Comments on EPA's Second External Draft of the Integrated Science Assessment For Particulate Matter (PM)

by

George T. Wolff, Dennis F. Kahlbaum and Jon M. Heuss

Air Improvement Resource, Inc.

Prepared for the

Alliance of Automobile Manufacturers

September 30, 2009

Executive Summary

The second draft of the Particulate Matter (PM) Integrated Science Assessment (ISA) differs from the first draft in a significant way. In the second draft, EPA appears to be building a case for the stringent regulation of coarse PM, $PM_{10-2.5}$, to replace the existing PM_{10} NAAQS. In addition, EPA appears to be focusing on all sources of $PM_{10-2.5}$, including crustal material.

The impetus for this change of emphasis is the publication of a June 2009 study by Zanobetti and Schwartz which associates current ambient concentrations of $PM_{10-2.5}$ to premature mortality and a re-evaluation of a number previous studies that focused on the same relationship but which EPA dismissed as having "mixed result" in the first draft of the ISA.

Air Improvement Resource, Inc. (AIR) examined the Zanobetti and Schwartz study and concluded that is not a reliable basis for evidence of coarse PM health effects. Zanobetti and Schwartz fail to demonstrate how the measure they use for $PM_{10-2.5}$ compares to the Federal Reference Method (FRM) and it is not clear how much data was included in their analysis. They do not consider or evaluate biases or uncertainty due to model selection or potential confounders. Although Zanobetti and Schwartz present the results of a two-pollutant model with fine and coarse PM, they do not report any information on the correlation between the two metrics. This limits the interpretation of the results since, as the ISA notes, models that include both $PM_{10-2.5}$ and $PM_{2.5}$ may suffer from instability due to colinearity. Although they analyzed data from 188

cities, they only provided $PM_{10-2.5}$/mortality associations for the pooled result. This prevents an analysis of individual city associations which is vital to determine the robustness of their conclusions or the spatial patterns of the associations.

In the first ISA, EPA concluded that the results were mixed and that "more data is needed." AIR examined the previously available studies on $PM_{10-2.5}$ mortality relationships and agrees with this conclusion.

In addition, the Second External Draft largely ignores the important issues of model selection and publication bias. When considered, these issues cast further doubt on the validity that coarse PM is a causal agent for mortality and the other health effects that EPA claims at the ambient $PM_{10-2.5}$ concentrations measured in the reported epidemiological studies. At this time the epidemiological evidence does not provide adequate justification for a new $PM_{10-2.5}$ NAAQS.

AIR also examined the current ambient air database for $PM_{10-2.5}$ and concluded that the existing $PM_{10-2.5}$ database is extremely sparse. The available data does not contain a sufficient number of measurements to characterize the degree of nonattainment that would exist if such a NAAQS were promulgated or to develop the required risk and exposure assessment. This is because there is no requirement for the routine deployment of the FRM sampler for $PM_{10-2.5}$, a dichotomous sampler. As a result, the only $PM_{10-2.5}$ data available is from the relatively few sites that deploy both PM_{10} and $PM_{2.5}$ monitors and the $PM_{10-2.5}$ concentrations are calculated by difference. However, this procedure is subjected to large uncertainties. The sparseness and uncertainty of the $PM_{10-2.5}$ data alone is sufficient reason for EPA not to mandate a $PM_{10-2.5}$ NAAQS until a reliable nationwide database exists.

Introduction

In January, 2006, at the conclusion of the last PM review by CASAC, the U.S. EPA proposed to replace the 24-hour PM_{10} NAAQS of 150 $\mu g/m^3$ with a new 98^{th} percentile 24-hour $PM_{10-2.5}$ NAAQS of 70 $\mu g/m^3$, but solicited comments from the public over a range from 50 to 70 $\mu g/m^3$. In addition, the focus of the $PM_{10-2.5}$ would be in urban areas where $PM_{10-2.5}$ concentrations were dominated by traffic-related and industrial sources rather than in rural areas where wind-blown soil dust is the dominate source of $PM_{10-2.5}$.[1] In September, 2006, after consideration of the public comments, EPA reversed its position; instead of creating a new $PM_{10-2.5}$ NAAQS, the Administrator retained the existing 24-hour PM_{10} NAAQS.[2]

[1] 40 CFR Part 50, National Ambient Air Quality Standards for Particulate Matter; Proposed Rule, pp. 2620-2708, January 17, 2006.

[2] 40 CFR Part 50, National Ambient Air Quality Standards for Particulate Matter; Final Rule, pp. 61144-61233, October 12, 2006.

In the first draft of the Integrated Science Assessment (ISA) published in December, 2008, as part of the present review, it appeared that EPA would continue its focus on PM_{10} Although EPA concluded that "the effect of short-term exposure to $PM_{10-2.5}$ on mortality is suggestive of a casual relationship at ambient concentrations," EPA stated that "the majority of studies that examined $PM_{10-2.5}$ reported mixed results in terms of the relative impact of $PM_{10-2.5}$ on mortality." EPA also stated that "more data is needed.[3]"

In contrast, the current second draft of the ISA states "the majority of studies evaluated in this review provide some evidence for mortality associations with $PM_{10-2.5}$."[4] The main reason for this change of position is EPA's inclusion of a new multi-city study[5] that was only published in June 2009 which was too late to be included in the first draft of the ISA. In addition, their characterization of the other studies as having "mixed results" appears to have changed. In this ISA, EPA now states:"Overall, the consistent positive association between short-term exposure to $PM_{10-2.5}$ and mortality observed in the U.S. and Canadian-based multicity studies, along with the positive associations from single-city studies conducted in these locations, provides evidence that is suggestive of a causal relationship between short-term exposures to $PM_{10-2.5}$ and mortality."[6]

Furthermore, it appears that EPA may be making a case for a new 24-hour $PM_{10-2.5}$ NAAQS in the vicinity of 12 – 14 $\mu g/m^3$. This is based on statements made in the ISA that "the associations observed between $PM_{10-2.5}$ and cardiovascular mortality in areas with similar 24-h avg $PM_{10-2.5}$ concentrations ranging from 6.1-16.4 $\mu g/m^3$, with effects becoming more precise and consistently positive in locations with mean $PM_{10-2.5}$ concentrations of 12 $\mu g/m^3$ and above."[7] EPA continues, "studies have provided evidence that is suggestive for relationships between short-term exposure to $PM_{10-2.5}$ and cardiovascular effects, respiratory effects, and mortality. Conclusions regarding causation for the various health effects and outcomes were made for $PM_{10-2.5}$ as a whole regardless of origin, since $PM_{10-2.5}$-related effects have been demonstrated for a number of different environments. These effects have been observed in locations with mean $PM_{10-2.5}$ concentrations ranging from 5.6 to 13 $\mu g/m^3$."[8] Thus, the present emphasis is broad-

[3] U.S EPA, Integrated Science Assessment for Particulate Matter, First External Review Draft, EPA/600/R-08/139, p. 6-242, December, 2008.

[4] U.S EPA, Integrated Science Assessment for Particulate Matter, Second External Review Draft, EPA/600/R-08/139B, p. 2-27, July, 2009

[5] Zanobetti A; Schwartz J. (2009). The effect of fine and coarse particulate air pollution on mortality: A national analysis. Environ Health Perspect, 117: 898-903.

[6] ISA, p. 2-28.

[7] ISA, p. 2-26.

[8] ISA, p. 2-28.

based as it appears to include all $PM_{10-2.5}$ (regardless of origin") including crustal and soil dust. However, because the 24-hour concentrations of concern are so low, most urban and rural sources will undoubtedly be targeted for control. Consequently, this review will focus on the studies that EPA uses to support its conclusion of a "suggestive causal relationship between short-term exposures to $PM_{10-2.5}$ and mortality."[9]

However, before the mortality $PM_{10-2.5}$ relationships are discussed, the existing $PM_{10-2.5}$ ambient database for the U.S. will be examined.

Existing Ambient $PM_{10-2.5}$ in the U.S.

The existing ambient database for $PM_{10-2.5}$ is quite sparse because when EPA decided to retain the PM_{10} NAAQS in lieu of a coarse standard, there was no requirement for states to routinely collect $PM_{10-2.5}$ data. Consequently, the only places where $PM_{10-2.5}$ concentrations can be estimated are at sites that have both PM_{10} and $PM_{2.5}$ samplers, in which case the $PM_{10-2.5}$ can be estimated from the differences. However, there are problems associated with this methodology, as discussed in chapter 3 of the ISA. In this chapter EPA presents a map of available $PM_{10-2.5}$ data for the 2005-2007 period, shown in Figure 1 below.

As a result of the sparse coverage for this time period, AIR gathered the existing data from EPA's database and computed site by site values for 2006-2008. It must be cautioned, however, that the PM_{10} concentrations are reported in EPA's database at standard condition while the $PM_{2.5}$ concentrations are reported in local conditions. A rigorous analysis would require converting the PM_{10} to local conditions using local meteorological conditions which was beyond the scope of this analysis. Since this was not done, the data in Figures 2 and 3 should be viewed with caution.

Figure 2 shows the estimated annual means for the 2006-2008 period, while Figure 3 displays the 98^{th} percentile. Although the data coverage is still sparse, Figure 2 gives an idea that there would be a high degree of non-attainment if EPA was to select a standard in the ranges mentioned in the Introduction (6.1 -14 and 5.6 - 13$\mu g/m^3$). The 98^{th} percentile is also presented because that is the form of the existing $PM_{2.5}$ standard.

[9] ISA, p. 2-28.

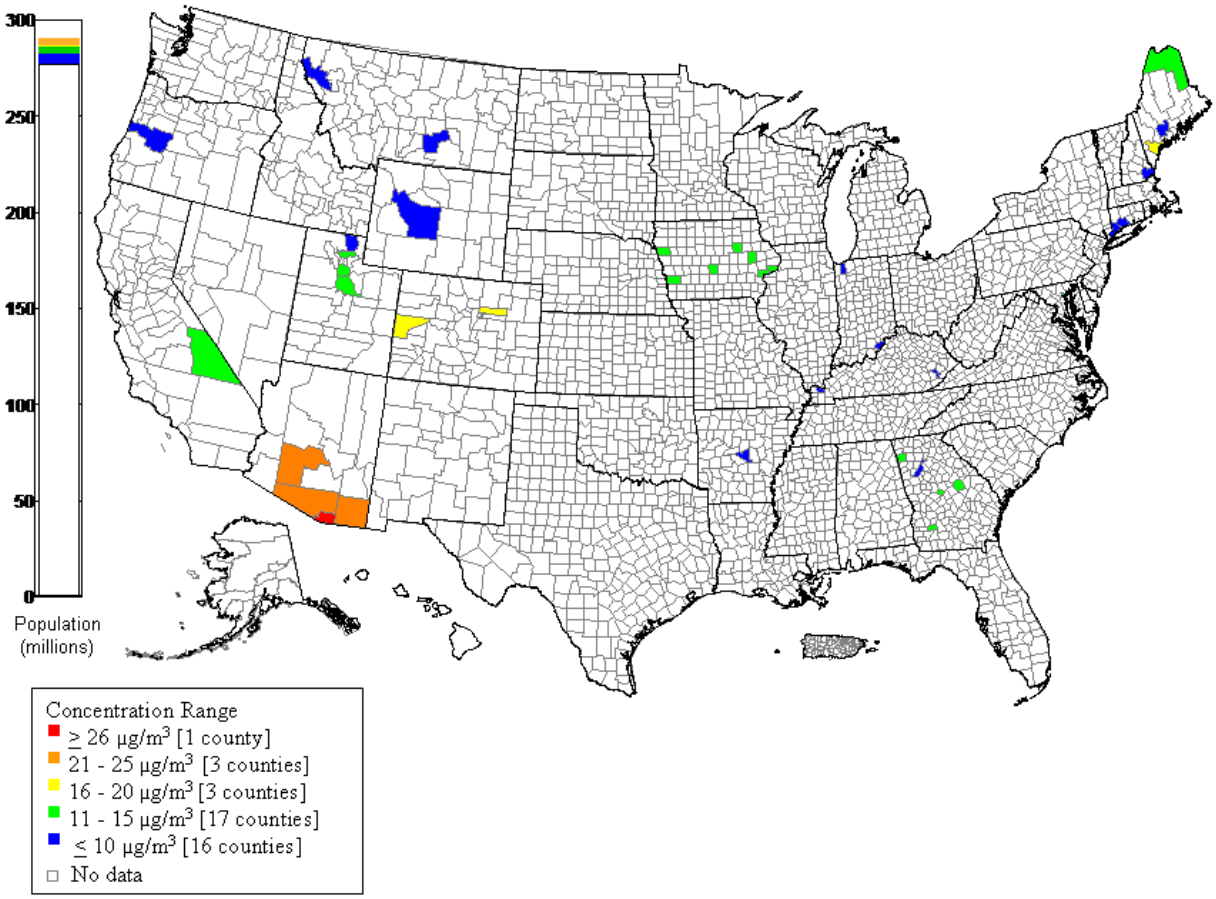

Figure 1: Three-yr avg 24-h $PM_{10-2.5}$ concentration by county derived from co-located low volume Federal Reference Method PM_{10} and $PM_{2.5}$ monitors, 2005-2007. The population bar shows the number of people residing within counties that reported county-wide average concentrations within the specified ranges

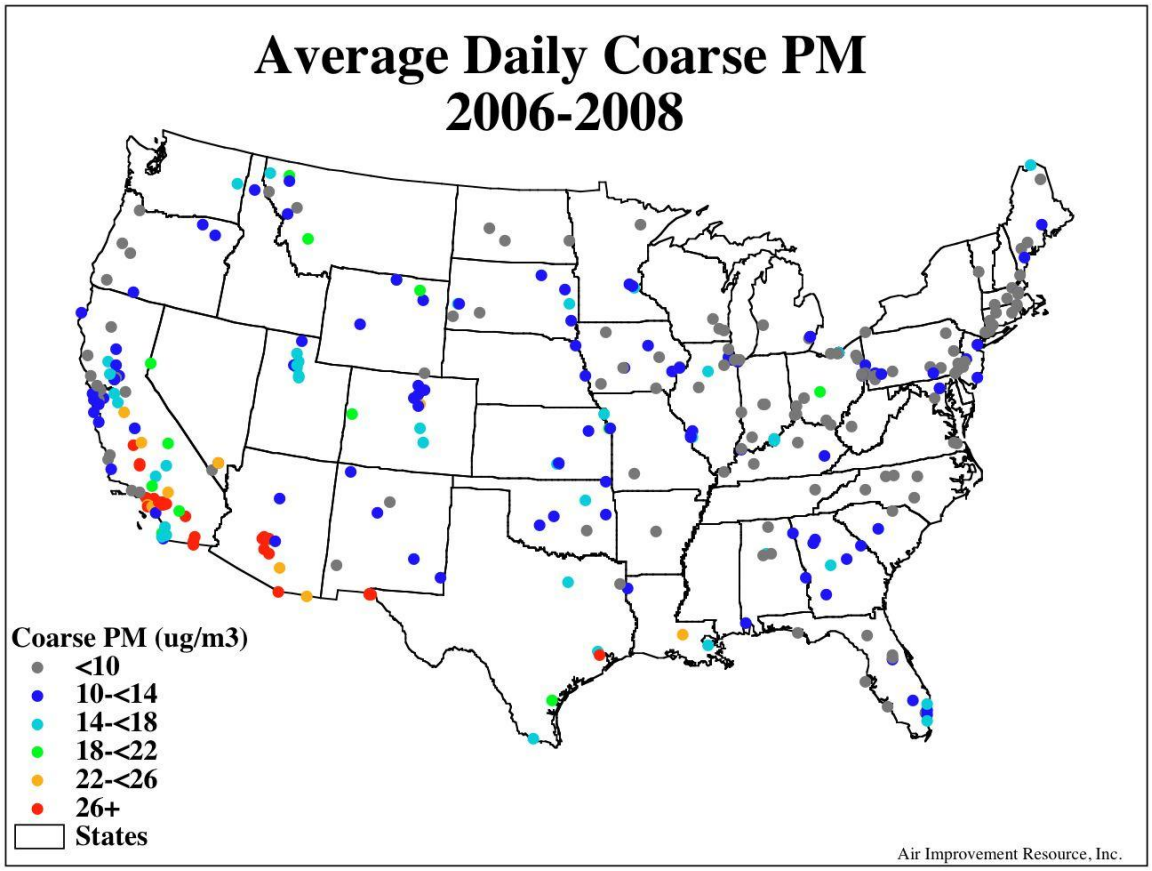

Figure 2: Annual mean $PM_{10-2.5}$ concentrations. Computed site by site values for 2006-2008 based on EPA's existing database.

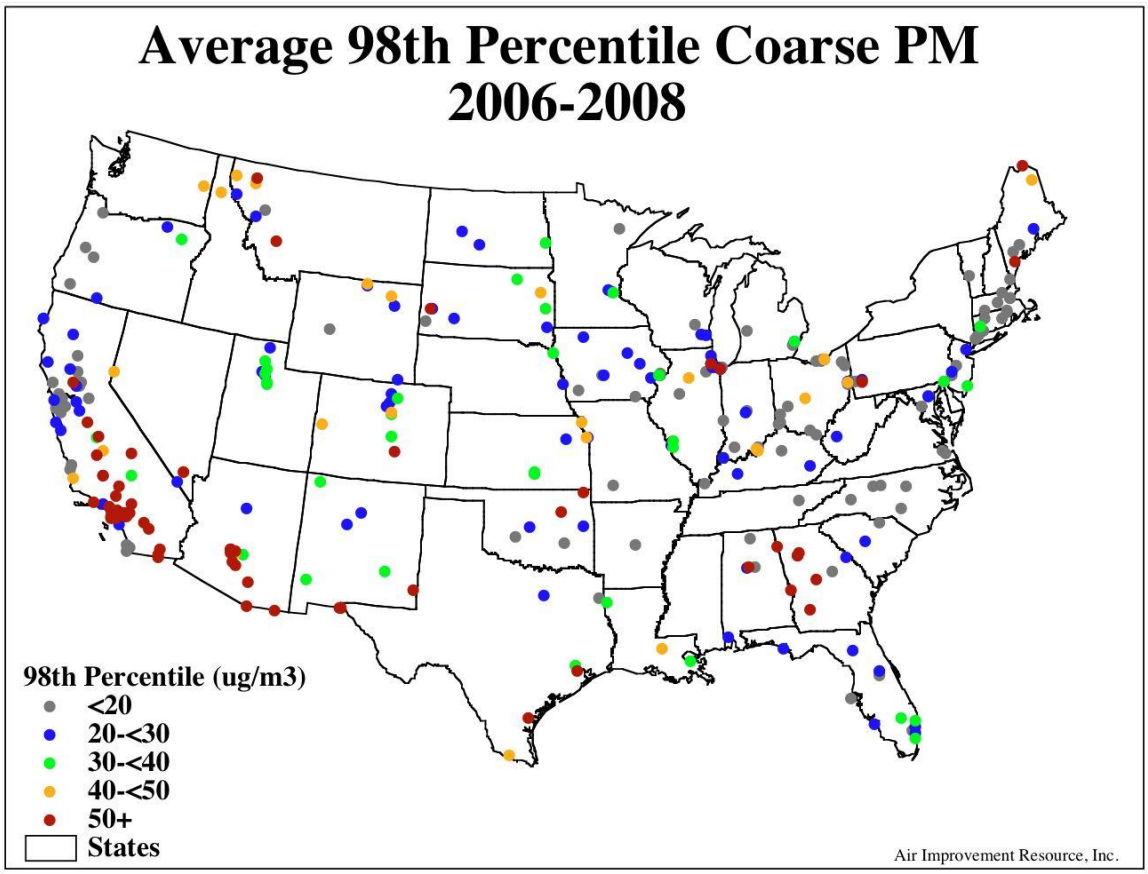

Figure 3: 98th percentile $PM_{10-2.5}$ concentrations. Computed site by site values for 2006-2008 based on EPA's existing database.

The New Multi-City Study

The discussion of coarse PM in the second draft PM ISA relies heavily on the recent Zanobetti and Schwartz (2009)[5] study ("the Z/S study"). Therefore, a detailed review and discussion of that analysis is appropriate. The draft ISA already notes a number of important issues and limitations in the interpretation of the reported results. The draft ISA particularly notes the lack of analyses of potential confounders of the $PM_{10-2.5}$-mortality relationship or of the influence of model specification on $PM_{10-2.5}$ risk estimates.[10] In addition, the concern is raised that models that include both $PM_{10-2.5}$ and $PM_{2.5}$ may suffer from instability due to collinearity. Questions and concerns arise at each stage of the analysis, starting with the exposure metric used for the analysis. Since the Z/S study estimated coarse PM by subtracting county average-fine PM from county average PM_{10}, the ISA notes, "Specifically, it is not clear how the computed $PM_{10-2.5}$ measurements used by Zanobetti and Schwartz (2009) compare with the $PM_{10-2.5}$ concentrations obtained by directly measuring $PM_{10-2.5}$ using a dichotomous sampler, or the $PM_{10-2.5}$ concentrations computed using the difference of PM_{10} and $PM_{2.5}$ measured at co-located samplers."[11]

The spatial and temporal variations in $PM_{10-2.5}$ associations reported by Z/S raise additional concerns. For example, finding no associations in regions where $PM_{10-2.5}$ levels are high but reporting associations with mortality in areas of the U.S. where $PM_{10-2.5}$ levels are not high was noted as not consistent with the findings of the $PM_{10-2.5}$ studies evaluated in the 2004 PM Criteria Document.[12] The finding of strong seasonality in the associations raised additional questions, leading the draft ISA to call for more data "to characterize the chemical and biological components that may modify the potential toxicity of $PM_{10-2.5}$."[13]

Each of the major concerns with the study is discussed in turn.

<u>Lack of validation of coarse PM metric</u>

The Z/S analysis should not be relied upon by the Agency until the measurement issues raised in these comments are addressed. As noted above, the ISA notes it is not clear how the computed $PM_{10-2.5}$ measurements used by Z/S compare with the $PM_{10-2.5}$ concentrations obtained by directly measuring $PM_{10-2.5}$ using a dichotomous sampler, or the $PM_{10-2.5}$ concentrations

[10] ISA, p. 6-301.

[11] ISA, p. 6-294.

[12] ISA, p. 6-301.

[13] ISA, p. 6-318.

computed using the difference of PM_{10} and $PM_{2.5}$ measured at co-located samplers. This is an issue because there are complications involved in dealing with the PM data obtained in the national monitoring network. For example, in evaluating the spatial distribution of coarse PM concentrations around the U. S., Chapter 3 of the draft ISA indicates:

> Since $PM_{10-2.5}$ is not routinely measured and reported to AQS, co-located PM_{10} and $PM_{2.5}$ measurements from the AQS network were used to investigate the spatial distribution in $PM_{10-2.5}$ Only low-volume FRM or FRM-like samplers were considered in calculating $PM_{10-2.5}$ to avoid complications with vastly different sampling protocols (e.g., flow rates) between the independent PM_{10} and $PM_{2.5}$ measurements. The same 11+ days per quarter completeness criterion discussed above was applied to the PM_{10} and $PM_{2.5}$ measurements. The $PM_{2.5}$ concentrations are reported to AQS at local conditions whereas the PM_{10} concentrations are reported at standard conditions. Therefore, prior to calculating $PM_{10-2.5}$ by subtraction, the PM_{10} AQS data were adjusted to local conditions on a daily basis using temperature and pressure measurements from the nearest National Weather Service station.[14]

The procedure described in Z/S for handling the data does not indicate that the PM_{10} data was corrected to local conditions or whether any seasonal completeness criterion was applied to the data. There is also no discussion of any complications due to using vastly different flow rates in the PM_{10} and $PM_{2.5}$ samplers. The complications arise because both PM measures are filter-based techniques that are prone to positive and negative artifacts due to pick-up of material absorbed from the gaseous state or loss of semi-volatile material. EPA limits the analysis of coarse PM concentrations in Chapter 3 to samplers with similar flow rates so that the impact of artifacts would be similar for both samplers.

The procedure explained by Z/S indicates that when more than one monitor was available in one county, the 24-hr integrated mass concentrations were averaged over the county, after first excluding any monitor that was not well correlated with the others ($r < 0.8$ for two or more monitor pairs within a county). It is not clear how much data was lost and potential bias introduced because of the exclusion of monitors.

Z/S required that at least 265 days of data in at least 1 year be available in order to be included in the study. They found 112 cities with at least 265 days of monitoring of $PM_{2.5}$ per year and at least 300 days of mortality data per year. Z/S indicate that PM coarse was estimated by differencing the countywide averages of PM_{10} and $PM_{2.5}$. They note that PM coarse was available for fewer locations, because less monitoring of PM_{10} is currently being done. They

[14] ISA, p. 3-64,65.

report that 47 locations met their criteria for PM coarse. In fact, the amount of data being gathered in the U. S. for $PM_{2.5}$ and PM_{10} are very similar in recent years with about 400,000 monitor-days of data for each metric. If Z/S finds fewer cities that meet their criteria it is due to either the requirement that the data be concurrent or the restrictions due to their exclusion procedure.

The database used to construct Figures 2 and 3 above consisted of 102,137 monitor-days that had co-located and concurrent $PM_{2.5}$ and PM_{10} data. Since $PM_{10-2.5}$ was determined by difference, the possibility exists that a negative $PM_{10-2.5}$ concentration might be calculated. In fact 9.5 % of the calculated values were negative. They were set to zero in the analysis. Negative values occur because of taking the difference between two measurements each of which has substantial uncertainty. The uncertainties and artifacts that lead to a substantial portion of negative numbers would also affect the full distribution of coarse measurements. Thus, there is greater uncertainty in the coarse PM data than in the $PM_{2.5}$ used by Z/S. There is also substantial uncertainty in the extreme values of $PM_{10-2.5}$ determined by difference that would hinder the interpretation of any study that used the difference method to measure $PM_{10-2.5}$.

Lack of information on extent of coarse PM data used

Since they report associations by season, it is not clear how much data was available in each season in each city. Although Z/S indicated that their analysis included data from 1999-2005, the limitations and exclusions they described suggest that much less than six years of data was involved for the analysis of coarse effects. Z/S also indicate that there were 5,609,349 total deaths in the 112 cities during the study period 1999-2005, leaving the impression that this was a large study with much statistical power. However, the extent of actual data used in the analysis of coarse effects is not reported.

Lack of consideration or evaluation of biases or uncertainty due to model choice

Given the knowledge that the choice of smoothing algorithm can substantially affect the results, Z/S either failed to evaluate alternative specifications or did evaluate alternative specifications and reported only one -- presumably the model that gave the largest effects. In either case, this is not following "best practices" for conducting and reporting such analyses.

The draft ISA notes the lack of studies regarding the influence of model specification on coarse PM health associations.[15] This is important because there is substantial evidence that the specification of the statistical model can either bias or add uncertainty to the results of time series

[15] ISA, p. 6-301.

studies. While this general issue is discussed in greater detail in a subsequent section of these comments, a couple of examples relevant to coarse PM associations are particularly relevant.

Klemm and Mason 2003[16] showed that the degree of association of coarse PM with mortality was substantially affected by the choice of temporal smoothing algorithm. Using the data from the Schwartz, Dockery and Neas 1996 analysis of six Eastern cities, Klemm and Mason showed that with a greater number of degrees of freedom, the association for fine PM was substantially reduced and the association for coarse PM became zero or negative. Klemm *et al*. 2004[17] in an analysis of fine and coarse PM and other air pollutants in Atlanta showed that are "substantial differences in terms of mean effects and statistical significance depending on the number of knots used to smooth time." Klemm *et al*. 2004 conclude that:

> Results can differ significantly across model specifications. We believe it is very important to consider a comprehensive set of models in future analyses, and the results of all analyses should be presented and considered in subsequent inferences.

Lack of consideration or evaluation of potential confounders

Z/S present the results of single-pollutant models for fine and coarse PM and one multipollutant model with both fine and coarse PM but they do not discuss the potential confounding by other co-existing pollutants. This omission is a severe limitation.

The ISA notes the lack of analysis of other air pollutants as potential confounders.[18] Klemm *et al.* 2004 note:

> It is axiomatic that effects attributed to a given pollutant based on a single-pollutant regression will include effects from any other pollutants with which the given pollutant may be correlated. Thus, single-pollutant regressions may be a useful screening tool but cannot provide valid judgments as to the relative importance of a given pollutant.

[16] Klemm RJ; Mason R. (2003). Replication of reanalysis of Harvard Six-City mortality study, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp.165-172..

[17] Klemm RJ; Lipfert FW; Wyzga RE; Gust C. (2004). Daily mortality and air pollution in Atlanta: two years of data from ARIES. Inhal Toxicol, 16 Suppl 1: 131-141.

[18] ISA, p. 6-301.

Lack of information on collinearity of fine and coarse PM

Although Z/S present the results of a two-pollutant model with fine and coarse PM, they do not report any information of the correlation between the two metrics. This limits the interpretation of the results since, as the ISA notes, models that include both $PM_{10-2.5}$ and $PM_{2.5}$ may suffer from instability due to collinearity.[19]

Lack of information on individual city–specific associations

Although Z/S estimated 188 city and season specific associations for each of several health outcomes, only pooled results are presented. It is particularly important that the pattern in the individual associations be available for inspection and analysis. This could have been done in the supplementary material without overburdening the main paper. The ISA includes Bayesian shrunken city-specific estimates in Figure 6-29 based on data obtained from the authors. If Bayesian-adjusted estimates are included in the final ISA, the un-adjusted city specific estimates must also be included. The range and pattern in individual-city associations in multi-city studies is highly relevant and analyses that omit such information are unacceptable.

Many other multi-city studies of PM and other air pollutants are available. Where individual-city results are included, they all report a very wide range of individual associations ranging from strongly positive to strongly negative. Such a range is biologically implausible and is an indication of a previously unacknowledged stochastic variability. Considerable detail on this point was included in AIR comments on the first draft PM ISA.[20] One of the implications of the stochastic variability is that any one individual-city result is unreliable.

Lack of adequate discussion of spatial and temporal pattern of associations

The pattern of associations reported in Z/S is not consistent with the mass of coarse PM, per se, causing or contributing to premature mortality. The Z/S study, if taken at face value, indicates that coarse PM is dangerous in the spring but benign in the winter. It indicates that coarse PM is dangerous in the East but benign in the West. The pattern does not demonstrate or provide support for a consistent coarse PM mortality association.

Z/S present combined results for coarse PM associations by season, by climatic region, and by

[19] ISA, p. 6-131.

[20] Heuss, J. M. and Wolff, G. T. (2009) Review and Critique of the U.S. Environmental Protection Agency's First External Review Draft of the "Integrated Science Assessment for Particulate Matter." Prepared for the Alliance of Automobile Manufacturers and the Engine Manufacturers Association, March 13, 2009.

cause of death. The pattern of associations is intriguing. In the analysis by seasons (Table 2), only spring has a significant positive combined association for all-cause mortality. The combined association in winter is actually negative. Summer and fall have smaller positive but non-significant associations. The overall association with a four-day distributed lag is actually less than for the average of day 0 and day 1. As shown in Figure 2, this arises because the only consistently positive association is on day 1; day 0 and day 3 actually have negative associations with all-cause mortality. The breakdown by cause of death suggests that the strongest positive associations are for respiratory deaths especially in the spring. However, there are many non-significant categories and seasons in the various combinations in Table 2.

With regard to region, the strongest positive associations are in the East particularly in the North Central and Northeast U. S. The associations in the two regions of the Southwest and along the West Coast, noted as the "Dry" and "Mediterranean" regions, are actually negative. The Western cities included in the coarse analysis are Los Angeles, Bakersfield, Sacramento, Seattle, Spokane, Phoenix, and Albuquerque. Z/S note these regional differences and posit "this suggests that there are regional variations in the toxicity of coarse particles that require further study." The ISA also notes the lack of association in high coarse PM cities in the Western U. S.[21] The mean and 98th percentile coarse PM concentrations (shown in Table 6-15 of the ISA) in the cities that, when pooled, show slight negative associations with mortality are as high as or higher than the concentrations in the cities in the Eastern U. S. where positive results are reported.

Lack of discussion of the biological plausibility of the pattern of coarse associations

There is little or no discussion in Z/S of the biological plausibility of the pattern of coarse associations they report. There is some general discussion of evidence for the toxicity of particles, but nothing persuasive that addresses the stark contrasts in result by season and region. Z/S offer some possible rationalizations for their findings but they are not persuasive. They note that "It may be possible that coarse particles are coated with different substances in different regions," without offering any evidence to support the hypothesis or how it could explain the stark seasonal and regional differences. They note that mild temperatures are associated with greater indoor penetration and that this may explain the maximum associations in spring. While there may be greater penetration indoors in spring, penetration issues cannot explain the lack of an association in winter or the regional differences. Penetration indoors may be slightly higher in some seasons compared to others but it occurs in all seasons and in all regions. Z/S note that one possible explanation for the lower associations in the Mediterranean region is greater measurement error due to the large size of counties in California where people may live farther away from the monitors. This hypothesis should be tested by evaluating the

[21] ISA, p. 6-294.

placement of monitors in relation to the population. In addition, Z/S excluded monitors that were not correlated so the likelihood of remote monitors dominating the data is very small. Event the argument that measurement error biases towards the null is suspect given the very wide range of individual associations in single-city estimates within multi-city studies. Thus, the rationalizations offered by Z/S are insufficient to explain the pattern of results.

Summary

Each of the concerns raised in the draft ISA (and additional concerns documented in these comments) are legitimate reasons to place less emphasis on the Zanobetti and Schwartz analysis in the final ISA. When combined, they document that the analysis is not a reliable basis for evidence of coarse PM health effects. At most it can be considered as suggestive but not sufficient to provide support for a coarse PM standard at this time.

Other Mortality Studies Used by EPA

The studies that EPA cites to support the Zanobetti and Schwartz claims are summarized in Figure 6-30 on page 6-300 in the ISA. However, accepting the results in Figure 6-30 at face value is misleading because in most cases the investigators haves reported a variety of results, both positive and negative, but EPA tends to select the one that shows the largest effect. This will be demonstrated below.

Excluding the Z/S results from Table 6-30, there are 20 additional results. Although 19 of these 20 results indicate a positive relationship between $PM_{10-2.5}$ and mortality, only 2, Ostro *et al.* (2003)[22] and Mar *et al*. (2003)[23] report results that are barely statistically significant at the 95th percentile confidence level. That means that for the other 18 results reported in Figure 6-30 a zero effect cannot be ruled out based on conventionally accepted statistical practices. However, there are reasons to believe that the uncertainties in Ostro *et al.* and Mar *et al.* are larger than the stated statistical uncertainty. Ostro *et al*. used estimated $PM_{10-2.5}$ concentrations for the 10-year record that were based on a statistical relationship developed from a 2.5 year period with both PM_{10} and $PM_{2.5}$ measurements. Therefore there is an unaccounted for uncertainty in the estimated $PM_{10-2.5}$ concentrations. Mar *et al*. only reported a statistically significant relationship

[22] Ostro BD; Broadwin R; Lipsett MJ. (2003). Coarse particles and daily mortality in Coachella Valley, California, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp.199-204.

[23] Mar TF; Norris GA; Larson TV; Wilson WE; Koenig JQ. (2003). Air pollution and cardiovascular mortality in Phoenix, 1995-1997, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp. 177-182.

for lag zero, but not for any longer lags. The existence of a lag zero relationship only is inherently problematic because the victim's exposure depends on the time of death.

The Klemm and Mason (2003)[24] study (which is erroneously cited as Mason *et al*. (2003) in Figure 6-30) is an example. EPA cites one value which is appears to be from the Generalized Linear Model for the pooled six city estimates. Klemm and Mason, however present the results of 42 different model runs. They show the results from six different statistical models for each of the six cities plus a pooled cities estimate. Of the 42 outcomes, only one outcome has a statistically significant positive result and 14 are negative. This paper is an excellent example of how model selection can greatly influence the sign and magnitude of the outcome. Model selection will be addressed later in these comments.

EPA is complimentary of the Burnett *et al*. (2004)[25] study stating, "one well conducted multicity Canadian study also provides evidence for an association between short-term exposure to $PM_{10-2.5}$ and risk estimates." However, upon close examination, the evidence is very weak. The single model $PM_{10-2.5}$ effect estimate (% mortality increase per 10 $\mu g/m^3$ increase in $PM_{10-2.5}$) is 0.65%, with a 95^{th} percentile confidence interval of -0.1 to 1.4 which is not statistically significant. When NO_2 is added into a two pollutant model, the effect estimate is further decreased to 0.31%, and becomes even more insignificant as the confidence interval becomes -0.49 to 1.1. In addition the coefficient for NO_2 is statistically significant. The authors dismiss any significance of a $PM_{10-2.5}$/mortality relationship and instead focus on the NO_2 relationship. In their conclusions the authors attribute the NO_2/mortality relationship not to cause and effect, but that NO_2 is a surrogate for all combustion emissions.

Burnett and Goldberg (2003)[26] present the pooled results for $PM_{10-2.5}$/mortality relationships using seven different models for eight Canadian cities. In this case, none of the results were statistically significant, but there were considerable model to model variations. The authors state that the association was sensitive to the method of statistical analysis, which underscores the importance of the model selection issue.

The remaining results in Figure 6-30 are all single city studies.

---

[24] Klemm RJ; Mason R. (2003). Replication of reanalysis of Harvard Six-City mortality study, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp.165-172.

[25] Burnett RT; Stieb D; Brook JR; Cakmak S; Dales R; Raizenne M; Vincent R; Dann T. (2004). Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities. Arch Environ Occup Health, 59: 228-236.

[26] Burnett RT; Goldberg MS. (2003). Size-fractionated particulate mass and daily mortality in eight Canadian Cities, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp. 85-89.

Villeneuve *et al*. (2003)[27] obtained data on a comprehensive suite of pollutants (seven PM parameters and six criteria gases) from Vancouver, Canada for a 13-year period. They explored single-pollutant models for each constituent for all-cause, cardiovascular and respiratory mortality using lags up to three days and a three-day average. Of the twelve models run, only cardiovascular mortality at lag zero produced a statistically significant positive result for $PM_{10-2.5}$.In fact, it was the only statistically significant model for any pollutant measurement and cardiovascular mortality. Further, they ran a total of 132 different single pollutant models. Using the 95% confidence interval as a criterion, the number statistically significant models that would be expected to be found just by chance is seven. Of the 132, only four actually produced a statistically significant result. Clearly the $PM_{10-2.5}$ relationship could have occurred by chance. A flag should also be raised since the significant result at lag zero is followed by a highly insignificant result for lag 1.

Wilson *et al*. (2007)[28] examined a number of single pollutant models to determine the $PM_{10-2.5}$ cardiovascular mortality relationship in Phoenix. They observed the relationship at lags from zero to five days plus a six-day distributive lag. In addition, they examined the effect as a function of distance from the central monitoring station by creating three rings based on zip codes. The central ring contained the monitor. Adjacent zip codes formed the middle ring while zip codes beyond the middle ring formed the outer ring. This gives a total of 21 models that were examined. Of the 21 models, only three produced a statistically significant positive relationship. All three were for the middle ring at lags 1, 2, and the 6-day distributive lag. Since the spatial distribution of $PM_{10-2.5}$ is known to be quite heterogeneous over urban areas, one would expect the strongest effect of $PM_{10-2.5}$ to be in the central ring where the monitor is located. Consequently these results are not logical and should be viewed with caution. Another issue raised by the authors that adds to the uncertainty of these results is the fact that in order to compute a 6-day distributive lag data was needed for every day. Since $PM_{10-2.5}$ was missing for 12% of the days, it had to be estimated from the $PM_{2.5}$ data which the authors admit was poorly correlated with $PM_{10-2.5}$.

In Atlanta, Klemm *et al*. (2004)[29] conducted an extensive analysis of many PM components including $PM_{10-2.5}$ and used various modeling methods. None of the mortality/$PM_{10-2.5}$ relationships were statistically significant.

---

[27] Villeneuve PJ; Burnett RT; Shi Y; Krewski D; Goldberg MS; Hertzman C; Chen Y; Brook J. (2003). A time series study of air pollution, socioeconomic status, and mortality in Vancouver, Canada. J Expo Sci Environ Epidemiol, 13: 427-435.

[28]Wilson WE; Mar TF; Koenig JQ. (2007). Influence of exposure error and effect modification by socioeconomic status on the association of acute cardiovascular mortality with particulate matter in Phoenix. J Expo Sci Environ Epidemiol, 17: S11-S19.

[29] Klemm RJ; Lipfert FW; Wyzga RE; Gust C. (2004). Daily mortality and air pollution in Atlanta: two years of data from ARIES. Inhal Toxicol, 16 Suppl 1: 131-141.

Ito (2003)[30] presents the results for mortality/$PM_{10-2.5}$ for total, respiratory and cardiovascular mortality from two different statistical models. None of the six model outcomes were statistically significant.

Fairley (2003)[31] use two different statistical models to measure the mortality/$PM_{10-2.5}$ relationship in Santa Clara County for lag zero and one for total, respiratory and cardiovascular mortality. None of the results were statistically significant and the lag 1 results were negative.

Chock *et al*. (2000)[32] examined the $PM_{10-2.5}$ relationship in Pittsburgh for two populations, 0 – 74 years of age and > 74 years. They explored both single-pollutant and multiple-pollutant models and the effect of season. None of the outcomes were statistically significant.

Lipfert *et al*. (2000)[33] also examined the relationship in Philadelphia for two populations, < 65 and 65+ years of age. He found no statistically significant results.

The ISA also cites three additional studies that examined the $PM_{10-2.5}$/mortality relationship but were not included in Figure 6-30. Two of the studies, Slaughter *et al*. (2005)[34] in Spokane, WA and Kettunen *et al*. (2009)[35] in Helsinki, Finland, showed no relationship or mostly negative relationships, respectively. The third study, Perez *et al*. (2008)[36], did find a positive and statistically significant relationship only on days when Barcelona, Spain was impacted by Saharan dust outbreaks. However, they caution that these effects may be due to the high content of active bioaerosols, which are known to be a component of Saharan dust, rather than the crustal

[30] Ito K. (2003). Associations of particulate matter components with daily mortality and morbidity in Detroit, Michigan, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp.143-156.

[31] Fairley D. (2003). Mortality and air pollution for Santa Clara County, California, 1989-1996, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp.97-106.

[32] Chock DP; Winkler SL; Chen C. (2000). A study of the association between daily mortality and ambient air pollutant concentrations in Pittsburgh, Pennsylvania. J Air Waste Manag Assoc, 50: 1481-1500.

[33] Lipfert FW; Morris SC; Wyzga RE. (2000). Daily mortality in the Philadelphia metropolitan area and size classified particulate matter. J Air Waste Manag Assoc, 50: 1501-1513.

[34] Slaughter JC; Kim E; Sheppard L; Sullivan JH; Larson TV; Claiborn C. (2005). Association between particulate matter and emergency room visits, hospital admissions and mortality in Spokane, Washington. J Expo Sci Environ Epidemiol, 15: 153-159.

[35] Kettunen J; Lanki T; Tiittanen P; Aalto PP; Koskentalo T; Kulmala M; Salomaa V; Pekkanen J. (2007). Associations of fine and ultrafine particulate air pollution with stroke mortality in an area of low air pollution levels. Stroke, 38: 918-922.

[36] Perez L; Tobias A; Querol X; Kunzli N; Pey J; Alastuey A; Viana M; Valero N; Gonzalez-Cabre M; Sunyer J. (2008). Coarse particles from Saharan dust and daily mortality, 19: 800-807.

materials themselves. The Alexis *et al.* (2006)[37] study discussed in the ISA provides confirming information that the biological components of coarse PM are responsible for macrophage-meditated responses to coarse PM. Thus, to the extent that coarse PM may have human health effects, biologic components that are not amenable to reduction by control of anthropogenic components may be responsible.

Summary

As with Z/S, the collection of $PM_{10-2.5}$/mortality studies do not make a compelling case to support a stringent $PM_{10-2.5}$ NAAQS. Figure 6-30 does not tell the complete story of the relationships between $PM_{10-2.5}$ and mortality. Most of the studies listed in the Figure presented more results than those contained in the Figure. When all of them are examined, few are statistically significant and many show no relationships and some show negative relationships. If all the results from single-pollutant studies were included and the range of individual-city estimates in the multi-city studies were included in the figure, a more complete picture of the distribution of results would be available, enabling a better determination of the consistency (or lack of consistency) in the data.

Publication Bias and Model Selection Issues

The draft ISA mischaracterizes the consistency of health effects as estimated from epidemiology. Model selection uncertainty, publication bias, and potential confounding cloud the interpretation of the epidemiological data

Model selection uncertainty impacts the appearance of consistency

In interpreting the epidemiological evidence, the draft ISA downplays major new findings concerning uncertainty due to model selection issues. Model selection uncertainty relates to confounding of air pollutant associations by temporal trends, weather and co-pollutants. During the last ozone review, EPA acknowledged that the uncertainties in the estimates of pollutant effects are understated by consideration of the statistical uncertainty of the fitted model alone. Much more uncertainty arises from the lack of information regarding the choice of appropriate models for adjusting confounding by other covariates, and the choice of appropriate lag structures. As Lumley and Sheppard (2003) point out:

[37] Alexis NE; Lay JC; Zeman K; Bennett WE; Peden DB; Soukup JM; Devlin RB; Becker S. (2006). Biological material on inhaled coarse fraction particulate matter activates airway phagocytes in vivo in healthy volunteers. J Allergy Clin Immunol, 117: 1396-1403.

> Estimation of very weak associations in the presence of measurement error and strong confounding is inherently challenging. In this situation, prudent epidemiologists should recognize that residual bias can dominate their results. Because the possible mechanisms of action and their latencies are uncertain, the biologically correct models are unknown. This model selection problem is exacerbated by the common practice of screening multiple analyses and then selectively reporting only a few important results.[38]

Others have also pointed out the critical importance of model choice, particularly when effect estimates are small. For example, Smith *et al*. caution:

> From a statistical point of view, the common epidemiological practice of choosing variables (including lagged variables, co-pollutants, etc.) that maximize the resulting effect estimates is a dangerous approach to model selection, particularly when the effect estimates are close to 0 (i.e., RR close to 1).[39]

Smith *et al.* note that Lumley and Sheppard (2000)[40] showed that the effect of choosing lags in this fashion has a bias which is of the same order of magnitude as the relative risk being estimated. Morris has also shown a similar result.[41] He showed using the theory of extreme value distributions that evaluating multiple lags and reporting the maximum effect, even when there is no underlying effect, can yield estimates of effect size with a magnitude similar to those routinely reported for particles.

The revised analyses necessitated by the problems with the commonly used software for time-series analyses clearly show that methods used for controlling temporal trends and weather can profoundly affect the results. To make matters worse, there appears to be no objective statistical test to determine whether these factors have been adequately controlled. The HEI Expert Panel[42] for the re-analysis states, "Ritov and Bickel (1990)[43] have shown, however, that for any

---

[38] T. Lumley and L. Sheppard, (2003).Time series analyses of air pollution and health: straining at gnats and swallowing camels? Epidemiology, 14, 13-14, 2003.

[39] R. Smith , P. Guttorp, L. Sheppard, T. Lumley, and N. Ishikawa. (2001) Comments on the Criteria Document for Particulate Matter Air Pollution, Northwest Research Center for Statistics and the Environment Technical Report Series No. 66, July 2001.

[40] T. Lumley and L. Sheppard. (2000). Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses, Environmetrics, 11, 705-717.

[41] R. Moris. (2001) Airborne Particulates and Hospital Admissions for Cardiovascular Disease: A Quantitative Review of the Evidence. Environ. Health Perspect, **109**, Supplement 4, 495-500.

[42] Health Effects Institute (2003) Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, Health Effects Institute, Cambridge, Massachusetts, at 267, 269.

[43] Y. Ritov and P. Bickel. (1990). "Achieving information bounds in non- and semi-parametric models," Ann. Stat., 18, 925-938.

continuous variable, no strictly data-based (i.e., statistical) method can exist by which to choose a sufficient number of degrees of freedom to insure that the amount of residual confounding due to that variable is small. This means that no matter what statistical method one uses to select the degrees of freedom, it is always logically possible that even if the true effect of pollution is null, the estimated effect is far from null due to confounding bias." The HEI Expert Panel concluded further, "Neither the appropriate degree of control for time, nor the appropriate specification of the effects of weather, has been determined for time-series analyses". In other words, it is impossible to adjust temporal trends without accurate information from external sources regarding the appropriate degrees of freedom to use. Such information, however, simply does not exist.

In agreement with these findings, as noted above, Klemm *et al.* 2004 concluded that:

> Results can differ significantly across model specifications. We believe it is very important to consider a comprehensive set of models in future analyses, and the results of all analyses should be presented and considered in subsequent inferences.

With regard to uncertainty due to model selection, the Koop and Tole 2004[44] Bayesian model averaging study, which thoroughly evaluated model selection in one city for many air pollution and meteorological variables, concludes:

> Point estimates of the effect of numerous air pollutants all tend to be positive, albeit small. However, when model uncertainty is accounted for in the analysis, measures of uncertainty associated with these point estimates became very large. Indeed they became so large that the hypothesis that air pollution has no effect on mortality is not implausible. On the basis of these results, we recommend against the use of point estimates from time-series data to set regulatory standards for air pollution exposure.

Koop and Tole showed that a single model based on a sequence of hypothesis tests will overestimate the certainty of the results. This is not a new finding in the statistical literature. The 2004 CD noted that "testing many models to identify the model with the best fit can lead to an underestimation of uncertainty" and "if the observed confidence intervals were arrived at by a number of prior model specification searches, eliminating some worse fitting models, the true interval may well be wider."[45]

---

[44] G. Koop and L. Tole, (2004) Measuring the Health Effects of Air Pollution: to What Extent Can We Really Say that People are Dying from Bad Air, J. of Environmental Economics and Management, 47, 30-54.

[45] PM CD at page 8-226.

Despite the issues concerning uncertainty due to model selection in the 2004 PM CD, in the HEI Special Panel report, and in the publications referenced above, the first draft ISA was essentially silent on this issue (and any changes in the relevant science) except to acknowledge[46] in the introductory section on methodology in Chapter 6 that, to date, a clear consensus as to the extent of modeling required to accurately control for weather or confounding by other pollutants or to measure PM-mortality/morbidity effects has not been reached. Unfortunately this section was removed in the second draft ISA. The final ISA must acknowledge and address the uncertainty due to model selection as it affects the interpretation of epidemiological results.

Publication bias inflates the magnitude of effects and impacts the appearance of consistency

Since there is substantial evidence that publication bias inflates the apparent magnitude and consistency of air pollution health effects in single-city studies, the final PM ISA must address and discuss the important impact of publication bias in the integrative sections, not only in the introduction.

Publication bias is another major issue in interpreting the epidemiology. The commentary by Goodman concerning meta-analyses is particularly insightful.[47] He noted a factor of at least three difference between the results of ozone meta-analyses and the NMMAPS data which are not affected by publication bias. Goodman concludes that the implications of an EPA-sponsored exercise of funding three separate meta-analyses "go far beyond the question of the ozone mortality effect." He cautions that "depending on published single-estimate, single-site analyses are an invitation to bias." He notes that "the most plausible explanation is the one suggested by the authors, that investigators tend to report, if not believe, the analysis that produces the strongest signal; and in each single-site analysis, there are innumerable model choices that affect the estimated strength of that signal." A separate review by a panel of ten knowledgeable scientists[48] concluded that "taken together, the meta-analyses provide evidence of a disturbingly large publication bias and model selection bias."

Similarly, Anderson *et al.* 2005[49] concluded that publication bias is present in single-city time series studies of ambient particles. After correcting for publication bias, they still report a positive association. However, they also note that the regression estimates from the multi-city

[46] PM ISA at pages 6-2 and 6-3.

[47] S. Goodman. (2005) ,The Methodologic Ozone Effect, Epidemiology, 16, 430-435.

[48] Report of a Working Conference, Critical Considerations in Evaluating Scientific Evidence of Health Effects of Ambient Ozone, held in Rochester, New York, June 2007.

[49] H. Anderson, R. Atkinson, J. Peacock, M. Sweeting, and L. Marston (2005). Ambient Particulate Matter and Health Effects: Publication Bias in Studies of Short-Term Associations, Epidemiology, **16**, 155-163.

studies (which are not prone to publication bias) and the corrected single-city studies are approximately half of the mortality estimates of the mid-1990's, that the correction for publication bias may not be complete, and that differential selection of positive lags may also inflate estimates.

Thus, publication bias is a major concern inflating the size of any potential effect. As EPA has reviewed other criteria pollutants, the Agency has acknowledged[50] that the summary of health effects evidence is vulnerable to the errors of publication bias and multiple testing. The only reference in the first draft IPM SA to publication bias was found on page 6-11 in a discussion of the heart rate variability findings. The second draft acknowledges in Chapter 1 that:

> Publication bias is a source of uncertainty regarding the magnitude of health risk estimates. It is well understood that studies reporting non-null findings are more likely to be published than reports of null findings, and publication bias can also result in overestimation of effect estimate sizes (Ioannidis, 2008). For example, effect estimates from single-city epidemiologic studies have been found to be generally larger than those from multicity studies.[51]

Ioannidis (2005)[52] points out that the smaller the effect sizes in a scientific field, the less likely the research findings are to be true. He notes that if the true effect sizes are very small in a scientific field, this field is likely to be plagued by almost ubiquitous false positive claims. Ioannidis indicates that the greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true. He points out that flexibility increases the potential for transforming what would be "negative" results into "positive" results, introducing bias, Although Ioannidis addresses general issues in scientific research, the concerns and cautions he draws attention to apply directly to air pollution epidemiology where effect sizes are very small and model selection uncertainty provides the flexibility that can introduce a positive bias in the results.

[50] U. S. EPA, Second External Review Draft of Integrated Science Assessment for Oxides of Nitrogen-Health Criteria, EPA 600/R-07/093aB, March 2008 at page 3-2; U. S. EPA, Integrated Science Assessment for Oxides of Sulfur-Health Criteria, EPA/600/R-07/047F, September 2008 at pages 3-1 and 3-48.

[51] ISA at page 1-31.

[52] J. Ioannidis, "Why most published research findings are false," PLoS Med. 2005 August; 2(8): e124.

Although multi-city studies avoid publication bias, there are additional issues and concerns in their interpretation

In contrast to the pattern of associations in publications reporting single-city time-series results, there is a biologically implausible, very wide range in the PM/mortality or PM/morbidity associations in the individual cities included in multi-city studies, including a substantial portion of negative associations between air pollutants and heath endpoints. Numerous examples of the wide range were included in the AIR-Alliance comments on the first draft PM ISA.

There are two important conclusions to draw from this discussion. First, the acknowledgement of substantial stochastic variability such that individual single-city estimates are not reliable indicates that single-city associations should not be used to establish levels for air quality standards. Searching for the strongest association in a city with the lowest air pollution concentrations will identify the outliers in a very wide distribution of associations, not real health effects. Second, even the combined association in a large multi-city study is subject to error and uncertainty if confounding and model misspecification are present in the underlying analysis.

Although a wide range of associations (both positive and negative) is clearly evident in systematic studies, the authors of the studies either do not mention the range or mention it only in regard to there being heterogeneity in the results. However, the presence of a substantial portion of actually negative associations in individual cities in the multi-city studies is evidence for a larger degree of stochastic variation than heretofore acknowledged. A fundamental question is whether the combined effect estimate is meaningful when the individual city results vary in a biologically implausible way, and that question has not been addressed. The pros and cons of combining such disparate results needs to be carefully considered and discussed by EPA, CASAC, and the scientific community, including in the ISA.

An example, from the Z/S multi-city study is instructive. The draft ISA indicates that the Bayesian shrunken estimates in Figure 6-24 were derived from Z/S using the method of Tertre et al. 2005. Tertre *et al.* make a number of points that are relevant to interpreting both the Z/S study in particular and the PM epidemiology in general. They point out that "In the case where there is no true heterogeneity, variations in the city-specific effect estimates about the overall mean are purely stochastic." They note that "estimates from city-specific models are more specific, but have greater uncertainty than those provided from multicity analyses." They argue for using pooled estimates in multi-city studies because "the use of the local estimate is subject to too much noise to be reliable." However, they also point out that "each of the methods described in this work depends heavily on the assumption that city-specific models from which city-specific estimates were derived have precluded at least the known sources of bias, such as confounding, model misspecification, etc."

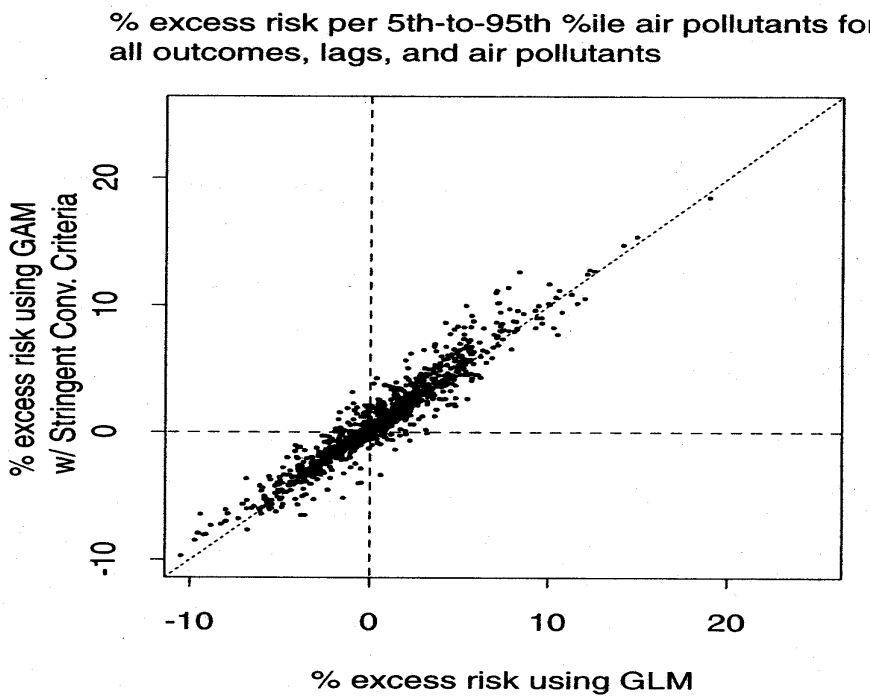

Figure 4: Figure 2 from Ito (2003).[54]

Additional evidence for substantial stochastic variation comes from an important new HEI study[53] that evaluated coherence between the time-series associations of mortality and hospital admissions in 14 cities. That study found little or no coherence between the PM_{10} mortality and morbidity associations and, importantly, found little or no correlation between the time series of health event counts (mortality and hospital admissions) in the various cities. As in other multi-city studies, the individual associations for mortality and morbidity covered a wide range from positive to negative. Given the substantial stochastic variation, the EPA needs to acknowledge and consider the wide range of associations with regard to both biological plausibility and the limitations on the use of time series and other epidemiological studies to set ambient standards. Another example of the wide-range of associations in systematic analyses comes from Ito 2003. When the statistical software issue noted in the ISA was raised and many time series studies were re-analyzed, Ito 2003 carried out a systematic re-analysis of the air pollution associations within a given city. Ito[54] re-analyzed the 1220 separate air pollution mortality and morbidity

[53] F. Dominici, *et al*. HEI Research Report 94, Part IV, 2005.

[54] Ito K. (2003). Associations of particulate matter components with daily mortality and morbidity in Detroit, Michigan, Health Effects Institute Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health, pp.143-156.

associations that were included in the original Lippmann et al. 2000 HEI study of Detroit. As shown in the figure above, there was a wide range of negative and positive risks in Detroit when all pollutants, lags, and endpoints were considered. Ito showed in separate figures that the wide range of associations occurred for each pollutant. Although the focus in the original Lippmann *et al.* study, as it is in almost all the published literature, was on the positive associations, Ito's plot shows that there are many negative associations in the data. Although there may be somewhat more positive associations than negative associations, there is so much noise or variability in the data, that identifying which positive associations may be real health effects and which are not appears beyond the capability of current methods. Moreover, in the Ito re-analysis, the overall pattern for each pollutant is similar so that one pollutant or one PM indicator is not implicated over any of the others.

Potential confounding by other pollutants is not adequately considered

The 2004 PM CD described a growing body of evidence from both epidemiological and toxicological studies supporting general conclusions of the form "PM (or one or more PM components), acting alone and/or in combination with gaseous co-pollutants, are likely causally related to observed ambient particle-associated health effects."[55]

The important qualifications "or one or more components " and "acting alone and/or in combination with co-pollutants" were included because of the large body of information showing that different particles have different toxicities and the continuing concern over discerning particle effects within the mix of correlated air pollutants. Despite the even larger number of studies related to PM health effects now available, neither of these concerns has been ameliorated. The pattern of epidemiological associations for coarse (as well as other PM metrics), as well as the findings from extensive toxicological studies, are not consistent with the assumption that all PM can be considered equally toxic. Similarly, the concern over confounding by other pollutants and the uncertainty in how to interpret single-pollutant model results persists. As Klemm *et al.* 2004 point out:

> It is axiomatic that effects attributed to a given pollutant based on a single-pollutant regression will include effects from any other pollutants with which the given pollutant may be correlated. Thus, single-pollutant regressions may be a useful screening tool but cannot provide valid judgments as to the relative importance of a given pollutant.

[55] PM CD at pages 8-338 and 9-79.

Therefore, all the determinations of causality in the final ISA must be qualified as they were in the 2004 Criteria Document to refer to "PM (or one or more PM component) acting alone and/or in combination with gaseous pollutants" rather than to PM mass alone.

A related concern is that the pattern of acute associations is remarkably similar for all the criteria pollutants, raising the issue of double or triple counting of health effects. For example, a similar pattern of associations was observed for all the major pollutants in single pollutant models in NMMAPS. For each pollutant, at each of the three lags evaluated, an implausibly wide range in individual-city associations from negative to positive was observed.[56]

As EPA has considered each criteria pollutant in turn, single-pollutant model results have been used to estimate the strength and consistency of association. Single-pollutant PM associations were used in the previous PM review as evidence of a causal relation between PM and respiratory endpoints.[57] In addition, single-pollutant ozone associations were used in the recent ozone review as evidence of a causal relation between ozone and the same respiratory endpoints.[58] The recently completed NOx and SOx ISAs[59] have also used selected single-pollutant model results as evidence of respiratory health effects from these pollutants.
In each case, the Agency has plotted selected individual city associations from the literature in the same manner and used the resulting figures to make the argument for respiratory health effects caused by the pollutant under consideration. Visual inspection of the figures referenced above reveals a remarkably similar pattern. This raises three issues:

First, as the air quality standard for each pollutant is reviewed in turn, the current practice of selecting specific studies and selecting specific single-pollutant associations for that pollutant results in a false appearance of consistency. If the various ISA documents for different pollutants are to be a scientifically sound basis for policy, more thorough analyses considering the full suite of pollutants is mandatory.

---

[56] While the full range of individual city results is presented in some multi-city studies, there has been a tendency to omit the individual city results in some recent publications. However, when the HEI sponsors requested that the individual city results from the re-analysis of NMMAPS be made available, the individual city results for PM_{10} and the various gases were posted on the Johns Hopkins website. The data show a remarkable similarity in that there was a biologically impossible wide range of associations from positive to negative for each pollutant on each lag that was evaluated. This data was also provided to EPA and CASAC during the PM review process; J. Heuss, Comments on the 4th Draft Criteria Document for Particulate Matter, AIR, Inc. comments prepared for the Alliance of Automobile Manufacturers, August 20, 2003.

[57] Figure 1 in proposed PM rule, 71 Federal Register 2620, January 17, 2006.

[58] Figure 1 in proposed ozone rule, 72 Federal Register 37818, July 11, 2007.

[59] July 2008 NOx ISA, EPA/600/R-08/071, at page 5-9, Figure 5.3-1; September 2008 SOx ISA, Figures 5-1 and 5-2 at pages 5-6 and 5-7.

Second, claiming health effects for each pollutant based on single-pollutant models raises the issue of double-, triple-, or even quadruple-counting of health effects.

Third, the remarkably similar pattern for each pollutant, together with the evidence of stochastic variability, model selection uncertainty, and publication bias, raise the concern that it is beyond the capability of current methods to identify which positive associations may be real health effects and which are not. Time-series epidemiology of air pollution associations is only capable of very blunt analysis. CASAC raised this issue in a June 2006 letter to the Administrator, noting that "because results of time-series studies implicate all of the criteria pollutants, findings of mortality time-series studies do not seem to allow us to confidently attribute observed effects specifically to individual pollutants."[60] The final ISA needs to acknowledge the stochastic variability in time series associations (both positive and negative) and consider the implications of that variability in both the interpretation of the epidemiology and its integration with results from controlled studies.

Summary and Conclusions

EPA appears to be laying the groundwork to establish a new 24-hour NAAQS for $PM_{10-2.5}$. It is doing this based on a new study based primarily on a new paper by Zanobetti and Schwartz supplemented by older papers that they previously characterized as having mixed results. EPA seems to favor a 24-hour NAAQS in the vicinity of 14 $\mu g/m^3$.

An examination of current ambient concentrations shows that the existing $PM_{10-2.5}$ database is extremely sparse and does not contain sufficient data to characterize the degree on nonattainment that would exist if such a NAAQS were promulgated or adequate to develop the required risk and exposure assessment. This is because there is no requirement for the routine deployment of the Federal Reference Method (FRM) sampler for, a dichotomous sampler. As a result, the only $PM_{10-2.5}$ data available is from the relatively few sites that deploy both PM_{10} and $PM_{2.5}$ monitors and the $PM_{10-2.5}$ concentrations are calculated by difference. However, this procedure is subjected to large uncertainties. The sparseness and uncertainty of the $PM_{10-2.5}$ data alone is sufficient reason for EPA not to mandate a $PM_{10-2.5}$ NAAQS until a reliable nationwide database exists.

In addition, the Zanobetti and Schwartz study is not a reliable basis for evidence of coarse PM health effects. They fail to demonstrate how the measure they use for $PM_{10-2.5}$ compares to the FRM and it is not clear how much data was included in their analysis. Zanobetti and Schwartz

[60] R. Henderson, CASAC letter, EPA-CASAC-06-07, June 5, 2006 at page 3.

do not consider or evaluate biases or uncertainty due to model selection or potential confounders. Although Zanobetti and Schwartz present the results of a two-pollutant model with fine and coarse PM, they do not report any information on the correlation between the two metrics. This limits the interpretation of the results since, as the ISA notes, models that include both $PM_{10-2.5}$ and $PM_{2.5}$ may suffer from instability due to colinearity. Although they analyzed data from 188 cities, Zanobetti and Schwartz only provide $PM_{10-2.5}$ mortality associations for the pooled result that prevents an analysis of individual city associations which is vital in determining the robustness of their conclusions or the spatial patterns of the associations.

An examination of the previously available studies on $PM_{10-2.5}$/mortality relationships indicates that EPA conclusions in the first draft of the ISA are justified. In the first ISA, EPA concluded that the results were mixed and that" more data is needed."

EPA largely ignores the issues of model selection and publication bias. When considered, these issues cast further doubt on the validity that coarse PM is a causal agent for mortality and the other health effects that EPA claims at the ambient $PM_{10-2.5}$ concentrations measured in the reported epidemiological studies. At this time there is no justification for a new $PM_{10-2.5}$ NAAQS.