Gradient Comments on Draft PM Policy Assessment

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

October 15, 2019

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

Table of Contents

Page

Executive Summary ES-1

1 Introduction 1

2 General approach (3.1.2) 2
- 2.1 Causal and likely causal determinations 2
- 2.2 The air quality distributions associated with mortality and morbidity have critical uncertainties. 2
- 2.3 Pseudo-design values increase the likelihood of study areas appearing to meet the current $PM_{2.5}$ standard. 5
- 2.4 Controlled human exposure studies are not informative regarding biological plausibility. 6

3 Evidence-based considerations – Mortality (3.2.1.1) 7
- 3.1 Long-term exposure studies do not reduce uncertainty. 7
  - 3.1.1 Cohort studies do not establish a linear no-threshold relationship between long-term $PM_{2.5}$ and total mortality. 7
  - 3.1.2 The draft PA does not adequately review accountability studies. 9
  - 3.1.3 Study quality and relevance are not considered in biological plausibility assessment. 9
- 3.2 Short-term exposure studies do not reduce uncertainty. 10
  - 3.2.1 Multi-city studies do not establish a linear no-threshold relationship between short-term $PM_{2.5}$ and total mortality. 10
  - 3.2.2 The draft PA does not adequately review causal inference/quasi-experimental studies. 11

4 Evidence-based considerations – Cardiovascular effects (3.2.1.2) 12
- 4.1 Long-term exposure 12
  - 4.1.1 Cohort studies do not establish a linear no-threshold relationship between long-term $PM_{2.5}$ and total CV-related mortality. 12
  - 4.1.2 Experimental studies are not sufficiently informative regarding biological plausibility. 12
- 4.2 New studies of short-term exposure do not reduce uncertainty regarding CV effects. 13

5 Evidence-based considerations – Cancer (3.2.1.4) 14
- 5.1 Epidemiology studies do not adequately consider latency, smoking, and family history of lung cancer. 14
- 5.2 Experimental studies do not provide evidence for biological plausibility. 14

6 Key studies (3.2.3) 15
6.1 Experimental studies do not reduce uncertainty. (3.2.3.1) 15
6.1.1 Controlled exposure studies not provide evidence of biological plausibility. 15
6.1.2 As stated in the draft PA, "animal toxicology studies are of limited utility" in informing the adequacy of the standard. 15
6.2 Epidemiology studies do not provide sufficient support for elevated health risks below the current standard. (3.2.3.2) 15
6.2.1 $PM_{2.5}$ air quality distributions are not evaluated systematically or comprehensively. 15
6.2.2 Pseudo-design value analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current standard. (3.2.3.2.2) 15
6.3 Preliminary conclusions from the evidence are overstated. (3.2.3.3) 16

7 Risk-based considerations (3.3) 17
7.1 Study area selection is questionable. (3.3.1) 17
7.2 Causal classifications are not supported for selected health outcomes. (3.3.1). 18
7.3 Key epidemiology studies underlying the concentration-response functions have several methodological limitations. (3.3.1) 18
7.3.1 Long-term mortality studies have substantial biases and uncertainties.. 19
7.3.2 Short-term mortality studies have substantial biases and uncertainties. 20
7.4 $PM_{2.5}$ air quality scenarios evaluated are an insufficient basis for a risk assessment. (3.3.1) 22
7.5 Risk assessment has significant variability and uncertainty. (3.3.1, 3.3.2)..... 24

8 Preliminary conclusions on the current standard (3.4.1) 28
8.1 Studies published since the last review do not reduce uncertainty. 28
8.2 Recent US and Canadian studies do not provide support for effects at concentrations lower than in the last review and are likely to be allowed by the current primary $PM_{2.5}$ standards. 28
8.3 Pseudo-design values do not support health effects at air quality levels likely to have met current standards. 29
8.4 The risk assessment estimates do not account for uncertainty. 29
8.5 Experimental studies do not provide support for lower standards. 29

9 Preliminary conclusions on alternative standards (3.4.2) 30
9.1 Long-term mean $PM_{2.5}$ concentrations in key epidemiology studies do not support a lower standard. 30
9.2 Long-term mean $PM_{2.5}$ concentrations and annual design values over-predict health impacts..... 30
9.3 $PM_{2.5}$ pseudo-design values in many study locations are above the standard.... 30
9.4 Controlled exposure studies are not informative..... 30
9.5 $PM_{2.5}$-associated risk estimates are not meaningful. 31

10 Alternative annual standard levels (3.4.2.4.1) 32

11 Alternative 24-hour standard levels (3.4.2.4.2) 33

12 Future research (3.5) 34

References 35

Executive Summary

The "Policy Assessment for the Review of the National Ambient Air Quality Standards for Particulate Matter, External Review Draft" (hereafter, the draft PA) addresses whether available scientific evidence and risk-based information support or call into question the adequacy of the public health protection afforded by the current annual and 24-hour fine particulate matter ($PM_{2.5}$) standards, and if so, what alternative standards are supported. As discussed below, new evidence and risk assessment results do not call into question the adequacy of the standard.

The 2009 and 2018 draft Integrated Science Assessments (ISAs) concluded that there are causal associations between long-term and short-term $PM_{2.5}$ exposure and mortality and cardiovascular effects, and the 2018 draft ISA concluded there is a "likely to be causal" association between long-term exposure and cancer (US EPA, 2009, 2018). However, the quality and results of studies in the 2018 draft ISA were not evaluated objectively using a detailed systematic review protocol. The new epidemiology studies have similar designs, and, hence, similar key methodological limitations. As such, new studies do not appreciably reduce uncertainty regarding these associations, particularly at exposure concentrations below the current standards. In addition, these studies do not provide evidence for non-threshold concentration-response functions, particularly in light of the lack of biological plausibility for this to be the case.

The draft PA indicates that new studies may support health effects at exposure concentrations below the current primary standards based on the mean $PM_{2.5}$ concentrations in key epidemiology studies and pseudo-design values calculated by the US Environmental Protection Agency (EPA). The mean $PM_{2.5}$ concentrations are impacted by potential exposure measurement error and the lack of near-road monitors. Exposure measurement error can be large as a result of irrelevant exposure measurement windows, unaccounted residential mobility or temporal variation, and poor prediction model performance. Near-road monitors tend to capture higher $PM_{2.5}$ concentrations than those in surrounding areas, so study-reported mean $PM_{2.5}$ concentrations would have been higher had near-road monitors been in place. Also, the draft PA focuses on the center of the $PM_{2.5}$ distributions, but associations observed at the center of the data are likely driven by the upper portion of the air quality distribution, assuming risk increases with exposure.

While pseudo-design values are more similar to design values than mean $PM_{2.5}$ concentrations, the pseudo-design values in the draft PA likely overestimated the extent to which study areas met the current standards because of the way they were calculated and the lack of near-road monitors. Even so, the observed distributions of pseudo-design values indicate that a large proportion of the study populations in the key epidemiology studies were likely in areas that did not meet the current standard. Results of pseudo-design analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current $PM_{2.5}$ standard.

The draft PA claims that the risk assessment shows substantial decreases in risk at lower standard levels (*i.e.*, 8, 9, 10, and 11 $\mu g/m^3$), but this does not account for the unquantified, high level of uncertainty in the modeling calculations. For example, the risk assessment was performed for a small fraction of the US that does not necessarily correlate with the areas in epidemiology studies from which concentration-response functions were derived. The air quality scenarios evaluated use unrealistic emissions scenarios and unvalidated linear extrapolation (for 9 $\mu g/m^3$) and interpolation (for 11 $\mu g/m^3$) from two data points. The risk assessment did not evaluate 8 $\mu g/m^3$ at all, but the draft PA discusses it as a possible alternative. Furthermore, uncertainty in the modeled $PM_{2.5}$ concentrations, unconventional rounding

choices, and compounding levels of conservativism in the modeling preclude drawing any distinctions between proposed alternative standards. Finally, and critically, the draft PA does not conduct a robust uncertainty analysis, and this effectively renders meaningless the comparisons between the risk assessment results.

The draft PA indicates that a conclusion that the current standards provide adequate public health protection would require that little weight be placed on epidemiology evidence, and greater weight on uncertainties and limitations in the evidence and analyses. It indicates that these uncertainties include the fact that studies that examine biological pathways do so at exposures well above those likely to occur in areas meeting the current standards, that accountability studies only evaluated exposures starting above the current standard, and that uncertainty in the risk assessment results from uncertainties in the underlying epidemiology studies, in the air quality adjustments, and in the application of study and air quality information to develop quantitative estimates of $PM_{2.5}$-associated mortality risk.

It is unclear how acknowledging these uncertainties places little weight on the epidemiology evidence, when all of these factors greatly impact the epidemiology evidence. Taken together, currently available scientific evidence and risk-based information does not call into question the adequacy of the public health protection afforded by the current annual and 24-hour $PM_{2.5}$ standards or indicate that alternative standards will increase public health protection.

1 Introduction

The "Policy Assessment for the Review of the National Ambient Air Quality Standards [NAAQS] for Particulate Matter, External Review Draft" (hereafter, the draft PA) addresses whether available scientific evidence and risk-based information support or call into question the adequacy of the public health protection afforded by the current annual and 24-hour fine particulate matter ($PM_{2.5}$) standards and if so, what alternative standards are supported (US EPA, 2019a). The draft PA (US EPA, 2019a) specifically poses the following questions:

- Does the currently available scientific evidence and risk-based information support or call into question the adequacy of the public health protection afforded by the current annual and 24-hour $PM_{2.5}$ standards?
- What is the range of potential alternative standards that could be supported by the available scientific evidence and risk-based information to increase public health protection against short- and long-term fine particle exposures?
- For primary $PM_{2.5}$ standards defined in terms of the current averaging times and forms, what potential alternative levels are appropriate to consider in order to increase public health protection against long- and short-term exposures to $PM_{2.5}$ in ambient air?

The draft PA states that a conclusion that the current standards provide adequate health protection "would place little weight on the broad body of epidemiologic evidence reporting generally positive and statistically significant health effect associations, particularly for $PM_{2.5}$ air quality distributions likely to have been allowed by current primary standards, or on the $PM_{2.5}$ risk assessment." Rather, it would place greater weight on uncertainties and limitations in the evidence and analyses.

The draft PA indicates that these uncertainties include the fact that studies that examine biological pathways do so at exposures well above those likely to occur in areas meeting the current standards, that accountability studies only evaluated exposures starting above the current standard, and that uncertainty in the risk assessment results from uncertainties in the underlying epidemiology studies, in the air quality adjustments, and in the application of study and air quality information to develop quantitative estimates of $PM_{2.5}$--associated mortality risk.

In fact, the conclusion that the current standards provide adequate health protection does not place "little weight" on epidemiology evidence. On the contrary, it is primarily based on epidemiology evidence, and the risk assessment based on this evidence, that this conclusion is reached. As discussed below, currently available scientific evidence and risk-based information does not call into question the adequacy of the public health protection afforded by the current annual and 24-hour $PM_{2.5}$ standards or indicate that alternative standards will increase public health protection.

2 General approach (3.1.2)

2.1 Causal and likely causal determinations

The US Environmental Protection Agency (EPA) developed a framework for conducting systematic reviews and making causal determinations, which it described in the 2018 draft Integrated Science Assessment (ISA) and related documents. However, as discussed in more detail in our comments (Gradient, 2018) and the Clean Air Scientific Advisory Committee (CASAC) letter regarding the 2018 draft ISA (US EPA, 2019b), this framework has several limitations related to the review protocol itself, the evaluation of study quality and relevance, and the causality framework.

One issue is that the protocol lacks sufficient detail. This led to an evaluation that was not conducted in a systematic, unbiased, or transparent manner. The protocol should have included well-developed methods for the literature search strategy; study inclusion and exclusion criteria; a process for data extraction and quality control; specific, prescriptive criteria for evaluating study quality; methods for data analyses; and PM-specific methods for evidence integration and causality determinations.

In addition, the framework should have adequately addressed study quality. While the 2018 draft ISA has a list of important study quality aspects for evaluating health effects in Appendix A, it is not complete or sufficiently detailed to allow for a consistent evaluation of individual study quality. Also, only high-quality studies should have been considered key studies (*i.e.*, given the most weight in analyses), and the quality of all studies, including new and previously evaluated studies, should have been considered for causal determinations. Quality criteria should have been provided for *in vitro* studies, as these studies informed causal conclusions. Importantly, while the 2018 draft ISA listed detailed quality aspects, it did not review and judge individual study quality in a consistent manner across studies. Finally, the 2018 draft ISA used a five-level causal framework for making causal determinations. This framework is structured in such a way that is biased towards causal conclusions.

As a result of these issues, some relevant literature was excluded, and the included literature was not evaluated in a consistent, objective manner. This review of the evidence resulted in biased conclusions regarding causation.

2.2 The air quality distributions associated with mortality and morbidity have critical uncertainties.

In considering the public health protection provided by the current $PM_{2.5}$ standard, the draft PA evaluated the $PM_{2.5}$ air quality distributions associated with mortality or morbidity in key epidemiology studies, with a focus on the center of the distributions. The justification in the draft PA is that the most confidence in the reported magnitude of exposure-response associations occurs around the center of the distribution because this corresponds to the bulk of the underlying data (as indicated by narrow confidence intervals).

However, statistically, influential points for the exposure-response association tend to be located at the data extremes (*i.e.*, outliers), where data are sparse and each data point is given a disproportionately large weight in least square fitting (Bollen and Jackman, 1985). Considering that the incidence of health outcomes increases with increasing exposure concentrations in cohort studies (as demonstrated by positive

associations in linear models), the observed associations at the center of the data are at least partially driven by the upper portion of the air quality distribution. In other words, setting aside methodological limitations (discussed below), while cohort studies report effects in *study populations* for which the average $PM_{2.5}$ concentrations are below the standard, they do not show effects in *individuals* below the standard. That is, the average $PM_{2.5}$ concentration does not provide any information about the locus of mortality or the effectiveness of the current standard in protecting public health.

It is also notable that the current PA focuses on studies with average $PM_{2.5}$ concentrations of ~10 μg/m^3, but most deaths occurred above this concentration, just as studies that relied on $PM_{2.5}$ data from the 1990s through the early 2000s evaluated studies where most mortality occurred at exposures above 12 μg/m^3. Given the lack of information on the exposures concentrations to which *individuals* were exposed in any of these studies (and other issues discussed below), the modeling in the draft PA has generated highly uncertain results.

To address this issue, the draft PA identified three studies that performed analyses restricted to long-term exposures below 12 μg/m^3 (Di *et al.*, 2017a) or 10 μg/m^3 (Shi *et al.*, 2016) and short-term exposures below 25 μg/m^3 (Di *et al.*, 2017b) or 30 μg/m^3 (Shi *et al.*, 2016). The draft PA indicated that the observed associations in these analyses are not disproportionately driven by the upper portion of the air quality distribution. However, these studies have several critical methodological limitations, as discussed below, and do not necessarily provide evidence for risks around the center of the subgroup data (*e.g.*, 9.6 μg/m^3 in Di *et al*., 2017a).

Furthermore, as stated by Papadogeorgou *et al*. (2019):

> [R]estricting the analysis to a subset of the data has some interpretational limitations. Considering a subgroup of the data effectively changes the population of interest. Specifically, it is likely that the subpopulation exposed to low levels of $PM_{2.5}$ does not have the same characteristics as the full study population. If the distribution of certain modifiers of the association between $PM_{2.5}$ and the outcome of interest is different among participants living in lower exposure levels (e.g., rural vs. urban residence, age, socioeconomic status, etc.) compared to the characteristics in the full population, then the effect estimates from the restricted analysis are not necessarily directly comparable to those of the full analysis.

The draft PA's assessment of the reported $PM_{2.5}$ distributions also does not adequately take into consideration potential exposure measurement error. In fact, the assessment of $PM_{2.5}$ concentrations in key epidemiology studies can be subject to considerable measurement error due to the use of irrelevant exposure measurement windows, unaccounted residential mobility or temporal variation, or poor prediction model performance (as reviewed by Gradient, 2018).

Another important source of exposure measurement error is the placement of $PM_{2.5}$ monitors. As noted in the draft PA (US EPA, 2019a), in response to a key change in the EPA's monitoring requirements, "the addition of $PM_{2.5}$ monitoring at near-road locations was phased in from 2015 to 2017," which is after the study periods covered by the key epidemiology studies. Since near-road monitoring sites tend to capture higher $PM_{2.5}$ concentrations than those in surrounding areas, had the near-road monitors been placed during the study periods, the study-reported mean $PM_{2.5}$ concentrations would have been higher.

In addition, long-term cohort studies of all-cause or nonaccidental mortality often do not assess exposure timing or duration during etiologically relevant periods within individuals' lifetimes. In most studies, ambient exposure is only measured for a few years, often contemporaneously with follow-up, leading to innumerable misalignments between exposure and disease processes that inevitably result in death. In effect, these ambient measurement periods are only a small window within an individuals' lifetimes that are

not matched with the natural history of any particular health condition that leads to death. Because cohort studies evaluate all-cause or nonaccidental mortality, and different causes of death have different etiologies, these different causes of death also have very different relevant exposure windows. Some causes of death are also more likely acute, and not chronic.

For example, the key Medicare studies included in the draft PA (Di *et al*., 2017a; Zanobetti *et al*., 2014) are based on data from this administrative cohort. There is no information on exposure (or many other relevant factors) before entry into the cohort at age 65. Di *et al.* (2017a) matched same-year exposures with the deaths during that year, which does not necessarily conform to any disease-mortality natural history.

Jerrett *et al.* (2016) included at most two years of exposure data for the most recent deaths in 2004. Deaths occurring before 2002 had no relevant exposure history recorded. Even under an untested assumption of a relative correspondence between the magnitude of the 2002-2004 exposures and historical exposures (*e.g.*, an exposure twice as high in 2002-2004 would have been twice as high decades ago), the failure to account for residential mobility during that time has obviously resulted in significant exposure misclassification for an unknown proportion of study participants.

Turner *et al.* (2016) has the natural history of cancer reversed. The latency between the relevant exposure and the clinical onset of lung cancer is generally considered to be 20-30 years. The cancer deaths attributed to $PM_{2.5}$ during the study period were matched to ambient-surrogate exposures in the early 2000s rather than the etiologically relevant exposures in the 1960s and 1970s. More than half of the measured concentrations occurred *after* the final deaths occurred.

Pope *et al.* (2015) used the American Cancer Society (ACS) cohort, with its attendant limitations, to assess long-term effects from $PM_{2.5}$ on cardiometabolic and cardiovascular mortality. Once again, the ambient exposures are chronic for a small proportion of deaths during the study period. Additionally, as in other ACS-based studies, about half of the ambient concentrations were recorded *after* all deaths had occurred.

Given the potential existence of multiple sources of exposure measurement error, it is possible that the observed associations with mortality or morbidity at lower mean $PM_{2.5}$ concentrations simply reflect true associations at higher $PM_{2.5}$ concentrations that were substantially underestimated in the studies. This is particularly important when considering there is very limited evidence regarding health effects at lower mean $PM_{2.5}$ concentrations. As noted in the draft PA, the mean concentrations in most of the studies were at or above 9.6 $\mu g/m^3$, and above 10.7 $\mu g/m^3$ in studies that used monitored $PM_{2.5}$ data alone.

Finally, more than one-third of the key epidemiology studies were excluded from the evaluation of air quality distributions, and the draft PA does not address how this may have impacted the interpretation of its evaluation. The draft PA also did not assess whether the observed associations were consistently stronger in areas with higher mean $PM_{2.5}$ concentrations, which would be expected if a concentration-response relationship truly exists. Therefore, the draft PA's statement (US EPA, 2019a) that the key epidemiology studies indicate statistically significant positive associations "consistently for distributions with long-term mean $PM_{2.5}$ concentrations at or above 8.1 $\mu g/m^3$" is not valid.

2.3 Pseudo-design values increase the likelihood of study areas appearing to meet the current $PM_{2.5}$ standard.

EPA calculates design values to determine whether an area is in compliance with the NAAQS. Design values are calculated at individual monitors meeting appropriate data quality and completeness criteria. As stated in the draft PA (US EPA, 2019a):

> For the annual $PM_{2.5}$ standard, design values are calculated as the annual arithmetic mean $PM_{2.5}$ concentration, averaged over 3 years. For the 24-hour standard, design values are calculated as the 98th percentile of the annual distribution of 24-hour $PM_{2.5}$ concentrations, averaged over three years. For an area to meet the NAAQS, all valid design values in that area, including the highest annual and 24-hour monitored values, must be at or below the levels of the standard.

Because design values are based on the monitor that meets data quality and completeness criteria with the *highest* monitored values, and not the *mean* value of monitors over an area, mean $PM_{2.5}$ concentrations reported in the key epidemiology studies are not directly comparable to the NAAQS, and thus are not as informative regarding whether areas would have met the current standard or not. Thus, the evaluation of air quality distributions discussed above in Section 2.2 does not provide direct evidence as to whether the current standard would have provided sufficient protection in any particular study population.

To address this, the draft PA calculated pseudo-design values, which are similar to $PM_{2.5}$ design values, for the locations and periods evaluated by the key studies. Pseudo-design values in the draft PA were initially calculated in a similar way for individual monitors; then, the highest annual and 24-hour $PM_{2.5}$ pseudo-design values for each 3-year period of the study were identified for each monitored study location (*e.g.*, county) and averaged over the study period. The calculated pseudo-design values are used in the draft PA to evaluate whether study areas would have met the current or alternative standard during the study periods and to what extent, based on percentage of the study area population (or health events) in locations meeting the current standard.

In considering whether the current standard (or an alternative standard) is providing sufficient protection in a population, the draft PA considers whether elevated mortality/morbidity risk is observed in key studies and to what extent areas in that study meet the standard. Based on the evaluation of the air quality distribution (discussed here in Section 2.2), the key epidemiology studies reported elevated risk under mean concentrations below the current standard. Under the entire air quality distributions, however, some study areas did not meet the current standard. The greater the extent to which these study areas met the current standard, the higher the certainty that the key epidemiology studies provide evidence indicating the current standard is not sufficient in protecting the population (not just that there is mortality/morbidity risk under lower $PM_{2.5}$ concentrations). Therefore, the percentage of study area population (or health events) in locations meeting the current standard plays a key role in determining the certainty of evidence provided by the key studies regarding the sufficiency of protection provided by the current standard.

However, the pseudo-design values calculated in the draft PA likely have overestimated the percentage of study area population (or health events) in locations meeting the current standard (*i.e.*, the extent to which study areas met the current standard). First, the highest annual and 24-hour pseudo-design values for each 3-year period were averaged over the entire study period for each study area. This likely masked some sporadic concentrations that exceeded the current standard, thus misclassifying some areas that violated the current standard over the study period as meeting the standard. For example, in Di *et al*. (2017a), of the counties that have study-period average pseudo-design values between the interval of (11.04, 12.04] (*i.e.*, they just meet the current standard), a substantial proportion (39.14%) of individual 3-year pseudo-design

values were above the standard. An estimated 8.97% were above the standard in the interval of (10.04, 11.04]. The underlying proportions of counties for which sporadic violations over the study period may have been masked was likely even larger.

In addition, since near-road monitors were not added until after the study periods evaluated in the key studies, the evaluation of pseudo-design values did not account for near-road $PM_{2.5}$ concentrations. As shown in Table 2-2 of the draft PA, the addition of near-road monitors between 2015 and 2017 captured higher daily and annual $PM_{2.5}$ design values in a large proportion of core-based statistical areas (CBSAs) (*e.g.*, 24% of daily and 52% of annual design values). Had the near-road monitors been placed during the study periods, a decent proportion of pseudo-design values would have been higher. This also contributed to an overestimate of study areas meeting the current $PM_{2.5}$ standard.

According to Figure 3-9 in the draft PA, based on an analysis of pseudo-design values, a portion of study area populations in all US studies included in the draft PA were in areas not meeting the current standard of 12 $\mu g/m^3$. In fact, the proportion of such study area populations was higher than 50% in 20 out of the 27 studies. For the two reasons mentioned above, the actual proportions not meeting the current standard is likely even higher.

In summary, even though the key epidemiology studies show elevated mortality/morbidity risks at lower mean $PM_{2.5}$ concentrations (as indicated in the draft PA's evaluation of air quality distribution), a large proportion of the study populations was likely in areas that did not meet the current standard. Results of pseudo-design analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current $PM_{2.5}$ standard.

2.4 Controlled human exposure studies are not informative regarding biological plausibility.

The draft PA indicates that in the evaluation of controlled human exposure studies, it considered the effects reported, the exposure concentrations/durations, and whether the concentrations evaluated would occur at current or alternative standards (US EPA, 2019a). The draft PA acknowledges that the exposure concentrations evaluated in most of these studies are well above ambient concentrations typically measured in locations meeting the current primary standards. In some sections, the draft PA indicates that these studies are not informative regarding ambient concentrations. However, elsewhere, the draft PA concludes that these studies provide biological plausibility for the occurrence of cardiovascular (CV) effects from short-term exposure at ambient concentrations (US EPA, 2019a).

In truth, several issues with these studies limit their relevance to ambient exposures. Many of the effects measured in controlled exposure studies have threshold modes of action and do not occur at lower concentrations. If the threshold is above ambient concentrations, then these studies do not provide support for these effects at ambient concentrations. Furthermore, it is not clear from the draft PA whether results from controlled human exposure studies reporting CV effects are consistent and coherent across exposure concentrations and outcomes. In light of these limitations, discussed in more detail below, the results from these studies are not relevant to potential CV effects at ambient $PM_{2.5}$ concentrations.

3 Evidence-based considerations – Mortality (3.2.1.1)

The 2009 ISA and 2018 draft ISA concluded that there are causal associations between long-term and short-term $PM_{2.5}$ exposure and mortality (US EPA, 2009, 2019). New studies evaluated in the 2018 draft ISA do not appreciably reduce uncertainty regarding this association and do not provide evidence for health effects below the current primary standards.

3.1 Long-term exposure studies do not reduce uncertainty.

3.1.1 Cohort studies do not establish a linear no-threshold relationship between long-term $PM_{2.5}$ and total mortality.

Based on analyses presented in the 2018 draft ISA, the draft PA states:

> Recent evidence further demonstrates that (1) associations with mortality remain robust in analyses of potential confounding by copollutants (U.S. EPA, 2018, section 11.2.3); (2) associations persist in analyses restricted to long-term exposures below 12 μg/m^3 (Di et al., 2017b) or 10 μg/m^3 (Shi et al., 2016) (i.e., indicating that risks are not disproportionately driven by the upper portions of the air quality distribution); and (3) concentration-response relationships remain linear over the distribution of ambient $PM_{2.5}$ concentrations with no evidence of a threshold, though uncertainty increases near the upper and lower ends of the $PM_{2.5}$ air quality distribution due to limited exposure and outcome data. (US EPA, 2019a)

However, the 2018 draft ISA did not sufficiently consider study quality. Compared to the studies reviewed in the 2009 ISA, recent cohort studies are subject to the same methodological limitations (*e.g.*, exposure measurement error, outcome measurement error, residual confounding, irrelevant exposure windows), and thus do not meaningfully reduce the uncertainty of evidence; this prevents causal inference at exposures below the current NAAQS. Although some recent studies have taken into consideration potential confounding by copollutants, others have not, and the evaluation of copollutants itself is subject to methodological issues, such as the mismatch of the copollutant exposure window and mortality, a lack of accounting for collinearity or a nonlinear relationship with $PM_{2.5}$, and a lack of accounting for temporal variation.

Regarding confounding, three assessments by different researchers (Janes *et al*., 2007; Greven *et al*., 2011; Pun *et al*., 2017) using the Medicare cohort at different time periods have each detected confounding, conferring doubt on the reliability/validity of national-level effect estimates derived from this cohort and effect estimates from like-type cohorts. Each of the three analyses have observed remarkable differences between the temporal (global) effect estimates and the spatiotemporal (local) estimates. In the absence of confounding by variables trending on the national level, these decomposed estimates would be approximately equal. Local effect estimates, not confounded by national trends such as healthcare and economic changes, have shown little to no evidence of an association between $PM_{2.5}$ and mortality. The draft PA makes no reference to this central issue on which EPA's criteria for causal and likely causal determinations are based.

In addition, the Cox Proportional Hazards model used in all cohort studies cannot adequately control for strong time-varying confounding. A recent simulation, based on a realistic cohort of 500,000 adults constructed using the National Cancer Institute Smoking History Generator, indicates that the Cox model poorly controls for a time-dependent strong risk factor (smoking in this simulation), yielding unreliable relative risk estimates unless detailed, time-varying information is incorporated in modeling. None of the studies identified as key in the draft PA incorporated these parameters in their modeling. As a result, the effect estimates from these studies are of questionable reliability given the relative modesty of $PM_{2.5}$ as a risk factor for mortality (Moolgavkar *et al.*, 2018).

It is also not valid to conclude that, as stated in the draft PA (US EPA, 2019a), "risks are not disproportionately driven by the upper portions of the air quality distribution" solely based on subgroup analysis results in only two studies (Di *et al*., 2017a; Shi *et al*., 2016), especially considering the above-mentioned methodological limitations. Furthermore, Rhomberg *et al.* (2011) showed that exposure measurement error can lead to an underestimate of risks at higher exposures and an overestimate of risk at low exposures.

Finally, the 2018 draft ISA does not take a consistent approach in its evaluation of the concentration-response relationships, so conclusions regarding a threshold are not warranted. Exposure measurement error can systematically preclude finding thresholds, even if they exist. The draft PA correctly states that, to date, no threshold below which health effects do not occur has been observed in epidemiology studies. However, measurement errors are pervasive in air pollution observational studies, ranging from instrument imprecision to the practice of serially averaging the measured values over time and space. These errors preclude the ability to detect a threshold, even if one were to exist. Given that such errors introduce a substantial level of uncertainty in determining the true shape of the concentration-response function, the risk assessment based on these curves is of dubious reliability.

As discussed in more detail in Gradient's comments on the 2018 draft ISA (Gradient, 2018), this is particularly the case for Di *et al*. (2017a) and Shi *et al*. (2016).

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

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

covariates, important confounders such as smoking and body mass index were not available for the Medicare cohort.

In light of these methodological limitations, and the lack of biological plausibility for a non-threshold relationship, the concentration-response relationships reported by Shi *et al*. (2016) and Di *et al*. (2017a) are not sufficiently robust in general and do not establish a linear, no-threshold relationship for total mortality.

3.1.2 The draft PA does not adequately review accountability studies.

The draft PA indicates that accountability studies support causation at exposures below the current standard. However, there are many more accountability studies available than the few mentioned in the draft PA, and those mentioned were not critically reviewed in a systematic manner.

Evidence from accountability studies absent in the 2018 draft ISA would have had a material impact on the conclusions. Accountability studies are best suited for evaluating *policy* effectiveness, as opposed to the direct measurement of PM effects on health, *per se*. From the Health Effects Institute alone, 14 out of 15 accountability studies were ommitted from the ozone ISA (Dan Greenbaum letter to EPA, 21 Feb, 2019). One in particular (Zigler *et al.*, 2016) is particularly relevant to the draft PA because it assessed the public health impact/effectiveness of a lower NAAQS. In this instance, the results for all-cause mortality and cardiovascular hospitalizations were uncertain and inconsistent.

These results are practically identical to an earlier nationwide NAAQS accountability study (Chay *et al.*, 2003) that assessed total suspended particulates. The authors found that regulatory compliance with the Total Suspended Particles (TSP) standard was associated with large TSP reductions but had little impact on either adult or elderly mortality. With few exceptions, other more limited pollution reduction actions in various areas also showed mixed results in separate studies. In all, these studies routinely demonstrated that policies and actions lowered either emissions or ambient concentrations (or both) as intended, but the public health impact was unclear.

More recently, a Cochrane Database Systematic Review of various interventions and their public health effectiveness showed mixed results (Burns *et al.*, 2019). After an exhaustive review of the worldwide literature in this area, the investigators were " not able to provide a simple answer regarding 'what works'."

The results from the collection of accountability studies, particularly the Zigler and Chay papers which are especially relevant to the draft PA, suggest that further reduction of the PM NAAQS is unlikely to achieve a measurable improvement in public health.

3.1.3 Study quality and relevance are not considered in biological plausibility assessment.

Coherence occurs when all of the known facts related to the observed association from the various evidence streams fit together in a logical manner (Hill, 1965). The ISA Preamble (US EPA, 2015) indicates coherence refers to the fact that "[a]n inference of causality from one line of evidence (*e.g.*, epidemiologic, controlled human exposure, animal, or ecological studies) may be strengthened by other lines of evidence that support a cause-and-effect interpretation of the association."

The draft PA states that the coherence of animal toxicological, controlled human exposure, and epidemiology studies supports the biological plausibility of $PM_{2.5}$-attributable mortality. It indicates that evidence supports the hypothesis that inhalation exposure of $PM_{2.5}$ could result in mortality from CV and respiratory endpoints and metabolic disease and diabetes (US EPA, 2019a).

The draft PA indicates that the supporting evidence includes epidemiology studies of morbidity effects, which are the largest contributors to total (nonaccidental) mortality (US EPA, 2019a). However, as discussed below in Section 4.1.1, these studies have several methodological limitations. Also, the draft PA does not compare exposure concentrations or magnitudes of effects across morbidity and mortality studies to determine whether the evidence collectively supports coherence across studies.

In addition, the 2018 draft ISA did not sufficiently consider study quality or relevance when evaluating experimental studies (discussed below in Section 4.1.2). Setting this aside, the 2018 draft ISA does not demonstrate a complete pathway for any effect. As noted in the draft PA on page 3-98 (US EPA, 2019a):

> Uncertainty in the biological pathways through which $PM_{2.5}$ exposures could cause serious health effects increases as the ambient concentrations being considered fall farther below the $PM_{2.5}$ exposure concentrations shown to cause effects in experimental studies. In the current review, such studies generally examine the occurrence of $PM_{2.5}$-attributable effects following exposures to $PM_{2.5}$ concentrations well-above those likely to occur in the ambient air in areas meeting the current primary $PM_{2.5}$ standards (i.e., discussed in section 3.2.3.1).

Finally, the draft PA indicates evidence is not sufficient to infer an association between long-term $PM_{2.5}$ exposure and metabolic effects. More specifically, it states that the results from experimental studies did not consistently support associations between long-term $PM_{2.5}$ and direct effects on metabolic function (*e.g.,* impaired blood glucose, insulin resistance, visceral adiposity, and inflammation) (US EPA, 2019a). It is unclear how insufficient evidence regarding metabolic diseases can provide evidence for biological plausibility of mortality.

3.2 Short-term exposure studies do not reduce uncertainty.

3.2.1 Multi-city studies do not establish a linear no-threshold relationship between short-term $PM_{2.5}$ and total mortality.

Referring to the conclusion from the 2018 draft ISA, the draft PA states that the body of evidence "is sufficient to conclude that a causal relationship exists between short-term $PM_{2.5}$ exposure and total mortality" (US EPA, 2019a). However, similar to mortality studies of long-term $PM_{2.5}$ exposure, mortality studies of short-term $PM_{2.5}$ exposure also did not sufficiently consider study quality. Despite some improvements in research methods in recent years, methodological limitations, especially the potential for exposure measurement error and residual confounding, remain. We demonstrated this in our comments on the 2018 draft ISA for the four studies reviewed in the 2018 draft ISA that evaluated the concentration-response relationship between short-term $PM_{2.5}$ exposure and total mortality (Section 11.1.10 of the 2018 draft ISA) (Gradient, 2018). Taken together, recent studies do not significantly reduce the uncertainty of evidence, so causal inference cannot be made at exposures below the current NAAQS.

In addition, as discussed above in Section 2.2, it is not appropriate to conclude that, as stated in the draft PA (US EPA, 2019a), "risks associated with short-term $PM_{2.5}$ exposures are not disproportionately driven by the peaks of the air quality distribution" solely based on subgroup analysis results in only two studies (Di *et al*., 2017b and Shi *et al*., 2016), especially considering the above-mentioned methodological limitations.

3.2.2 The draft PA does not adequately review causal inference/quasi-experimental studies.

The draft PA indicates that accountability studies support causation at exposures below the current standard (US EPA, 2019a). However, there are many more accountability studies available than the few mentioned in the draft PA, and those mentioned were not critically reviewed in a systematic manner.

4 Evidence-based considerations – Cardiovascular effects (3.2.1.2)

The 2009 ISA and 2018 draft ISA concluded that there is a causal association between long-term and short-term $PM_{2.5}$ exposure and CV effects (US EPA, 2009, 2018). New studies evaluated in the 2018 draft ISA do not appreciably reduce uncertainty regarding this association and do not provide evidence for health effects below the current primary standard.

4.1 Long-term exposure

4.1.1 Cohort studies do not establish a linear no-threshold relationship between long-term $PM_{2.5}$ and total CV-related mortality.

The draft PA indicates that the evidence from multiple recent epidemiology studies support the causal relationship between long-term $PM_{2.5}$ and CV-related mortality (US EPA, 2019a). However, studies of CV morbidity and mortality have similar methodological limitations as the total mortality studies, with the most bias and uncertainty in the exposure assessments and confounding adjustment. For example, several studies relied on exposures estimated based on residential zip codes, but these do not account for air pollution exposures that occur outside of the residence (*e.g.*, during the commute, at work, or at school). In some other studies, there was a mismatch between exposures and follow-up periods that could have biased the effect estimates of long-term $PM_{2.5}$ on CV effects. Furthermore, few studies rigorously accounted for confounding at both the individual level (*e.g.*, smoking, physical activity, body mass index) and the community/ecological level (*e.g.*, socioeconomic, access to health care, violence).

In addition, statistical analyses were generally insufficient to establish the shape of concentration-response curves. This is because, as indicated in the 2018 draft ISA, most studies did not conduct a thorough evaluation of alternatives to linearity (US EPA, 2018). Furthermore, similar to studies of all-cause mortality, there is uncertainty in the shape of the concentration-response curve at the lower (and upper) ends of the curves. In light of these limitations, epidemiology studies do not provide strong evidence for any specific concentration-response relationships between long-term $PM_{2.5}$ exposure and CV effects, particularly at lower $PM_{2.5}$ concentrations.

4.1.2 Experimental studies are not sufficiently informative regarding biological plausibility.

The 2018 draft ISA did not sufficiently consider study quality or relevance when evaluating experimental studies. It also did not sufficiently consider coherence (*i.e.*, whether different outcomes across studies fit together in a logical manner and with a specific biological mechanism). In addition, providing support for the possibility of effects is not equivalent to evidence that indicates an effect is likely to occur in humans. The draft PA appears to agree with when it states (US EPA, 2019a):

> Thus, while controlled human exposure studies support the plausibility of the serious cardiovascular effects that have been linked with ambient $PM_{2.5}$ exposures (U.S. EPA, 2018, Chapter 6), the $PM_{2.5}$ exposure concentrations evaluated in most of these studies are well-above the ambient concentrations typically measured in locations meeting the current

> primary standards. Therefore, controlled human exposure studies provide limited insight into the occurrence of cardiovascular effects following $PM_{2.5}$ exposures likely to occur in the ambient air in areas meeting the current primary $PM_{2.5}$ standards and are of limited utility in informing conclusions on the public health protection provided by the current standards.

Furthermore, the draft PA states that controlled human exposure studies found less consistent evidence for CV effects at lower concentrations (US EPA, 2019a). Based on this evidence, it is unclear how any conclusion can be drawn except that controlled exposure studies do not provide sufficient evidence regarding biological plausibility for CV effects at ambient $PM_{2.5}$ concentrations.

4.2 New studies of short-term exposure do not reduce uncertainty regarding CV effects.

The draft PA has not provided any evidence that new epidemiology studies provide additional evidence of a causal association for short-term exposure, as these studies use similar methodologies, and thus have similar limitations, to studies that were available at the time of the 2009 ISA. The draft PA acknowledges that "there are inconsistencies in results across some animal toxicological, controlled human exposure, and epidemiologic panel studies", and indicates this could be due to differences in study design and/or study populations (US EPA, 2019a). This appears to be a *post hoc* explanation for inconsistent and incoherent findings which, according to EPA's causal framework (US EPA, 2015, 2018), should not provide evidence for causation.

5 Evidence-based considerations – Cancer (3.2.1.4)

The 2009 ISA indicated that the evidence was "suggestive of, but not sufficient to infer" a causal relationship between long-term $PM_{2.5}$ and cancer (US EPA, 2009), but the 2018 draft ISA concluded that there is a "likely to be causal" association (US EPA, 2018). As CASAC (US EPA, 2019b) stated:

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

5.1 Epidemiology studies do not adequately consider latency, smoking, and family history of lung cancer.

Referring to the conclusion from the 2018 draft ISA, the draft PA stated that the evidence "is sufficient to conclude that a causal relationship is likely to exist between long-term $PM_{2.5}$ exposure and cancer" (US EPA, 2019a). The draft PA relies heavily on studies of the ACS cohort, and particularly a re-analysis published by Turner *et al*. (2016) to support its position.

As discussed in considerable detail in Section 5 of our 2018 draft ISA comments, the epidemiology studies on which this conclusion is based are undermined by considerable methodological limitations; most critically, they do not, or do not adequately, account for latency, smoking, and family history of lung cancer (Gradient, 2018). Specific issues with the ACS studies are described in Section 7.3.1 below.

5.2 Experimental studies do not provide evidence for biological plausibility.

The draft PA indicates that animal inhalation studies reviewed in the 2009 ISA "found little evidence of tumor formation in response to chronic exposures," and states that evidence from more recent studies support biological plausibility of an effect of $PM_{2.5}$ on cancer (US EPA, 2019a). While the draft PA lists a range of cancer-related effects investigated with $PM_{2.5}$ exposure, there was no indication that more recent studies have found evidence to link $PM_{2.5}$ exposure to tumor formation. The strongest evidence that the draft PA has put forth is that $PM_{2.5}$ can "lead to a range of effects indicative of mutagenicity, genotoxicity, and carcinogenicity, as well as epigenetic effects" (US EPA, 2019a). There is no discussion of the quality, consistency, or coherence of this evidence or whether the doses from the experimental studies are relevant to ambient $PM_{2.5}$ exposures. Even setting this aside, these newer studies do not provide a complete pathway for cancer and thus do not provide support for biological plausibility.

6 Key studies (3.2.3)

6.1 Experimental studies do not reduce uncertainty. (3.2.3.1)

6.1.1 Controlled exposure studies not provide evidence of biological plausibility.

The draft PA only discusses controlled human exposure studies of CV endpoints. As discussed above in Section 4.1.2, and as indicated in the draft PA itself, these studies do not provide evidence regarding biological plausibility because the exposure concentrations are not relevant to ambient $PM_{2.5}$ concentrations. Even setting this aside, it does not appear that the reported outcomes were coherent across studies. The draft PA states that controlled human exposure studies found less consistent evidence for CV effects at lower concentrations (US EPA, 2019a).

6.1.2 As stated in the draft PA, "animal toxicology studies are of limited utility" in informing the adequacy of the standard.

In light of the uncertainties of extrapolating the health effects observed in animals to human populations, we agree with the draft PA that "the animal toxicology studies are of limited utility in informing conclusions on the public health protection provided by the current or alternative primary $PM_{2.5}$ standards" (US EPA, 2019a).

6.2 Epidemiology studies do not provide sufficient support for elevated health risks below the current standard. (3.2.3.2)

6.2.1 $PM_{2.5}$ air quality distributions are not evaluated systematically or comprehensively.

As discussed in Section 2.2, the draft PA evaluated the $PM_{2.5}$ air quality distributions associated with mortality and morbidity in key epidemiology studies but could not rule out that either (1) the observed associations at mean $PM_{2.5}$ concentrations below the current standard are driven by the upper portion of the air quality distribution or that (2) the apparent associations with mortality or morbidity at lower $PM_{2.5}$ concentration were due to substantial underestimation of $PM_{2.5}$ exposure (*i.e.*, exposure measurement error). Although the draft PA reviewed 67 epidemiology studies, it did not do so in a comprehensive or systematic manner. Therefore, the analysis of current epidemiology literature in the draft PA does not provide sufficient support for the existence of elevated mortality/morbidity risk at $PM_{2.5}$ concentrations below the current standard.

6.2.2 Pseudo-design value analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current standard. (3.2.3.2.2)

As discussed above in Section 2.3, the draft PA calculated pseudo-design values to evaluate whether study areas would have met the current or alternative standard during the study periods and to what extent, based

on percentage of the study area population (or health events) in locations meeting the current standard. However, the calculated pseudo-design values likely overestimated the extent to which study areas met the standard, given the way in which pseudo-design values were estimated and the lack of near-road monitors.

Even though the key epidemiology studies show elevated mortality/morbidity risks at lower mean $PM_{2.5}$ concentrations (as indicated in the draft PA's evaluation of air quality distribution), a large proportion of the study populations were in areas that did not meet the current standard. Consequently, results of pseudo-design value analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current $PM_{2.5}$ standard.

6.3 Preliminary conclusions from the evidence are overstated. (3.2.3.3)

We agree with the draft PA conclusion (US EPA, 2019a) that experimental studies "provide limited insight into the occurrence of effects following $PM_{2.5}$ exposures likely to occur in the ambient air in areas meeting the current primary $PM_{2.5}$ standards."

We also agree with the draft PA that the evidence for concentrations < 8 $\mu g/m^3$ is limited (US EPA, 2019a). However, as discussed in Section 2.1, the epidemiology evidence also does not support associations between $PM_{2.5}$ and adverse health outcomes between 8 and 12 $\mu g/m^3$. Results of pseudo-design analyses also indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current $PM_{2.5}$ standard.

7 Risk-based considerations (3.3)

The draft PA seeks to address whether the current $PM_{2.5}$ standards are sufficient (*i.e.*, neither more nor less stringent than necessary) to protect public health and welfare (US EPA, 2019a). A risk assessment was performed to estimate population-level health risks to determine (1) what the estimated health risks are for air quality just meeting the current standards and (2) the extent to which risks are expected to decline when air quality is adjusted to meet potential alternative scenarios (US EPA, 2019a).

The draft PA uses the Community Multiscale Air Quality Modeling (CMAQ) program in conjunction with the Environmental Benefits Mapping and Analysis Program (BenMAP) to estimate health impacts under different long-term $PM_{2.5}$ standards for nine concentration-response functions from eight studies (US EPA, 2019a). Air quality data from 2015 was simulated using CMAQ and compared with measured values (US EPA, 2019a). To simulate air quality "just meeting" a particular standard, modeled emissions were scaled down or up as necessary until air quality was at a particular standard (US EPA, 2019a). Model results are calculated for simulated air quality just meeting standards of 12 $\mu g/m^3$ and 10 $\mu g/m^3$ (US EPA, 2019a). Linear interpolation[1] is used to estimate model results for air quality just meeting a standard of 11 $\mu g/m^3$. Linear extrapolation[2] is used to estimate model results for air quality just meeting a standard of 9 $\mu g/m^3$ (US EPA, 2019a). No other standard levels (*e.g.*, 13 $\mu g/m^3$, 8 $\mu g/m^3$) are evaluated. Results are presented in tables and charts showing predicted mortality rates due to $PM_{2.5}$ exposures (US EPA, 2019a).

As discussed below in Section 7.1, a limited number of study sites were selected, and these are not nationally representative and do not necessarily correlate with epidemiology study data. Sections 7.2 and 7.3 discuss issues with the underlying concentration-response functions used in the modeling. Section 7.4 describes why the scenarios evaluated by the modeling are insufficient to support the conclusions in the risk assessment. Section 7.5 elaborates on issues with the model approach, the insufficiency of model uncertainty quantification, and interpreting the model results. The lack of a quantitative uncertainty analysis in the modeling and risk assessment, both of which contain many sources of uncertainty, effectively renders the results meaningless, since an appropriate quantification of uncertainty could show the ranges of values to be significantly larger or smaller than the ranges presented in the draft PA. These discussions demonstrate that the modeling does not present a reasonable case for lowering the $PM_{2.5}$ standard.

7.1 Study area selection is questionable. (3.3.1)

The draft PA indicates it focuses on areas that represent a variety of regions across the US and includes a substantial portion of the US population for risk modeling (US EPA, 2019a). However, the draft PA also indicates it only includes urban areas, and Figure 3-10 in the draft PA (US EPA, 2019a), reproduced below as Figure 7.1, shows that many regions in the US (*i.e.*, more than half of the States) are not included. The draft PA excludes areas that were more than 2 $\mu g/m^3$ below the current annual standard or 5 $\mu g/m^3$ below the current 24-hour standard from the analysis (US EPA, 2019a). Adjusting the standard lower by 2 $\mu g/m^3$ for the annual standard or 5 $\mu g/m^3$ for the 24-hour standard would have no impact on these excluded areas.

[1] Linear interpolation is a method to determine unknown values between two known values by assuming that intermediate values fall on a straight line between the two known values.

[2] Linear extrapolation is a method to determine unknown values beyond a range of known values by assuming that the unknown values fall on a straight line defined by the known values. Linear extrapolation is more uncertain than linear interpolation, since "as one moves outside the range of the original data, the reliability of the linear approximation as an empirical model of the true relationship will deteriorate" (Montgomery *et al*., 2010, p. 294).

In addition, the draft PA selects key epidemiology studies that cover large areas, sometimes the entire US, to derive concentration-response functions (US EPA, 2019a), and then applies the derived concentration-response functions to the narrower $PM_{2.5}$ study areas shown in Figure 7.1. The selected study areas (where $PM_{2.5}$ data were extracted for the risk assessment) are not necessarily consistent with the areas in the epidemiology studies from which the concentration-response functions were selected. This calls into question whether the concentration-response functions are applicable to the $PM_{2.5}$ concentrations extracted from the particular study areas selected in the risk assessment.

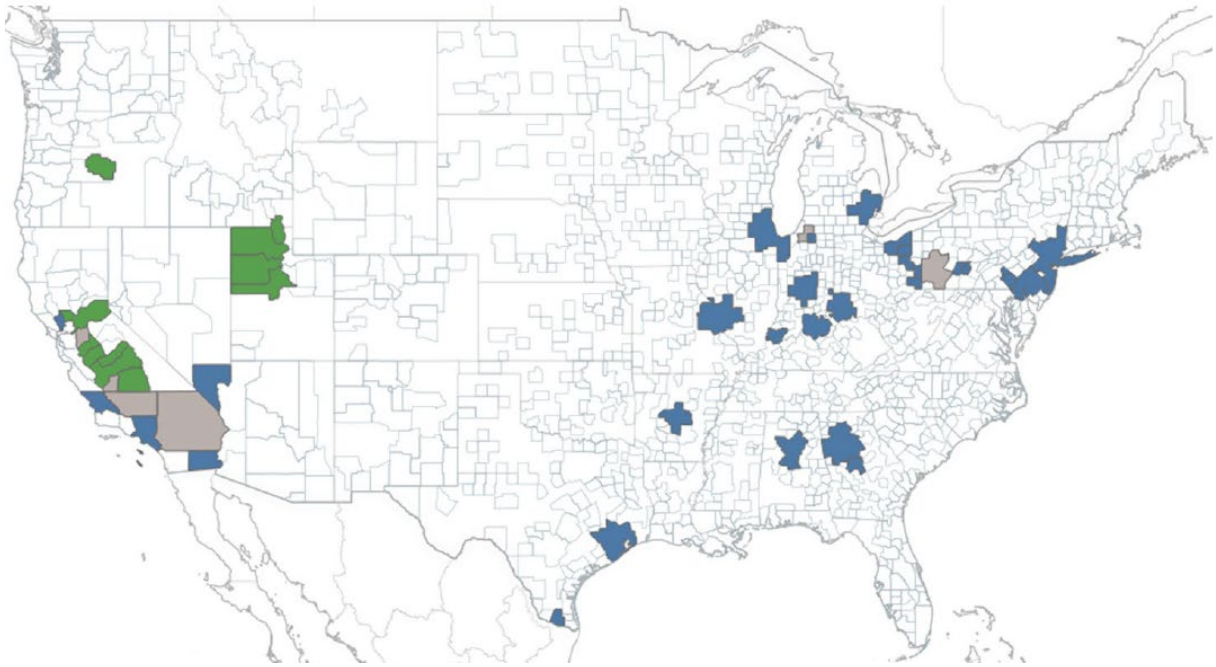

Figure 7.1 Map of Study Areas Used in the Risk Modeling. Source: US EPA (2019a).

7.2 Causal classifications are not supported for selected health outcomes. (3.3.1)

As discussed in Sections 2-6 and in Gradient's comments on the 2018 draft ISA (Gradient, 2018), the scientific evidence does not support the causal classifications for the health outcomes evaluated in the risk assessment, particularly at ambient exposures.

7.3 Key epidemiology studies underlying the concentration-response functions have several methodological limitations. (3.3.1)

The draft PA states on page 3-81 (US EPA, 2019a):

> The selection of specific epidemiologic studies and concentration-response functions for use in modeling risk is based on criteria that take into account factors such as study design, geographic coverage, demographic groups evaluated, and health endpoints examined.

Further details are provided in Appendix C of the draft PA. However, the draft PA is not clear on exactly what was required to meet these criteria. The eight studies that were used in the risk assessment are shown in Table 7.1.

Table 7.1 Epidemiology studies used to estimate $PM_{2.5}$-associated risk

Epidemiology Study	Study Population	Age Range (years)	Mortality Categories Covered
Long-term mortality studies			
Jerrett *et al*., 2016	ACS	30+	IHD
Pope *et al*., 2015	ACS	30+	All-cause, IHD
Turner *et al*., 2016	ACS	30+	Lung cancer
Thurston *et al*., 2016	AARP	55-85	All-cause
Di *et al*., 2017a	Medicare	65+	All-cause
Short-term mortality studies			
Baxter *et al*., 2017	77 cities	All ages	Nonaccidental
Ito *et al*., 2013	NPACT	All ages	All-cause
Zanobetti *et al*., 2014	121 communities	65+	All-cause

Notes:
ACS = American Cancer Survey; AARP = American Association of Retired Persons; IHD = Ischemic Heart Disease; NPACT = National Particle Components Society.
Table recreated from Table 3-4 of PM PA (US EPA, 2019a).

Each of these studies had methodological limitations that impacted the derivation and interpretation of concentration-response functions. As discussed in Sections 4.1.1, 4.2, and 5.1, these studies do not provide sufficient evidence against a nonlinear threshold model, particularly in light of the lack of biological plausibility of effects down to zero. Study limitations and their impact on the risk assessment are discussed below.

7.3.1 Long-term mortality studies have substantial biases and uncertainties.

The draft PA indicates that it considers eight key epidemiology studies (listed above in Table 7.1 and reviewed in Table 7.2) investigating the health effects associated with either long- or short-term $PM_{2.5}$ exposure in the risk assessment (US EPA, 2019a). These studies were conducted using data from across the US, and the draft PA indicates that the information from these studies was used to derive concentration-response functions and estimate population-level health risks at current and alternative standards. The health outcomes considered from the long-term $PM_{2.5}$ epidemiology studies included total mortality (all-cause and nonaccidental), ischemic heart disease mortality, and lung cancer.

As shown in Table 7.2, long-term mortality studies included several large multi-city studies with a robust prospective cohort study design. However, the studies have many methodological limitations that likely led to substantial biases and/or uncertainties of the findings. These methodological limitations can be broadly categorized into four categories: exposure assessment, covariate adjustment, evaluation of copollutants, and statistical analyses. Overall, we found issues with exposure assessment that likely resulted in considerable exposure measurement error. For many studies, there was a lack of information regarding the validity of the models used to estimate $PM_{2.5}$ concentrations; this was particularly the case for studies that used land use regression modeling.

Furthermore, more than half the studies did not use time-varying exposure estimates to account for temporal variability and did not account for residential mobility, both of which likely resulted in considerable exposure measurement error. In some cases, fixed exposure time windows were assigned to participants that did not account for when the person died. As a result, some analyses might have incorrectly considered

$PM_{2.5}$ exposures that occurred after a person's death. Temporality, defined as the exposure occurring before the health outcome, is a key criteria for determining causality.

For many studies, there were clear issues with the adjustment of covariates and consideration of copollutants. Individual-level variables collected *via* questionnaires at baseline are subject to information bias and do not reflect changes in measured characteristics over the course of the study. In addition, these studies did not account for temporal variation and residential mobility for the ecological covariates. While copollutants were considered in some of the studies, the exposure measurements of copollutants were often subject to the same methodological issues as the $PM_{2.5}$ measurements. As a result, it is uncertain whether the results seen in the studies reflect the independent effects of $PM_{2.5}$.

Finally, with regard to statistical analyses, most studies did not test model assumptions or statistically assess potential nonlinearity. In addition, some studies did not show concentration-response curves or present analyses demonstrating they settled on the most appropriate statistical model.

7.3.2 Short-term mortality studies have substantial biases and uncertainties.

The draft PA included three epidemiology studies focusing on total mortality associated with short-term $PM_{2.5}$ exposure in their risk assessment. Similar to the long-term studies, these were large multi-city studies.

While most short-term studies collected $PM_{2.5}$ exposure data from multiple monitors, suggesting a high spatial resolution, and used time-varying estimates, two of the three short-term studies did not account for residential mobility. In addition, most studies did not evaluate multiple exposure windows to determine which was most relevant.

One of the three short-term mortality studies had a case-crossover design, making adjustments for many confounders unnecessary (Zanobetti *et al*., 2014). In addition, one study conducted a time-series analysis of mortality and $PM_{2.5}$ exposure using population-level data, which also made adjusting for confounders unnecessary. While these study designs lend certainty to the results, there were still issues with confounding. For example, Zanobetti *et al*. and Baxter *et al*. did not consider the role of copollutants in their models (Zanobetti *et al*., 2014; Baxter *et al*., 2017). As a result, it is still unclear whether the results in the studies reflect the independent effects of $PM_{2.5}$ on mortality.

Finally, similar to the long-term studies, with regard to statistical analyses, two of the three studies did not test model assumptions or statistically assess potential nonlinearity. In addition, concentration-response curves were not reported.

Table 7.2 Sources of Bias and Uncertainty in Epidemiology Studies of Long- and Short-term PM2.5 Exposure and Total Mortality

Sources of Bias and Uncertainty		Long-term $PM_{2.5}$					Short-term PM2.5		
		Jerrett *et al.*, 2016	Pope *et al.*, 2015	Turner *et al.*, 2016	Thurston *et al.*, 2016	Di *et al.*, 2017a	Baxter *et al.*, 2017	Ito *et al.*, 2013	Zanobetti *et al.*, 2014
$PM_{2.5}$ Exposure Assessment	Central site monitoring (low spatial resolution)								
	No validation for $PM_{2.5}$ data		X	X	X				
	Temporal variation not accounted for	X	X	X					
	Residential mobility not accounted for	X	X	X			X		X
	No evaluation on multiple exposure windows	X	X	X	X	X	X	X	
	Personal activity no accounted for (*e.g.*, time spent indoors)	X		X	X	X			X
	Mismatch of $PM_{2.5}$ exposure window and mortality	X	X			X			
Individual Covariates	No adjustment of individual covariates						X		
	Information bias (*e.g.*, self-reported covariates)	X	X	X	X	X			
	Temporal variation not accounted for	X	X	X	X	X			
	Unmeasured confounding	X		X	X	X			
Ecological Covariates	No adjustment of ecological covariates						X*		
	Temporal variation not accounted for	X	X	X	X				
	Residential mobility not accounted for	X	X	X					
	Unmeasured confounding (*e.g.*, access to health care, violence)	X	X	X	X	X			
Evaluation of Copollutants	No adjustment of copollutants	X	X				X		X
	Central site monitoring (low spatial resolution)				X			X	
	No validation for copollutants data			X	X				
	Temporal variation not accounted for			X	X				
	Residential mobility not accounted for			X	X				
	Personal activity not accounted for (*e.g.*, time spent indoors)			X	X	X			
	Collinearity/nonlinear relationship with $PM_{2.5}$ not addressed/accounted for				X				
	Mismatch of copollutants exposure windows and mortality				X	X			
Statistical Analyses	Model assumptions not tested/relaxed	X	X	X	X	X	X	X	
	C-R curves sensitive to *df* (natural splines)	X	X	X	X		X	X	X
	Nonlinearity not assessed statistically	X	X	X	X	X	X		X
	Threshold not assessed	X	X	X	X	X	X	X	X

Notes:

C-R = Concentration-response; *df* = Degrees of Freedom; $PM_{2.5}$ = Particulate Matter Less Than 2.5 Microns in Diameter.

*= study design does not consider standard confounders.

This table summarizes several broad methodological categories where biases and uncertainties could arise in long-term and short-term $PM_{2.5}$ exposure and total mortality studies that were used in risk assessment analyses. Variables considered include exposure assessment, individual-level covariates, ecological covariates, evaluation of copollutants, and statistical analyses. Red shading with an "X" indicates the potential for bias and/or the presence of uncertainty with regard to specific methodological characteristics but does not reflect the magnitude of such a bias/uncertainty on study results. Unshaded cells indicate there are no apparent biases/uncertainties. For example, Zanobetti *et al.* (2014) did not account for potential confounding by copollutants when assessing $PM_{2.5}$ exposure; thus the red shading with an "X" reflects the potential for bias and the presence of uncertainty in this aspect.

7.4 $PM_{2.5}$ air quality scenarios evaluated are an insufficient basis for a risk assessment. (3.3.1)

Uncertainty in the draft PA modeled $PM_{2.5}$ concentrations is not sufficiently quantified, despite significant potential sources of error, and thus these modeled concentrations should not be used as the basis for a risk assessment. The draft PA directly evaluated "current" (2015) air quality conditions, a hypothetical air quality scenario in which air quality just meets the current annual standard of 12 μg/m^3, and a hypothetical air quality scenario in which air quality just meets an alternative annual standard of 10 μg/m^3 (US EPA, 2019a). EPA uses an unvalidated interpolation/extrapolation approach to evaluate air quality scenarios just meeting alternative standards of 11 μg/m^3 and 9 μg/m^3 (US EPA, 2019a); a demonstration that modeled $PM_{2.5}$ concentrations scale linearly between modeled alternative standards should have been performed to validate that the data were linear (*i.e.*, by including more than two directly evaluated scenarios). No scenarios for higher standard levels (*e.g.*, 13 μg/m^3) or below 9 μg/m^3 (*e.g.*, 8 μg/m^3) were evaluated. No discussion of background $PM_{2.5}$ concentrations was included in the assessment.

There was no evaluation of 13 μg/m^3, which means the draft PA did not demonstrate that the current 12 μg/m^3 standard is not more stringent than necessary to protect public health. It is possible a standard for 13 μg/m^3 would be equally protective. Further, the draft PA did not address risks that might occur at a standard of 8 μg/m^3 and cannot be used to support lowering the standard to this level.

Linear interpolation and extrapolation from the two modeled scenarios (12 μg/m^3 and 10 μg/m^3) are used to estimate risks under other standards (11 μg/m^3 and 9 μg/m^3). These estimated risks directly inform the conclusions of the draft PA but are calculated using an unvalidated and highly uncertain approach, since the modeled $PM_{2.5}$ concentrations are not shown to scale linearly with changes in selected alternative standards and projected emissions.

The draft PA qualitatively assesses the influence of different uncertainties in inputs and approaches on the final risk assessment results, categorizing each uncertainty by whether it is expected to have a "low," "medium," or "high" impact on the risk assessment results (US EPA, 2019a). Quantitative uncertainty analyses are largely missing from the draft PA. Using linear interpolation/extrapolation from the modeling results at 12 μg/m^3 and 10 μg/m^3 is assessed to have a "medium" magnitude impact on the risk assessment (US EPA, 2019a) (*i.e.*, to have "the potential to affect…conclusions [from the risk assessment]" [US EPA, 2019a]). Linear extrapolation is more uncertain than linear interpolation, since "as one moves outside the range of the original data, the reliability of the linear approximation as an empirical model of the true relationship will deteriorate" (Montgomery *et al.*, 2010), and the draft PA acknowledges that the modeling results for 9 μg/m^3 are thus even more uncertain than the results for 11 μg/m^3 (US EPA, 2019a). The risk assessment should not be used to evaluate risks at a standard of 9 μg/m^3 due to uncertainties in linearly extrapolating results without any demonstration that the results are linear.

Data from 2015 were used as the baseline scenario from which emissions were adjusted (US EPA, 2019a). As mentioned in Section 2.2 and 2.3, the addition of $PM_{2.5}$ monitoring at near-road locations started at the beginning of 2015, which is after the study periods in the key epidemiology studies. This means that, in the draft PA's risk assessment, $PM_{2.5}$ data accounted for near-road monitoring (since it used 2015 data) whereas the concentration-response functions did not (since they are based on key epidemiology studies of periods before 2015). This discrepancy further undermines the validity of the risk assessment results.

The air quality scenarios evaluated also did not include consideration of policy-relevant background (PRB). Background $PM_{2.5}$ was previously determined to contribute 1-3 μg/m^3 to ambient exposures, and the draft PA indicated this background contribution was likely unchanged under current conditions (US EPA, 2019a). None of the modeled air quality scenarios, or adjustments made to emissions as part of the

modeling, included an assessment of background $PM_{2.5}$ contributions, however. The draft PA draws quantitative mortality conclusions from the risk assessment without acknowledging that some $PM_{2.5}$-related impacts are caused by background. PRB is mentioned in the document exclusively in the list of acronyms.

In every modeled scenario where air quality just meets a particular standard, the draft PA assumes that all emitters adjust their emissions linearly (US EPA, 2019a). This is an inappropriate assumption because emitters below the revised standard will likely not adjust emissions. Further, distributed sources (*e.g.*, highway vehicles) may not scale emissions the same way as point sources (*e.g.*, power plants). The draft PA acknowledges that "[m]ore refined emission scenarios could be beneficial for projections in areas with relatively large seasonal and/or spatial variability in $PM_{2.5}$" (US EPA, 2019a). As shown in Table C-6 (US EPA, 2019a), the seasonal variability in average concentrations for every region from which data was included was significant at 2 $\mu g/m^3$ or more (*i.e.*, of the same magnitude as proposed changes in the annual standard). More refined emissions scenarios would thus be beneficial in every modeled area. The errors in the current modeling show that observed concentrations may be significantly lower than the modeled concentrations, and the disconnect between the models and the data is not quantified in the risk assessment. The results of the risk assessment are thus uncertain to an unquantified extent, and speculative.

The draft PA used only two emissions projection cases to model just meeting different air quality standards: one in which primary-emitted $PM_{2.5}$ was preferentially reduced and one in which secondarily produced $PM_{2.5}$ from NO_x and SO_2 was preferentially reduced (US EPA, 2019a). The draft PA notes that the potential projection cases "span a wide range of possible conditions, but these cases are necessarily a subset of the full set of possible projection cases" (US EPA, 2019a). The subset of emissions projection cases selected is unrealistic and, given spatial variation in emissions sources, may not represent bounding conditions for the risk assessment (*e.g.*, if primary-emitted $PM_{2.5}$ is risk-driving in one area and secondarily produced $PM_{2.5}$ is risk-driving in another area). In reality, a combination of adjustments to primarily emitted and secondarily produced $PM_{2.5}$ would occur, and evidence is not presented to indicate that the two emissions projection cases are appropriate or bounding. Overall, using simulations to project the impacts of meeting alternative standards was assessed to have a "medium" magnitude impact on the risk assessment in the draft PA (US EPA, 2019a) (*i.e.*, could affect the conclusions of the risk assessment). Again, the impact of this approach on the results of the risk assessment was not quantified appropriately, and the results of the risk assessment are more highly uncertain than presented in the draft PA.

A conclusion the draft PA draws from the risk assessment is that, for every alternative standard, "most of the risk remaining is estimated at annual average $PM_{2.5}$ concentrations that fall somewhat below the standard level" (US EPA, 2019a). This is based, however, on the inappropriate modeling assumption that emissions sources scale linearly to meet each modeled standard. As a result of the linear scaling, the shape of the curve will always be generally the same, regardless of whether the data are scaled to meet an alternative standard of 15 $\mu g/m^3$ or 5 $\mu g/m^3$ (see, *e.g.*, Figure 3-12 of US EPA, 2019a and Figure 7.3 in this report). In reality, the shape of the emissions curve may adjust if the standard is lowered, since not all emitters will adjust emissions linearly. Thus, the conclusion in the draft PA that "most of the risk remaining is estimated at annual average $PM_{2.5}$ concentrations that fall somewhat below the standard level" (US EPA, 2019a,) is derived from the modeling assumptions and is not an appropriate conclusion to draw from the available data.

In summary, the $PM_{2.5}$ air quality scenarios in the draft PA:

- are not sufficient to determine whether the current long-term NAAQS is more stringent than necessary;
- cannot be used to support an alternative standard of 8 $\mu g/m^3$ because there are no data there to assess;

- cannot be used to support an alternative standard of 9 μg/m^3 due to uncertainties in linearly extrapolating from two data points;
- do not correlate with the concentration-response curves upon which the health effect estimates are based; and
- do not use realistic emissions scenarios.

The $PM_{2.5}$ air quality scenarios are insufficient to support the conclusions drawn in the risk assessment.

7.5 Risk assessment has significant variability and uncertainty. (3.3.1, 3.3.2)

The draft PA draws precise, quantitative conclusions from the risk assessment (*e.g.*, p. 3-97), but limited quantitative uncertainty analyses were performed due to "limitations in data, methods and/or resources" (US EPA, 2019a). However, the conservativism and uncertainty in the modeling preclude drawing any precise, quantitative conclusions from the risk assessment. That is, drawing conclusions about the difference in risk between various standards is meaningless without appropriate quantification of uncertainty, since the results cannot be shown to be statistically significantly different without such a quantification.

Models are used in the draft PA to create estimates of air quality conditions for scenarios just meeting the current standard and meeting alternative standards (see discussion of scenarios in Section 7.4 above). Air quality conditions for 2015 are also modeled, and are compared to measured data in Table C-6 (US EPA, 2019a). The differences between the average modeled and average measured concentrations are frequently reported to be more than 1 μg/m^3, indicating that the modeling precision is insufficient to distinguish between concentrations at this magnitude (*i.e.*, the modeling should not be used to address differences between air quality just meeting standards at 12 μg/m^3 *vs*. 11 μg/m^3 or between 11 μg/m^3 and 10 μg/m^3). These differences between modeled and measured data are not reflected in the main text of the report (*i.e.*, the confidence intervals reported in the risk assessment results tables in Section 3.3.2 do not reflect this uncertainty in the true concentrations), and thus present overconfident representations of the results.

Pearson correlation coefficients measuring the linear correlation between modeled and measured values are reported in Table C-6 (US EPA, 2019a). A model that perfectly reproduced the measured concentrations would have a Pearson correlation coefficient of 1, while a Pearson correlation coefficient of 0 would indicate no linear correlation between the modeled and measured values. The correlation coefficients for annual data by region range from 0.34 to 0.67; correlation coefficients for the West, Northwest, Northwest Rockies and Plains, and the South regions were closer to 0 than 1, indicating a poor linear correlation between the modeled and measured values. Figure 7.2 shows an example of data with Pearson correlation coefficients (r) of 0, 0.34, 0.67, and 1. The correlation coefficients of 0.34 and 0.67 show significant deviation from linearity. Lack of correlation between modeled and measured values was not one of the quantitatively evaluated sources of uncertainty in the draft PA (US EPA, 2019a), and the lack of a strong linear correlation between modeled and measured values indicates that quantitative conclusions regarding the number of deaths caused by $PM_{2.5}$ per year (*e.g.*, p. 3-97) are uncertain to an extent beyond the level of confidence specified in the draft PA.

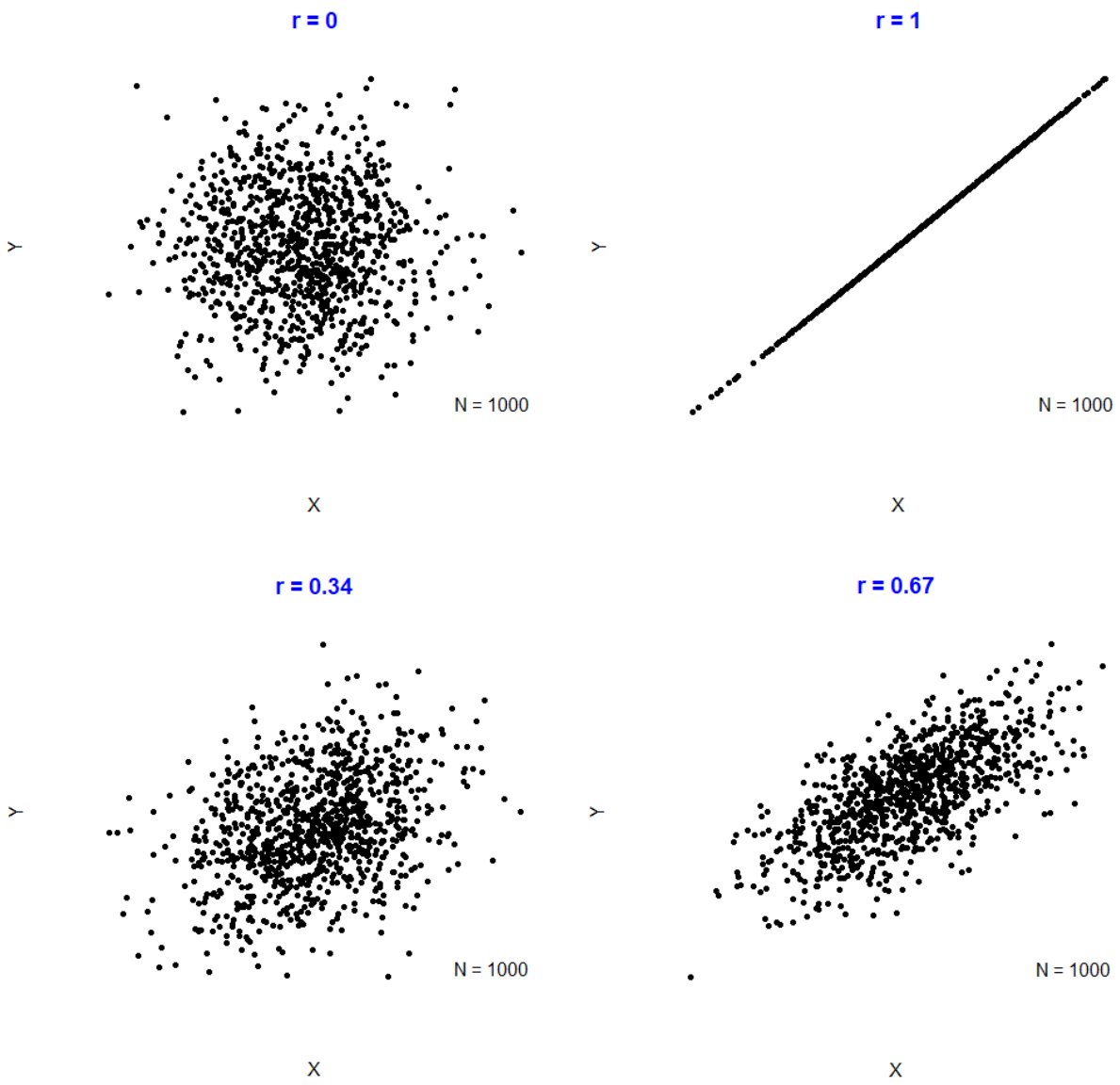

Figure 7.2 Example Correlation Coefficient Plots. The draft PA model deviates from reality at a rate represented by the lower two plots.

In addition, the draft PA employs an unusual rounding choice that overstates the $PM_{2.5}$ risk associated with a particular standard. In the draft PA, "risk estimates are rounded toward zero into whole $PM_{2.5}$ concentration values (*e.g.*, risk estimate at 10 μg/m^3 includes risk occurring at 10.0-10.9 μg/m^3)" (US EPA, 2019a). This inflates the risks occurring at "10 μg/m^3," or any other specified value, since risks associated with concentrations up to 10.9 μg/m^3 are included. A standard of "10 μg/m^3" should then permit actual concentrations up to 10.9 μg/m^3. This is inconsistent with the final rule rounding convention specified in 40 CFR Appendix N to Part 50 §4.3, which specifies that "[a]nnual $PM_{2.5}$ NAAQS DVs shall be rounded to the nearest tenth of a μg/m^3 (decimals x.x5 and greater are rounded up to the next tenth, and any decimal lower than x.x5 is rounded down to the nearest tenth)" (US EPA, 2013).

Using the draft PA's modeling approach with the deficiencies discussed in the preceding sections, the shape of the results for a modeled standard of 11 μg/m^3 best matches "current" (2015) conditions as shown in

Figure C-27 in the draft PA (US EPA, 2019a), reproduced in this report as Figure 7.3. The various levels of conservativism and uncertainty in modeling $PM_{2.5}$ risks under the current standard effectively result in a full 1 µg/m^3 overestimate of current risks (*i.e.*, current risks are best represented by a model for air quality just meeting an annual standard of 11 µg/m^3) rather than the current standard of 12 µg/m^3.

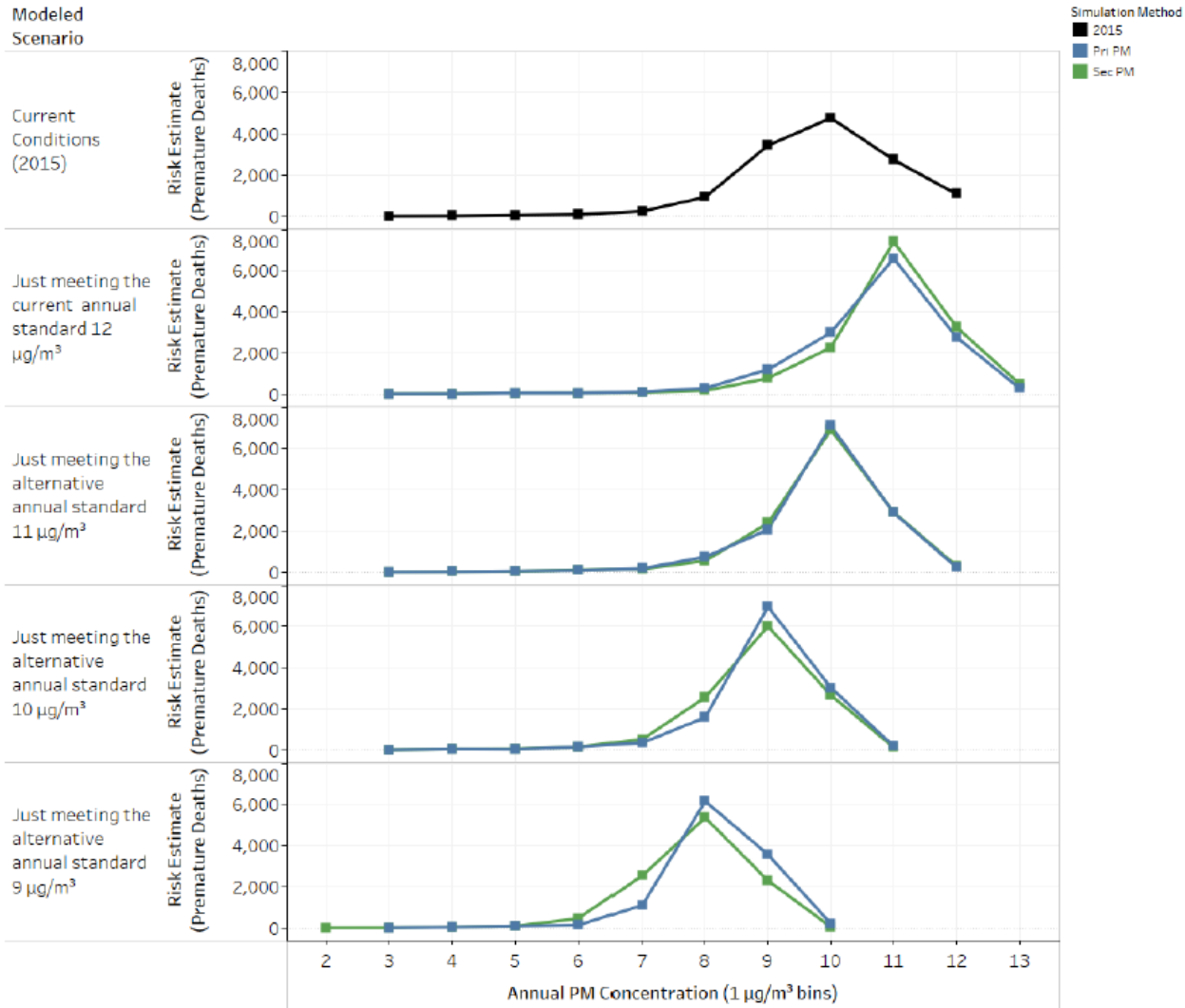

Figure 7.3 Distribution of estimated $PM_{2.5}$-associated mortality for current conditions (2015), the current annual standard (12 µg/m^3), and alternative standards (9.0, 10.0, and 11.0 µg/m^3) simulated for the 30 annual-controlled urban study areas. Reproduced from Figure C-27 (US EPA, 2019a).

The draft PA quantitatively evaluates three specific sources of uncertainty in the risk assessment (US EPA, 2019a), but there are many other model input parameters that require evaluation. A full uncertainty analysis was not performed as it should have been according to WHO guidelines: "[t]he overall analysis in the PM NAAQS risk assessment is relatively complex, thereby warranting consideration of a full probabilistic (WHO Tier 3) uncertainty analysis. However, limitations in available information prevent this level of analysis from being completed at this time…One approach that has been taken in such cases is expert elicitation; however, this approach is resource- and time-intensive and consequently, it was not feasible" (US EPA, 2019a). Given the magnitude of the uncertainty in modeled *vs*. measured values (US EPA, 2019a), the significant variation in estimates of risk presented in Table 3-5 (US EPA, 2019a), and the

number of potential sources of uncertainty discussed in this section, a more complete uncertainty analysis as described by WHO guidelines should be conducted before determining whether an alternative standard level is appropriate. The failure of the draft PA to include the necessary uncertainty analyses renders the results of the risk assessment meaningless. The results of the risk assessment cannot be used to support adjustment of the current standard.

8 Preliminary conclusions on the current standard (3.4.1)

Regarding current $PM_{2.5}$ standards, the draft PA states:

- There is a long-standing body of strong health evidence demonstrating relationships between long- or short-term $PM_{2.5}$ exposures and a variety of outcomes, including mortality and serious morbidity effects. Studies published since the last review have reduced key uncertainties and broadened our understanding of the health effects that can result from exposures to $PM_{2.5}$.
- Recent US and Canadian epidemiologic studies provide support for generally positive and statistically significant health effect associations across a broad range of ambient $PM_{2.5}$ concentrations, including for air quality distributions with overall mean concentrations lower than in the last review and for distributions likely to be allowed by the current primary $PM_{2.5}$ standards.
- Analyses of $PM_{2.5}$ pseudo-design values additionally support the occurrence of positive and statistically significant health effect associations based largely on air quality likely to have met the current annual and 24-hour primary standards.
- The risk assessment estimates that the current primary $PM_{2.5}$ standards could allow a substantial number of $PM_{2.5}$-associated deaths in the US. The large majority of these estimated deaths are associated with the annual average $PM_{2.5}$ concentrations near (and above in some cases) the average concentrations in key epidemiologic studies reporting positive and statistically significant health effect associations. (US EPA, 2019a)

These conclusions are not supported by the scientific evidence.

8.1 Studies published since the last review do not reduce uncertainty.

The studies published since the last review have the same limitations as those discussed in the 2009 ISA and thus have the same uncertainties (discussed above in Section 3).

8.2 Recent US and Canadian studies do not provide support for effects at concentrations lower than in the last review and are likely to be allowed by the current primary $PM_{2.5}$ standards.

New studies available since the last review have limitations that preclude a conclusion that there are health effects at lower concentrations than indicated in the last review. Furthermore, as discussed on page 3-98 of the draft PA, there is uncertainty in the public health impacts as concentrations fall further below those in accountability studies (US EPA, 2019a). These studies show decreased health impacts when starting concentrations are 13 to > 20 $\mu g/m^3$.

8.3 Pseudo-design values do not support health effects at air quality levels likely to have met current standards.

As discussed in Section 2.3, analyses with pseudo-design values indicate that a large proportion of the study populations in the key epidemiology studies was likely in areas not meeting the current standard. Consequently, results of pseudo-design analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current $PM_{2.5}$ standard.

8.4 The risk assessment estimates do not account for uncertainty.

EPA claims (US EPA, 2019a) that the risk assessment shows substantial decreases in risk at lower standard levels; as discussed in Section 7, these statements do not account for the high level of uncertainty in the modeling calculations supporting the risk assessment. The risk assessment includes several areas of uncertainty not appropriately quantified in the draft PA.

- The risk assessment was performed for a small fraction of the US that does not necessarily correlate with the modeled concentration-response functions (see Section 7.1);
- The air quality scenarios evaluated use unrealistic emissions scenarios and unvalidated linear extrapolation from two data points (see Section 7.4); and
- Uncertainty in the modeled $PM_{2.5}$ concentrations, unconventional rounding choices, and compounding levels of conservativism in the modeling preclude drawing any distinctions between proposed alternative standards (see Section 7.5).

The draft PA did not conduct a full uncertainty analysis as recommended in WHO guidelines, and the lack of appropriate accounting for uncertainty effectively renders meaningless the comparisons between the risk assessment results.

8.5 Experimental studies do not provide support for lower standards.

As discussed in Sections 3.1.3, 4.1.2, 5.2, and 6.1, experimental studies do not provide support for lower standards. Study quality was not consistently addressed, doses were often too high to be relevant for ambient exposures, and results were not always coherent across studies.

9 Preliminary conclusions on alternative standards (3.4.2)

9.1 Long-term mean $PM_{2.5}$ concentrations in key epidemiology studies do not support a lower standard.

As discussed in Section 2.2, the draft PA evaluated the $PM_{2.5}$ air quality distributions associated with mortality and morbidity in key epidemiology studies but could not rule out that either (1) the observed associations at mean $PM_{2.5}$ concentrations below the current standard are driven by the upper portion of the air quality distribution or that (2) the apparent associations with mortality or morbidity at lower $PM_{2.5}$ concentrations were due to substantial underestimation of $PM_{2.5}$ exposure (*i.e.*, exposure measurement error). Although the draft PA reviewed 67 epidemiology studies, it did not do so in a comprehensive or systematic manner. Therefore, the analysis of current epidemiology literature in the draft PA does not provide sufficient support for the existence of elevated mortality/morbidity risk at $PM_{2.5}$ concentrations below the current standard.

9.2 Long-term mean $PM_{2.5}$ concentrations and annual design values over-predict health impacts.

The draft PA indicates that design values are 10-20% higher than long-term mean $PM_{2.5}$ concentrations in epidemiology studies (US EPA, 2019a). It is not clear why the current and alternative standards are compared to mean concentrations in these studies when this type of analysis over-predicts health impacts.

9.3 $PM_{2.5}$ pseudo-design values in many study locations are above the standard.

A large proportion of the study populations was in areas that did not meet alternative standards. Consequently, results of pseudo-design analyses indicate that the key studies do not provide sufficient evidence against the adequacy of protection provided by the current $PM_{2.5}$ standard.

9.4 Controlled exposure studies are not informative.

The draft PA acknowledges that the $PM_{2.5}$ concentrations in most controlled exposure studies of cardiovascular effects are well above ambient concentrations typically measured in areas meeting the current primary standards, and it indicates these studies are not informative for assessing lower concentrations (US EPA, 2019a). We also note that the draft PA does not adequately evaluate consistence or coherence of effects across studies.

9.5 $PM_{2.5}$-associated risk estimates are not meaningful.

The draft PA presents several proposed alternative annual standard levels in Section 3.4.2, noting that lower alternative standards can be supported to the extent that "less weight is placed on the limitations in the evidence that contribute to greater uncertainty at lower concentrations" (US EPA, 2019a). This statement is made based on the incomplete uncertainty analyses conducted in the draft PA. As discussed in Section 7, the risk assessment includes several additional areas of uncertainty not appropriately quantified in the draft PA.

- The risk assessment was performed for a small fraction of the US that does not necessarily correlate with the modeled concentration-response functions (see Section 7.1);
- The air quality scenarios evaluated use unrealistic emissions scenarios and unvalidated linear extrapolation from two data points (see Section 7.4); and
- Uncertainty in the modeled $PM_{2.5}$ concentrations, unconventional rounding choices, and compounding levels of conservativism in the modeling preclude drawing any distinctions between proposed alternative standards (see Section 7.5).

The draft PA did not conduct a full uncertainty analysis as recommended in WHO guidelines, and the lack of appropriate accounting for uncertainty effectively renders meaningless the comparisons between the risk assessment results.

10 Alternative annual standard levels (3.4.2.4.1)

As discussed above, new evidence does not support lowering the standards. We also note that the draft PA states:

- Few key epidemiology studies (and only one key US study) report positive and statistically significant health effect associations for $PM_{2.5}$ air quality distributions with overall mean concentrations below 9.6 μg/m^3, and areas meeting a standard with a level of 10.0 μg/m^3 would generally be expected to have lower long-term mean $PM_{2.5}$ concentrations (and potentially around 8.0 μg/m^3 in some areas) (section 3.2.3.2.1; Appendix B, section B.7).
- There is increasing uncertainty in $PM_{2.5}$ exposure estimates in some of the largest key studies at lower ambient concentrations (*i.e.*, those that use hybrid model predictions to estimate exposures), given the more limited information available to develop and validate model predictions (sections 2.3.3 and 3.2.3.2.1).
- Pseudo-design values corresponding to the 50th percentiles of study area populations (or health events) are ≥ about 10.0 μg/m^3 for almost all key studies, particularly those conducted in the US (section 3.2.3.2.2).
- There is a high degree of uncertainty in the quantitative estimates of $PM_{2.5}$-associated mortality risk presented in the draft PA, and the uncertainty increases for standard levels below 10.0 μg/m^3, given that a substantial proportion of the risk reductions estimated for lower standard levels occur at annual average $PM_{2.5}$ concentrations below 8 μg/m^3, and thus below the lower end of the range of overall mean $PM_{2.5}$ concentrations in key epidemiologic studies that consistently report positive and statistically significant associations (section 3.3.2). (US EPA, 2019a)

As noted in Section 7.4, the conclusion in the draft PA that, for every alternative standard, "most of the risk remaining is estimated at annual average $PM_{2.5}$ concentrations that fall somewhat below the standard level" (US EPA, 2019a), is based on the inappropriate modeling assumption that emissions sources scale linearly to meet each modeled standard. Nevertheless, as discussed throughout Section 7, the risk assessment does not include a full uncertainty analysis, and this effectively renders meaningless the comparisons between the risk assessment results.

11 Alternative 24-hour standard levels (3.4.2.4.2)

As stated in the draft PA, "[R]ecent air quality analyses indicate that almost all CBSAs with maximum annual $PM_{2.5}$ design values at or below 12.0 µg/m^3 also have maximum 24-hour design values below 35 µg/m^3 (and below 30 µg/m^3 in most areas) (Chapter 2, Figure 2-11). The exceptions are a few CBSAs in the western U.S." (US EPA, 2019a). Given the conservative nature of the risk assessment, it is unlikely that a revised 24-hour standard will have any effect on public health, even in the few CBSAs with higher 24-hour design values.

12 Future research (3.5)

We generally agree with all areas of future research proposed. We think these areas are particularly important:

- Improving our understanding of the $PM_{2.5}$ concentration-response relationships near the lower end of the $PM_{2.5}$ air quality distribution, including the shapes of concentration-response functions and the uncertainties around estimated functions for various health outcomes and populations (*e.g.*, older adults, people with preexisting diseases, children).
- Improving our understanding of the potential for particle characteristics, other than size-fractionated mass, to influence PM toxicity (*e.g.*, composition, oxidative potential, *etc.*) and the PM health effect associations observed in epidemiology studies.
- Improving our understanding of the uncertainties inherent in the various approaches used to estimate $PM_{2.5}$ exposures in epidemiology studies, including how those uncertainties may vary across space and time, and over the $PM_{2.5}$ air quality distribution. Approaches to incorporating these uncertainties into quantitative estimates of $PM_{2.5}$ concentration-response relationships should also be explored.

In addition, causal associations should be reevaluated after doing the following:

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

Future risk assessment research should focus on quantitatively assessing uncertainty from sources outlined in Section 7 and reducing uncertainty as feasible to provide a measureable degree of confidence in the risk assessment results. Additional work to develop appropriate scaling methods for $PM_{2.5}$ emissions to evaluate alternative standards should also be undertaken.

References

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

Bollen, KA; Jackman, RW. 1985. "Regression diagnostics: An expository treatment of outliers and influential cases." *Sociol. Methods Res.* 13(4):510-542. doi: 10.1177/0049124185013004004.

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

Chay, K; Dobkin, C; Greenstone, M. 2003. "The Clean Air Act of 1970 and adult mortality." *J. Risk Uncertain.* 27(3):279-300.

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

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

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

Greenbaum, D. 2019. Letter to EPA. February 21.

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.

Hill, AB. 1965. "The environment and disease: Association or causation?" *Proc. R. Soc. Med.* 58(5):295-300.

Ito, K; Ross, Z; Zhou, J; Nadas, A; Lippmann, M; Thurston, GD. 2013. "NPACT Study 3. Time series analysis of mortality, hospitalizations, and ambient PM2.5 and its components." In *National Particle Component Toxicity (NPACT) Initiative: Integrated Epidemiologic and Toxicologic Studies of the Health Effects of Particulate Matter Components*. Health Effects Institute, Boston, MA. HEI Research Report 177 Part 3, p95-125. Accessed at https://www.healtheffects.org/publication/national-particle-component-toxicity-npact-initiative-integrated-epidemiologic-and.

Janes, H; Dominici, F; Zeger, SL. 2007. "Trends in air pollution and mortality: an approach to the assessment of unmeasured confounding." *Epidemiology.* 18(4):416-23.

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

Montgomery, DC; Runger, GC; Hubele, NF. 2010. *Engineering Statistics (Fourth Edition).* John Wiley, Hoboken, NJ. 487p.

Moolgavkar, SH; Chang, ET; Watson, HN; Lau, EC. 2018. "An Assessment of the Cox Proportional Hazards Regression Model for Epidemiologic Studies." *Risk Anal.* 38(4):777-794.

Papadogeorgou, G; Kioumourtzoglou, MA; Braun, D; Zanobetti, A. 2019. "Low levels of air pollution and health: Effect estimates, methodological challenges, and future directions." *Curr. Environ. Health Rep.* 6(3):105-115. doi: 10.1007/s40572-019-00235-7.

Pope, CA III; Turner, MC; Burnett, RT; Jerrett, M; Gapstur, SM; Diver, WR; Krewski, D; Brook, RD. 2015. "Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality." *Circ. Res.* 116(1):108-115. doi: 10.1161/CIRCRESAHA.116.305060.

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.

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.

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.

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

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

US EPA. 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. January 15.

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

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

US EPA. 2019a. "Policy Assessment for the Review of the National Ambient Air Quality Standards for Particulate Matter (External Review Draft)." Office of Air Quality Planning and Standards, Health and Environmental Impacts Division. EPA-452/P-19-001, 457p., September.

US EPA. 2019b. Letter Report to A. Wheeler (US EPA) re: CASAC Review of the EPA's Integrated Science Assessment for Particulate Matter (External Review Draft - October 2018). Clean Air Scientific Advisory Committee (CASAC). EPA-CASAC-19-002, 205p, April 11.

Zanobetti, A; Dominici, F; Wang, Y; Schwartz, JD. 2014. "A national case-crossover analysis of the short-term effect of PM2.5 on hospitalizations and mortality in subjects with diabetes and neurological disorders." *Environ. Health* 13(1):38. doi: 10.1186/1476-069X-13-38.

Zigler, CM; Kim, C; Choirat, C; Hansen, JB; Wang, Y; Hund, L; Samet, J; King, G; Dominici, F; HEI Health Review Committee. 2016. "Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations." *Res. Rep. Health Eff. Inst.* (187):5-49.