Kevin Buchan
Senior Manager
Bay Area Region Regulatory Affairs

November 29, 2022

Sent via email: methodfeedback@baaqmd.gov

Chair Solomon and Members of the Advisory Council
Bay Area Air Quality Management District
375 Beale Street, Suite 600
San Francisco, CA 94105

Re: WSPA Comments on the BAAQMD Proposed Methodology for Determining Local Health Risks from Fine Particulate Matter (PM2.5)

Dear Chair Solomon and Members of the Advisory Council,

The Western States Petroleum Association (WSPA) is a non-profit trade association representing twenty-six companies that explore for, produce, refine, transport and market petroleum, petroleum products, natural gas and other energy supplies in California, Arizona, Nevada, Oregon, and Washington. Our members in the Bay Area have operations and facilities regulated by the Bay Area Air Quality Management District (BAAQMD or District).

WSPA submits these comments to the Advisory Council (Council) for review and consideration of the BAAQMD staff proposal to estimate source specific PM2.5 concentrations and exposures.

EXECUTIVE SUMMARY

The staff proposed model predicts source direct contributions PM2.5 concentrations at a particular location. Concentration-response functions from epidemiology studies are then used in combination with modeled exposures to determine increases in mortality in adults and workers or asthma onset in children associated with a local source.

All of the evidence regarding effects associated with ambient PM2.5 are based on epidemiology studies. These studies have major limitations, primarily due to potential biases in the study designs, co-pollutants or other risk factors not being adequately considered, and the misclassification of estimated PM2.5 exposures not being accurate.

Although these studies show correlations between ambient PM2.5 and mortality in adults and asthma in children, these correlations do not provide evidence for causation, particularly at low exposure concentrations.

Staff used risk estimates of 1.01 and 1.045 per 1 μg/m^3 for mortality in older adults and pediatric asthma, respectively, in its model. These small increases in risk could have a large impact on a large population, only if these risk estimates are accurate.

Risk estimates this close to 1, particularly when based on studies with major methodological limitations, are not supportive of causal associations. When calculating a risk, if the estimated increase is large, a true risk is more likely indicated. If it is small however, it more likely reflects study design flaws.

If a risk estimate does not support causation, then it is inappropriate to use it in calculating risks.

Epidemiology studies do not provide reliable risk estimates at ambient PM2.5 exposure concentrations. As such, estimated incremental risk increases associated with 0.001-0.3 μg/m3 PM2.5 increments are even less reliable.

The model uses a cancer-based equation, which assumes that every exposure no matter how small can contribute to cancer risk. This is not a valid assumption for non-cancer endpoints.

Incremental differences of 0.001-0.3 μg/m3 PM2.5 are negligible compared to actual PM2.5 concentrations and fluctuations in the Bay Area. It is not possible to estimate any actual changes in risk associated with such small increments, given the large incremental PM2.5 fluctuations seen on hourly and daily bases.

Hundreds of epidemiology studies have evaluated PM2.5 associations with morbidity and mortality. While statistical associations have been reported, these associations do not provide evidence for causation at ambient PM2.5 concentrations. These issues and others have resulted in staff's overestimation of incremental risks associated with the very small increases (0.001-0.3 μg/m3) in PM2.5.

WSPA appreciates the opportunity to provide input on this important matter to the Council. We look forward to your December 15th meeting where WSPA will present our findings and seek meaningful dialog with the Council members.

Sincerely,

Enclosure: "Review of the Bay Area Air Quality Management District Report, Modeling Local Sources of Fine Particulate Matter (PM2.5) for Risk Management", (Gradient, 2022)

cc: Dr. David Holstius, Senior Advanced Projects Advisor, BAAQMD

Review of the Bay Area Air Quality Management District Report, "Modeling Local Sources of Fine Particulate Matter ($PM_{2.5}$) for Risk Management"

Prepared for
Kevin Buchan
Western States Petroleum Association
1320 Willow Pass Road, Suite 600
Concord, CA 94520

November 28, 2022

GRADIENT
www.gradientcorp.com
One Beacon Street, 17th Floor
Boston, MA 02108
617-395-5000

Table of Contents

Page

Overview 1

1 Introduction 3

2 Mortality 4
- 2.1 Exposure Measurement Error 4
- 2.2 Confounding 6
- 2.3 Low Exposure Concentrations 6
- 2.4 Risk Magnitude 6
- 2.5 Biological Plausibility 7
- 2.6 Thresholds 8

3 Pediatric Asthma Onset 9
- 3.1 BenMAP-CE 9
- 3.2 Tetreault *et al*. (2016) 10
- 3.3 Risk Magnitude 10

4 Context 11

5 Conclusions 12

References 13

List of Tables

Table 1 Daily Average $PM_{2.5}$ Concentrations ($\mu g/m^3$) in the Bay Area Air Quality Management District (BAAQMD) – October 2022

Table 2 Hourly Average $PM_{2.5}$ Concentrations ($\mu g/m^3$) in the Bay Area Air Quality Management District (BAAQMD) – October 1, 2022

Table 3 Difference Between Average $PM_{2.5}$ Concentration ($\mu g/m^3$) at the Same Hour on October 1 and September 31, 2022, in the Bay Area Air Quality Management District (BAAQMD)

Overview

The Bay Area Air Quality Management District (BAAQMD) proposes to estimate annual average source-specific $PM_{2.5}$ concentrations using a dispersion model that considers site and meteorological conditions. This model predicts a source's direct contribution to the total fine particulate matter ($PM_{2.5}$) concentration at a particular location. Concentration-response functions (which are equations that describe how risks change with increases in $PM_{2.5}$ concentrations) from epidemiology studies are then used in combination with modeled exposures to determine incremental increases in mortality in older adults and workers or asthma onset in children associated with a local source (BAAQMD, 2022a).

There are many issues with the BAAQMD (2022a) model, including:

- While experimental studies in humans and other animals have provided some evidence regarding causation at very high exposure concentrations, all of the evidence regarding effects associated with ambient $PM_{2.5}$ are based on epidemiology studies. These studies have major limitations, primarily due to potential biases in the study designs, the fact that co-pollutants or other risk factors were not adequately considered, and exposure measurement error or misclassification (meaning that the estimated $PM_{2.5}$ exposures in these studies are not accurate). Although these studies show correlations between ambient $PM_{2.5}$ and mortality in adults and asthma in children, these correlations do not provide evidence for causation, particularly at low exposure concentrations.
- BAAQMD (2022a) used risk estimates of 1.01 and 1.045 per 1 $\mu g/m^3$ for mortality in older adults and pediatric asthma, respectively, in its model. These small increases in risk could have a large impact on a large population, but *only* if these risk estimates are accurate. Risk estimates this close to 1, particularly when they are based on studies with major methodological limitations, are not supportive of causal associations, because they could be the result of large sample sizes (*i.e.*, very small risk estimates are statistically significant only because the of the large study population) and even very small amounts of bias. In other words, when calculating a risk, if the estimated increase is large, it is more likely to indicate a true risk. If it is small, it is less likely to indicate a true increase in risk, and more likely to reflect study design flaws. If a risk estimate does not support causation, then it is not appropriate to use it to calculate risks.
- Because epidemiology studies do not provide reliable risk estimates even at ambient $PM_{2.5}$ exposure concentrations, estimated incremental increases in risk associated with 0.001-0.3 $\mu g/m^3$ $PM_{2.5}$ increments are even less reliable.
- The model uses a cancer-based equation, which assumes that every exposure, no matter how small, can contribute to cancer risk. This is not a valid assumption for non-cancer endpoints, as pollutants can only cause adverse effects if they overwhelm the body's natural defenses.
- The Benefits Mapping and Analysis Program – Community Edition (BenMAP-CE) model (US EPA, 2022) is used for asthma risk estimates. The BenMAP-CE model has methodological limitations that result in risk estimates that are not likely to be reliable.
- Incremental differences of 0.001-0.3 $\mu g/m^3$ $PM_{2.5}$ are negligible compared to actual $PM_{2.5}$ concentrations and fluctuations in $PM_{2.5}$ in the Bay Area. It is not possible to estimate any actual changes in risk associated with such small changes, given the large incremental $PM_{2.5}$ fluctuations seen on hourly and daily bases.

Considering the limitations of the BAAQMD (2022a) model and the fact that $PM_{2.5}$ concentrations can vary considerably, it is clear that the BAAQMD model highly overestimates incremental increases in risk associated with $PM_{2.5}$ increments of 0.001-0.3 $\mu g/m^3$.

1 Introduction

As part of the Air Toxics Control Program, BAAQMD conducts local-scale modeling and sets corresponding source-specific thresholds for maximum contributions to a lifetime risk of cancer (BAAQMD, 2021). BAAQMD (2022a) developed a similar methodology for modeling non-cancer health risks attributable to local sources of $PM_{2.5}$ based both on models of source-specific contributions to local elevations of $PM_{2.5}$ that it developed in 2019 and epidemiology data (BAAQMD, 2022b).

BAAQMD (2022a) proposes to estimate annual average source-specific $PM_{2.5}$ concentrations using a dispersion model that considers site and meteorological conditions. This model predicts a source's direct contribution to the total $PM_{2.5}$ concentration at a particular location. Concentration-response functions (which are equations that describe how risks change with increases in $PM_{2.5}$ concentrations) from epidemiology studies are then used in combination with modeled exposures to determine incremental increases in mortality in older adults and workers, or asthma exacerbations in children, associated with 0.001-0.3 $\mu g/m^3$ $PM_{2.5}$ from a local source (BAAQMD, 2022a).

As discussed in more detail below, epidemiology studies do not accurately predict mortality or morbidity risks, particularly at low $PM_{2.5}$ concentrations. The BenMAP-CE model (US EPA, 2022) used for asthma risk estimates also has major limitations that are not fully considered by BAAQMD. As such, BAAQMD's proposed model for local sources of $PM_{2.5}$ is not accurate. Also, $PM_{2.5}$ concentrations can vary considerably (*i.e.*, far more than 0.001-0.3 $\mu g/m^3$ $PM_{2.5}$) over the course of a day and from day to day in the Bay Area. Because of these issues, the BAAQMD model highly overestimates incremental increases in risk associated with $PM_{2.5}$ increments of 0.001-0.3 $\mu g/m^3$.

2 Mortality

The BAAQMD (2022a) model is based on the premise that long-term ambient $PM_{2.5}$ concentrations are associated with increased premature mortality in adults. While experimental studies in humans and other animals have provided some evidence regarding causation at very high exposure concentrations, all evidence regarding effects associated with ambient $PM_{2.5}$ are based on epidemiology studies.

As noted by the National Academies of Sciences, Engineering, and Medicine (NASEM),

> [R]andomized experiments are strong in terms of their ability to isolate causal effects, but they generally are not conducted in the full populations of relevance for an [Integrated Science Assessment]. On the other hand, large-scale epidemiologic studies will provide data on broad populations and exposures, but will require strong non-experimental methods to generate robust conclusions. Each study needs to be assessed in terms of its design and analysis strengths and limitations, and how its design and results fit with other knowledge and data. (NASEM, 2022)

NASEM (2022) further discussed "emerging approaches for exposure assessment, methods for confounder selection and control, recent approaches for estimation of causal effects, how to deal with posttreatment variables and unmeasured confounders, and how to handle multiple exposures." This is because while epidemiology studies show correlations between ambient $PM_{2.5}$ and mortality in adults, they have major methodological limitations, primarily due to potential biases in the study designs, the fact that co-pollutants or other risk factors were not adequately considered, and exposure measurement error or misclassification (meaning that estimated $PM_{2.5}$ exposures in these studies are not accurate). As a result, correlations in these studies do not provide evidence for causation, particularly at very low exposure concentrations. These issues are discussed in a recent publication by Prueitt *et al*. (2021) (Attachment A) and are briefly summarized below.

2.1 Exposure Measurement Error

NASEM (2022) noted:

> Advances in techniques to measure, store, combine, harmonize, process, and analyze exposure data with high temporal and spatial resolutions are revolutionizing exposure assessment and resultant air pollution related studies for both health and welfare effects. The exposure assessment methods and data used are important aspects of overall study quality and merit in relation to causal assessments, including considerations of "biological gradient" (i.e., exposure-response relationship) and strength of the observed association.

To date, measurement error has been a major issue in air pollution epidemiology studies. This can occur when the measurement of the exposure itself is not correct, an obvious challenge when trying to accurately and fully measure an individual's exposure to something in their environment. This is particularly challenging when trying to capture an exposure that may be coming from multiple sources, and may vary over days, months, years, or decades. My colleagues and I discussed this issue at length in Rhomberg *et al*. (2011). As we stated:

> [T]he levels of individual exposure, whether estimated by measurement or modeling, inevitably have some statistical error. This results for a number of reasons: the estimates may be based on measurements at central air monitoring stations that only approximate the personal exposures of the nearby population, people move about in varying ways and hence experience individual histories of varying exposure that are not easily related to fixed-site measurements, people may have local sources of exposures in their homes and workplaces, and others. (Rhomberg *et al*., 2011)

Another contributor to exposure measurement error relates to the timing of exposure measurements. To assess a causal relationship between an exposure and an outcome the exposure needs to precede the outcome in a temporally meaningful way. For example, one would not expect cancer to be caused by an exposure that occurred the day prior to diagnosis, because cancer is generally a disease that takes years to manifest/develop. As demonstrated by Smith and Chang (2020) in recent comments to United States Environmental Protection Agency (US EPA), associations with recent $PM_{2.5}$ distributions have been reported when they are clearly not causal (*e.g*., because they occur after deaths occur), and this is at least partially attributable to the correlation of more recent exposures with earlier, higher exposures.

As my colleagues and I concluded in our 2011 publication:

> In nonlinear regression, independent variable error results in biased parameter estimates and the masking of true features, as well as the loss of statistical power. It also tends to smooth out and flatten, essentially linearize, all of these curves, including threshold functions. Even when a threshold is detected, it is likely to be biased based on the independent variable error. Importantly, it has been shown that the degree of bias known to apply to actual studies is sufficient to produce a false linear result, and that although nonparametric smoothing and other error-mitigating techniques may assist in identifying a threshold, they do not guarantee detection of a threshold. **Thus, exposure measurement errors as practically encountered in real environmental epidemiology data can result in biases that can affect the interpretation and use of the apparent exposure-response shapes in risk assessment applications. These errors result in an overestimation of risk at low exposures and an underestimation of risks at high exposures.** The consequences of this could be great, as it could lead to a misallocation of resources towards regulations that do not offer any benefit to public health, and may in fact cause harm owing to the underestimation of risks at higher exposures. (Rhomberg *et al*., 2011 [emphasis added])

Errors measuring exposure are very common. While most studies have used and validated $PM_{2.5}$ exposure estimates at a relatively high spatial resolution, the potential for exposure measurement error was likely high for several other aspects of the exposure assessment. A striking limitation of most of these studies is a mismatch between the $PM_{2.5}$ exposure period and the follow-up period for mortality. For at least some of the participants, the $PM_{2.5}$ exposure periods included time after death, which violates the temporality rule in causality (*i.e*., the cause has to occur before the effect). In addition, several studies did not account for changes in $PM_{2.5}$ exposure over time, using only one or a few exposure estimates for each person in the analyses. Also, more than half of the studies did not account for people moving throughout the day or moving their place of residence, likely resulting in considerable exposure measurement error. Most studies did not assess $PM_{2.5}$ in multiple time periods to identify the most relevant exposure window.

2.2 Confounding

Confounding is another major issue in these studies. Confounding occurs when a factor is correlated with both the exposure and the outcome, and when it is not fully accounted for, it can distort the perceived association between the exposure and the outcome (Gordis, 2009). A classic example of a confounder of the relationship between air pollution and respiratory health is smoking. Smoking is associated with both high concentrations of indoor particulate matter and cancer. If it is not fully controlled for in an epidemiology investigation, it can distort the relationship between indoor air pollution and cancer.

$PM_{2.5}$ is correlated with many factors, including atmospheric conditions, other co-pollutants, and socioeconomic status (SES), for example (Valberg, 2003; Bukowski, 2007, 2008a,b; Goldberg *et al*., 2008; Glymour and Greenland, 2008). For example, recent studies have shown that both individual and community SES have a considerable impact on mortality (Stringhini *et al*., 2017; Steel *et al*., 2018). Although most epidemiology studies often adjust for some socioeconomic factors at the individual and/or community level when evaluating long-term $PM_{2.5}$ exposure and mortality, measurement of these socioeconomic factors can be crude and likely does not entirely account for the effects of individual and community SES on mortality. In addition, few studies account for individual smoking, diet, and exercise, or community-level confounders such as access to and quality of health care and violence. It is also notable that most studies do not assess or adjust for co-pollutants. Thus, the observed associations in these studies may not reflect the effects of $PM_{2.5}$. That is, the lack of full consideration of these factors in statistical analyses significantly increased the uncertainty in the true relationship between $PM_{2.5}$ and mortality.

According to Boffetta *et al.* (2008):

> Although the importance of residual confounding and unmeasured confounders as a source of bias in epidemiological studies has been downplayed by many, a recent statistical simulation study showed that with plausible assumptions, effect sizes on the order of 1.5-2.0, which is a magnitude frequently reported in epidemiology studies, can be generated by residual and/or unmeasured confounding.

The mortality risk estimates used in the BAAQMD (2022a) model is 1.01 (and it is 1.045 for asthma in children). These values are far below 1.5-2 and are likely "noise" associated with confounding.

2.3 Low Exposure Concentrations

BAAQMD (2022a) indicates that $PM_{2.5}$ has marginal impacts on mortality at concentrations below the current National Ambient Air Quality Standard (NAAQS), and states, "In the Bay Area, about 98% of the residential population lives where a modeled annual average $PM_{2.5}$ concentration is less than 12 $\mu g/m^3$, and 75% where it is less than 10 $\mu g/m^3$." However, the reliability of epidemiology studies at lower levels of $PM_{2.5}$ has not been established. US EPA and the US EPA Clean Air Scientific Advisory Committee (CASAC) have both acknowledged that, at concentrations below 5-8 $\mu g/m^3$, the relationship between $PM_{2.5}$ and mortality is not clear.

2.4 Risk Magnitude

As noted above BAAQMD (2022a) selected a risk estimate of 1.01 per 1 $\mu g/m^3$ for premature adult mortality. An incremental increase in risk of 0.01 (1%) could have a large impact on a large population, but *only* if the risk estimate is accurate. A risk estimate that is this close to 1, particularly when it is based

on studies with major methodological limitations, is not supportive of a causal association because it could be the result of confounding (as discussed above), biases, or simply the fact that studies evaluate such large numbers of people. If a risk estimate does not support causation, then it is not appropriate to use it to calculate risks.

2.5 Biological Plausibility

Toxicology is the study of the potentially adverse health effects of chemicals and other substances on living organisms (Hayes and Kruger, 2014). It combines information from studies of human populations, experimental animals, isolated cells, and isolated molecules. An understanding of toxicology is necessary for evaluating health effects from exposure to chemicals and for determining how much of a chemical one can be exposed to, and under what conditions, without the likelihood of harm.

To evaluate whether exposure to a chemical may be associated with potential health effects, it is necessary to understand the critical health effects caused by the chemical of interest and the exposure levels, or doses, at which these critical effects occur. Evaluating the relationship between health effects and exposure is referred to as a dose-response assessment. Although every substance is capable of producing toxic (or adverse) effects at some dose, the range of doses necessary to produce adverse effects, injury, or death varies widely among chemicals (Aleksunes and Eaton, 2019; Faustman, 2019). The body has many biochemical and physiological processes that allow it to counteract a chemical's adverse effects, and most chemicals do not cause adverse effects unless they are at a dose sufficient to overwhelm the body's normal processes for a certain period of time. As such, many chemicals are not harmful when one is exposed to low doses (see, for example, Paustenbach and Madl [2014] and Aleksunes and Eaton [2019]). In other words, there is a threshold dose below which adverse health effects are not observed.

The nature and severity of effects observed with exposure to a chemical can vary with dose, and some chemicals are actually beneficial at low doses. Aspirin, for example, provides pain relief and may help prevent cardiovascular disease at or below the recommended dose of two tablets per day, but increasingly higher doses of aspirin may cause adverse effects ranging from fever and acidosis to convulsions and respiratory failure (Ellenhorn and Barceloux, 1988; Grosser *et al*., 2011).

The frequency and duration of exposure to a chemical are also critical factors for determining toxicity, and the adverse effects of a chemical can differ depending on whether exposure is to a single, large dose (acute exposure) or to lower doses over a long period of time (chronic exposure). For example, in the case of ethyl alcohol, a single acute dose can cause severe adverse effects in the central nervous system, whereas chronic exposure to lower doses can damage the liver and cardiovascular system (Bruckner *et al*., 2019). For most chemicals, the severity of health effects is typically much greater for acute, single-dose exposures. With chronic exposure to sufficiently low doses, the body has numerous defense mechanisms and is able to eliminate the dose *via* excretion and repair any damage that may have occurred, or find other means of responding to or adapting to the chemical exposure (Aleksunes and Eaton, 2019; Calabrese, 2014). Even if the frequency and duration of exposure are the same, the severity of adverse effects resulting from chemical exposure can vary among individuals. Factors that may influence the severity of effects in individuals include genetic background, sex, age, health status, behavioral traits (*e.g.*, smoking and alcohol use), diet, and nutritional status (Aleksunes and Eaton, 2019).

As noted by Prueitt *et al*. (2021):

> While some controlled human exposure and experimental animal studies provide evidence for certain morbidity endpoints with exposure to $PM_{2.5}$, the evidence is not strong nor consistent across studies and the effects are reported almost exclusively at high exposures

> (US EPA 2020) and therefore do not support biological plausibility for more serious effects at ambient exposures. Many of the adverse health effects reported in these experimental studies also have thresholds and do not occur at lower concentrations; for example, Green et al. (2002) reported that various chronic exposure studies in rats with different compositions of $PM_{2.5}$ indicate that concentrations of 100–200 $\mu g/m^3$ must be exceeded before potentially adverse changes occur. As this threshold is above ambient concentrations, these experimental studies do not provide support for adverse effects at ambient concentrations. Thus, while there is evidence in the literature for a variety of potential biological mechanisms for the underlying health effects that contribute to total mortality, the experimental studies of adverse health effects with $PM_{2.5}$ exposure do not provide evidence of biological plausibility for mortality associated with ambient $PM_{2.5}$ exposures.

That is, there is no evidence that it is biologically plausible that adverse effects, such as mortality, could be caused by low $PM_{2.5}$ exposures.

2.6 Thresholds

Also, the BAAQMD (2022a) model uses a cancer-based equation, which assumes that every exposure, no matter how small, can contribute to cancer risk. That is, BAAQMD (2022a) compares estimated excess mortality and asthma risks to those used US EPA's Air Toxics Program for regulatory decision making for cancer (1×10^{-4} to 1×10^{-6}) for interpretation. It is not appropriate to compare the health or regulatory risk of non-cancer health effects to standards based on cancer because the body has biological and physiological processes to respond to pollutant exposures. It is only when exposures overwhelm the body's natural defenses that adverse health effects may occur.

* * *

Collectively, the epidemiology studies of long-term $PM_{2.5}$ and total mortality have many limitations that likely had a substantial impact on the validity of the study results. This in turn impacts the validity of the output of the BAAQMD model.

3 Pediatric Asthma Onset

BAAQMD (2022a) also modeled risks of pediatric asthma onset in the same way as it modeled mortality risks in adults. However, most of the weaknesses of the mortality epidemiology studies also exist in epidemiology studies of $PM_{2.5}$ and asthma. This includes all of the challenges with regard to measurement of air pollution exposures and, in addition to the confounding variables noted for mortality, such as copollutants, there are also many well-known factors associated with asthma, including allergies, respiratory infections, and exposure to second-hand smoke, which are often not fully accounted for in epidemiology studies (Castro-Rodriguez *et al*., 2016). Also, while mortality tends to be well captured in vital records, misclassification of asthma is more common. There is less reliability in the results of breathing function tests in children under 5, the diagnosis of asthma is often based on symptoms alone (which may be from viral infections and not asthma) or a family member with asthma (Pedersen *et al*., 2011), and hospital and clinic records have been shown to have coding errors and inconsistencies (De Coster *et al*., 2006; Shiff *et al*., 2014).

3.1 BenMAP-CE

In contrast to the mortality estimates, the BAAQMD (2022a) model relied on the BenMAP-CE model (US EPA, 2022) to estimate the risk of asthma due to $PM_{2.5}$ in children. BenMAP-CE is a software program that was developed by US EPA to allow users to evaluate the health and economic burden of air-pollution changes (US EPA, 2022). While the BenMAP-CE tool (US EPA, 2022) is useful for researchers to be able to easily assess changes in health outcomes associated with changes in air pollution, there are key limitations to the tool that limit a user's ability to understand the uncertainties and limitations in the model and its output.

The BenMAP-CE model (US EPA, 2022) does not allow for an understanding or evaluation of how accurate or reliable the estimates it generates are. It emphasizes the statistical uncertainty generated by the model itself rather than the larger issues caused by limitations in the underlying epidemiology and environmental input data including the risk estimates themselves, the shape of the risk function, and the risk of the individual $PM_{2.5}$ constituents (Smith and Gans, 2015; NRC, 2002).

The BenMAP-CE tool (US EPA, 2022) has only limited built-in capabilities that allow the user to assess the sensitivities of the risk estimates. As an example, it has been demonstrated that the sensitivity tools available to assess the mortality risk estimates are not sufficient, and that the actual range of mortality risk estimates supported by the literature are more variable, and the range is much wider than the tool allows the user to model (Smith and Gans, 2015). This results in the inability to model the impact of mortality on $PM_{2.5}$ with an understanding of how reliable that estimate is. To our knowledge, the impact of model limitations on pediatric asthma risk estimates has not been assessed, but due to the difficulties and limitations described above in estimating this risk in epidemiology studies, it can be reasonably assumed that the tool is not likely capturing the accuracy or reliability of this outcome either. Without the ability to assess the impact of these issues through sensitivity analyses built into the BenMAP-CE tool (US EPA, 2022), it is difficult to interpret the model results, and using an estimate from a this model in a subsequent calculation, such as proposed by BAAQMD (2022a), compounds these issues. Further, the BAAQMD (2022a) model does not provide additional sensitivity analyses to demonstrate impacts of uncertainties in the BenMAP-CE input values (US EPA, 2022).

3.2 Tetreault *et al*. (2016)

In addition to the general limitations of the BenMAP-CE model (US EPA, 2022), BAAQMD (2022a) suggests that the basis for the pediatric asthma risk in the BenMAP-CE model (US EPA, 2022) are based on a single study examining the association between pediatric asthma and air pollution in Quebec, Canada (Tetreault *et al*., 2016). The Quebec study had major methodological limitations that impact the ability to draw causal conclusions with certainty. Exposure was not modeled locally, and a single $PM_{2.5}$ value reflecting the average value estimated from satellites over a 5-year period was considered the exposure value for the entire 15-year period. Despite evaluating for and finding associations between asthma and exposures to ozone (O_3) and nitrogen dioxide (NO_2) as well, there were no adjustments for these pollutants used in the model assessing the risk of $PM_{2.5}$ on asthma. Individual level SES, smoking in the home, and family history of asthma also were not accounted for, which also could have impacted results.

In addition to issues with the Tetreault *et al*. (2016) study estimates due to limitations and biases in the study itself, the estimates from Quebec reflect a composite of both urban and very rural Canadian populations, which may not represent the pediatric population in the US that is being modeled in the BenMAP-CE tool (US EPA, 2022). Beyond the challenges in generalizing the Canadian data to the US, the national level data modeled out of the BenMAP-CE tool (US EPA, 2022) is then used as an input with unknown representativeness to the Bay Area population in the BAAQMD (2022a) model. Further, as noted above, there is no consideration for the statistical uncertainty of the risk estimates, either in the inputs into the BAAQMD (2022a) model (no range of inputs), or in the output (no confidence intervals).

3.3 Risk Magnitude

Similar to what was discussed previously for mortality, the risk estimate of 1.045 per 1 $\mu g/m^3$ for pediatric asthma that BAAQMD (2022a) chose could have a large impact on a large population, but *only* if the risk estimate is accurate. This risk estimate is even smaller (*i.e*., closer to 1) than that estimated for mortality. Again, a risk estimate that is this close to 1, particularly one based on studies with major methodological limitations, is not supportive of a causal association because it could be the result of large numbers and a small amount of bias. If a risk estimate does not support causation, then it is not appropriate to use it to calculate risks.

* * *

Similar to the uncertainties described above for the mortality estimates, the asthma risk estimates, based on limited epidemiology data, are subject to the same methodological and statistical limitations, resulting in significant uncertainties in the point estimates provided by the BAAQMD (2022a) model.

4 Context

BAAQMD (2022a) calculates incremental increases in risk associated with 0.001-0.3 μg/m^3 increases in $PM_{2.5}$ exposures, but does not put these numbers in context. These incremental increases are extremely small compared to $PM_{2.5}$ concentrations in the Bay Area, and are also much smaller than $PM_{2.5}$ fluctuations throughout the day or day-to-day in the Bay Area.

To demonstrate this, I have tabulated $PM_{2.5}$ air quality data from October 2022, as this is the most recent full month for which data are available (BAAQMD, 2022c). As shown in Table 1, daily average concentrations ranged from 0.1 to 21.9 μg/m^3. The difference between the minimum and maximum average daily $PM_{2.5}$ concentration at a given monitor over the month of October 2022 ranged from 9.1 to 19.4 μg/m^3. Even looking on one particular day (October 1, 2002, for example), average hourly $PM_{2.5}$ concentrations ranged from 0-57 μg/m^3 throughout the Bay Area, and differences between minimum and maximum average hourly concentrations at each monitor varied from 4-54 μg/m^3 (Table 2). BAAQMD (2022d) also provided data on the difference between each hourly average $PM_{2.5}$ concentration and the hourly average $PM_{2.5}$ concentration at the same hour the day before. Using September 31 and October 1 as an example, the differences between hourly concentrations 24 hours apart ranged from -23 to 37 μg/m^3 (Table 3).

These data demonstrate that incremental differences of 0.001-0.3 μg/m^3 are negligible compared to actual $PM_{2.5}$ concentrations and fluctuations in $PM_{2.5}$ in the Bay Area, and indicate that it is not possible to estimate any actual changes in risk associated with such small changes, given the large incremental $PM_{2.5}$ fluctuations seen on hourly and daily bases.

5 Conclusions

Hundreds of epidemiology studies have evaluated $PM_{2.5}$ associations with morbidity and mortality. While statistical associations have been reported, these associations do not provide evidence for causation at ambient $PM_{2.5}$ concentrations, because of study limitations (*e.g.*, exposure measurement error, confounding). In addition, the reliability of the results from the BenMAP-CE model (US EPA, 2022) used for asthma risk estimates by BAAQMD (2022a) has not been demonstrated. These issues led to the overestimation of incremental risks associated with the very small increases (0.001-0.3 $\mu g/m^3$) in $PM_{2.5}$ that BAAQMD models. When considering that $PM_{2.5}$ concentrations can vary considerably over the course of a day and from day to day, it is clear that the BAAQMD model highly overestimates incremental increases in risk associated with $PM_{2.5}$ increments of 0.001-0.3 $\mu g/m^3$.

References

Aleksunes, LM; Eaton, DL. 2019. "Principles of toxicology." In *Casarett and Doull's Toxicology: The Basic Science of Poisons (Ninth Edition)*. (Ed.: Klaassen, CD), McGraw-Hill Education, New York, NY, p25-64.

Bay Area Air Quality Management District (BAAQMD). 2021. "Air Toxics Control Programs Health Risk Assessment Guidelines." 9p., December.

Bay Area Air Quality Management District (BAAQMD). 2022a. "Modeling Local Sources of Fine Particulate Matter (PM2.5) for Risk Management (Draft v.1.1)" 33p., October.

Bay Area Air Quality Management District (BAAQMD). 2022b. "West Oakland Community Action Plan." April 1. Accessed at https://www.baaqmd.gov/community-health/community-health-protection-program/west-oakland-community-action-plan.

Bay Area Air Quality Management District (BAAQMD). 2022c. "Air Quality Data for Fine Particulate Matter (PM-2.5) (Daily)." October. Accessed at https://www.baaqmd.gov/about-air-quality/current-air-quality/air-monitoring-data/#/airp?date=2022-10-01&id=316&view=daily&style=table.

Bay Area Air Quality Management District (BAAQMD). 2022d. "Air Quality Data for Fine Particulate Matter (PM-2.5) (Hourly)." October. Accessed at https://www.baaqmd.gov/about-air-quality/current-air-quality/air-monitoring-data/#/airp?date=2022-10-01&id=316&view=hourly&style=table.

Boffetta, P; McLaughlin, JK; La Vecchia, C; Tarone, RE; Lipworth, L; Blot, WJ. 2008. "False-positive results in cancer epidemiology: A plea for epistemological modesty." *J. Natl. Cancer Inst.* 100:988-995.

Bruckner, JV; Anand, SS; Warren, DA. 2019. "Toxic effects of solvents and vapors." In Casarett and Doull's Toxicology: The Basic Science of Poisons (Ninth Edition). (Ed.: Klaassen, CD), McGraw-Hill Education, New York, NY, p1163-1255.

Bukowski, JA. 2007. "Do time-series studies contain residual confounding by risk factors for acute health events? (Letter)" *Epidemiology* 18:839.

Bukowski, J. 2008a. "Do pollution time-series studies contain uncontrolled or residual confounding by risk factors for acute health events?" *Regul. Toxicol. Pharmacol.* 51:135-140.

Bukowski, JA. 2008b. "Response to counterpoint: Time-series studies of acute health events and environmental conditions are not confounded by personal risk factors by Mark S. Goldberg, Richard T. Burnett, Jeffrey R. Brook." *Regul. Toxicol. Pharmacol.* 51:148-150.

Calabrese, EJ. 2014. "Dose-response: A fundamental concept in toxicology." In *Hayes' Principles and Methods of Toxicology (Sixth Edition)*. (Eds.: Hayes, AW; Kruger, CL), CRC Press, Boca Raton, FL, p89-139.

Castro-Rodriguez, JA; Forno, E; Rodriguez-Martinez, CE; Celedon, JC. 2016. "Risk and protective factors for childhood asthma: What is the evidence?" *J. Allergy Clin. Immunol. Pract.* 4(6):1111-1112. doi: 10.1016/j.jaip.2016.05.003.

De Coster, C; Quan, H; Finlayson, A; Gao, M; Halfon, P; Humphries, KH; Johansen, H; Lix, LM; Luthi, JC; Ma, J; Romano, PS; Roos, L; Sundararajan, V; Tu, JV; Webster, G; Ghali, WA. 2006. "Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: Report from an international consortium." *BMC Health Serv. Res.* 6:77. doi: 10.1186/1472-6963-6-77.

Ellenhorn, MJ; Barceloux, DG. 1988. "Salicylates." In *Medical Toxicology: Diagnosis and Treatment of Human Poisoning*. Elsevier, New York, 3p.

Faustman, EM. 2019. "Risk assessment." In *Casarett and Doull's Toxicology: The Basic Science of Poisons (Ninth Edition)*. (Ed.: Klaassen, CD), McGraw-Hill Education, New York, NY, p127-155.

Glymour, MM; Greenland, S. 2008. "Causal diagrams." In *Modern Epidemiology (Third Edition)*. (Eds.: Rothman, KJ; Greenland, S; Lash, TL), Lippincott Williams & Wilkins, Philadelphia, PA, p183-209.

Goldberg, MS; Burnett, RT; Brook, JR. 2008. "Counterpoint: Time-series studies of acute health events and environmental conditions are not confounded by personal risk factors." *Regul. Toxicol. Pharmacol.* 51:141-147.

Gordis, L. 2009. *Epidemiology (Fourth Edition)*. Saunders Elsevier, Philadelphia, PA, 375p.

Grosser, T; Smyth, E; FitzGerald, GA. 2011. "Anti-inflammatory, antipyretic, and analgesic agents; Pharmacotherapy of gout." In *Goodman and Gilman's The Pharmacological Basis of Therapeutics (Twelfth Edition)*. (Eds.: Brunton, LL; Chabner, BA; Knollman, BC), McGraw-Hill, New York, NY, p959-1004.

Hayes, AW; Kruger, CL; eds.. 2014. *Hayes' Principles and Methods of Toxicology (Sixth Edition)*. CRC Press, Boca Raton, FL, 2157p.

National Academies of Sciences, Engineering, and Medicine (NASEM). 2022. "Advancing the Framework for Assessing Causality of Health and Welfare Effects to Inform National Ambient Air Quality Standard Reviews." National Academies Press, Washington, DC, 220p. doi: 10.17226/26612. Accessed at https://nap.nationalacademies.org/catalog/26612/advancing-the-framework-for-assessing-causality-of-health-and-welfare-effects-to-inform-national-ambient-air-quality-standard-reviews.

National Research Council (NRC). 2002. "Estimating the Public Health Benefits of Proposed Air Pollution Regulations." Committee on Estimating the Health-Risk-Reduction Benefits of Proposed Air Pollution Regulations. National Academies Press, Washington, DC, 170p. Accessed at http://www.nap.edu/catalog.php?record_id=10511.

Paustenbach, DJ; Madl, AK. 2014. "The practice of exposure assessment." In *Hayes' Principles and Methods of Toxicology (Sixth Edition)*. (Eds.: Hayes, AW; Kruger, CL), CRC Press, Boca Raton, FL, p453-525.

Pedersen, SE; Hurd, SS; Lemanske, RF Jr.; Becker, A; Zar, HJ; Sly, PD; Soto-Quiroz, M; Wong, G; Bateman, ED; Global Initiative for Asthma. 2011. "Global strategy for the diagnosis and management of asthma in children 5 years and younger." *Pediatr. Pulmonol.* 46(1):1-17. doi: 10.1002/ppul.21321.

Prueitt, RL; Li, W; Edwards, L; Zhou, J; Goodman, JE. 2021. "Systematic review of the association between long-term exposure to fine particulate matter and mortality." *Int. J. Environ. Health Res.* doi: 10.1080/09603123.2021.1901864.

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

Shiff, NJ; Jama, S; Boden, C; Lix, LM. 2014. "Validation of administrative health data for the pediatric population: A scoping review." *BMC Health Serv. Res.* 14:236. doi: 10.1186/1472-6963-14-236.

Smith, AE; Gans, W. 2015. "Enhancing the characterization of epistemic uncertainties in PM2.5 risk analyses." *Risk Anal.* 35(3):361-378. doi: 10.1111/risa.12236.

Smith, AE; Chang, W. [NERA Economic Consulting]. 2020. "Exposure Window Uncertainty in Chronic Effects Associations of PM2.5 and Health Risks." Report to NAAQS Regulatory Review and Rulemaking (NR3) Coalition. Submitted to US EPA Docket. EPA-HQ-OAR-2015-0072-0708. 34p., June 28. Accessed at https://www.regulations.gov/document?D=EPA-HQ-OAR-2015-0072-0708.

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

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

Tetreault, LF; Doucet, M; Gamache, P; Fournier, M; Brand, A; Kosatsky, T; Smargiassi, A. 2016. "Childhood exposure to ambient air pollutants and the onset of asthma: An administrative cohort study in Quebec." *Environ. Health Perspect.* 124(8):1276-1282. doi: 10.1289/ehp.1509838.

US EPA. 2022. "Environmental Benefits Mapping and Analysis Program - Community Edition (BenMAP-CE)." October 4. Accessed at https://www.epa.gov/benmap.

Valberg, PA. 2003. "Possible noncausal bases for correlations between low concentrations of ambient particulate matter and daily mortality." *Nonlinearity Biol. Toxicol. Med.* 1(4):521-530. doi: 10.1080/15401420390271137.

Tables

Table 1 Daily Average $PM_{2.5}$ Concentrations (µg/m^3) in the Bay Area Air Quality Management District (BAAQMD) – October 2022

Monitor	Date in October 2022																															Average	Minimum	Maximum	Maximum-Minimum
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31				
San Rafael	5.2	6.1	6	5.7	4.9	4.2	3.2	2	2.1	3.6	9.2	8.2	2.3	5.5	8	10	7	10	8.2	11	12	5.8	4.6	4.7	4.4	6.1	5.7	6.8	8.1	9.7	12	6.6	2.0	12.4	10.4
Sebastopol	4.2	5	5.2	5	3.5	2.7	2	1.9	1	2.4	5	4.4	1.7	2.2	6.5	8.7	6.2	7.5	5.9	6.7	8.8	4.7	4	4.3	4.3	4.3	6.4	5.7	11	9.1	8.7	5.1	1.0	11.1	10.1
Vallejo	3.9	4.8	3.3	6.7	5.7	4.5	2.6	3	2.6	4	9.2	7.2	1.5	4.7	6.3	9.4	6.4	11	9.5	8.9	11	2.7	5.7	2.1	4.5	5	5.7	8.2	7.6	9.7	9.5	6.0	1.5	10.6	9.1
Laney College	5	8.2	7.7	8.8	8	6.1	5.5	6.4	5.9	7.2	15	11	5.5	8.9	11	13	12	15	17	13	16	7.2	3.8	6.5	7.7	10	9.6	9.5	10	13	13	9.6	3.8	16.8	13.0
Oakland East	6.5	5.2	6	7.6	7	7.4	5.9	7.5	5.1	7.7	13	10	3.9	7.6	8.5	11	9.7	13	11	13	15	7.2	4.3	5.5	5.1	7.1	7.5	8	9.2	9.6	11	8.3	3.9	15.2	11.3
Oakland West	4	4.1	5.6	8.1	4.6	4.5	2.6	3.5	2	4.8	12	8.5	2.3	5.1	9.1	9.6	9	12	12	10	13	3.7	2	3.7	6.6	7.7	8.5	7.4	9.6	8.8	11	6.9	2.0	12.9	10.9
San Francisco - Arkansas St.	2	3.4	4.5	4.6	3.9	2.5	1.1	0.1	0.7	1.6	12	7.5	2.4	3.6	7.7	9.8	9.2	12	11	10	12	3.1	3.7	2.3	2.6	7.3	7.5	5.5	5.6	7.8	11	5.7	0.1	12.3	12.2
San Pablo - Rumrill	14	13	7.6	9.2	6.5	4.6	5.5	4.4	7.6	6.6	13	9.8	3.7	10	22	20	9.4	9.9	11	11	19	5.2	2.5	6.6	7.8	6.7	7.8	9.4	11	12	14	9.7	2.5	21.9	19.4
Concord	4.6	4.7	4.6	7	8.1	7.3	8.1	6	4.6	5.7	9.5	8.2	4.1	6.2	8.7	9.7	7.6	9.2	5.7	7.1	13	3	3	2.6	3.5	6	5	6.5	9.6	8.3	11	6.7	2.6	12.5	9.9
Livermore - Rincon Ave.	5.9	5.3	5.3	6	7	8	9.9	7.1	5.8	6.9	9.8	12	6.1	7.5	10	9	7	9.5	7.6	7.4	13	6.1	3.1	3.1	4.8	5.2	5.2	5.5	8	8.9	11	7.3	3.1	12.9	9.8
Pleasanton - Owens Ct.	5.8	4.2	5.7	8.1	7.9	8	8.2	6.6	6.9	7.3	12	12	6.6	9.5	9.8	8.7	8.6	10	7.9	8.5	16	5.1	2.1	4.1	6.6	7.4	7.5	4.7	7.2	11	12	7.9	2.1	15.7	13.6
Redwood City	3	4	4.4	5.8	6.8	5	4.3	3.7	4.2	6.5	11	9.9	3.3	5.9	5.8	7.3	5	8.9	10	8.7	13	5	1.9	1.6	5.7	3.3	5.3	5.3	6	9.2	10	6.1	1.6	13.2	11.6
Gilroy	2.9	5.7	6.1	8.9	8.3	9.3	9.3	7.3	6.7	8.6	9.1	9.5	7.4	9	6.5	11	6.3	9.3	7.6	11	17	9.4	4.5	5	6.4	6.3	5.8	6.2	8.6	11	13	8.2	2.9	16.9	14.0
San Jose - Jackson St.	9	5.4	7.1	9.4	13	11	11	11	11	12	19	16	7.9	12	9.6	9.6	8.5	16	17	15	18	11	4.9	7.3	6.9	6.5	8.1	8.5	11	15	17	11.1	4.9	19.2	14.3
San Jose - Knox Ave.	5.9	4.3	6.6	7.6	11	9.4	8.6	7.5	7.7	10	16	14	7.2	11	6.2	9.1	8	16	13	14	18	9.4	5	6.3	6.3	6.8	8.5	8.9	12	12	16	9.7	4.3	17.6	13.3

Notes:

$PM_{2.5}$ = Fine Particulate Matter.

Source: BAAQMD (2022c).

Table 2 Hourly Average $PM_{2.5}$ Concentrations (µg/m³) in the Bay Area Air Quality Management District (BAAQMD) – October 1, 2022

Monitor	Hour																								Average	Minimum	Maximum	Maximum-Minimum
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23				
San Rafael	2	1	4	3	0	2	4	3	7	6	4	3	1	2	5	6	16	11	10	10	9	6	7	5	5.3	0	16	16
Sebastopol	0	0	0	0	0	2	1	0	0	2	2	2	5	4	26	7	6	5	6	8	6	8	5	6	4.2	0	26	26
Vallejo	3	-1	0	2	5	5	3	2	2	6	6	5	6	6	3	4	6	3	5	7	5	4	4	3	3.9	-1	7	8
Laney College	5	5	7	8	4	4	4	5	6	5	4	5	5	1	2	5	4	5	6	7	8	5	4	6	5.0	1	8	7
Oakland East	8	7	8	13	7	7	7	4	5	6	7	7	4	3	6	5	7	6	6	9	9	6	6	5	6.6	3	13	10
Oakland West	3	4	4	3	3	2	1	7	5	6	5	2	1	1	1	3	5	5	8	7	7	7	4	3	4.0	1	8	7
San Francisco - Arkansas St.	-1	3	4	3	3	-1	-1	3	1	0	1	3	2	1	0	-1	3	3	4	5	4	4	3	2	2.0	-1	5	6
San Pablo - Rumrill	7	7	4	4	4	3	4	18	21	57	42	18	24	20	18	11	12	13	11	9	9	6	11	7	14.2	3	57	54
Concord	4	2	5	5	5	8	4	3	3	4	5	4	2	5	6	4	3	6	4	7	5	4	8	5	4.6	2	8	6
Livermore - Rincon Ave.	6	4	7	8	7	7	5	11	8	5	9	8	5	8	7	4	3	1	6	5	2	7	5	4	5.9	1	11	10
Pleasanton - Owens Ct.	1	3	6	5	9	8	9	6	9	9	8	7	4	1	5	4	9	7	2	2	9	6	7	5	5.9	1	9	8
Redwood City	-3	-3	-4	-2	4	7	6	10	9	4	6	8	7	5	3	5	6	3	4	3	1	-1	-2	-2	3.1	-4	10	14
Gilroy	7	9	5	3	2	1	2	3	2	1	-1	1	2	1	2	3	3	2	0	1	8	6	4	3	2.9	-1	9	10
San Jose - Jackson St.	6	19	5	7	8	10	10	9	18	18	9	10	13	13	10	5	7	7	4	4	8	8	5	5	9.1	4	19	15
San Jose - Knox Ave.	6	6	6	6	6	5	3	7	5	5	4	6	6	7	7	6	6	7	7	6	6	7	5	7	5.9	3	7	4

Notes:

$PM_{2.5}$ = Fine Particulate Matter.

Source: BAAQMD (2022d).

Table 3 Difference Between Average $PM_{2.5}$ Concentration (µg/m³) at the Same Hour on October 1 and September 31, 2022, in the Bay Area Air Quality Management District (BAAQMD)

Monitor	Hour																							Average	Minimum	Maximum
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23			
San Rafael	-3	-4	0	0	-6	-7	-3	-4	-6	-7	-5	-5	-7	-3	0	0	8	5	4	5	7	1	4	4	-0.9	-7	8
Sebastopol	-4	-4	-4	-3	-2	-1	-1	-3	-5	-2	-3	-2	1	0	21	4	2	0	2	3	1	4	4	6	0.6	-5	21
Vallejo	-1	-9	-6	-2	0	0	-1	-8	-8	-3	0	-3	1	3	0	-1	-2	-5	-1	5	5	3	0	-2	-1.5	-9	5
Laney College	-6	-3	3	1	-6	-7	-5	-1	-14	-13	-14	-8	-18	-10	-7	0	1	2	1	2	4	2	-1	1	-4.0	-18	4
Oakland East	2	-2	1	6	0	0	-1	-5	-5	-3	-3	1	0	-9	-7	-7	0	-1	-1	4	4	-2	1	1	-1.1	-9	6
Oakland West	-6	-7	-4	-4	-3	-4	-5	-12	-6	-14	-14	-16	-16	-13	-12	-6	0	-1	0	0	-1	2	0	-1	-6.0	-16	2
San Francisco - Arkansas St.	-11	-6	-2	-2	-4	-12	-15	-13	-21	-19	-22	-18	-14	-5	-4	-2	3	3	2	1	2	3	-2	0	-6.6	-22	3
San Pablo - Rumrill	-3	-3	-2	2	-6	-4	-2	13	4	37	16	-2	7	10	7	2	5	7	7	5	6	2	8	5	5.0	-6	37
Concord	2	2	3	2	2	5	1	-2	-3	1	2	0	-2	-1	1	0	-1	1	1	5	2	-1	5	1	1.1	-3	5
Livermore - Rincon Ave.	-1	1	5	4	2	2	1	4	0	0	5	4	3	6	5	-2	-8	-7	0	-1	-5	-1	-1	-2	0.6	-8	6
Pleasanton - Owens Ct.	-4	-2	1	1	7	3	1	6	-1	0	3	4	0	-7	-2	-3	3	5	-5	-1	9	4	5	4	1.3	-7	9
Redwood City	-4	-6	-7	-3	4	7	2	3	-6	-14	-13	-4	-2	-2	-8	-7	-5	-4	2	3	0	-2	-2	-2	-2.9	-14	7
Gilroy	0	3	2	0	-2	-5	-2	0	-4	-10	-10	-5	-4	-4	-3	-1	-4	-9	-9	-7	3	1	-2	-2	-3.1	-10	3
San Jose - Jackson St.	-3	-1	-7	-5	-2	-8	8	4	10	8	-3	-1	5	5	5	1	1	2	-3	-23	-4	-1	3	1	-0.3	-23	10
San Jose - Knox Ave.	-7	-4	0	2	-3	-4	-4	4	2	2	-4	-2	0	1	1	-1	1	2	0	-2	1	5	-1	1	-0.4	-7	5

Notes:

$PM_{2.5}$ = Fine Particulate Matter.

Source: BAAQMD (2022d).

Red text indicates that concentrations were higher on October 1, or the same on both days if "0." Green text indicates that concentrations were lower on October 1.

Attachment A

"Systematic Review of the Association Between Long-Term Exposure to Fine Particulate Matter and Mortality" (Prueitt *et al*., 2021)

INTERNATIONAL JOURNAL OF ENVIRONMENTAL HEALTH RESEARCH
https://doi.org/10.1080/09603123.2021.1901864

Taylor & Francis
Taylor & Francis Group

Systematic review of the association between long-term exposure to fine particulate matter and mortality

Robyn L. Prueitt[a], Wenchao Li[b], Lariah Edwards[b], Jieqiong Zhou[b] and Julie E. Goodman[b]

[a]Gradient, Seattle, WA, USA; [b]Gradient, Boston, MA, USA

ABSTRACT

We used a transparent systematic review framework based on best practices for evaluating study quality and integrating evidence to conduct a review of the available epidemiology studies evaluating associations between long-term exposure to ambient concentrations of $PM_{2.5}$ and mortality (all-cause and non-accidental) conducted in North America. We found that while there is some consistency across studies for reporting positive associations, these associations are weak and several important methodological issues have led to uncertainties with regard to the evidence from these studies, including potential confounding by measured and unmeasured factors, exposue measurement error, and model misspecification. These uncertainties provide a plausible, alternative explanation to causality for the weakly positive findings across studies. Using a causality framework that incorporates best practices for making causal determinations, we concluded that the evidence for a causal relationship between long-term exposure to ambient $PM_{2.5}$ concentrations and mortality from these studies is inadequate.

ARTICLE HISTORY
Received 4 December 2020
Accepted 8 March 2021

KEYWORDS
Causal framework; mortality; particulate matter

Introduction

Particulate matter (PM) is the generic term for a mixture of solid particles and liquid droplets in various size fractions in ambient air that comprises the particle phase of air pollution. PM originates from numerous primary sources, including industrial activities, fossil fuel combustion, motor vehicles, crustal material, and burning of natural materials (e.g. forest fires) (US EPA 2019). Secondary PM can be formed in ambient air from chemical reactions of gaseous pollutants such as nitrogen oxides, sulfur oxides, and volatile organic compounds (US EPA 2019). As a consequence of this wide variety of sources, PM has a variable chemical composition and particle size distribution.

While the toxicity of PM is dependent on the chemical composition of the particles, particle size is also an important characteristic with respect to potential health effects from exposure to PM (Miller 2014; US EPA 2019). Different sized particles can penetrate into different regions of the respiratory tract, with potential for different health outcomes. The United States Environmental Protection Agency (US EPA) evaluates the potential health effects of exposure to three main size fractions of PM, classified according to the aerodynamic diameter of particles (US EPA 2019). Coarse or thoracic coarse PM ($PM_{10\text{-}2.5}$) has a nominal mean aerodynamic diameter > 2.5 μm and ≤ 10 μm, and is largely comprised of particles such as soil and street dust, road wear debris, fly ash,

CONTACT Julie E. Goodman jgoodman@gradientcorp.com Gradient, One Beacon Street, 17th Floor, Boston, MA 02108, USA

Present affiliation of Lariah Edwards: Department of Environmental and Occupational Health, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052

Supplemental data for this article can be accessed here.

oxides of crustal elements, sea salt, nitrates, sulfates, and biological aerosols (e.g. pollen, fungal spores, mold). Fine PM ($PM_{2.5}$) has a nominal mean aerodynamic diameter ≤ 2.5 μm and is typically comprised of water; elemental carbon; low and moderate volatility organic compounds; metal compounds; and sulfate, nitrate, ammonium, and hydrogen ions. Ultrafine particles (UFPs) are generally considered to have a diameter ≤ 0.1 μm based on physical size, thermal diffusivity, or electrical mobility, and are commonly comprised of elemental carbon, low volatility organic compounds, metal compounds, and sulfate. In addition, PM with a nominal mean aerodynamic diameter of ≤ 10 μm (which includes all of the above PM size fractions) is referred to as PM_{10} or thoracic PM, though US EPA does not focus on this particular size fraction for evaluations of health effects.

Particulate matter is one of six criteria air pollutants for which the Clean Air Act (CAA) mandates the US EPA set health-based National Ambient Air Quality Standards (NAAQS). In 2012, the US EPA established a new annual $PM_{2.5}$ primary NAAQS of 12 $\mu g/m^3$ (annual mean averaged over 3 years) and retained the 24-hour $PM_{2.5}$ NAAQS of 35 $\mu g/m^3$ (98th percentile averaged over 3 years) previously set in 2006 (US EPA 2013). The CAA mandate also requires that the NAAQS for each criteria air pollutant be reviewed every 5 years. As part of this review process, the US EPA develops an Integrated Science Assessment (ISA) for each criteria air pollutant, in which causal relationships between criteria air pollutant exposures and various human health and welfare effects are evaluated using a framework that US EPA developed specifically for this purpose (US EPA 2015). The most recent ISA for PM was finalized in 2019 (US EPA 2019) and the human health effects evaluation focused on studies of short- and long-term exposures to PM at concentrations relevant to the range of human ambient exposures that were published after those reviewed in the previous PM ISA, which was finalized in 2009 (US EPA 2009).

Over the last several decades, the epidemiology literature has evaluated associations between $PM_{2.5}$ exposure and mortality. US EPA (2019) conducted a comprehensive review of this literature in the PM ISA and concluded that there is a causal relationship between long-term (i.e. one month or longer) exposure to $PM_{2.5}$ and total (nonaccidental) mortality. The Preamble to the ISAs (US EPA 2015) describes the general framework for evaluating scientific evidence (referred to herein as the 'NAAQS framework') and the Appendix of the PM ISA (US EPA 2019) provides aspects for assessing the quality of studies of PM exposure, but neither document provides detailed guidance on evidence evaluation and causal determinations. We have recently developed a more transparent systematic review and causality framework that is based on the NAAQS framework but is modified to incorporate best practices for evaluating study quality, evaluating and integrating evidence, and making causal determinations, to allow for a scientifically sound assessment of the evidence (Goodman et al. 2020). Here, we evaluate the epidemiology literature on the association between long-term exposure to ambient concentrations of $PM_{2.5}$ and mortality (all-cause and non-accidental) using our modified framework. To be consistent with the evaluation in the most recent PM ISA, we limit our analysis to epidemiology studies published after those included in the 2009 PM ISA. We also limit our analysis to studies conducted in North America, as these are most generalizable to the US population and therefore are most relevant to the PM NAAQS. We contrast our analysis to that conducted by US EPA in the PM ISA and consider whether and how differences between the NAAQS framework and our modified framework led to different conclusions regarding causality.

Methods

Literature searches and study selection

The principal question of our evaluation is whether the available evidence supports a causal relationship between long-term exposure to $PM_{2.5}$ and mortality (all-cause or non-accidental) at ambient concentrations. We searched the PubMed and Scopus databases for epidemiology studies

published between 1 January 2009, and 1 January 2020, using the following terms: (PM2.5 OR 'PM2.5' OR 'particulate matter 2.5') AND (exposure OR exposures OR exposed) AND (mortality OR death) AND ('all cause' OR 'total mortality' OR 'long term'). We also cross-referenced the PM ISA (US EPA 2019) and the bibliographies of relevant review articles to identify additional studies that were not included in the literature search results.

We included peer-reviewed, observational studies that evaluated the association between long-term exposure to $PM_{2.5}$ (defined by US EPA as one month or longer in duration; US EPA 2019) and all-cause or non-accidental mortality. We excluded studies that met any of the following criteria: laboratory animal, *in vitro*, experimental, or controlled human exposure studies; studies that were not published in English; studies that evaluated constituents of $PM_{2.5}$ but did not include any evaluation of total $PM_{2.5}$; studies that did not evaluate long-term, ambient $PM_{2.5}$ exposure (*i.e.* studies that evaluated short-term, indoor, or source specific $PM_{2.5}$ exposures); studies that only evaluated cause-specific mortality and not all-cause mortality; studies that used relative risk estimates or concentration-response information from other epidemiology studies; reviews; editorials; commentaries; correspondence/communications; letters to the editor; studies reviewed in the 2009 PM ISA (US EPA 2009), and studies conducted outside of North America.

After identifying studies that met our inclusion and exclusion criteria, we further narrowed down the list of studies to focus on in our evaluation. If we identified more than one study of the same cohort, we included only the most recent study or the one or two studies reporting the most informative data regarding the $PM_{2.5}$-mortality association in the cohort (*e.g.* greater population coverage, improved PM exposure estimates, and/or improved statistical analysis with copollutant adjustments, non-linearity examination, or additional confounder adjustments). We excluded ecological studies, because such studies are subject to ecological bias. We also excluded studies that used new or causal modeling approaches, because those approaches have not been widely applied or accepted.

Study quality criteria

The appendices of the most recent ISAs for criteria pollutants, including the PM ISA (US EPA 2019), provide a discussion of study quality aspects to consider for evaluating epidemiology studies of the respective pollutant. As some of these aspects are either lengthier or less detailed than others, we previously compiled the aspects into a table that included succinct criteria for what is indicative of higher quality for each aspect, and recommended additional aspects and criteria based on our survey of best practices for evaluating study quality (Goodman et al. 2020). Here, we further modified these aspects and criteria to be specific to epidemiology studies of $PM_{2.5}$ and mortality (Table 1).

The study quality criteria include a total of 36 specific aspects of epidemiology studies, grouped into seven general categories (study design, study population, pollutant specification, $PM_{2.5}$ exposure assessment, mortality outcome assessment, confounding, and statistical methods), that are informative of potential bias and uncertainties. While the majority of these aspects assess important dimensions of study conduct (i.e. those in bold font in Table 1), some assess the clarity of study reporting (e.g. those regarding study objectives, participant characteristics, inclusion/exclusion criteria, pollutant description, descriptive statistics, and univariate analyses). Because aspects of $PM_{2.5}$ specification (i.e. pollutant description and pollutant source) were incorporated in the inclusion/exclusion criteria, and the assessment of outcomes (i.e. all-cause or non-accidental mortality) is much less subjective to misclassification compared to other disease endpoints (e.g. incidence or cause-specific mortality), we focused more on the other five categories of quality criteria, and particularly on the aspects related to study conduct within these categories, in our evaluation of study quality.

Several details with respect to the study quality criteria are worth noting. Regarding sample size, we considered cohorts with sample sizes ≥1 million to be sufficient without power calculations.

Table 1. Quality criteria for epidemiology studies of $PM_{2.5}$ and mortality.

Category	Aspect	Criteria for Higher Quality
Study Design	Study Objectives	Objectives/hypotheses are clearly described
	Study Design	**Cohort studies and nested case-control studies are considered better than cross-sectional, other case-control, or ecologic studies**
	Study Location	Multiple cities rather than a single city
	Sample Size	**Power calculation is presented to indicate sufficient sample size**
	Study Duration	Conducted over multiple years
Study Population	Participant Characteristics	Characteristics (*e.g.* age, race, sex, eligibility criteria) are reported
	Inclusion/Exclusion Criteria	Inclusion/exclusion criteria clearly reported and consistent with study objectives
	Recruitment -Representativeness of Source	**Representative of the source population**
	Participation Rate	**High participation rate, not dependent on exposure or outcome**
	Underlying Health Conditions	**Underlying conditions ascertained by independent clinical assessment or self-report of physician diagnosis**
	Follow-up	**Minimal or non-differential loss to follow-up**
Pollutant	Pollutant Description	Pollutant clearly described (*e.g.* size of PM fraction)
	Pollutant Source	Pollutant source-related indicators evaluated
Exposure Assessment	Measurement Methods	Utilized and compared more than one exposure assessment method
	Measurement Validity	**Used well-established, sensitive methods: Direct measurements of exposure or indirect measurements that have been validated**
	Spatial Variability	**Sufficiently captured the spatial variability of the exposure: When only using monitoring data, exposure was estimated from the closest central site monitor (within ≤5km) or from averaging concentrations from multiple monitors; when models were used, exposure was estimated with sufficient spatial resolution (e.g. at 10km^2 grid)**
	Temporal Variability	**Used time-varying or multiple lags of exposure estimates**
	Temporality	**Exposure occurred before the outcome**
	Assignment to Participants' Locations	**Measured or estimated ambient air pollutant data were assigned to participants' locations (e.g. ZIP codes, census tracts, counties) recorded in the same period**
	Residential Mobility	**Residential mobility was accounted for**
	Personal Activities	**Personal activities (e.g. time spent indoors) were accounted for**
Outcome Assessment	Blinding	Assessors of outcome were blinded to exposure levels
	Time Points	Time points of outcome evaluation are consistent with study objectives

(*Continued*)

Table 1. (Continued).

Category	Aspect	Criteria for Higher Quality
Confounding	**Key Confounders**	**Identified and adjusted for potential confounders and primary covariates (*e.g.* socioeconomic status, age, sex, race, body mass index, diet, physical activity, temperature, relative humidity, medication use, smoking status, and other chemical exposures)**
	Proper Adjustment	**Did not adjust for inappropriate covariates (*e.g.* mediators).**
	Temporal Conceptualization	Appropriately conceptualized temporality and accounted for temporal variation (e.g. when examining time-varying PM, also appropriately adjusted for time-varying covariates. Note: this does not apply to characteristics that in general do not change over time, such as sex, race, education, etc.)
	Copollutant Adjustment	**Copollutant (*e.g.* two-pollutant) modeling was conducted**
	Copollutant Correlations	**Correlations between the pollutant of interest and copollutants were considered**
	Copollutant Measurement	Appropriately accounted for temporal variation, spatial variation, residential mobility, and personal activities
Statistical Methods	Descriptive Statistics	Summary statistics for the study population are presented
	Univariate Analyses	Univariate analyses with pollutant of interest, covariates, and copollutants were conducted and results are presented
	Multivariate– Model Specification	**Employed appropriate statistical models for multivariate analyses (e.g. appropriately accounted for correlated data, complex survey design, time-varying data)**
	Multivariate– Model Assumptions	**Model assumptions were tested and satisfied**
	Multivariate– Nonlinearity	**Nonlinearity was assessed statistically**
	Multivariate– Multiple Comparison	**Multiple comparison-corrected, if applicable**
	Sensitivity Analyses	Sensitivity analyses were conducted

Notes:
Bold font indicates the aspects and criteria that assess important dimensions of study conduct.
PM = Particulate Matter.

Because of the large number (e.g. 10+ or even 20+) of potential confounders that are usually adjusted for in air pollution studies, we considered all other sample sizes to require justification based on a power calculation.

Regarding recruitment/participation, if a study was a secondary analysis of data from an existing cohort that was initially recruited for research questions unrelated to $PM_{2.5}$ or mortality, we considered the criteria related to whether the study population is representative of the source population and the participation rate as not applicable, because such studies usually had to exclude participants for logistic reasons (*e.g.* not having data for $PM_{2.5}$ or mortality) rather than through a recruitment/participation process. Similarly, for secondary analyses, we considered the criterion for follow-up as not applicable if the authors used linkage to conveniently identify mortality outcomes in existing cohorts.

Regarding exposure assessment, we considered either a comparison of modeled *vs.* monitored $PM_{2.5}$ or multiple $PM_{2.5}$ modeling approaches as utilizing and comparing more than one exposure assessment method. Regarding spatial variability, we considered 10 km^2 or more refined grids as sufficient for modeled $PM_{2.5}$, as is generally accepted; we considered 5 km or smaller buffers as sufficient for direct site measurements of $PM_{2.5}$ (i.e. 'monitored $PM_{2.5}$'), in order to reduce the potential for measurement error.

With regard to confounding, we considered the criterion for adjustment for potential confounders to be met only if all the listed key confounders in Table 1 were adjusted for. It is worth noting that this is not an exhaustive list of important potential confounders. The $PM_{2.5}$-mortality association could also be confounded by factors that are not typically measured in epidemiology studies of $PM_{2.5}$ and mortality, such as stress and noise (Clougherty and Kubzansky 2009; Stansfeld 2015; US EPA 2019). There could also be residual confounding due to incomplete adjustment of covariates (e.g. socioeconomic status [SES]) and/or lack of adjustment for confounding by secular trend and unknown confounders.

Regarding statistical methods, we focused only on key, testable assumptions (e.g. proportional hazards assumption for Cox proportional hazards model) for the criterion regarding model assumptions, and we considered five or more comparisons based on the same model/analysis pertaining to the $PM_{2.5}$-mortality association of interest to be subject to the multiple comparison issue and thus needing correction (*e.g.* Bonferroni correction).

We tabulated whether each of the included studies met each of the criteria listed in Table 1. This tabulation allowed for a consistent evaluation of study quality across all studies, by considering whether certain studies met more of the criteria for higher quality than other studies. We used the study quality criteria to identify the strengths and limitations of the studies, and used these to evaluate the study results, as discussed below.

Evidence integration

We assessed the results of the studies in the context of their methodological strengths and limitations (as determined from the analysis of study quality aspects and criteria) and evaluated the reliability of each study's results to inform potential causality. We then integrated the evidence across studies using Bradford Hill aspects (Hill 1965) modified from those listed in the Preamble to the ISAs (US EPA 2015) to be more succinct, as described by Goodman et al. (2020) (Supplemental Table S1). We did not use the Bradford Hill aspects as a checklist, as not meeting one or more of the aspects should not automatically preclude a conclusion of causality; rather, the aspects were used to provide a framework to systematically evaluate the weight of the evidence for making causal determinations. It is difficult to imagine a situation in which an association is not causal if every one of these aspects is met, however. Thus, if all of the Bradford Hill aspects are met, we concluded that the evidence as a whole supports causation. By contrast, it may be difficult to conclude that observed associations are causal if most or all of the aspects are not met. Thus, if not all of the Bradford Hill aspects were met, we determined whether it is more likely that the evidence as a whole

supports causation (i.e. we provided likely explanations for any aspect that was not met), is suggestive of causation, is inadequate to determine causation, or supports no causation, as described below.

Causal conclusion

To form a conclusion regarding causality, we used a four-tiered framework for causality that is consistent with other causal frameworks, such as that defined in the Institute of Medicine (IOM) report *Improving the Presumptive Disability Decision-Making Process for Veterans* (IOM 2008) (Supplemental Table S2). This differs from the current NAAQS framework, which uses five categories for causation (*causal, likely causal, suggestive, inadequate*, and *not likely causal*). As discussed by Goodman et al. (2020), US EPA's definitions of these categories preclude the need for a *likely causal* category, which can instead be represented by the *suggestive* category in a four-tiered framework.

Consistent with the four-tiered framework, if all the modified Bradford Hill aspects were met, we concluded that the relationship between long-term exposure to $PM_{2.5}$ and mortality is *causal*. If most of the aspects were met and there is a likely explanation for each that was not met, we also concluded that the relationship is *causal*. If there was inadequate information to assess some of the modified Bradford Hill aspects and all other aspects were met, we concluded that the evidence for a causal relationship is *suggestive*. If there was inadequate information to assess some of the Bradford Hill aspects and there was a likely explanation for each of the other aspects that was not met, we also concluded that the evidence for a causal relationship is *suggestive*. If there was inadequate information to assess most or all of the modified Bradford Hill aspects, we concluded that the evidence for a causal relationship is *inadequate*. If most or all of the aspects were not met and there is no likely explanation for why they were not met, we also concluded that the evidence for a causal relationship is *inadequate*. If the overall evidence indicated there is no causal relationship based on the modified Bradford Hill aspects (*e.g.* there was a consistent lack of an association in robust epidemiology studies), we concluded that the relationship between long-term exposure to $PM_{2.5}$ and mortality is *not causal*.

Results

Literature selection

Our literature search for epidemiology studies evaluating the association between long-term $PM_{2.5}$ exposure and all-cause or non-accidental mortality yielded 360 studies in PubMed and 115 studies in Scopus. We also reviewed the reference lists of three relevant reviews identified in our PubMed and Scopus searches, which contained 321 studies, and the section of the PM ISA that evaluated 34 North American studies of long-term exposure to $PM_{2.5}$ and mortality. After a review of titles and abstracts, we identified 127 studies from the PubMed search, 6 studies from the Scopus search, 3 studies from cross-referencing the PM ISA, and 1 study from the reference lists of reviews for full text review. After a full text review, we identified 46 studies that met our inclusion and exclusion criteria. Two of these studies only assessed $PM_{2.5}$ exposure for 30 days; as these studies were outliers compared to the majority of studies that assessed $PM_{2.5}$ exposure for multiple years, we excluded these two studies. We further narrowed down the study selection by excluding ecological studies and the least recent or least informative studies of cohorts examined in multiple studies (as discussed above in the Methods section). The results for this study selection and detailed exclusion rationales are shown in Table 2.

The results of our literature search and study selection are summarized in Supplemental Figure S1. Overall, 23 studies representing 20 underlying cohorts were included in the present review. One study was selected for each cohort, except for the Canadian Census Health and Environment

Table 2. Final study selection rationales.

Cohort	Included Studies ($PM_{2.5}$ Measurement Approach)[a]	Excluded Studies of the Same Cohort	Exclusion Rationale
AHS	Weichenthal et al. (2014) (Modeled)	NA	NA
CanCHEC 1991	Crouse et al. (2015) (Modeled); Weichenthal et al. (2016) (Monitored)	Crouse et al. (2012)	Updated by Weichenthal et al. (2016) and Crouse et al. (2015) with additional follow-up, greater population coverage, improved PM measurement, and further copollutant adjustments.
		Crouse et al. (2016)	Focused on combined effects of PM concentration and particulate composition. Only one reported estimate is relevant to this review, and it did not adjust for copollutants.
CanCHEC 2001	Weichenthal et al. (2017) (Modeled)	Pinault et al. (2017)	PM estimation was very similar to the other two studies, but did not exclude outliers. The study also did not adjust for copollutants.
		Crouse et al. (2019)	PM estimation was very similar to the other two studies, but did not examine non-linearity.
CCHS	Pinault et al. (2016) (Modeled)	NA	NA
CCR-HCC	Deng et al. (2017 (Modeled)	NA	NA
Cleveland Clinic GeneBank Study	Hartiala et al. (2016) (Modeled)	NA	NA
CNBSS	Villeneuve et al. (2015) (Modeled)	NA	NA
CPSII	Jerrett et al. (2009) (Monitored); Turner et al. (2016) (Modeled)	Turner et al. (2017)	The study only compared high (75th percentile) vs. low (25th percentile) levels of modeled PM, which is less comparable to the exposure contrasts in other studies in this review. The study also did not adjust for copollutants.
CTS	Ostro et al. (2010) (Monitored); Ostro et al. (2015) (Modeled)	Lipsett et al. (2011)	PM estimation had lower temporal and spatial resolution compared to that by Ostro et al. (2015).
EFFECT	Chen et al. (2016) (Modeled)	NA	NA
Harvard Six Cities Study	Lepeule et al. (2012) (Monitored)	NA	NA
HPFS	Puett et al. (2011) (Modeled)	NA	NA
Medicare	Di et al. (2017) (Monitored & Modeled)	Greven et al. (2011)	Ecological study, using data at each monitor location. The study used a new statistical model to investigate confounding.
		Kioumourtzoglou et al. (2015)	The study population is a subset of that in the study by Kioumourtzoglou et al. (2016), which used the same source of monitored PM data and statistical analysis approach. The study also used a causal modeling approach.
		Kioumourtzoglou et al. (2016)	The study population is a subset of that in the study by Di et al. (2017), which used the same source of monitored PM data but with improved spatial resolution. The study also used a causal modeling approach.
		Makar et al. (2017)	The study used a causal modeling approach.
		Pun et al. (2017)	Ecological study, using data at each monitor location. Part of the analysis followed the statistical methods described by Greven et al. (2011).
		Shi et al. (2016)	Ecological study, using data at ZIP code or county level.
		Wang et al. (2017a)	The majority of the study population (i.e. Medicare 2000–2013) is a subset of that in the study by Di et al. (2017) (e.g. Medicare 2000–2012), which used improved PM estimate and statistical analyses (i.e. non-linearity, copollutant and confounder adjustment).
		Wang et al. (2017b)	The study applied a new causal modeling approach to the analysis of data from Wang et al. (2017a).
		Wu et al. (2019)	Ecological study, using data at ZIP code/area level. The study also used a new causal modeling approach.
		Yitshak-Sade et al. (2019)	The study was built on the causal modeling approach used by Wang et al. (2016) with extended confounding adjustment.

(*Continued*)

Table 2. (Continued).

Cohort	Included Studies ($PM_{2.5}$ Measurement Approach)[a]	Excluded Studies of the Same Cohort	Exclusion Rationale
NHIS	Lefler et al. (2019) (Modeled)	Pope et al. (2017)	The study population for the fully adjusted analysis is a subset of that in the study by Lefler et al. (2019), which had additional follow-up, more comprehensive PM estimates for copollutant adjustments, and additional confounding adjustments.
NHS	Hart et al. (2015) (Monitored & Modeled)	Puett et al. (2009)	Updated by Hart et al. (2015), with additional follow-up, greater population coverage, and improved PM estimates and statistical analysis.
		Liao et al. (2018)	The study applied a new method to the analysis of Hart et al. (2015) for PM measurement error adjustment.
NHS & NHSII	DuPre et al. (2019) (Modeled)	NA	NA
NIH-AARP	Thurston et al. (2016) (Modeled)	NA	NA
NJ Residents	NA	Wang et al. (2016)	Ecological study, using data at census tract level. The study also used a causal modeling approach.
Older CA Residents	NA	Garcia et al. (2016)	Ecological study, using data at ZIP code level.
TrIPS	Hart et al. (2011) (Monitored)	NA	NA
TRIUMPH & PREMIER	Malik et al. (2019) (Modeled)	NA	NA
VCMS	Lipfert and Wygza (2019) (Modeled)	NA	NA

Notes:

(a) We considered direct site measurements of $PM_{2.5}$ without any statistical modeling as 'monitored.'

AHS = Agricultural Health Study; CA = California; CanCHEC = Canadian Census Health and Environmental Cohort; CCHS = Canadian Community Health Survey; CCR-HCC = California Cancer Registry patients with hepatocellular carcinoma; CNBSS = Canada National Breast Screening Study; CPSII = Cancer Prevention Study II; CTS = California Teachers Study; EFFECT = Enhanced Feedback for Effective Cardiac Treatment; HPFS = Health Professionals Follow-up Study; NA = Not Applicable; NHIS = National Health Interview Survey; NHS = Nurses' Health Study; NIH-AARP = National Institutes of Health-AARP; NJ = New Jersey; PM = Particulate Matter; PREMIER = Prospective Registry Evaluating Myocardial Infarction: Events and Recovery; TrIPS = Trucking Industry Particle Study; TRIUMPH = Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status; VCMS = Veterans Cohort Mortality Study.

Cohort (CanCHEC) 1991, Cancer Prevention Study (CPS) II, and California Teachers Study (CTS) cohorts, where two studies were selected for each cohort to represent distinct $PM_{2.5}$ measurement approaches (*i.e.* 'monitored' *vs.* 'modeled'), of which each approach has its own strengths and limitations and is not necessarily considered better than the other approach. In addition, the studies by Hart et al. (2015) and DuPre et al. (2019) both evaluated the Nurses' Health Study (NHS) cohort, but the latter study was restricted to female nurses with breast cancer and additionally included the NHS II cohort, so the overlap in the study population was not substantial and we included both studies in this review.

The characteristics of the 23 included studies are summarized in Table 3. All of these studies used a cohort study design, with follow-up periods generally from the 1980s to the 2000s and follow-up time ranging from 5 to 35 years. Seventeen studies were conducted in US populations whereas the other six studies were conducted in Canadian populations. Within either country, most of the studies were conducted across multiple cities. In general, the studies analyzed individuals from three types of source population: (1) the general population, including both males and females; (2) individuals in specific professions (*e.g.* veterans, trucking industry workers, teachers, farmers, and health professionals), mostly limited to only males or only females; and (3) patients with underlying health conditions (*e.g.* hepatocellular cancer, myocardial infarction [MI]), including both males and females. Participants were mostly middle-aged or older, and only a few studies also included younger individuals. The sample size of the studies varied substantially, from as low as a few thousand (e.g. Malik et al. 2019) to as high as tens of millions (e.g. Di et al. 2017).

All but one study (Villenueve et al. 2015) examined only one type of mortality outcome, with half of the studies examining all-cause mortality and the other half examining non-accidental mortality. While seven studies measured $PM_{2.5}$ concentrations directly from monitoring sites, 18 studies estimated $PM_{2.5}$ concentrations using modeling approaches. Only two studies (Hart et al. 2015; Di et al. 2017) examined both direct site measured and modeled $PM_{2.5}$ in relation to mortality. As shown in Table 3, the $PM_{2.5}$ modeling approaches varied among the studies that examined modeled $PM_{2.5}$, with GEOS-Chem chemical transport model (CTM) and the Geographic Information System (GIS)-based smoothing model being the most commonly used techniques. The reported mean $PM_{2.5}$ concentration also varied among the studies, ranging from 6.32 to 10.7 $\mu g/m^3$ in Canadian studies and from 9.52 to 18.2 $\mu g/m^3$ in US studies.

Study quality evaluation

The results of our study quality evaluation are presented in Table 4. If a study met a specific criterion, the column for that study shows a '+' in the row for that criterion. If a study did not meet a specific criterion, the column for that study is blank in the row for that criterion. If a criterion is not applicable to a particular study (as discussed above in the Methods section), the column for that study shows 'NA' in the row for that criterion.

With regard to the study reporting aspects, all 23 studies clearly described the study objectives and the size of PM fraction, reported participant characteristics, and presented descriptive statistics. All of the studies clearly reported the inclusion/exclusion criteria that were also consistent with study objectives except the study by Deng et al. (2017), which did not report inclusion/exclusion criteria. No study presented univariate analyses with $PM_{2.5}$, covariates, and copollutants, although it was not uncommon for the studies to instead present analyses that adjusted for a minimum set of confounders.

While the specifics related to study conduct varied among the studies, they all share many common strengths and limitations. With regard to the study design category, all of the studies used a cohort study design with long study duration (i.e. multiple years). Most of the studies were conducted in multiple cities across multiple states/provinces, except for five studies where participants were from a single state/province (Ostro et al. 2010, 2015; Hartiala et al. 2016; Chen et al. 2016; Deng et al. 2017). None of the studies presented a power calculation to indicate sufficient

Table 3. Characteristics of epidemiology studies of long-term $PM_{2.5}$ exposure and all-cause or non-accidental mortality.

Study	Cohort Name	Location	Source Population	Age (years)	Sex	Sample Size	Study Period	Mortality Type	Exposure Measurement	Air Model	$PM_{2.5}$ Concentration ($\mu g/m^3$)	
											Mean (SD)	Median (IQR)
Weichenthal et al. (2014)	AHS	US, Iowa and North Carolina	Commercial pesticide applicators, farmers, and their families	46–47 on average	Both	83,378	1993–2009	Non-accidental	Modeled	GEOS-Chem CTM	9.52 (1.66)	NR
Crouse et al. (2015)	CanCHEC 1991	Canada, multiple cities	General population	25–89	Both	2,521,525	1991–2006	Non-accidental	Modeled	LUR	8.9 (3.4)	8.6 (5.8)
Weichenthal et al. (2016)	CanCHEC 1991	Canada, multiple cities	General population	25–89	Both	193,300	1991–2009	Non-accidental	Monitored	NA	9.81(1.59)	10.09 (2.17)
Weichenthal et al. (2017)	CanCHEC 2001	Canada, multiple cities	General population	25–89	Both	2,448,500	2001–2011	Non-accidental	Modeled	GEOS-Chem CTM + geographically weighted regression	7.37 (NA)	7.12 (3.7)
Pinault et al. (2016)	CCHS	Canada, multiple cities	General population	25–90	Both	299,500	2000–2011	Non-accidental	Modeled	GEOS-Chem CTM + geographical weighted regression	6.32 (2.54)	5.9 (4.1)
Deng et al. (2017)	CCR-HCC	US, California	Patients with hepatocellular cancer	63.7 ± 12.4	Both	20,221	2000–2011	All-cause	Modeled	Inverse distance-squared weighing model	13.3 (5.0)	NR
Hartiala et al. (2016)	Cleveland Clinic GeneBank study	US, Ohio	Patients undergoing elective diagnostic coronary angiography	64 ± 11	Both	6,575	2001–2010	All-cause	Modeled	Inverse distance-squared weighing model	14.6 (1.1)	NR
Villeneuve et al. (2015)	CNBSS	Canada, multiple cities	General population	40–59	Females	89,248	1980–2005	All-cause; Non-accidental	Modeled	GEOS-Chem CTM	NR	9.1 (6)
Jerrett et al. (2009)	CPS II	US, multiple cities	Residents living in MSAs	≥30	Both	448,850	1982–2000	All-cause	Monitored	NA	13.76 (2.71)[a]	NR
Turner et al. (2016)	CPS II	US, multiple cities	General population	≥30	Both	669,046	1982–2004	All-cause	Modeled	CMAQ + LUR + BME HBM	LURBME: 12.6 (2.9) HBM 12.1 (2.6)	LURBME 12.5 (3.9) HBM 12.1 (3.6)
Ostro et al. (2010)	CTS	US, California	Female teachers in California	30+	Females	8 km buffer: 7,888 30 km buffer: 44,847	2002–2007	Non-accidental	Monitored	NA	8-km buffer: 17.0 (NR) 30-km buffer: 17.5 (NR)	NR (6.1)
Ostro et al. (2015)	CTS	US, California	Female teachers in California	30+	Females	101,884	2001–2007	Non-accidental	Modeled	UCD/CIT SOCTM	18.2 (9.6)[a]	18.2 (9.6)
Chen et al. (2016)	EFFECT	Canada, Ontario	Newly admitted AMI patients	35+	Both	8,873	1999–2011	Non-accidental	Modeled	GEOS-Chem CTM	10.7 (2.2–16.5) (4)	NR
Lepeule et al. (2012)	Harvard Six Cities Study	US, multiple cities	White adults	25–74	Both	8,096	1974–2009	All-cause	Monitored	NA	15.9 (NA)	NR
Puett et al. (2011)	HPFS	US, multiple cities	Male health professionals	43–78[a]	Males	17,545	1989–2003	Non-accidental	Modeled	GIS-based spatial smoothing model	17.8 (3.4)	NR (4.3)
Di et al. (2017)	Medicare	US, multiple cities	Medicare beneficiaries	65+	Both	60,925,443	2000–2012	All-cause	Modeled and monitored	GEOS-Chem CTM + geographical linear regression	Modelled: 11 (NR)	NR

(*Continued*)

Table 3. (Continued).

Study	Cohort Name	Location	Source Population	Age (years)	Sex	Sample Size	Study Period	Mortality Type	Exposure Measurement	Air Model	$PM_{2.5}$ Concentration ($\mu g/m^3$)	
											Mean (SD)	Median (IQR)
Lefler et al. (2019)	NHIS	US, multiple cities	General population	18–84	Both	635,539	1987–2015	All-cause	Modeled	GEOS-Chem CTM + LUR + v1 Empirical model by the center for Air, Climate, and Energy Solutions	10.67 (2.37)	NR (3.12)
Hart et al. (2015)	NHS	US, multiple cities	Female nurses	54–79[a]	Females	108,767	2000–2006	Non-accidental	Modeled and monitored	GIS-based spatial smoothing model	Modeled: 12.0 (2.8) Monitored: 12.7 (3.1)	NR
DuPre et al. (2019)	NHS & NHSII	US, multiple cities	Female nurses with breast cancer	64.4 ± 8.3[a] NHS; 47.1 ± 5.7[a] NHS II	Females	8,936	1988–2014	All-cause	Modeled	GIS-based spatial smoothing model	Pooled 13.19 (3.40)[a]	NR
Thurston et al. (2016)	NIH-AARP	US, multiple cities	AARP members	50–71	Both	517,041	2000–2009	Non-accidental	Modeled	GEOS-Chem CTM + LUR + BME	12.2 (3.4)	NR
Hart et al. (2011)	TrIPS	US, multiple cities	Male trucking industry workers	15–84	Males	53,814	1985–2000	All-cause	Monitored	NA	14.1 (4.0)	NR (4)
Malik et al. (2019)	TRIUMPH & PREMIER	US, multiple cities	Patients with MI	60.6 ± 12.2	Both	5,650	2003–2013	All-cause	Modeled	Bayesian space-time downscaling fusion model	11.96 (2.11)	12.05 (2.6)
Lipfert and Wyzga (2019)	VCMS	US, multiple cities	Male ostensibly hypertensive veterans	~51 on average	Males	70,000	1976–2001	All-cause	Modeled	GIS-based smoothing models	White Vets: 13.9 Black Vets: 15.7	NR

Notes:

(a) Estimates based on data shown in the paper.

AHS = Agricultural Health Study; AMI = Acute Myocardial Infarction; BME = Bayesian Maximum Entropy; CA = California; CanCHEC = Canadian Census Health and Environmental Cohort; CCHS = Canadian Community Health Survey; CCR-HCC = California Cancer Registry patients with hepatocellular carcinoma; CMAQ = Community Multiscale Air Quality; CNBSS = Canada National Breast Screening Study; CPSII = Cancer Prevention Study II; CTM = Chemical Transport Model; CTS = California Teachers Study; EFFECT = Enhanced Feedback for Effective Cardiac Treatment; GIS = Geographic Information System; HBM = Hierarchical Bayesian Model; HPFS = Health Professionals Follow-up Study; LUR = Land Use Regression; MI = Myocardial Infarction; MSA = Metropolitan Statistical Area; NA = Not Applicable; NHIS = National Health Interview Survey; NHS = Nurses' Health Study; NIH-AARP = National Institutes of Health-AARP; NJ = New Jersey; PM = Particulate Matter; PREMIER = Prospective Registry Evaluating Myocardial Infarction: Events and Recovery; SOCTM = Source Oriented Chemical Transport Model; TrIPS = Trucking Industry Particle Study; TRIUMPH = Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status; UCD/CIT = University of California Davis/California Institute of Technology; VCMS = Veterans Cohort Mortality Study.

Table 4. Study quality evaluation of epidemiology studies of $PM_{2.5}$ and mortality.

Category	Criteria	Weichenthal et al. 2014	Crouse et al. 2015	Weichenthal et al. 2016	Weichenthal et al. 2017	Pinault et al. 2016	Deng et al. 2017	Hartiala et al. 2016	Villeneuve et al. 2015	Jerrett et al. 2009	Turner et al. 2016	Ostro et al. 2010	Ostro et al. 2015	Chen et al. 2016	Lepeule et al. 2012	Puett et al. 2011	Di et al. 2017	Lefler et al. 2019	Hart et al. 2015	DuPre et al. 2019	Thurston et al. 2016	Hart et al. 2011	Malik et al. 2019	Lipfert and Wyzga 2019
Study Design	Study Objectives	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Study Design	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Study Location	+	+	+	+	+			+	+	+				+	+	+	+	+	+	+	+	+	+
	Sample Size		+		+												+							
	Study Duration	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Study Population	Participant Characteristics	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Inclusion/Exclusion Criteria	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Representativeness of Source	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Participation Rate	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Underlying Health Conditions	NA	NA	NA	NA	NA	+	+	NA	NA	NA	NA	NA	+	NA	NA	NA	NA	NA	+	NA	NA	+	+
	Follow-up	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pollutant	Pollutant Description	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Pollutant Source										+	+	+					+						
Exposure Assessment	Multiple Methods										+						+		+					
	Validity	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+		+	+	+	+	+	+
	Spatial Variability	+	+	+	+	+			+	+	+		+	+		NR	+	+	+	+	+	NR		
	Temporal Variability		+		+	+							+	+	+	+	+	+	+	+	+			
	Temporality	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Assignment to Participants' Locations				+		+	+				+	+	+		+	+	+	+				+	
	Residential Mobility	+	+		+							+	+	+		+	+		+	+				
	Personal Activities	+																						
Outcome Assessment	Blinding	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Time Points	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Confounding	**Key Confounders**																							
	Proper adjustment	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+
	Temporal conceptualization	+	+	+	+		+	+	+	+	+	+				+	+	+	+	+		+	+	+
	Copollutant adjustment		+	+	+					+	+					+	+	+			+		+	
	Copollutant correlations	NA	+	+	+	NA	NA	NA	NA	+	+	NA	NA	NA	NA		+	+	NA	NA		NA	+	NA
	Copollutant Measurement	NA				NA	NA	NA	NA			NA	NA	NA	NA				NA	NA		NA		NA
Statistical Methods	Descriptive Statistics	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Univariate Analyses																							
	Model Specification	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	Model Assumptions													+	+					+			+	
	Nonlinearity		+		+	+	+		+					+	+		+		+		+		+	
	Multiple Comparison				NA			NA						NA					NA				NA	
	Sensitivity Analyses	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+

Criteria correspond to the aspects and criteria listed in Table 1. Bold font indicates the criteria that assess important dimensions of study conduct.
Columns for studies that meet a specific criterion are indicated with a '+' in the row for that criterion, while columns for studies that do not meet a specific criterion are blank in the row for that criterion. NA = Not Applicable.

Table 5. Results from linear association analyses.

Study	Mortality Type	Grouping	Exposure Metric	Exposure Contrast (μg/m^3)	Statistical Model	Measure of Association	Risk Estimate (95% Confidence Interval)	Copollutant Adjustment	Other Confounding Adjustment
Weichenthal et al. (2014)	Non-accidental	Overall	6-year average	10	Cox	HR	0.95 (0.76, 1.20)	None	Age; sex; state of enrollment; birth year category; pack years of smoking; BMI; marital status; education level; alcoholic drinks per month; vegetable intake.
		Males				HR	1.05 (0.80, 1.39)		
		Females				HR	0.78 (0.51, 1.19)		
Crouse et al. (2015)	Non-accidental	Overall	7-year moving average	5	Cox	HR	1.011 (1.003–1.020)	O_3, NO_2	Sex; 5-year age group; aboriginal ancestry; visible minority status; immigrant status; marital status; highest level of education; employment status; occupational classification; quintiles of household income; census division and census tract-census division; percent adults without a high school diploma; percent individuals in the lowest income quintile; percent unemployed adults.
							1.035 (1.029, 1.041)	None	
Weichenthal et al. (2016)	Non-accidental	Overall	Annual average	2.17	Cox	HR	1.025 (1.012,1.039)	OP^{GSH}	Age at entry; sex; aboriginal ancestry; visible minority status; immigrant status; marital status; highest level of education; employment status; occupational classification; household income.
							1.026 (1.012–1.039)	None	
Weichenthal et al. (2017)	Non-accidental	Overall	3-year moving average	3.858	Cox	HR	1.041 (1.029–1.052)	Ox (combined oxidant capacity of O_3 and NO_2)	Sex; 5-year age group; airshed; population center size; visible minority status; aboriginal status; marital status; educational attainment; income quintile; labor force status; percent unemployed (aged 25 or older); percent not graduated from high school (aged 25 or older); percent low income status within census divisions; population density within dissemination areas.
							1.073 (1.062–1.083)	None	
Pinault et al. (2016)	Non-accidental	Overall	3-year moving average	10	Cox	HR	1.26 (1.19, 1.34)	None	Age; sex; immigrant Status; visible minority status; aboriginal status; marital status; educational attainment; income adequacy quintile; employment status; smoking; BMI; fruit and vegetable consumption; alcohol consumption; proportion of recent immigrants; educational attainment; low income.
		Males				HR	1.344 (1.239, 1.457)		Age; immigrant status; visible minority status; aboriginal status; marital status; educational attainment; income adequacy quintile; employment status; smoking; BMI; fruit and vegetable consumption; alcohol consumption.
		Females				HR	1.181 (1.088, 1.282)		Age; immigrant status; visible minority status; aboriginal status; marital status; educational attainment; income adequacy quintile; employment status; smoking; BMI; fruit and vegetable consumption; alcohol consumption.

(*Continued*)

Table 5. (Continued).

Study	Mortality Type	Grouping	Exposure Metric	Exposure Contrast (μg/m^3)	Statistical Model	Measure of Association	Risk Estimate (95% Confidence Interval)	Copollutant Adjustment	Other Confounding Adjustment
Deng et al. (2017)	All-cause	Overall	Survival-period average	5	Cox	HR	1.18 (1.16–1.20)	None	Age; sex; race/ethnicity; marital status; SES; rural-urban commuting area; distance to primary interstate highway; distance to primary US and state highways; month of diagnosis; year of diagnosis; initial treatments.
Hartiala et al. (2016)	All-cause	Overall	36-month average	2.2	Cox	HR	1.16 (0.96–1.41)	None	Age; sex; current smoking; education level; Framingham Adult Treatment Panel III risk score.
Villeneuve et al. (2015)	All-cause	Overall	Long-term (9-year) average	10	Cox	HR	1.10 (1.03–1.17)	None	Age at entry; marital status; occupation; attained education; cigarette smoking; BMI; mean income; proportion with high school education; percentage of low income households; unemployment rate.
	Non-accidental	Overall				HR	1.12 (1.04–1.19)		
Jerrett et al. (2009)	All-cause	Overall	2-year average (as proxy of study period average)	10	Cox	HR	1.080 (1.048–1.113)	O_3	Age; sex; race; education; marital status; BMI; smoking status; alcohol consumption; level of dietary-fat consumption; level of dietary-fiber consumption; level of occupational exposure to PM; self-reported exposure to dust or fumes; % nonwhite race; percent homes with air conditioning; percent high school education or greater; percent unemployment rate; Gini coefficient of income inequality; proportion of population with income <125% of poverty line; annual household income.
							1.048 (1.024–1.071)	None	
Turner et al. (2016)	All-cause	Overall	6-year average (as proxy of study period average)	10	Cox	HR	1.06 (1.04–1.08) [HBM PM2.5]	O_3	Age; sex; race; education; marital status; BMI; BMI squared; cigarette smoking status; cigarettes per day or cigarettes per day squared; years smoked and years smoked squared; started smoking at younger than 18 years of age; passive smoking; vegetable, fruit, fiber, and fat intake; alcohol consumption; occupational exposures; "occupational dirtiness index"; median household income; percent African American residents; percent Hispanic residents; percent adults with post-secondary education; percent unemployment; percent poverty.
							1.06 (1.04–1.08) [HBM $PM_{2.5}$]	None	
							1.07 (1.06–1.09) [$PM_{2.5}$ LURBME]	None	
Ostro et al. (2010)	Non-accidental	Females	Survival period average	6.1	Cox	HR	Within 8 km buffer: 1.49 (1.28, 1.74)	None	Age; race/ethnicity; smoking status; total pack-years; body mass index; marital status; alcohol consumption; secondhand smoke exposure at home; dietary fat, fiber, and caloric intake; physical activity; menopausal status; hormone therapy use; family history of myocardial infarction, stroke; hypertension medication/aspirin use; income; income inequality; population size; education; racial composition; percent unemployed.
						HR	Within 30 km buffer: 1.45 (1.36–1.55)		

(*Continued*)

Table 5. (Continued).

Study	Mortality Type	Grouping	Exposure Metric	Exposure Contrast (μg/m^3)	Statistical Model	Measure of Association	Risk Estimate (95% Confidence Interval)	Copollutant Adjustment	Other Confounding Adjustment
Ostro et al. (2015)	Non-accidental	Females	Long-term average (up to each death)	9.6	Cox	HR	1.01 (0.98, 1.05)	None	Age; race/ethnicity; smoking status; smoking pack-years; secondhand smoke exposure; BMI; marital status; alcohol consumption; physical activity; menopausal status and hormone replacement therapy use combined; family history of heart disease; hypertension medication/aspirin use; dietary fat, fiber, and caloric intake.
Chen et al. (2016)	Non-accidental	Overall	Cumulative time-weighted exposure	10	Cox	HR	1.22 (1.03, 1.45)	None	Age; sex; region; marital status; employment status; smoking status; family history of coronary artery disease; DM; hyperlipidemia; hypertension; stroke; previous AMI; previous percutaneous coronary intervention; AMI type; acute pulmonary edema; preexisting angina, cancer; dementia; dialysis; chronic obstructive pulmonary disease; cardiovascular medications at hospital discharge (statins, aspirin, ACE inhibitors, and beta-blockers); ST elevation myocardial infarction (STEMI/non-STEMI); length of hospital stay in days; characteristics of attending physicians and hospitals; Global Registry of Acute Cardiac Events risk score (severity of AMI event); census division level unemployment rate, education and annual household income (plus the subtraction of these variables at the census-tract level from their census division mean).
Lepeule et al. (2012)	All-cause	Overall	Annual moving average	10	Cox	HR	1.14 (1.07–1.22)	None	Sex; age; time in the study; smoking history; education; BMI.
Puett et al. (2011)	Non-accidental	Overall	12-month moving average	4	Cox	HR	0.94 (0.87–1.00) 0.94 (0.87–1.02)	None $PM_{10-2.5}$	Age; state of residence; year; season; hypertension; DM; hypercholesterolemia; physical activity; alcohol consumption; BMI; smoking status; smoking pack-years; diet; family history of MI.
Di et al. (2017)	All-cause	Overall	Annual average	10	Cox	HR	Modeled 1.073 (1.071–1.075) Monitored 1.061 (1.059–1.063) 1.084 (1.081–1.086) Modeled	O_3 None	Age; race; sex; Medicaid eligibility; percent Hispanic; percent black; median household income; median value of owner occupied housing; percent above 65 years living below poverty; percent above 65 with less than a high school education; percent owner occupied housing units; population density; county level BMI; percent ever smokers; percent with blood lipid test; percent with HbA1c test; ≥1 ambulatory visit to a primary care clinician; temperature; humidity.
		Males					1.087 (1.083, 1.090)	O_3	
		Females					1.060 (1.057, 1.063)	O_3	
		Whites					1.063 (1.060, 1.065)	O_3	

(*Continued*)

Table 5. (Continued).

Study	Mortality Type	Grouping	Exposure Metric	Exposure Contrast (μg/m^3)	Statistical Model	Measure of Association	Risk Estimate (95% Confidence Interval)	Copollutant Adjustment	Other Confounding Adjustment
Lefler et al. (2019)	All-cause	Overall	Annual average	3.12	Cox	HR	1.045 (1.030–1.061) Complex model 1.045 (1.032–1.058) Basic model	$PM_{2.5-10}$, SO_2, NO_2, O_3, and CO $PM_{2.5-10}$, SO_2, NO_2, O_3, and CO	Age; sex; race-ethnicity; marital status; inflation-adjusted household income; education; smoking status; US Census region; urban vs. rural designation; survey year; BMI.
Hart et al. (2015)	Non-accidental	Overall	12-month moving average	10	Cox	HR	Modeled 1.13 (1.05–1.22) Monitored 1.12 (1.05–1.21)	None	Age; race; region; year; season; smoking status; smoking pack years; family history of MI; BMI; hypercholesterolemia; median family income in census tract of residence; median house value in census tract of residence; physical activity; AHEI (includes alcohol consumption); individual-level SES (nurses' education level, occupation of both parents, marital status, and husband's education).
DuPre et al. (2019)	All-cause	Overall	2-year moving average	10	Cox	HR	Pooled 1.12 (0.96–1.30)	None	Age; race/ethnicity; census tract median income; breast cancer diagnosis date, stage, ER status, treatments, hormones, grade; region of residence at diagnosis; pre- to post-diagnosis weight change categories; post-diagnostic physical activity, aspirin use, and smoking.
Thurston et al. (2016)	Non-accidental	Overall	Annual average	10	Cox	HR	1.03 (1.00, 1.05) [Time-independent] 1.03 (0.99, 1.05) [Time-dependent]	None None	Age; sex; region; race/ethnicity; education; marital status; BMI; alcohol consumption; smoking history; median census tract income; percent of census tract population with less than a high school education.
Hart et al. (2011)	All-cause	Males	Annual average	4	Cox	HR	1.04 (1.01, 1.07)	None	Age; race; census region of residence; calendar year and decade of hire; healthy worker survivor effect; years of work for each group (long-haul drivers, pick-up and delivery drivers, dockworkers, combination workers, mechanics, hostlers, clerks, and other workers).
Malik et al. (2019)	All-cause	Overall	12-month average	2.11	Cox	HR	1.13 (1.07–1.20)	O_3	Age; sex; race; SES (education, insurance status, history of avoiding care because of cost, and end-of-the-month financial resources); date of enrollment; smoking status; comorbidities including hypertension, DM, CKD, and heart failure before MI; left ventricular function; Global Registry of Acute Coronary Events Mortality risk score.
Lipfert and Wyzga (2019)	All-cause	Black White	Annual average	NR (annual average mean-min)	Cox	HR HR	0.817 (0.750, 0.891) 1.051 (1.005, 1.100)	None	Age; race; smoking status; height; diastolic and systolic blood pressure; BMI; percent of ZIP code residents below 75% poverty level; zip-code average education; heating degree-days; percent blacks in ZIP code.

ACE = Angiotensin Converting Enzyme; AHEI = Alternate Healthy Eating Index; AMI = Acute Myocardial Infarction; BMI = Body Mass Index; CKD = Chronic Kidney Disease; CO = Carbon Monoxide; DM = Diabetes Mellitus; ER = Estrogen Receptor; HbA1c = Hemoglobin A1c; HR = Hazard Ratio; MI = Myocardial Infarction; NO_2 = Nitrogen Dioxide; NR = Not Reported; O_3 = Ozone; OP^{GSH} = Glutathione-related Oxidative Potential; PM = Particulate Matter; SES = Socioeconomic Status; SO_2 = Sulfur Dioxide; US = United States.

Table 6. Results of non-linear association analyses.

Study	Mortality Type	Splines (df)	Shape of E-R Curve	Threshold Identification	Threshold Estimate	Confounding Adjustment
Weichenthal et al. 2014	Non-accidental	Natural (2 df)	NR	NA	NA	Same as for linear analysis
Crouse et al. 2015	Non-accidental	Restricted cubic (2 df)	Supralinear	NT	NT	Same as for linear analysis
Weichenthal et al. 2017	Non-accidental	NA (Joint non-linear Ox – $PM_{2.5}$ Relative Risk Model)	Supralinear (at Ox = 37.60 ppb or Ox = 20.26 ppb)	NT	NT	Same as for linear analysis
Pinault et al. 2016	Non-accidental	Splines (NR)	NR; appears supralinear	Hockey-stick threshold model (starting from 1–10 μg/m^3, at 1-μg/m^3 increment; minimizing −2LL)	0 μg/m^3, upper 95% CI = 4.5 μg/m^3	Same as for linear analysis of males and females combined
Deng et al. 2017	All-cause	Natural cubic (2 df)	J-shaped	NT	NT	Same as for linear analysis
Villenueve et al., 2015	Non-accidental	Natural cubic (3 df)	Nonlinear V-shaped	Hockey stick threshold model (starting from 2–14 μg/m^3, at 1 μg/m^3 increment; minimizing −2LL)	11 μg/m^3 (p = 0.004)	Same as for linear analysis
Chen et al. 2016	Non-accidental	Natural cubic (≤4 df)	Linear	NA	NA	Same as for linear analysis
Lepeule et al. 2012	All-cause	Penalized cubic (NA)	Linear (Wald test of non-linearity p-value NR)	NR[a]	No threshold (down to 8 μg/m^3)	Same as for linear analysis
Di et al. 2017	All-cause	Penalized (NR)	Roughly linear	NR[a]	No threshold (down to approximately 5 μg/m^3)	Same as for linear analysis
Hart et al. 2015	Non-accidental	Stepwise restricted cubic (NR)	Approximately linear (both modeled and monitored)	NT	NT	Same as for linear analysis[b]
Thurston et al. 2016	Non-accidental	Natural (4 df)	Monotonically increasing	NA	NA	Same as for linear analysis
Malik et al. 2019	All-cause	Smoothing or restricted cubic (NA)	Linear (test of non-linearity p = 0.59)	NT	NT	Same as for linear analysis

(a) No threshold model was specified, but the authors concluded that there was no threshold based on the approximately linear curve.
(b) The paper reports that adjustment is different by hypertension and diabetes, but this appears to be a typographical error.
df = Degrees of Freedom; E-R = Exposure-response; LL = Log Likelihood; NA = Not Applicable; NR = Not Reported; NT = Not Tested.

sample size, however, so the three studies considered to have met the sample size criterion had sample sizes that were greater than 1 million (Crouse et al. 2015; Weichenthal et al. 2017; Di et al. 2017).

With regard to the study population category, all six studies that were conducted among patients with underlying health conditions (Hartiala et al. 2016; Chen et al. 2016; Deng et al. 2017; DuPre et al. 2019; Malik et al. 2019; Lipfert and Wygza 2019) ascertained these conditions by independent clinical assessment or self-report of physician diagnosis. Because all 23 studies were secondary analyses of existing cohorts for which members were initially recruited for research questions unrelated to $PM_{2.5}$ or mortality, and all the studies used linkage to conveniently identify mortality outcomes, we considered the criteria related to representativeness of source, participation rate, and follow-up as not applicable (as discussed above).

With regard to the exposure assessment category, most of the studies used well-established, sensitive methods and sufficiently captured the spatial variability of $PM_{2.5}$, and all studies estimated participants' $PM_{2.5}$ exposures before the outcome. While half of the studies accounted for temporal variability of $PM_{2.5}$, fewer accounted for residential mobility and only one study (Weichenthal et al. 2014) accounted for personal activities by performing a stratified analysis by estimated time spent outdoors. The majority of the studies also did not compare more than one exposure assessment method. Importantly, half of the studies did not assign measured or estimated ambient $PM_{2.5}$ data to participants' locations from the same time period. Specifically, eight studies assigned $PM_{2.5}$ data from as long as 10+ years later to participants' locations (Jerrett et al. 2009; Hart et al. 2011; Lepeule et al. 2012; Villeneuve et al. 2015; Crouse et al. 2015; Turner et al. 2016; Thurston et al. 2016; Lipfert and Wyzga 2019); three studies assigned $PM_{2.5}$ data from as far as 5+ (but <10) years later to participants' locations (Weichenthal et al. 2014, 2016; Pinault et al. 2016); and one study assigned to $PM_{2.5}$ data from as far as 5+ (but <10) years earlier to participants' locations (DuPre et al. 2019).

With regard to the confounding category, none of the studies adjusted for all of the key potential confounders. Specifically, very few (n = 1–2) studies adjusted for relative humidity or other chemical exposures; and only a few studies adjusted for temperature (n = 4), medication use (n = 5), physical activity (n = 6), and diet (n = 8). A small number of studies also did not adjust for race, body mass index (BMI), or smoking status (n = 3–5). Nonetheless, the confounders that were included in most of studies were adjusted for properly. Copollutants were not adjusted for in more than half of the studies. In the studies that accounted for copollutant exposures, most of these examined the correlations between $PM_{2.5}$ and the copollutants; however, the measurements of copollutants in these studies were subject to errors, as they did not properly account for temporal variation, spatial variation, residential mobility, or personal activities.

With regard to the statistical methods category, all studies employed appropriate statistical models (i.e. Cox proportional hazards model) for multivariate analyses, but only four studies (Lepeule et al. 2012; Chen et al. 2016; DuPre et al. 2019; Malik et al. 2019) indicated that key model assumptions (i.e. proportional hazards assumptions) were tested and satisfied. All but five studies are subject to the multiple comparison issue (with the number of comparisons as high as approximately 60), but none of these studies performed any correction to address this issue. While the primary objectives of the studies are variable, all but one study (Malik et al. 2019) assessed the robustness of the $PM_{2.5}$-mortality risk estimates and half of the studies assessed potential non-linearity of the $PM_{2.5}$-mortality relationship.

With regard to outcome assessment, in all studies the assessments of outcome were at time points consistent with study objectives and were blinded to exposure levels. With regard to $PM_{2.5}$ specification, only four studies (Ostro et al. 2010, 2015; Turner et al. 2016; Lefler et al. 2019) additionally evaluated $PM_{2.5}$ source-related indicators.

Evaluation of study results

The linear and non-linear study results are summarized in Tables 5and 6, respectively. Regarding linear results, we included the fully adjusted result of the $PM_{2.5}$-mortality association reported for each study in Table 5. If the fully adjusted result was adjusted for copollutants, we further included the result without copollutant adjustment, if available, for comparison purposes. When statistically significant effect modification on the $PM_{2.5}$-mortality association was reported, we also included stratum-specific results, if available. If a study reported results for multiple $PM_{2.5}$ indicators (e.g. modeled and monitored, generated from different prediction models, within different buffers), mortality indicators (i.e. all-cause and non-accidental), or statistical analyses (e.g. weighted *vs*. non-weighted, time-dependent *vs*. time-independent), we included all such results for comparison purposes. We included all non-linear results reported in the studies in Table 6. Below, we present and discuss results by type of study population (i.e. general population, occupation-specific cohorts, and patients with underlying health conditions), as the results for one of type of study population cannot necessarily be applied to another type of study population.

General population

Eleven of the reviewed studies were conducted in the general population (Jerrett et al. 2009; Lepeule et al. 2012; Villeneuve et al. 2015; Crouse et al. 2015; Weichenthal et al. 2016, 2017; Pinault et al. 2016; Thurston et al. 2016; Turner et al. 2016; Di et al. 2017; Lefler et al. 2019). All of these studies reported a risk estimate for the $PM_{2.5}$-mortality association assuming linearity. Seven of the eleven studies (Lepeule et al. 2012; Villeneuve et al. 2015; Crouse et al. 2015; Pinault et al. 2016; Thurston et al. 2016; Weichenthal et al. 2017; Di et al. 2017) also evaluated potential non-linearity of the association.

Linear Results. All studies in the general population without copollutant adjustment reported a statistically significant, positive association between $PM_{2.5}$ exposure and mortality (either all-cause or non-accidental), with the exception of the study by Thurston et al. (2016), which reported a statistically non-significant, positive association between $PM_{2.5}$ and non-accidental mortality (hazard ratio [HR] = 1.03, 95% CI: 1.00–1.05 in a time-independent analysis; HR = 1.03, 95% CI: 0.99–1.05 in a time-dependent analysis). The magnitude of the HR estimates in these studies ranged from 1.026 (95% CI: 1.012–1.039) in the study by Weichenthal et al. (2016) to 1.26 (95% CI: 1.19–1.34) in the study by Pinault et al. (2016), although the corresponding exposure metric, exposure contrast, and adjustment of other confounders (i.e. other than copollutants) varied. The HR estimates in 8 of the 11 studies fell under 1.10, indicating weak associations. The width of 95% CIs in the largest study (Di et al. 2017; n = 60,925,443; HR = 1.084, 95% CI: 1.081–1.086) is substantially narrower than that in the smallest study (Lepeule et al. 2012; n = 8,096; HR = 1.14, 95% CI: 1.07–1.22). Although a larger sample size increases the statistical power of a study to detect an effect, when the sample size is too large (such as in the millions in the studies by Di et al. 2017; Crouse et al. 2015;; Weichenthal et al. 2017), statistically significant findings could be artifacts due to inflated statistical power and extremely narrow confidence intervals rather than reflecting a true underlying association, so the results from such studies should be interpreted with caution. Results did not appear to differ substantially between studies of all-cause *vs*. non-accidental mortality, modeled *vs*. monitored $PM_{2.5}$, or US *vs*. Canadian populations.

Statistically significant effect modification by sex was identified by Pinault et al. (2016), where males (HR = 1.344, 95% CI: 1.239–1.457, per 10 $\mu g/m^3$ increment of $PM_{2.5}$) had a higher risk of mortality (non-accidental) than females (HR = 1.181, 95% CI: 1.088–1.282, per 10 $\mu g/m^3$ increment of $PM_{2.5}$). The latter risk estimate is slightly higher than what was reported in the female-only study by Villenueve et al. (2015) (HR = 1.10, 95% CI: 1.03–1.17 for all-cause mortality; HR = 1.12, 95% CI: 1.04–1.19 for non-accidental mortality), which may be attributable to differences in study design.

Eight of the studies in the general population estimated the $PM_{2.5}$-mortality association with further copollutant adjustment. Specifically, four studies further adjusted for ozone (O_3) alone

(Jerrett et al. 2009; Thurston et al. 2016; Turner et al. 2016; Di et al. 2017); one study further adjusted for glutathione-related oxidative potential (OP^{GSH}) alone (Weichenthal et al. 2016); two studies further adjusted for both O_3 and nitrogen dioxide (NO_2) (Crouse et al. 2015; Weichenthal et al. 2017); and one study further adjusted for $PM_{2.5-10}$, O_3, NO_2, sulfur dioxide (SO_2), and carbon monoxide (CO) (Lefler et al. 2019). Compared to the risk estimate without copollutant adjustment within the same study, the risk estimate with further adjustment for copollutants was slightly attenuated (*i.e.* closer to the null) in five of the eight studies (Crouse et al. 2015; Thurston et al. 2016; Weichenthal et al. 2016, 2017; Di et al. 2017). This attenuation is expected, as copollutant concentrations tend to be positively associated with $PM_{2.5}$ and mortality (WHO 2006; US EPA 2019). By contrast, the risk estimate with further adjustment for copollutants remained the same in one study (Turner et al. 2016) and was slightly exaggerated (i.e. further away from the null) in two studies (Jerrett et al. 2009; Lefler et al. 2019). This variation in results could be due to variation in copollutant adjustments or errors in copollutant measurements that are of similar sources as $PM_{2.5}$ measurement errors. However, it is worth noting that the copollutant adjustments in these studies are likely ineffective, as in none of the eight studies were copollutants measured at both the same temporal and spatial scales as $PM_{2.5}$ to fully and accurately capture how the different pollutants were correlated with each other.

With the adjustment of O_3, Di et al. (2017) identified statistically significant effect modification by sex. Similar to the study by Pinault et al. (2016), which did not adjust for copollutants, Di et al. (2017) reported that males (HR = 1.087, 95% CI: 1.083–1.090, per 10 $\mu g/m^3$ increment of $PM_{2.5}$) were at higher risk of mortality (all-cause) than females (HR = 1.060, 95% CI: 1.057–1.063, per 10 $\mu g/m^3$ increment of $PM_{2.5}$).

The seemingly consistent linear results in the studies should be interpreted with caution, considering the large variations across studies in terms of participants' characteristics (e.g. location, age, sex, race), exposure assessment (e.g. measurement, metric, contrast), outcome type (all-cause *vs*. non-accidental), and confounder adjustments. In fact, heterogeneity underlying the consistent linear results in recent studies of long-term $PM_{2.5}$ and mortality has been reported by Di et al. (2017). Specifically, these authors compiled the results of 22 studies (including studies published prior to 2009) that reported HR estimates ranging from 1.01 to 1.26, which are very similar to the HR estimates from the studies reviewed here. Di et al. (2017) performed a meta-analysis of these studies using a random-effect model and reported a meta-HR of 1.11 (95% CI: 1.08–1.15). A heterogeneity test indicated a high degree of heterogeneity (I-squared = 95.9%, tau-squared = 0.0035, $p < 0.0001$) among the study results, however. While it is possible that the large variations in study design aspects across studies have only small impacts on the magnitude of risk estimates, one cannot rule out that the impact of this variation could also be large but masked by other factors that are consistently and potentially substantially influencing the studies and their risk estimates, as discussed below in the evaluation of study quality.

Non-linear Results. In the evaluation of potential non-linearity of the $PM_{2.5}$-mortality association, six of the seven studies (Lepeule et al. 2012; Villeneuve et al. 2015; Crouse et al. 2015; Pinault et al. 2016; Thurston et al. 2016; Di et al. 2017) used spline techniques, although with varied types of spline, degrees of freedom, and confounding adjustments. Unlike the linear results summarized above, the observed shapes of the $PM_{2.5}$-mortality curves are inconsistent across the studies. Two studies reported a linear shape for the $PM_{2.5}$-mortality (all-cause) curve with no apparent threshold (Lepeule et al. 2012; Di et al. 2017). Three studies reported a supralinear shape for the $PM_{2.5}$-mortality (non-accidental) curve (Crouse et al. 2015; Pinault et al. 2016; Weichenthal et al. 2017), among which Pinault et al. (2016) further estimated a threshold $PM_{2.5}$ concentration of 0 $\mu g/m^3$ (+95% CI = 4.5 $\mu g/m^3$). Villeneuve et al. (2015) reported the $PM_{2.5}$-mortality (non-accidental) curve to be V-shaped, with an estimated threshold at 11 $\mu g/m^3$ ($p = 0.004$), and Thurston et al. (2016) reported the shape of the $PM_{2.5}$-mortality (non-accidental) curve to be monotonically increasing.

While all the studies in the general population estimated linear associations between $PM_{2.5}$ and mortality, the observed non-linear curves in the studies above indicate that linearity may not be

a valid modeling assumption. The contrast between highly consistent linear results and highly inconsistent non-linear results in these studies also indicates that the linearity assumption, although straightforward, may have masked important heterogeneity and details of the underlying $PM_{2.5}$-mortality relationships, especially considering the variations in $PM_{2.5}$ assessment approach (e.g. prediction model, exposure metric, exposure contrast, and exposure window or lag time), $PM_{2.5}$ concentration distribution, and confounding adjustment across the studies. It is also possible that the different non-linear modeling techniques used in the studies could contribute to the variations in the observed shapes of the $PM_{2.5}$-mortality association across studies.

Study Quality. The studies conducted in the general population share certain strengths and limitations. All 11 studies were conducted in multiple cities, so the study results have higher generalizability across North American populations. Nine of the eleven studies had a sample size of 100,000 or greater, indicating these studies have greater statistical power to detect an underlying $PM_{2.5}$-mortality association, if it exists. Specifically, the studies in the general population included three of the largest studies in this review, with sample sizes in the millions (Crouse et al. 2015; Weichenthal et al. 2017; Di et al. 2017). As discussed above, however, the extremely large sample sizes of these three studies can inflate statistical power such that the weak but statistically significant findings reported in these studies may be artifacts rather than a representation of a true underlying association.

In general, all 11 studies assessed each participant's exposure to $PM_{2.5}$ by assigning to his/her location (primarily residential location) an ambient $PM_{2.5}$ concentration that was either from direct measurements at one or a few nearby stationary monitoring sites or estimates from prediction models. This approach for exposure assessment does not account for individual factors, such as time spent indoors or at non-residential locations and personal activities, that vary among participants and can greatly affect their actual $PM_{2.5}$ exposures. Further, while 10 of the 11 studies meet our quality criterion for spatial variability and 7 of the 11 studies meet the criterion for temporal variability, only three studies meet the criterion for assignment to participants' locations, three studies meet the criterion for residential mobility, and none of the 11 studies meet the criterion for personal activities. These indicate that the results of all studies are subject to substantial exposure measurement error, though the associated overestimation or underestimation of $PM_{2.5}$ exposure and the direction of bias to the study results are difficult to anticipate.

It is important to note that for the eight studies that did not assign measured or estimated ambient $PM_{2.5}$ data to participants' locations in the same time period (Jerrett et al. 2009; Lepeule et al. 2012; Villeneuve et al. 2015; Crouse et al. 2015; Weichenthal et al. 2016; Turner et al. 2016; Pinault et al. 2016; Thurston et al. 2016), the reported distribution of $PM_{2.5}$ concentrations was likely not representative of the distribution of participants' actual $PM_{2.5}$ exposure. Considering that ambient $PM_{2.5}$ concentrations are generally decreasing over time due to the implementation of more stringent regulations, and that all eight studies that do not meet the 'assignment to participants' locations' criterion assigned $PM_{2.5}$ data from as long as 5+ to 10+ years later to participants' locations, these studies likely have underestimated the participants' actual $PM_{2.5}$ exposure concentration and overestimated the mortality rate associated with lower $PM_{2.5}$ exposures. It is only from the three studies that meet this criterion (Weichenthal et al. 2017; Di et al. 2017; Lefler et al. 2019) that an inference can confidently be made regarding the $PM_{2.5}$ concentration under which an association was observed with mortality (mean $PM_{2.5}$ concentration was 7.37 $\mu g/m^3$ in the study by Weichenthal et al. 2017; 10.67 $\mu g/m^3$ in the study by Lefler et al. 2019; and 11 $\mu g/m^3$ in the study by; Di et al. 2017). Still, in making such an inference, the other potential sources of exposure measurement error mentioned above, as well as other sources of bias and confounding, also need to be taken into consideration.

While 8 of the 11 studies adjusted for copollutants, none adjusted for physical activity or medication use, and few studies adjusted for diet, humidity, temperature, or other chemical exposures as potential confounders or primary covariates. Thus, the results of these studies, even those that are the largest and less subject to exposure measurement error (i.e. by meeting our

criteria for all aspects of the exposure assessment category except for personal activities and multiple methods) (Weichenthal et al. 2017; Di et al. 2017), are still subject to residual confounding by these and other unmeasured and unknown factors. Moreover, 7 of the 11 studies examined nonlinearity, although their findings are inconsistent, as discussed above.

Occupation-specific cohorts

Six of the reviewed studies were conducted in occupation-specific cohorts without known underlying health conditions (Ostro et al. 2010, 2015; Puett et al. 2011; Hart et al. 2011, 2015; Weichenthal et al. 2014). By contrast, two studies of occupation-specific cohorts that focus only on individuals with health conditions are included below in the evaluation of studies of patients with underlying health conditions (DuPre et al. 2019; Lipfert and Wyzga 2019). Similar to the studies conducted in the general population, the six studies conducted in occupation-specific cohorts all reported a risk estimate (HR) for the $PM_{2.5}$-mortality association assuming linearity. Two of the six studies (Weichenthal et al. 2014; Hart et al. 2015) also evaluated potential non-linearity of the association.

Linear Results. Among the six studies in occupation-specific cohorts, three studies included females only (teachers in the studies by Ostro et al. 2010, 2015; nurses in the study by Hart et al. 2015), two studies included males only (health professionals in the study by Puett et al. 2011; trucking industry workers in the study by Hart et al. 2011), and only one study included both males and females (commercial pesticide applicators, farmers, and their families in the study by Weichenthal et al. 2014). All studies reported results without copollutant adjustment, and only one study (Puett et al. 2011) further reported copollutant-adjusted results.

The three studies among females do not report consistent results. Although both Ostro et al. (2010) and Ostro et al. (2015) examined $PM_{2.5}$-mortality (non-accidental) associations among participants of the CTS, the former study reported statistically significantly positive associations (within 8 km buffer, HR = 1.49, 95% CI: 1.28–1.74; within 30 km buffer, HR = 1.45, 95% CI: 1.36–1.55) whereas the latter study reported no association (HR = 1.01, 95% CI: 0.98–1.05). A key difference between the two studies is that Ostro et al. (2010) examined direct site measured $PM_{2.5}$ and restricted the analyses to subjects whose residences were within 8 km and 30 km of a monitor, respectively, whereas Ostro et al. (2015) examined modeled $PM_{2.5}$ and included CTS participants regardless of their distance to monitors. As a result, the participants in the study by Ostro et al. (2010) (n = 7,888 within 8 km buffer; n = 44,847 within 30 km buffer) are largely a non-representative subsample of the participants in the study by Ostro et al. (2015) (n = 101,884) and the results of the two studies are not directly comparable. Other differences between the two studies that could have partly contributed to the difference in observed results may be related to the follow-up period, as well as the exposure metric, temporal scale, and contrast. The study among female nurses by Hart et al. (2015) reported a positive $PM_{2.5}$-mortality (non-accidental) association (HR = 1.13, 95% CI: 1.05–1.22 for modeled $PM_{2.5}$; HR = 1.12, 95% CI: 1.05–1.21 for monitored $PM_{2.5}$) that is of similar magnitude to the female-specific results reported by Pinault et al. (2016) and Villeneuve et al. (2015) in studies conducted in the general population.

The magnitude of the results of the two male-only studies conducted in occupation-specific cohorts, without copollutant adjustment, are weaker than the male-specific result in the general population reported by Pinault et al. (2016). Specifically, Puett et al. (2011) reported no $PM_{2.5}$-mortality (non-accidental) association (HR = 0.94, 95% CI: 0.87–1.00) and Hart et al. (2011) reported a very weak, positive $PM_{2.5}$-mortality (all-cause) association (HR = 1.04, 95% CI: 1.01–-1.07), whereas Pinault et al. (2016) reported an HR of 1.344 (95% CI: 1.239–1.457). While the healthy worker effect is often a possible explanation for weaker associations observed in occupation-specific cohorts compared to the general population, such speculation should be made with caution in this case because Pinault et al. (2016) reported an association that is much stronger than all the other studies conducted in the general population and, therefore, could be an outlier. With copollutant adjustment, Puett et al. (2011) still reported no $PM_{2.5}$-mortality (non-accidental) association (HR = 0.94, 95% CI: 0.87–1.02), as opposed to the male-specific result of a weak positive

association in the general population with copollutant adjustment reported by Di et al. (2017) (HR = 1.087, 95% CI: 1.083–1.090).

Weichenthal et al. (2014) reported no $PM_{2.5}$-mortality (non-accidental) association, either overall or in sex-specific subgroups, although the exact P-value was not reported for the test of effect modification by sex. These null findings are consistent with the null results reported by Puett et al. (2011) and Ostro et al. (2015), although the studies vary by occupation of participants and many other aspects of study design.

Non-linear Results. Hart et al. (2015) used stepwise restricted cubic spline techniques (degree of freedom not reported) to evaluate potential non-linearity of the $PM_{2.5}$-mortality (non-accidental) association and reported an approximately linear shape of the curve for both direct site measured $PM_{2.5}$ and modeled $PM_{2.5}$, similar to the non-linear results reported in the studies by Di et al. (2017) and Lepeule et al. (2012) that were conducted in the general population. A potential threshold for the $PM_{2.5}$-mortality curve was not examined by Hart et al. (2015).

In the study by Weichenthal et al. (2014), the authors stated that 'concentration–response functions were graphed using natural splines for $PM_{2.5}$ with two degrees of freedom using adjusted Cox survival models.' However, non-linear results were only reported for cardiovascular-specific mortality, the other health outcome of interest in the study, and not for non-accidental mortality.

Study Quality. The studies conducted in occupation-specific cohorts share certain strengths and limitations. In general, these studies have smaller sample sizes than the studies conducted in the general population. The two largest studies have sample sizes just above 100,000, which we considered insufficient without justification from power calculation in our study quality evaluation. Because of the particular characteristics of workers and the limited geographic locations within which some of the studies were conducted (e.g. Ostro et al. 2010, 2015; Puett et al. 2011; Weichenthal et al. 2014), the results of these studies have limited generalizability.

Similar to the studies conducted in the general population, the six studies conducted in occupation-specific cohorts all assessed each participant's exposure to $PM_{2.5}$ by assigning to his/her location (primarily residential location) an ambient $PM_{2.5}$ concentration that was either from direct site measurements at one or a few nearby stationary monitoring sites or estimates from prediction models; this methodology is subject to substantial exposure measurement error. Yet, most of the occupation-specific studies meet our criteria for assignment to participants' locations and residential mobility and are therefore less subject to exposure measurement error associated with these aspects, which is a clear strength compared to the studies conducted in the general population.

The occupation-specific studies also, in general, adjusted for a larger number of key confounders, particularly individual-level behavioral factors (including diet, physical activity, and medication use), than the studies conducted in the general population. The results of the occupation-specific studies are still subject to residual confounding by other key confounders (particularly temperature, relative humidity, and other chemical exposures), as well as unmeasured and unknown confounders, however. Five of the six studies conducted in occupation-specific cohorts, including two studies that are less subject to exposure measurement error (i.e. by meeting our criteria for all aspects of the exposure assessment category except for personal activities and multiple methods) (Hart et al. 2015; Ostro et al. 2015), did not adjust for copollutants, indicating the results of these studies likely do not reflect the independent association of $PM_{2.5}$ with mortality. This is a clear limitation compared to the studies conducted in the general population. In the only study that did adjust for copollutants (Puett et al. 2011), the correlation between $PM_{2.5}$ and copollutants was not examined (which undermines the effectiveness of copollutant adjustment) and thus the study does not meet the quality criterion for copollutant measurement.

As mentioned above, nonlinearity was not examined in most of the studies conducted in occupation-specific cohorts, which is a clear limitation compared to the studies conducted in the general population. In addition, because non-linear results were not reported for non-accidental mortality by Weichenthal et al. (2014), we did not consider this study as meeting the nonlinearity

criterion in the study quality evaluation, although it is possible that the authors examined the $PM_{2.5}$-mortality (non-accidental) curve but did not report the results.

Patients with underlying health conditions

Six of the reviewed studies were conducted in patients with underlying health conditions (Hartiala et al. 2016; Chen et al. 2016; Deng et al. 2017; Malik et al. 2019; DuPre et al. 2019; Lipfert and Wyzga 2019). As noted above, these include two studies where patients were also from occupation-specific cohorts (DuPre et al. 2019; Lipfert and Wyzga 2019). Similar to the studies conducted in the general population and in occupation-specific cohorts, the studies conducted in patients with underlying health conditions all reported a risk estimate (HR) for the $PM_{2.5}$-mortality association assuming linearity. Three of the six studies (Chen et al. 2016; Deng et al. 2017; Malik et al. 2019) also evaluated potential non-linearity of the association.

Linear Results. Of the six studies in patients with underlying health conditions, four included patients with cardiovascular disease (CVD) or CVD risk factors (e.g. myocardial infarction [MI] in the studies by Malik et al. 2019; Chen et al. 2016; undergoing elective diagnostic coronary angiography in the study by Hartiala et al. 2016; male ostensibly hypertensive veterans in the study by Lipfert and Wyzga 2019) and two studies included cancer patients (e.g. female nurses with breast cancer in the study by DuPre et al. 2019; hepatocellular cancer in the study by Deng et al. 2017). One of the six studies (Malik et al. 2019) only reported copollutant-adjusted results, whereas the other five studies only reported results without copollutant adjustment.

Both studies conducted among MI patients reported statistically significant positive associations between $PM_{2.5}$ and mortality (HR = 1.13, 95% CI: 1.07–1.20 in the study by Malik et al. 2019; HR = 1.22, 95% CI: 1.03–1.45 in the study by Chen et al. 2016), which are stronger than most of the associations reported in the general population. It is possible that MI patients are more susceptible to the impact of $PM_{2.5}$ exposure, but this contrast in magnitude of association could also be at least partly attributable to differences in $PM_{2.5}$ assessment, adjustments of confounders and copollutants, and other study design aspects. Chance findings also cannot be ruled out for the observed stronger association among MI patients, particularly because of the very small number of studies of these patients.

On the contrary, the two studies conducted among patients with CVD risk factors reported mixed results, with either weaker positive, null, or negative associations. Specifically, Hartiala et al. (2016) reported no association between $PM_{2.5}$ and mortality (all-cause) in patients undergoing elective diagnostic coronary angiography (HR = 1.16, 95% CI: 0.96–1.41). Lipfert and Wyzga (2019) examined the $PM_{2.5}$-mortality (all-cause) association among male ostensibly hypertensive veterans and reported a very weak, positive association among whites (HR = 1.051, 95% CI: 1.005–1.100) and a statistically significant inverse association among blacks (HR = 0.817, 95% CI: 0.750–0.891). It is possible that patients with CVD risk factors, similar to the general population, are less susceptible to the impact of $PM_{2.5}$ exposures compared to MI patients, but, given the large variations in study design aspects and the very small number of studies available, it is impossible to rule out other possible explanations, such as confounding, bias, or chance.

DuPre et al. (2019) reported no $PM_{2.5}$-mortality (all-cause) association in female nurses with breast cancer (HR = 1.12, 95% CI: 0.96–1.30); whereas Deng et al. (2017) reported a positive $PM_{2.5}$-mortality (all-cause) association in patients with hepatocellular cancer (HR = 1.18, 95% CI: 1.16–-1.20). The magnitude of this association is similar to those reported in MI patients and greater than most of the associations reported in the general population. Although it is possible that hepatocellular cancer patients are also more susceptible to the impact of $PM_{2.5}$ exposure compared to the general population, it cannot be ruled out that the observed contrast is attributable to confounding, bias, or chance, given the large variations of study design aspects and the very small number of studies available.

Non-linear Results. All three studies that evaluated potential non-linearity of the $PM_{2.5}$-mortality association used cubic spline techniques, although the degree of freedom and

confounding adjustments varied. Both of the studies conducted among MI patients (Chen et al. 2016; Malik et al. 2019) reported a linear shape for the $PM_{2.5}$-mortality curve, similar to the studies by Di et al. (2017) and Lepeule et al. (2012) that were conducted in the general population, and to the study by Hart et al. (2015) that was conducted in an occupation-specific cohort. In the study by Deng et al. (2017) that was conducted among patients with hepatocellular cancer, a J-shaped $PM_{2.5}$-mortality (all-cause) curve was reported. Potential thresholds for the $PM_{2.5}$-mortality curve were not examined in the studies among patients with underlying health conditions.

Study Quality. The six studies conducted among patients with underlying health conditions share certain strengths and limitations. In general, these studies have smaller sample sizes than the studies conducted in the general population and in occupation-specific cohorts, with four studies having sample sizes below 10,000, where statistical power is very limited considering the large number of potential confounders adjusted for. All underlying health conditions were ascertained by independent clinical assessment or self-report of physician diagnosis and as such, all six studies meet our study quality criterion for underlying health conditions. Because of the particular characteristics of patients and the limited geographic location within which some of the studies were conducted (*e.g.* Hartiala et al. 2016; Chen et al. 2016; Deng et al. 2017), however, the results of these studies have limited generalizability across populations. Three of the six studies tested model assumptions in their statistical analyses to ensure that they were satisfied, which is a strength compared to the studies conducted in the general population and occupation-specific cohorts where almost none of the studies did such testing.

Similar to the studies conducted in the general population and in occupation-specific cohorts, the six studies conducted in patients with underlying health conditions all assessed each participant's exposure to $PM_{2.5}$ by assigning to his/her location (primarily residential location) an ambient $PM_{2.5}$ concentration that was either from direct site measurements at one or a few nearby stationary monitoring sites or estimates from prediction models; this methodology is subject to substantial exposure measurement error. As with the studies conducted in occupation-specific cohorts, most of the studies among patients meet our study quality criterion for assignment to participants' locations, which is a clear strength compared to the studies conducted in the general population. Similar to the studies conducted in the general population, most of the studies among patients do not meet the criterion for residential mobility, which is a clear limitation compared to occupation-specific cohorts. Further, most of the studies among patients do not meet the criteria for spatial or temporal variabilities, which is a clear limitation compared to studies conducted in the general population and occupation-specific cohorts. As a result, the studies among patients are also subject to exposure measurement error due to a lack of accounting for residential mobility or spatial or temporal variabilities.

As with the studies conducted in occupation-specific cohorts, the studies conducted in patients with underlying health conditions are more likely to have adjusted for individual-level behavioral factors, such as physical activity and medication use, than the studies conducted in the general population. However, most of the studies conducted in patients did not adjust for at least one of the key confounders that were typically adjusted for in the studies conducted in the general population, including race, BMI, and smoking. As such, the results of these studies are still subject to residual confounding by many key, unmeasured, and unknown confounders. Similar to the studies conducted in occupation-specific cohorts, five of the six studies conducted in patients, including the study that is less subject to exposure measurement error (i.e. by meeting our criteria for all aspects of the exposure assessment category except for personal activities and multiple methods) (Chen et al. 2016), did not adjust for copollutants, indicating the results of these studies likely do not reflect the independent association of $PM_{2.5}$ with mortality. This is a clear limitation compared to the studies conducted in the general population. The only study that did adjust for copollutants (Malik et al. 2019) did not meet our study quality criterion for copollutant measurement, which undermines the effectiveness of copollutant adjustment.

Evidence integration

We integrated the evidence across the epidemiology studies using modified Bradford Hill aspects (Supplemental Table S1) as a framework. These aspects were originally developed to answer the question 'is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?' (Hill 1965). Thus, the aspects should be used as guides for evaluating alternative explanations of the observed patterns in the study results and to assess whether they are a more compelling explanation of the results at hand than the explanation of causality (Rhomberg et al. 2013). Although evidence integration is typically conducted by assigning greater weight to higher quality studies and less weight to lower quality studies, the delineation of studies into higher and lower quality groups was not done in this review, considering the shared key strengths and limitations (e.g. with respect to exposure assessment, confounding, and statistical methods) and apparent consistency of the linear results across studies. In addition, some of the shared strengths and limitations go beyond the study quality criteria (e.g. using ambient $PM_{2.5}$ concentration to estimate individual $PM_{2.5}$ exposure, assuming linearity of the $PM_{2.5}$-mortality relationship), and as discussed above, could have consistently and more substantially affected the studies and their risk estimates. Therefore, we incorporated the overall study strengths and limitations into the integration of evidence, particularly where they are relevant to the evaluation of alternative explanations of the results.

Consistency

Evidence for causality is stronger if consistent effects are observed among studies of different designs, populations, locations, circumstances, and time periods. The studies reviewed here were conducted in various locations across the US and Canada and evaluated different types of populations (general, occupational, or patients with underlying health conditions). All used a cohort study design, but there were many differences among studies with regard to specific aspects of study conduct. Despite the differences in these factors across studies, the majority of studies (particularly those in the general population) reported weak, positive associations that were statistically significant.

Null associations were reported more often in occupational populations compared to the general population, which is not surprising given that occupational populations tend to be healthier than the general population (Li and Sung 1999; Chowdhury et al. 2017). Studies of MI patients reported stronger positive associations than most of those reported for the general population, whereas studies of patients with CVD risk factors or cancer patients reported mixed results, with some positive, null, or negative $PM_{2.5}$-mortality associations. One would expect that patients with underlying health conditions would be more susceptible and thus would have a greater risk of mortality from $PM_{2.5}$ exposure, but this was only the case for the studies of MI patients and not patients with other health conditions. Given the small number of studies of patients with each particular underlying health condition however, it cannot be ruled out that the observations from these studies may be attributable to chance, bias, or confounding. Regardless, most of the studies conducted in the general population, as well as some of those conducted in occupational and patient populations, reported risk estimates of a similar magnitude, indicating that there is some consistency for weak, positive associations between long-term exposure to $PM_{2.5}$ and total (all-cause or non-accidental) mortality across studies.

Strength of association

Large and precise risk estimates for an exposure-outcome association are less likely to be due to bias, confounding, or chance and, therefore, are more indicative of an underlying causal relationship than risk estimates that are small and imprecise. Although the HRs for the $PM_{2.5}$-mortality association reported in the studies in this review are mostly of high precision (and of extremely high precision in the studies with extremely large sample sizes), their magnitudes mostly indicate a weak

association. Considering the substantial extent of potential bias and confounding that these HR estimates are subject to based on the methodology of the studies, the weak associations do not support a causal $PM_{2.5}$-mortality relationship.

A key source of bias to the reported weak associations is $PM_{2.5}$ exposure measurement error, which could be substantial. As discussed above, all studies assessed each participant's exposure to $PM_{2.5}$ by assigning to his/her location (primarily residential location) an ambient $PM_{2.5}$ concentration that was either from direct measurements at one or a few nearby stationary monitoring sites or estimates from prediction models, which does not account for individual factors that vary among participants (such as time spent indoors or at non-residential locations, and personal activities) and can greatly affect their actual $PM_{2.5}$ exposures. Moreover, almost none of the studies reviewed here accounted for personal activities, and many of the studies did not assign ambient $PM_{2.5}$ data to participants' locations from the same time period and did not account for temporal variability or residential mobility.

Another potential important source of bias is model misspecification. As discussed above, all studies calculated a risk estimate for the $PM_{2.5}$-mortality association assuming linearity, but the shapes of the $PM_{2.5}$-mortality curves varied across the studies that also evaluated potential non-linearity, indicating that linearity may not be a valid modeling assumption. In calculating the risk estimate under a linear assumption, all studies also used a Cox proportional hazards regression model, which relies on a key assumption of proportional hazards, yet very few studies tested the proportional hazards assumption to ensure that it was satisfied, leaving biased modeling results unidentified.

The reported weak associations are also subject to confounding by copollutants, unmeasured confounders (e.g. diet, physical activity, temperature, relative humidity, medication use, other chemical exposures, stress, and noise), and unknown confounders (Clougherty and Kubzansky 2009; Stansfeld 2015; US EPA 2019). As discussed above, none of the studies meet the criterion for key confounders and many of the studies did not adjust for any copollutant exposure. Further, in the studies that meet our criterion for copollutant adjustment, only one or a few select copollutants were adjusted for and none of the studies meet the criterion for copollutant measurement, indicating that the copollutant adjustments are likely ineffective and the results likely do not reflect the independent association of $PM_{2.5}$ with mortality. Residual confounding could also exist when covariate adjustment is incomplete or secular trend is not sufficiently adjusted for (Cox 2017).

The above-mentioned universal sources of bias and confounding could have systematically shifted the study results and artificially created consistency of weak, positive associations. Given this consistency across studies, chance is less likely as a possible non-causal explanation compared to bias and confounding. Nonetheless, it is worth noting that the majority of the HR estimates from the studies are subject to the multiple comparison issue, so chance findings are still possible. Overall, the aspect of strength for $PM_{2.5}$-mortality associations is not met.

Coherence

Coherence occurs when all of the known facts related to an observed association that come from various realms of evidence fit together in a logical manner (Hill 1965). Coherence is difficult to assess for the evaluation of associations between long-term $PM_{2.5}$ exposure and mortality. Controlled human exposure studies are conducted with short exposure durations and evaluate health outcomes of generally low adversity for ethical reasons. Experimental animal studies can be conducted with longer exposure durations and can evaluate more severe health effects, but the available chronic studies of $PM_{2.5}$ exposure in experimental animals used $PM_{2.5}$ concentrations that are much higher than ambient concentrations (US EPA 2019, 2020), so any health effects reported in these studies are not informative regarding potential human health effects at lower $PM_{2.5}$ concentrations. It is notable, however, that in a review of multiple morbidity studies of rodents with lifetime inhalation exposures to various forms of $PM_{2.5}$ (such as diesel exhaust, carbon black, and coal dust), there was no increase in mortality for any exposure level compared to controls, even

when exposures were so high as to produce lung overload (Gamble 1998). Similarly, in studies evaluating atherosclerotic changes in apolipoprotein E-null mice (which are susceptible to atherosclerosis due to their high plasma levels of low-density lipoprotein and very low-density lipoprotein) with chronic exposures to $PM_{2.5}$, such as those reviewed by Prueitt et al. (2015), mortality was not increased with exposure to $PM_{2.5}$ concentrations ranging from 85–138 $\mu g/m^3$ compared to controls. The lack of increased mortality in experimental animal studies of long-term $PM_{2.5}$ exposure, even at very high concentrations that induce other adverse effects and in an animal model that is susceptible to cardiovascular morbidity, does not provide support for a causal relationship between long-term $PM_{2.5}$ exposure at lower, ambient concentrations and mortality.

Biological plausibility

Evidence for a plausible biological mechanism for an effect can contribute to a scientifically defensible determination of causation. Agencies such as US EPA consider the underlying morbidities for cardiovascular-, respiratory-, and metabolic disease-specific mortality (which contribute largely to total mortality) as support for the plausibility of associations with all-cause mortality (US EPA 2019). Several biological mechanisms have been proposed for these underlying morbidities, based on evidence from experimental animal, controlled human exposure, and epidemiology studies (US EPA 2019). Although we did not systematically review this evidence, we provide a high level review of the proposed mechanisms below, based on other comprehensive reviews in the peer-reviewed literature.

Two well-studied mechanistic pathways involve induction of oxidative stress and inflammation in the respiratory tract after inhalation of $PM_{2.5}$, leading to lung cell injury (Xing et al. 2016; Li et al. 2018; US EPA 2019; Yu et al. 2020). Release of inflammatory mediators, as well as direct translocation of $PM_{2.5}$ particles into the systemic circulation, can contribute to local oxidative stress and inflammation at extrapulmonary sites, resulting in cardiovascular effects (e.g. arrhythmia, atherosclerotic plaque instability) that increase the risk of cardiovascular disease (US EPA 2019; Yitshak-Sade et al. 2019; Miller 2020; Yu et al. 2020), or metabolic effects such as insulin resistance and metabolic syndrome comorbidities (US EPA 2019). The oxidative stress induced by $PM_{2.5}$ in the respiratory tract can also disrupt calcium homeostasis by increasing intracellular calcium concentrations, which can further activate inflammatory reactions and lead to cell damage or cell death (Xing et al. 2016). There is also evidence from a few experimental animal studies that $PM_{2.5}$ can modulate the autonomic nervous system, potentially by binding to receptors on lung or nerve cells, resulting in changes in heart rate (US EPA 2019; Yang et al. 2020). Such changes could potentially lead to cardiovascular outcomes such as hypertension, arrhythmia, and cardiovascular diseases such as ischemic heart disease or heart failure (US EPA 2019).

Despite the available mechanistic evidence, the epidemiology evidence for associations between $PM_{2.5}$ exposure and cardiovascular, respiratory, and metabolic disease morbidity has similar issues (such as potential exposure measurement error and confounding) as the mortality evidence reviewed here, as epidemiology studies for morbidity and mortality are conducted in a generally similar manner. Morbidity evidence that is subject to such uncertainty does not provide strong support for biological plausibility of associations between $PM_{2.5}$ exposure and mortality. Further, because morbidity associated with air pollution is less severe than mortality and, thus, is a more sensitive indicator of adverse health effects than death, morbidity should show stronger associations than mortality (Gamble 1998). This is not observed for $PM_{2.5}$, however, as the evidence reported in the PM ISA indicates that $PM_{2.5}$ associations are similar or weaker, but not stronger, as the health effects become less severe (US EPA 2019). For example, the evidence is stronger (i.e. effect estimates are higher and positive results are more consistent) for cause-specific mortality compared to underlying morbidity outcomes such as adult asthma prevalence, ischemic heart disease, myocardial infarction, or stroke (US EPA 2019, 2020).

While some controlled human exposure and experimental animal studies provide evidence for certain morbidity endpoints with exposure to $PM_{2.5}$, the evidence is not strong nor consistent

across studies and the effects are reported almost exclusively at high exposures (US EPA 2020) and therefore do not support biological plausibility for more serious effects at ambient exposures. Many of the adverse health effects reported in these experimental studies also have thresholds and do not occur at lower concentrations; for example, Green et al. (2002) reported that various chronic exposure studies in rats with different compositions of $PM_{2.5}$ indicate that concentrations of 100–200 $\mu g/m^3$ must be exceeded before potentially adverse changes occur. As this threshold is above ambient concentrations, these experimental studies do not provide support for adverse effects at ambient concentrations. Thus, while there is evidence in the literature for a variety of potential biological mechanisms for the underlying health effects that contribute to total mortality, the experimental studies of adverse health effects with $PM_{2.5}$ exposure do not provide evidence of biological plausibility for mortality associated with ambient $PM_{2.5}$ exposures, so the aspect of biological plausibility is only partially met.

Biological gradient

An association is more likely to be causal when a well-characterized exposure-response relationship exists (e.g. disease risk increases with greater exposure intensity and duration). The studies in this review were generally consistent in reporting weak but statistically significant associations that indicate an increasing exposure-response relationship with increasing $PM_{2.5}$ exposure, but this relationship is not well characterized and therefore may not be reliable. While all the studies reported a risk estimate for the $PM_{2.5}$-mortality association assuming linearity, as discussed above, a linear $PM_{2.5}$-mortality relationship with no threshold is not biologically plausible for the underlying morbidity that contributes to the outcome of mortality. Among the studies that also evaluated potential non-linearity of the association, the reported shape varied substantially, from approximately linear to supralinear to V-shaped, J-shaped, or monotonically increasing. Among the two studies that formally evaluated potential thresholds for the $PM_{2.5}$-mortality curve, the estimated thresholds varied drastically, from 11 $\mu g/m^3$ to 0 $\mu g/m^3$ (Villeneuve et al. 2015; Pinault et al. 2016).

Although a few studies reported an approximately linear shape of the exposure-response curve, the degree of potential bias in those studies due to exposure measurement error (as discussed above) may have been sufficient to produce a false linear result and prevent the detection of a threshold (Rhomberg et al. 2011). As discussed above, the reported variation in non-linear shapes across studies also indicates that linearity may not be a valid modeling assumption. In fact, the linear assumption may have masked important heterogeneity and details of the underlying $PM_{2.5}$-mortality relationships.

Before the $PM_{2.5}$-mortality curve can be well characterized and contribute to an evaluation of causation with confidence, a number of other issues need to be addressed. For example, the studies in this review rarely used the same non-linear modeling techniques to evaluate the $PM_{2.5}$-mortality exposure-response curves, so it is unclear as to the extent that this affects the comparability of the non-linear results. The available data at lower levels of $PM_{2.5}$ (e.g. below the current standard of 12 $\mu g/m^3$) are sparse, limiting the ability to characterize the curve at lower ambient $PM_{2.5}$ levels with confidence (Smith and Gans 2015). As $PM_{2.5}$ refers to a heterogeneous mixture of constituents that may vary greatly from one location to the other, and mortality (either all-cause or non-accidental) entails a variety of cause-specific deaths that have different etiologies, it is important to develop methods to account for these heterogeneities when characterizing the $PM_{2.5}$-mortality curve in a multi-city or even nationwide study (Cox 2017). Overall, the aspect of biological gradient is not met, as these issues need to be addressed before the $PM_{2.5}$-mortality exposure-response relationship can be considered to be well characterized.

Temporality

For a causal relationship to exist, exposure must precede the occurrence of disease with sufficient lag time, if any is expected. Because all the studies in this review were cohort by design, our study quality criterion for temporality is considered as being met in all studies. However, this is not

without caveats that undermine the establishment of temporality and, thus, affect a judgment of causality.

As discussed above, all studies in this review were secondary analyses of data from existing cohorts that were initially recruited for research questions unrelated to $PM_{2.5}$ or mortality. Although the conceptualized study baseline clearly preceded mortality follow-up in each study, ambient $PM_{2.5}$ data, unlike data for participants' locations where ambient $PM_{2.5}$ data were assigned to, were often unavailable at the exact time of baseline (or the time of address update during the follow-up), as the ratings of studies for our 'assignment to participants' locations' study quality criterion show. Of the studies that do not meet the this criterion, all but one study assigned $PM_{2.5}$ data from as long as 5+ to 10+ years later to participants' locations, thus underestimating the participants' actual $PM_{2.5}$ exposure concentration and overestimating the mortality rate associated with lower $PM_{2.5}$ exposures.

Another caveat is that the temporality criterion used in this review does not enforce any lag time between $PM_{2.5}$ exposure and mortality, as such a lag time is largely unknown. However, the $PM_{2.5}$ exposure windows examined in the studies were often within a period of five years before mortality, which is unlikely to be the most relevant exposure window considering the chronic pathological changes and disease processes that have been proposed as potential underlying causal mechanisms for mortality (US EPA 2019). Even in the studies where longer lag times were examined, $PM_{2.5}$ exposure was only measured for a short period of time when, in fact, $PM_{2.5}$ exposure persists throughout an individual's lifetime (even though the concentrations can change over time) and unmeasured historical $PM_{2.5}$ exposures can be substantially higher than exposures measured in the studies. Thus, even though all of the studies in this review were designed to allow for exposure to precede the outcome, these caveats undermine the full establishment of temporality; therefore, this aspect is only partially met.

Specificity

Causal inference is strengthened when there is evidence that links a specific exposure to a specific health outcome, although any health outcome may have multiple causes. Mortality and the underlying morbidity associated with it have multiple causes and thus are not specific effects of $PM_{2.5}$ exposure. Other risk factors for mortality include many of the key confounders that should be identified and adjusted for in epidemiology studies examining associations between air pollutants and mortality, such as SES, BMI, physical activity, temperature, relative humidity, medication use, smoking status, and other chemical exposures. As discussed above, other potential confounders not typically measured in air pollution epidemiology studies, such as stress and noise, are also risk factors for mortality (Clougherty and Kubzansky 2009; Stansfeld 2015; US EPA 2019).

It is of note that $PM_{2.5}$ itself is not a 'specific' chemical but rather is comprised of many different solid and liquid constituents that vary in their presence and concentrations across locations and time periods due to variation in their sources. As discussed below, if ambient $PM_{2.5}$ is causally associated with various health effects, including mortality, the specific constituents responsible are unknown. Overall, the aspect of specificity for $PM_{2.5}$-mortality associations is not met.

Analogy

The evidence for causality is stronger when a similar substance is an established causal factor for a similar effect. A comparison of $PM_{2.5}$ to other types of ambient particulates is difficult, as all such particulates in the $PM_{2.5}$ size fraction are included as $PM_{2.5}$ components. However, exposures to other size fractions of PM ($PM_{10\text{-}2.5}$ and UFPs) are not established causal factors for mortality, due to limited available data or uncertainties associated with the epidemiology studies of these PM size fractions (US EPA 2019).

$PM_{2.5}$ composition varies from one location to another, and the specific constituents potentially responsible for the reported associations between long-term $PM_{2.5}$ exposure and mortality are unknown. For example, US EPA recently concluded that the pattern of results across studies of

particular components or sources of $PM_{2.5}$ 'demonstrate that no individual $PM_{2.5}$ component or source is a better predictor of mortality than $PM_{2.5}$ mass' (US EPA 2019). It is notable that experimental studies in both humans and animals indicate that exposures to nonacidic, soluble sulfates and nitrates, which make up sizable mass fractions of ambient PM, are associated with little to no adverse effects (as reviewed by Green et al. 2002). Further, exposures to strongly acidic sulfates induce adverse respiratory effects in humans or experimental animals only at high exposure levels (> 100 $\mu g/m^3$), but such constituents are typically present in ambient air at concentrations below 5 $\mu g/m^3$ (Green et al. 2002). Thus, it is unlikely that lower exposures to these constituents in ambient air are associated with morbidity, let alone mortality.

Environmental tobacco smoke (ETS) is a source of $PM_{2.5}$ that is itself a mixture of thousands of constituents (Rojas-Rueda et al. 2021). Multiple studies have reported statistically significant associations between ETS exposure and all-cause mortality, with the magnitude of associations being similar to or slightly higher than those reported for long-term $PM_{2.5}$ exposure and all-cause mortality (Lv et al. 2015; Diver et al. 2018; Pelkonen et al. 2019). The concentration of $PM_{2.5}$ particles in ETS is much higher (up to an order of magnitude) than that of $PM_{2.5}$ in indoor and outdoor environments where smoking does not occur (Van Deusen et al. 2009; Ruprecht et al. 2016), so ETS can be considered an analogous substance to $PM_{2.5}$ exposures well above the $PM_{2.5}$ NAAQS, but not to lower, ambient concentrations near the $PM_{2.5}$ NAAQS. Overall, we did not identify any particulate substances similar to $PM_{2.5}$ that are established causal factors for all-cause mortality at low, ambient concentrations.

Experiment

Natural experiments can provide strong evidence for causation when an intervention or cessation of exposure results in decreased health risks. $PM_{2.5}$ concentrations have decreased in the US over time as the PM NAAQS have been revised and reduced, but even the epidemiology studies with the most recent $PM_{2.5}$ data continue to report positive associations between $PM_{2.5}$ exposure and mortality. Most of the exposure data measured or modeled in the studies reviewed here is from 1990 to 2010, with only one study (Lefler et al. 2019) including exposure data after 2013, when the impact of the most recent lowering of the $PM_{2.5}$ NAAQS (implemented in 2013) can be assessed. It is likely that even if future studies include $PM_{2.5}$ exposure data from after 2013, they would continue to report positive associations with mortality or other health endpoints at lower and lower exposure concentrations. This is because when annual average $PM_{2.5}$ concentrations decline during the study period to a similar degree across study locations, it is possible that the distribution of $PM_{2.5}$ concentrations that occurred in any particular year is associated with mortality that was at least partially attributable to the higher $PM_{2.5}$ exposures that occurred in earlier years (Smith and Chang 2020). In addition, if most studies continue to use similar exposure assessment approaches (e.g. using ambient $PM_{2.5}$ to estimate individual $PM_{2.5}$ exposure), the degree of potential bias due to exposure measurement error may produce a false linear result and obscure any thresholds.

Several interventional and 'accountability' studies have examined past reductions in ambient $PM_{2.5}$ and the degree to which those reductions have resulted in decreased health risks by using causal modeling approaches, which are not within the scope of this review. Two recent, comprehensive reviews of air pollution interventional and accountability studies reported mixed results across studies, indicating that measures to reduce $PM_{2.5}$ have not clearly reduced mortality risks, particularly when confounding was well controlled (Henneman et al. 2017; Burns et al. 2019a). Even in the studies that showed an association between $PM_{2.5}$ reduction and mortality reduction, one cannot directly attribute the mortality reduction to a decrease in $PM_{2.5}$ concentrations, as these studies primarily evaluated the effectiveness of policies that could lower ambient $PM_{2.5}$ concentrations but could also affect other risk factors for mortality. Conversely, for studies reporting no association between $PM_{2.5}$ reduction and mortality reduction, one can conclude that similar policy changes do not lead to a reduction in mortality, even though they may have led to a reduction in

$PM_{2.5}$ concentrations. Overall, these studies do not provide any compelling evidence that a reduction in ambient $PM_{2.5}$ concentrations is associated with a reduction in mortality.

Causal conclusion

We evaluated the potential causal relationship between long-term $PM_{2.5}$ exposure and mortality using the four-tiered causal framework shown in Supplemental Table S2. The only Bradford Hill aspect that is fully met for the studies in this review is that of consistency, as there is some consistency across studies for reporting weak, positive associations. In addition, the aspects of temporality and biological plausibility are only partially met. All studies in this analysis are cohort by design and thus allow for exposure to precede the outcome; however, several caveats undermine the full establishment of temporality, as discussed above. Although there is evidence for a variety of potential biological mechanisms for the underlying health effects that contribute to total mortality, experimental studies of these effects do not provide evidence of biological plausibility for mortality associated with ambient $PM_{2.5}$ exposures.

The other Bradford Hill aspects are either not met or there is inadequate information for their full evaluation. The aspect of strength of association is not met, as all reported associations are very weak, and there are many alternative explanations for such small risk estimates, including bias attributable to exposure measurement error or model misspecification, and substantial confounding by copollutants and unmeasured or unknown confounders. The aspect of coherence is not met due to inadequate evidence. The available animal studies of $PM_{2.5}$ were only conducted at very high concentrations and are not informative regarding potential human health effects at lower $PM_{2.5}$ concentrations (although increased mortality was not even observed in animals exposed to high concentrations of $PM_{2.5}$ and thus is not likely to be observed at lower concentrations). The aspect of biological gradient is also not met due to inadequate evidence; although the studies indicate an exposure-response relationship, there are several issues that need to be addressed before it can be well characterized and, thus, reliable (as discussed above).

The aspect of specificity is not met because $PM_{2.5}$ exposure is not specific to mortality, and $PM_{2.5}$ is not a specific chemical but is made up of varying constituents depending on the location and time period. The aspect of analogy is also not met, because there are no particulate substances similar to $PM_{2.5}$ that are established causal factors for all-cause mortality at low, ambient concentrations. Finally, the aspect of experiment is not met due to inconsistent evidence. Although the evidence from interventional and accountability studies does not indicate that reductions of $PM_{2.5}$ concentrations have clearly reduced mortality risks, these studies only evaluated the effects of policy changes that may have reduced $PM_{2.5}$ concentrations but could also affect other risk factors for mortality.

Overall, our evaluation of causality using the Bradford Hill aspects indicates that there is some consistency across studies for reporting positive associations, but these associations are very weak and explanations other than causality, such as bias and confounding, cannot be ruled out. There is no coherence with the available experimental evidence and there is no clear evidence for a biological mechanism for $PM_{2.5}$ to cause mortality at ambient concentrations, and several caveats undermine the full establishment of the aspects of temporality and biological gradient. Exposure to $PM_{2.5}$ is not specific to mortality, there is no evidence to show that reductions in $PM_{2.5}$ have clearly reduced mortality risks, and there are no substances similar to $PM_{2.5}$ that are established causes of mortality. For these reasons, our evaluation supports a conclusion that the evidence for a causal relationship between long-term exposure to ambient $PM_{2.5}$ and mortality (all-cause or non-accidental) from epidemiology studies published since the 2009 PM ISA is inadequate.

Discussion

We used a transparent systematic review framework based on best practices for evaluating study quality and integrating evidence to conduct a review of the available epidemiology studies evaluating associations between long-term exposure to ambient concentrations of $PM_{2.5}$ and mortality (all-cause and non-accidental) conducted in North America and published after those included in the 2009 PM ISA. Using a causality framework that incorporates best practices for making causal determinations, we concluded that the evidence for a causal relationship between long-term exposure to ambient $PM_{2.5}$ concentrations and mortality from these studies is inadequate.

Our conclusion differs from US EPA's conclusion in the most recent PM ISA that there is a causal relationship between long-term exposure to $PM_{2.5}$ and total (non-accidental) mortality (US EPA 2019). Our review includes all of the North American studies of long-term $PM_{2.5}$ exposure and all-cause or non-accidental mortality included in the most recent PM ISA (but not also included in the 2009 PM ISA), with the exception of four studies that we excluded because they were ecological studies (Garcia et al. 2016; Shi et al. 2016; Wang et al. 2016; Pun et al. 2017); four studies that we excluded because they were the least recent or least informative studies of cohorts examined in more than one study (Lipsett et al. 2011; Crouse et al. 2012; Kioumourtzoglou et al. 2016; Wang et al. 2017a); and one study that we excluded because it did not present relevant effect estimates for associations with mortality (Cox and Popken 2015). Our review also includes seven studies that were not included in the evaluation of mortality in the PM ISA, likely because most were published after the cutoff date for the literature searches conducted for the PM ISA (Hartiala et al. 2016; Deng et al. 2017; Weichenthal et al. 2017; DuPre et al. 2019; Lefler et al. 2019; Lipfert and Wyzga 2019; Malik et al. 2019). Altogether, there are 16 studies included in both our review and the most recent PM ISA. While our conclusion is solely based on the evidence published since the 2009 PM ISA, it is worth noting that US EPA's conclusion in the 2019 PM ISA, although mainly focused on the most recent studies published since the 2009 PM ISA, also relied on the evidence evaluated in the 2009 PM ISA and the associated conclusions.

Although it is possible that the difference in conclusions regarding causality between our review and that in the PM ISA may be partly attributable to the differences in the specific studies included in each review, it is likely that the difference is also attributable to the methodologies used to evaluate the evidence. In the PM ISA, US EPA (2019) did not evaluate and integrate the evidence for causality in a transparent or systematic manner, as the overall process lacks a detailed protocol to ensure that the evaluation is consistent across studies. The PM ISA also lacks an explanation for how the study quality aspects provided in its Appendix were used in the evaluation and integration of the evidence, as it is clear that these aspects were not applied consistently across studies. The study quality aspects should be included in the discussion of study results so they can be considered in the evaluation (including an evaluation of alternative explanations) and appropriate conclusions with regard to causality can be drawn. While US EPA discussed some of the study quality issues (e.g. exposure measurement error, confounding) in the PM ISA, it did not fully consider their impact on the study results and their implications for causality.

US EPA also uses a five-level causal framework that is prone to bias toward causal conclusions. In this framework, the evidence is considered sufficient to conclude a causal relationship if chance, confounding, and other biases can be ruled out with 'reasonable confidence' but does not include guidance for what constitutes 'reasonable confidence.' In addition, US EPA's causal framework requires only one high-quality study for evidence of a causal relationship to be deemed as suggestive, rather than requiring an equivalent review of all studies under the same criteria. The lack of consistent application of study quality aspects to the evaluation and integration of evidence can lead to causal conclusions that are biased and not fully supported by the evidence as a whole.

For our review, when a particular cohort was evaluated in more than one study, we excluded studies if they were less recent or less informative than other studies of the same cohort, even if they met our initial study selection criteria (as described above). It is unlikely that our causal conclusion

would be different if we had included these studies, however, as they had similar methodologies (and thus similar strengths and limitations) and reported similar results as the other studies of the same cohort, though we did exclude some of these studies based on additional limitations with regard to exposure assessment, statistical analyses, and confounder adjustment compared to the included studies of the same cohort. For example, the studies of the CanCHEC 1991 general population cohort reviewed here (Crouse et al. 2015; Weichenthal et al. 2016) reported weak, positive associations with nonaccidental mortality, as did the two studies of this cohort that we excluded (Crouse et al. 2012, 2016). Similarly, the study of female nurses in the NHS cohort by Hart et al. (2015) reviewed here reported a weak association with mortality (HR = 1.13 for nonaccidental mortality), as did the two other studies of this cohort that we excluded (Puett et al. 2009, who reported an HR of 1.29; Liao et al. (2018), who reported an HR of 1.18, both for all-cause mortality). In addition, the study of female teachers in the CTS cohort by Ostro et al. (2015) reviewed here reported no association (HR = 1.01, 95% CI: 0.98–1.05) with non-accidental mortality, as did the study of this cohort that we excluded (Lipsett et al. 2011; HR = 1.01, 95% CI: 0.95–1.09). The results are highly similar among other studies that we excluded compared to the studies of the same cohort that we included in this review.

There are several key uncertainties related to the available epidemiology evidence for associations between exposure to ambient $PM_{2.5}$ and mortality that are primarily due to potential confounding by copollutants and unmeasured/unknown confounders, exposure measurement error, model misspecification, and a limited understanding of risks related to relatively low $PM_{2.5}$ concentrations. As studies begin to address these key uncertainties more, future studies may be better able than the current literature to improve our understanding of potential causal relationships between $PM_{2.5}$ and mortality or other adverse health effects. Burns et al. (2019b) recently developed a matrix for communicating risk assessment 'asks' of epidemiology research that describes characteristics of epidemiology studies that should be considered when they are used for risk assessment and decision making. These characteristics include confirming exposures and outcomes and determining the direction and magnitude of error surrounding exposure and dose-response assessments, for example. Most of the recent epidemiology studies of $PM_{2.5}$ exposure and mortality do not fully meet these 'asks' of risk assessors or appreciably reduce uncertainty regarding associations between ambient $PM_{2.5}$ concentrations and mortality and, thus, are of limited use for risk assessment; therefore, the 'asks' could be an important tool for consideration in future epidemiology publications to improve their value for use in decision making.

Conclusions

We conducted a review of the epidemiology studies of long-term exposure to ambient $PM_{2.5}$ and mortality using a transparent systematic review framework based on best practices for evaluating study quality and integrating evidence. There is some consistency across studies for reporting positive associations, but these associations are weak and several important methodological issues have led to uncertainties with regard to the evidence from these studies, including potential confounding by measured and unmeasured factors, exposure measurement error, and model misspecification. Because these uncertainties provide a plausible, alternative explanation to causality for the weakly positive findings across studies, we concluded that the evidence for a causal relationship between long-term exposure to ambient $PM_{2.5}$ concentrations and mortality (all-cause or non-accidental) from these studies is inadequate. Our review shows that a relatively consistent pattern of weak, positive associations does not necessarily lead to a conclusion of causality when study quality is incorporated into the evaluation and integration of evidence in a consistent manner and alternative explanations for the evidence are explored. Our conclusion that the evidence for a causal relationship between long-term ambient $PM_{2.5}$ exposure and mortality is inadequate is based on the many study limitations and uncertainties associated with the evidence, and indicates

that the epidemiology studies of $PM_{2.5}$ and mortality should be interpreted with caution, particularly if they are to be used for regulatory decision making.

Acknowledgments

We thank Ms. Carla Walker for her assistance on this manuscript.

Funding

This work was supported by the American Petroleum Institute [No Grant].

Declaration of interest

All of the authors are employees of Gradient, an independent environmental risk science consulting firm. The work reported in this paper was conducted during the normal course of employment, with financial support by the American Petroleum Institute (API). J.E. Goodman and R.L. Prueitt have previously given presentations or testimony on topics related to air pollution at scientific conferences and meetings with regulatory agencies, with funding provided by API. All other authors declare that they have not been involved in any regulatory activities related to the contents of this paper. This manuscript is the professional work product of the authors, and the opinions and conclusions offered within are not necessarily those of their employers or the financial sponsor of the work.

References

Burns CJ, LaKind JS, Mattison DR, Alcala CS, Branch F, Castillo J, Clark A, Clougherty JE, Darney SP, Erickson H, et al. 2019b. A matrix for bridging the epidemiology and risk assessment gap. Glob Epidemiol. 1:100005. doi:10.1016/j.gloepi.2019.100005

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, Van Erp AM, Turley R, Rehfuess E. 2019a. Interventions to reduce ambient air pollution and their effect on health. Cochrane Database Syst Rev. 5(5):CD010919. doi:10.1002/14651858.CD010919.pub2.

Chen H, Burnett RT, Copes R, Kwong JC, Villeneuve PJ, Goldberg MS, Brook RD, Van Donkelaar A, Jerrett M, Martin RV, et al. 2016. Ambient fine particulate matter and mortality among survivors of myocardial infarction: population-based cohort study. Environ Health Perspect. 124(9):1421–1428. doi:10.1289/EHP185.

Chowdhury R, Shah D, Payal AR. 2017. Healthy worker effect phenomenon: revisited with emphasis on statistical methods - A review. Indian J Occup Environ Med. 21(1):2–8. doi:10.4103/ijoem.ijoem_53_16.

Clougherty JE, Kubzansky LD. 2009. A framework for examining social stress and susceptibility to air pollution in respiratory health. Environ Health Perspect. 117(9):1351–1358. doi:10.1289/ehp.0900612.

Cox LA Jr. 2017. Do causal concentration-response functions exist? A critical review of associational and causal relations between fine particulate matter and mortality. Crit Rev Toxicol. 47(7):609–637. doi:10.1080/10408444.2017.1311838.

Cox LA Jr., Popken DA. 2015. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? Ann Epidemiol. 25(3):162–173. doi:10.1016/j.annepidem.2014.11.006.

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

Crouse DL, Peters PA, Van Donkelaar A, Goldberg MS, Villeneuve PJ, Brion O, Khan S, Atari DO, Jerrett M, Pope CA, et al. 2012. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. Environ Health Perspect. 120(5):708–714. doi:10.1289/ehp.1104049.

Crouse DL, Philip S, Van Donkelaar A, Martin RV, Jessiman B, Peters PA, Weichenthal S, Brook JR, Hubbell B, Burnett RT. 2016. A new method to jointly estimate the mortality risk of long-term exposure to fine particulate matter and its components. Int J Environ Health Re. 26(2):145–157. doi:10.1038/srep18916.

Crouse DL, Pinault L, Balram A, Brauer M, Burnett RT, Martin RV, Van Donkelaar A, Villeneuve PJ, Weichenthal S. 2019. Complex relationships between greenness, air pollution, and mortality in a population-based Canadian cohort. Environ Int. 128:292–300. doi:10.1016/j.envint.2019.04.047

Deng H, Eckel SP, Liu L, Lurmann FW, Cockburn MG, Gilliland FD. 2017. Particulate matter air pollution and liver cancer survival. Int J Cancer. 141(4):744–749. doi:10.1002/ijc.30779.

Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Choirat C, Dominici F, Schwartz JD. 2017. Air pollution and mortality in the Medicare population. N Engl J Med. 376(26):2513–2522. doi:10.1056/NEJMoa1702747.
Diver WR, Jacobs EJ, Gapstur SM. 2018. Secondhand smoke exposure in childhood and adulthood in relation to adult mortality among never smokers. Am J Prev Med. 55(3):345–352. doi:10.1016/j.amepre.2018.05.005.
DuPre NC, Hart JE, Holmes MD, Poole EM, James P, Kraft P, Laden F, Tamimi RM. 2019. Particulate matter and traffic-related exposures in relation to breast cancer survival. Cancer Epidemiol Biomarkers Prev. 28(4):751–759. doi:10.1158/1055-9965.EPI-18-0803.
Gamble JF. 1998. PM2.5 and mortality in long-term prospective cohort studies: cause-effect or statistical association? Environ Health Perspect. 106(9):535–549. doi:10.1289/ehp.98106535.
Garcia CA, Yap PS, Park HY, Weller BL. 2016. Association of long-term PM2.5 exposure with mortality using different air pollution exposure models: impacts in rural and urban California. Int J Environ Health Res. 26 (2):145–157. doi:10.1080/09603123.2015.1061113.
Goodman JE, Prueitt RL, Harbison RD, Johnson GT. 2020. Systematically evaluating and integrating evidence in National Ambient Air Quality Standards (NAAQS) reviews. Glob Epidemiol. 2:1000019. https://www.sciencedirect.com/science/article/pii/S2590113320300031
Green LC, Crouch EAC, Ames MR, Lash TL. 2002. What's wrong with the national ambient air quality standard (NAAQS) for fine particulate matter (PM2.5)? Regul Toxicol Pharmacol. 35:327–337. doi:10.1006/rtph.2002.1548.
Greven S, Dominici F, Zeger S. 2011. An approach to the estimation of chronic air pollution effects using spatio-temporal information. J Am Stat Assoc. 106(494):396–406. doi:10.1198/jasa.2011.ap09392.
Hart JE, Garshick E, Dockery DW, Smith TJ, Ryan L, Laden F. 2011. Long-term ambient multipollutant exposures and mortality. Am J Respir Crit Care Med. 183(1):73–78. doi:10.1164/rccm.200912-1903OC.
Hart JE, 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(1):38. doi: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(8):e003947. doi:10.1161/JAHA.116.003947.
Henneman LRF, Liu C, Mulholland JA, Russell AG. 2017. Evaluating the effectiveness of air quality regulations: a review of accountability studies and frameworks. J Air Waste Manag Assoc. 67(2):144–172. doi:10.1080/10962247.2016.1242518.
[IOM] Institute of Medicine, 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. Washington (DC):National Academies Press.
Hill AB. 1965. The environment and disease: association or causation? Proc R Soc Med. 58(5):295–300.
Jerrett M, Burnett RT, Pope CA, Ito K, Thurston G, Krewski D, Shi Y, Calle E, Thun M. 2009. Long-term ozone exposure and mortality. N Engl J Med. 360(11):1085–1095. doi:10.1056/NEJMoa0803894.
Kioumourtzoglou MA, Austin E, Koutrakis P, Dominici F, Schwartz J, Zanobetti A. 2015. PM2.5 and survival among older adults: effect modification by particulate composition. Epidemiology. 26(3):321–327. doi:10.1097/EDE.0000000000000269.
Kioumourtzoglou MA, Schwartz J, James P, Dominici F, Zanobetti A. 2016. PM2.5 and mortality in 207 US cities: modification by temperature and city characteristics. Epidemiology. 27(2):221–227. doi:10.1097/EDE.0000000000000422.
Lefler JS, Higbee JD, Burnett RT, Ezzati M, Coleman NC, Mann DD, Marshall JD, Bechle M, Wang Y, Robinson AL, et al. 2019. Air pollution and mortality in a large, representative U.S. cohort: multiple-pollutant analyses, and spatial and temporal decompositions. Environ Health. 18(1):101. doi:10.1186/s12940-019-0544-9.
Lepeule J, Laden F, Dockery D, Schwartz J. 2012. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009. Environ Health Perspect. 120(7):965–970. doi:10.1289/ehp.1104660.
Li CY, Sung FC. 1999. A review of the healthy worker effect in occupational epidemiology. Occup Med. 49 (4):225–229. doi:10.1093/occmed/49.4.225.
Li R, Zhou R, Zhang J. 2018. Function of PM2.5 in the pathogenesis of lung cancer and chronic airway inflammatory diseases. Oncol Lett. 15(5):7506–7514. doi:10.3892/ol.2018.8355.
Liao X, Zhou X, Wang M, Hart JE, Laden F, Spiegelman D. 2018. Survival analysis with functions of mismeasured covariate histories: the case of chronic air pollution exposure in relation to mortality in the nurses' health study. J R Stat Soc Ser C Appl Stat. 67(2):307–327. doi:10.1111/rssc.12229.
Lipfert FW, Wyzga RE. 2019. Environmental predictors of survival in a cohort of U.S. military veterans: a multi-level spatio-temporal analysis stratified by race. Environ Res. 183. doi:10.1016/j.envres.2019.108842.
Lipsett MJ, Ostro BD, Reynolds P, Goldberg D, Hertz A, Jerrett M, Smith DF, Garcia C, Chang ET, Bernstein L. 2011. Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort. Am J Respir Crit Care Med. 184(7):828–835.

Lv X, Sun J, Bi Y, Xu M, Lu J, Zhao L, Xu Y. 2015. Risk of all-cause mortality and cardiovascular disease associated with secondhand smoke exposure: a systematic review and meta-analysis. Int J Cardiol. 199:106–115. doi:10.1016/j.ijcard.2015.07.011

Makar M, Antonelli J, Di Q, Cutler D, Schwartz J, Dominici F. 2017. Estimating the causal effect of low levels of fine particulate matter on hospitalization. Epidemiol. 28(5):627–634. doi:10.1097/EDE.0000000000000690.

Malik AO, Jones PG, Chan PS, Peri-Okonny PA, Hejjaji V, Spertus JA. 2019. Association of long-term exposure to particulate matter and ozone with health status and mortality in patients after myocardial infarction. Circ Cardiovasc Qual Outcomes. 12(4):e005598. doi:10.1161/CIRCOUTCOMES.119.005598.

Miller MR. 2014. The role of oxidative stress in the cardiovascular actions of particulate air pollution. Biochem Soc Trans. 42(4):1006–1011. doi:10.1042/BST20140090.

Miller MR. 2020. Oxidative stress and the cardiovascular effects of air pollution. Free Radic Biol Med. 151:69–87. doi:10.1016/j.freeradbiomed.2020.01.004

Ostro B, Hu J, Goldberg D, Reynolds P, Hertz A, Bernstein L, Kleeman MJ. 2015. Associations of mortality with long-term exposures to fine and ultrafine particles, species and sources: results from the California Teachers Study Cohort. Environ Health Perspect. 123(6):549–556. doi:10.1289/ehp.1408565.

Ostro B, Lipsett M, Reynolds P, Goldberg D, Hertz A, Garcia C, Henderson KD, Bernstein L. 2010. Long-term exposure to constituents of fine particulate air pollution and mortality: results from the California Teachers Study. Environ Health Perspect. 118(3):363–369. doi:10.1289/ehp.0901181.

Pelkonen MK, Laatikainen TK, Jousilahti P. 2019. The relation of environmental tobacco smoke (ETS) to chronic bronchitis and mortality over two decades. Respir Med. 154:34–39. doi:10.1016/j.rmed.2019.06.006

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

Pinault LL, Weichenthal S, Crouse DL, Brauer M, Erickson A, Donkelaar AV, Martin RV, Hystad P, Chen H, Fines P, et al. 2017. Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort. Environ Res. 159:406-415. doi:10.1016/j.envres.2017.08.037

Pope CA III, Ezzati M, Cannon JB, Allen RT, Jerrett M, Burnett RT. 2017. Mortality risk and PM2.5 air pollution in the USA: an analysis of a national prospective cohort. Air Qual Atmos Health. 11(3):245–252. doi:10.1007/s11869-017-0535-3.

Prueitt RL, Cohen JM, Goodman JE. 2015. Evaluation of atherosclerosis as a potential mode of action for cardiovascular effects of particulate matter. Regul Toxicol Pharmacol. 73(Suppl(2)):S1–S15. doi:10.1016/j.yrtph.2015.09.034.

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(8):1130–1135. doi:10.1289/ehp.1002921.

Puett RC, Hart JE, Yanosky JD, Paciorek C, Schwartz J, Suh H, Speizer FE, Laden F. 2009. Chronic fine and coarse particulate exposure, mortality, and coronary heart disease in the Nurses' Health Study. Environ Health Perspect. 117(11):1697–1701. doi:10.1289/ehp.0900572.

Pun VC, Kazemiparkouhi F, Manjourides J, Suh HH. 2017. Long-term PM2.5 exposure and respiratory, cancer, and cardiovascular mortality in older US adults. Am J Epidemiol. 186(8):961–969. doi:10.1093/aje/kwx166.

Rhomberg LR, Chandalia JK, Long CM, Goodman JE. 2011. Measurement error in environmental epidemiology and the shape of exposure-response curves. Crit Rev Toxicol. 41(8):651–671. doi:10.3109/10408444.2011.563420.

Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, Honeycutt M, Kaminski NE, Paoli G, Pottenger LH, et al. 2013. A survey of frameworks for best practices in weight-of-evidence analyses. Crit Rev Toxicol. 43(9):753–784. doi:10.3109/10408444.2013.832727.

Rojas-Rueda D, Morales-Zamora E, Alsufyani WA, Herbst CH, AlBalawi SM, Alsukait R, Alomran M. 2021. Environmental risk factors and health: an umbrella review of meta-analyses. Int J Environ Res Public Health. 18(2):704. doi:10.3390/ijerph18020704.

Ruprecht AA, De Marco C, Pozzi P, Mazza R, Munarini E, Di Paco A, Paredi P, Invernizzi G, Boffi R. 2016. Outdoor second-hand cigarette smoke significantly affects air quality (Letter). Eur Respir J. 48(3):918–920. doi:10.1183/13993003.00064-2016.

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

Smith AE, Chang W 2020. Exposure window uncertainty in chronic effects associations of PM2.5 and health risks. Report to NR3 Coalition. Submitted to US EPA Docket. EPA-HQ-OAR-2015-0072.

Smith AE, Gans W. 2015. Enhancing the characterization of epistemic uncertainties in PM2.5 risk analyses. Risk Anal. 35(3):361–378. doi:10.1111/risa.12236.

Stansfeld SA. 2015. Noise effects on health in the context of air pollution exposure. Int J Environ Res Public Health. 12(10):12735–12760. doi:10.3390/ijerph121012735.

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

Turner MC, Cohen A, Burnett RT, Jerrett M, Diver WR, Gapstur SM, Krewski D, Samet JM, Pope CA III. 2017. Interactions between cigarette smoking and ambient PM2.5 for cardiovascular mortality. Environ Res. 154:304–310. doi:10.1016/j.envres.2017.01.024

[US EPA] United States Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment (NCEA) - RTP Division. 2009. Integrated science assessment for particulate matter (Final). EPA/600/R-08/139F.

[US EPA] United States Environmental Protection Agency. 2013. National ambient air quality standards for particulate matter (Final rule). Fed Reg. 78(10):3086–3287. 40 CFR 50, 51, 52, 53 and 58.

[US EPA] United States Environmental Protection Agency. 2015. Preamble to the integrated science assessments. EPA/600/R-15/067.

[US EPA] United States Environmental Protection Agency. 2019. Integrated science assessment for particulate matter (Final). EPA/600/R-19/188.

[US EPA] United States Environmental Protection Agency, Office of Air Quality Planning and Standards, Health and Environmental Impacts Division. 2020. Policy assessment for the review of the National Ambient Air Quality Standards for particulate matter (Final). EPA-452/R-20-002.

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

Van Deusen A, Hyland A, Travers MJ, Wang C, Higbee C, King BA, Alford T, Cummings KM. 2009. Secondhand smoke and particulate matter exposure in the home. Nicotine Tob Res. 11:635–641. doi:10.1093/ntr/ntp018.

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

Wang Y, Kloog I, Coull BA, Kosheleva A, Zanobetti A, Schwartz JD. 2016. Estimating causal effects of long-term PM2.5 exposure on mortality in New Jersey. Environ Health Perspect. 124(8):1182–1188. doi:10.1289/ehp.1409671.

Wang Y, Lee M, Liu P, Shi L, Yu Z, Abu Awad Y, Zanobetti A, Schwartz JD. 2017b. Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. Epidemiology. 28(6):771–779. doi:10.1097/EDE.0000000000000742.

Wang Y, Shi L, Lee M, Liu P, Di Q, Zanobetti A, Schwartz JD. 2017a. Long-term exposure to PM2.5 and mortality among older adults in the Southeastern US. Epidemiology. 28(2):207–214. doi:10.1097/EDE.0000000000000614.

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

Weichenthal S, Pinault LL, Burnett RT. 2017. Impact of oxidant gases on the relationship between outdoor fine particulate air pollution and nonaccidental, cardiovascular, and respiratory mortality. Sci Rep. 7(1):16401. doi:10.1038/s41598-017-16770-y.

[WHO] World Health Organization. 2006. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Global update 2005: Summary of risk assessment. WHO/SDE/PHE/OEH/06.02.

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

Wu X, Braun D, Kioumourtzoglou MA, Choirat C, Di Q, Dominici F. 2019. Causal inference in the context of an error prone exposure: air pollution and mortality. Ann Appl Stat. 13(1):520–547. doi:10.1214/18-AOAS1206.

Xing YF, Xu YH, Shi MH, Lian YX. 2016. The impact of PM2.5 on the human respiratory system. J Thorac Dis. 8(1): E69–E74. doi:10.3978/j..2072-1439.2016.01.19.

Yang JW, Shen YC, Lin KC, Cheng SJ, Chen SL, Chen CY, Kumar PV, Lin SF, Lu HE, Chen GY. 2020. Organ-on -a-chip: opportunities for assessing the toxicity of particulate matter. Front Bioeng Biotechnol. 8:519. doi:10.3389/fbioe.2020.00519

Yitshak-Sade M, Kloog I, Zanobetti A, Schwartz JD. 2019. Estimating the causal effect of annual PM2.5 exposure on mortality rates in the Northeastern and mid-Atlantic states. Environ Epidemiol. 3(4):e052. doi:10.1097/EE9.0000000000000052.

Yu W, Guo Y, Shi L, Li S. 2020. The association between long-term exposure to low-level PM2.5 and mortality in the state of Queensland, Australia: a modelling study with the difference-in-differences approach. PLoS Med. 17(6): e1003141. doi:10.1371/journal.pmed.1003141.

Supplemental Material

Table S1. Modified Bradford Hill Aspects for Use in Evidence Integration

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

Note:
The aspect descriptions are modified from those presented by US EPA (2015).

Table S2. Framework for Reaching a Causal Conclusion

Conclusion	**Considerations for Reaching Conclusion**
Causal	All modified Bradford Hill aspects are met, or most are met and there is a likely explanation for each that is not met
Suggestive	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
Inadequate	Most or all of the modified Bradford Hill aspects have inadequate information or are not met and there is no likely explanation for each that is not met
Not Causal	Evidence indicates no causal relationship based on modified Bradford Hill aspects not being met and there is no likely explanation for not being met

Figure S1. Literature search and study selection flowchart. PM ISA = Particulate Matter Integrated Science Assessment.