validity (Were statistical methods appropriate? Did researchers adhere to study protocol?)	--	--
Model Specification (Were statistical models evaluated for validity of underlying assumptions or was model validity assessed with exposure ranges outside the range of interest for the research question?)	High risk of bias	Use of a linearized model to detect threshold concentrations is biased by exposure response relationships that fall outside the range of interest, but remain in the model. In this case, inferences drawn on the lower range of exposures are informed to some degree by exposure response relationships that occur at the higher end of the exposure range in terms of magnitude of effect and overall model validity. This type of bias is likely to be in a positive direction for low exposure associations.
External evidence of bias (does external research indicate there may be unmeasured uncertainty, bias, or confounding present in the study design or dataset?)	--	--

Summary of Risk of Bias Findings

As indicated in the protocol, studies will be classified as Tier 2 when the risk of bias domains are rated as “high” or “probably high” for either exposure assessment or confounding. Given the ranking of exposure assessment as “probably high” due to the indirect application of central monitoring to specific geographical locations (e.g. ecological exposure assessment), Cesaroni et al. 2014 would be classified as a Tier 2. This designation is further supported by a finding of high risk of bias from model specification, which is indicated as having a high potential for bias when attempting to draw inference on the stated research question for this systematic review.

Summary of Evidence Integration

Given that this appendix serves as an example for the treatment of a specific article under the proposed systematic review protocol and not a full implementation of the protocol, a full evidence integration analysis is not possible. However, using examples from the current Draft ISA, it is possible to place Cesaroni et al. 2014 within the broader context relevant literature to explore its potential impact in forming a conclusion on the charge question of the proposed systematic review. The Draft ISA identified Cesaroni et al. 2014 as the strongest epidemiological study to inform on IHD, with 9 others from the current review cycle that potentially supported this conclusion (Hart et al. 2015, Lipsett et al. 2011, Puett et al. 2011, Madrgano et al. 2013, Hartiala et al. 2016, Hoffman et al. 2015, Atkinson et al. 2013, Tonne et al. 2015, Koton et al. 2013) and one from the previous review cycle (Miller et al. 2007).

However, as we begin to consider evidence integration criteria (Table S.3), it is notable that low magnitudes of association exist in these studies (e.g. all but one are lower than 1.5) and several are statistically non-significant. Most studies are characterized by mixed exposures, making the element of specificity difficult to demonstrate, and most studies fail to demonstrate increased hazard ratios as exposure concentration increases in a consistent manner (e.g. dose response).

Table B.3 Cesaroni et al. 2014 Criteria for Evidence Integration

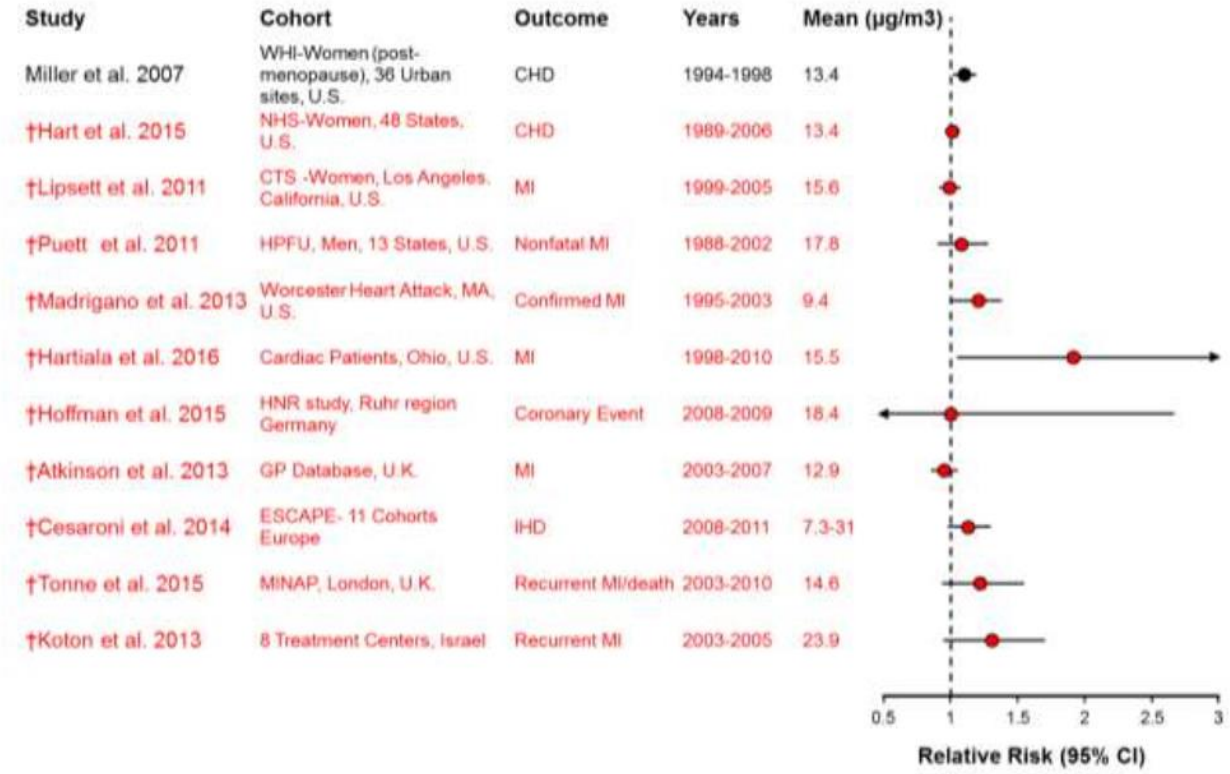
	Criteria for Evidence Integration
Aspect	**Explanation**
Strength of Association	Large and precise risk estimates are less likely to be due to chance, bias, or other factors
Consistency	Evidence is stronger if consistent effects are observed among studies of different designs, people, places, circumstances, and times
Specificity	Evidence is stronger when disease is specific to an exposure or exposure is specific to disease
Temporality	Exposure must precede the occurrence of disease
Exposure-response	Evidence is stronger when a well-characterized exposure-response relationship exists (*e.g.*, disease risk increases with greater exposure intensity and/or duration) -– at the relevant exposure levels (in this case below the current NAAQS)
Biological Plausibility	Evidence on the biological mechanism of an effect allows a scientifically defensible determination for causation at relevant exposure levels
Coherence	All of the known facts related to the observed association from the various evidence streams fit together in a coherent manner
Experiment	"Natural experiments" can provide strong evidence when an intervention or cessation of exposure results in a change in disease risks

Analogy	Evidence is stronger when a similar substance is an established causal factor for a similar effect

Considering the 11 total articles (Fig B.1) that inform on this issue, none would have superior risk of bias ratings for confounding, exposure characterization, or model specification than Cesaroni et al. 2014. This places all human health evidence for the outcome of IHD at policy relevant PM2.5 concentrations in either Tier 2 or Tier 3. In the uncertainty portion of a systematic review, this finding would substantially lower the confidence of an conclusions drawn from this analysis.

Further, statistical significance is broadly lacking across studies for measures of association between PM2.5 exposure and IHD. This is an example of unexplained inconsistency that impacts the integration of evidence for causal inference and reduces the confidence of a causal conclusion. Using the prescribed integration approach described in the primary review protocol, the lack of consistency, low magnitude of association, and high risk of bias would not lead to a causal conclusion for this outcome at policy relevant levels of PM2.5 exposure.

Figure B.1: Summary of IHD findings evaluated in the Draft ISA*



*Adapted from the Draft PM ISA

Summary of Findings

Coming to a specific conclusion regarding the indications of systematically integrated evidence analysis of this literature would require a full literature search and analysis as prescribed by the protocol. However, by examining some sentinel features of the literature relied upon in the Draft ISA to draw inferences on the relationship between IHD and PM2.5 exposure, it is clear that a systematic evaluation of this literature could potentially provide for alternate conclusions than those reached in the Draft ISA. Evaluating Cesaroni et al. 2014 for risk of bias and contextualizing the

findings with evidence integration criteria demonstrates how a systematic approach may provide a more detailed ranking and weighting of information and provide a better characterization of confidence in conclusions than a less systematic approach to literature review.

As well, this approach demonstrates the importance of specificity in the research question when drafting the criteria for a systematic review. The proposed protocol refers to a research question that targets a very specific range of exposure, and subsequently, study quality criteria reflect the need to accurately characterize that range of exposure. As a result, study quality evaluation, the informative value of a particular study, and the confidence evaluation could be substantially different under this protocol than one that asked a less specific research question or one that refers to a different set of exposure metrics. The specific question asked under this protocol provides an evaluation that may be more informative to policy decision makers that are charged with critically evaluating the efficacy of current NAAQS concentrations for protecting public health.

References

Atkinson, RW; Carey, IM; Kent, AJ; van Staa, TP; Anderson, HR; Cook, DG. (2013). Long-term exposure to outdoor air pollution and incidence of cardiovascular diseases. Epidemiology 24: 44-53. http://dx.doi.org/10.1097/EDE.0b013e318276ccb8

Cesaroni, G; Forastiere, F; Stafoggia, M; Andersen, ZJ; Badaloni, C; Beelen, R; Caracciolo, B; de Faire, U; Erbel, R; Eriksen, KT; Fratiglioni, L; Galassi, C; Hampel, R; Heier, M; Hennig, F; Hilding, A; Hoffmann, B; Houthuijs, D; Jöckel, KH; Korek, M; Lanki, T; Leander, K; Magnusson, PK; Migliore, E; Ostenson, CG; Overvad, K; Pedersen, NL; J, JP; Penell, J; Pershagen, G; Pyko, A; Raaschou-Nielsen, O; Ranzi, A; Ricceri, F; Sacerdote, C; Salomaa, V; Swart, W; Turunen, AW; Vineis, P; Weinmayr, G; Wolf, K; de Hoogh, K; Hoek, G; Brunekreef, B; Peters, A. (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. B M J 348: f7412.

Hart, JE; Liao, X; Hong, B; Puett, RC; Yanosky, JD; Suh, H; Kioumourtzoglou, MA; Spiegelman, D; Laden, F. (2015). The association of long-term exposure to PM2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environ Health 14: 38. http://dx.doi.org/10.1186/s12940-015-0027-6

Hartiala, J; Breton, CV; Tang, WH; Lurmann, F; Hazen, SL; Gilliland, FD; Allayee, H. (2016). Ambient air pollution is associated with the severity of coronary atherosclerosis and incident myocardial infarction in patients undergoing elective cardiac evaluation. J Am Heart Assoc 5. http://dx.doi.org/10.1161/JAHA.116.003947

Hoffmann, B; Weinmayr, G; Hennig, F; Fuks, K; Moebus, S; Weimar, C; Dragano, N; Hermann, DM; Kaelsch, H; Mahabadi, AA; Erbel, R; Joeckel, KH. (2015). Air Quality, Stroke, and Coronary Events Results of the Heinz Nixdorf Recall Study From the Ruhr Region. Deutsches Ärzteblatt International 112: 195-U123. http://dx.doi.org/10.3238/arztebl.2015.0195

Koton, S; Molshatzki, N; Yuval, N; Myers, V; Broday, DM; Drory, Y; Steinberg, DM; Gerber, Y. (2013). Cumulative exposure to particulate matter air pollution and long-term post-myocardial infarction outcomes. Prev Med 57: 339-344. http://dx.doi.org/10.1016/j.ypmed.2013.06.009

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: 828-835. http://dx.doi.org/10.1164/rccm.201012-2082OC

Madrigano, J; Kloog, I; Goldberg, T; Coull, BA; Mittleman, MA; Schwartz, J. (2013). Long-term exposure to PM2.5 and incidence of acute myocardial infarction. Environ Health Perspect 121: 192-196. http://dx.doi.org/10.1289/ehp.1205284

Miller, KA; Siscovick, DS; Sheppard, L; Shepherd, K; Sullivan, JH; Anderson, GL; Kaufman, JD. (2007). Longterm exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med 356: 447-458. http://dx.doi.org/10.1056/NEJMoa054409

Puett, RC; Hart, JE; Suh, H; Mittleman, M; Laden, F. (2011). Particulate matter exposures, mortality and cardiovascular disease in the health professionals follow-up study. Environ Health Perspect 119: 1130-1135. http://dx.doi.org/10.1289/ehp.1002921

Tonne, C; Halonen, JI; Beevers, SD; Dajnak, D; Gulliver, J; Kelly, FJ; Wilkinson, P; Anderson, HR. (2015). Longterm traffic air and noise pollution in relation to mortality and hospital readmission among myocardial infarction survivors. Int J Hyg Environ Health 219: 72-78. http://dx.doi.org/10.1016/j.ijheh.2015.09.003