

DETAILED EXXONMOBIL COMMENTS ON FIRST EPA FIRST DRAFT INTEGRATED SCIENCE ASSESSMENT FOR PARTICULATE MATTER

Chapter 1: Introduction

1.2. History of Reviews of the NAAQS for PM

In this section, EPA notes that they finalized the PM Air Quality Criteria Document (AQCD) for PM in 2004, and announced their final decision to revise the PM NAAQS on September 21, 2006. However, EPA fails to note that when serious problems with the convergence criteria used in the statistical programs commonly used in time series studies were identified by the Health Effects Institute (Greenbaum, 2002), the schedule to complete both the AQCD and the Staff Paper were impacted (Grant, 2002). As a result, while the cutoff period for including studies in the AQCD for the last review occurred in 2002, EPA clearly recognized and considered many studies published since 2002 in their last review of PM, which was completed four years later in 2006.

“The EPA is aware that a number of new scientific studies on the health effects of PM have been published since the 2002 cutoff date for inclusion in the Criteria Document. As in the last PM NAAQS review, EPA intends to conduct a review and assessment of any significant new studies published since the close of the Criteria Document, including studies submitted during the public comment period in order to ensure that, before making a final decision, the Administrator is fully aware of the new science developed since 2002. In this assessment, EPA will examine these new studies in light of the literature evaluated in the Criteria Document. This assessment and a summary of the key conclusions will be placed in the rulemaking docket. A preliminary list of potentially significant new studies identified to date has been compiled and placed in the rulemaking docket for this proposal, and EPA solicits comment on other relevant studies that may be added to this list.” (Environmental Protection Agency 40 CFR Part 50, National Ambient Air Quality Standards for Particulate Matter, Proposed Rule, Federal Register Vol. 71, No.10, January 17, 2006, page 2625)

EPA again repeated this theme in the final rule (FR Vol 21 No. 200, October 17, 2006) wherein they stated.

“In the proposal, EPA recognized that there were a number of new scientific studies on the health effects of PM that had been published recently and therefore were not included in the Criteria Document. The EPA committed to conduct a review and assessment of any significant “new” science before making a final decision on whether to revise the current PM NAAQS. The EPA screened and surveyed the recent literature, including studies submitted during the public comment period, and conducted a provisional assessment (EPA, 2006) that places the results of those studies of potentially greatest policy relevance in the context of the findings of the criteria document.” The provisional assessment found that the “new” studies expand the scientific information and provide important insights on the relationship between PM exposure and health effects of PM.”

Therefore, all of the studies that EPA summarized and included in the compilation of new studies (EPA, 2006) were considered in the last review, and should not be considered as new scientific data for the current review. Some notable examples include the update of the Six Cities study (Laden et al. 2006), the study of spatial analysis of air pollution and mortality in Los Angeles (Jerrett et al, 2005), the subchronic exposure study in a mouse model of

atherosclerosis (Sun et al., 2006), the NMMAPS study on hospital admissions (Dominici et al., 2006), the study on effects of air pollution on lung development in children (Gauderman et al., 2004), and the study on ambient air pollution and atherosclerosis in Los Angeles (Kunzli et al., 2005). These studies were considered numerous times by CASAC and others in public meetings during the last NAAQS review. The Administrator was also fully aware and considered these studies in making his decision to revise the PM NAAQS during the last review. Therefore, EPA should clearly state in Chapter 1 that studies included in the EPA 2006 new study compilation do not constitute new information available since the last review of the PM NAAQS.

Chapter 6: Integrated Health Effects of Short Term Exposure

6.1.1 Methodological Considerations

In this section, EPA includes a subsection entitled “**Human Clinical Studies, Advantages and Limitations.**” We note that human clinical studies are the only study type for which EPA includes this type of information. In our view, this demonstrates EPA bias towards systematically discounting the results of human clinical studies in favor of observational epidemiology studies, or, explaining away the results of human clinical study results when they report no effects or findings that fail to support hypothesis raised in observational epidemiology studies. We suggest including a section that presents the “advantages and limitations” of time series and field observational epidemiology studies. We strongly recommend that this section is developed by an epidemiologist without vested interest in promoting the results of observational studies. We present below some of the factors that we consider to be advantages and limitations of observation epidemiology studies of air pollution.

Advantage and Limitations of Short Term Observational Epidemiology Studies

Observational epidemiology studies evaluating the short-term health effects of air pollution include time-series morbidity and mortality studies and field or panel studies. One of the advantages of time-series studies is the unit of study is the general population, including the full range of potentially susceptible groups, thereby facilitating estimates of population risk. Also, many of the various socioeconomic factors that may confound the potential effects air pollution and that need to be addressed in chronic cohort studies do not need to be controlled in time series studies, if these factors do not vary on a daily basis or correlate with air pollution.

One of the main challenges and potential limitations of time-series studies is use of a population or group exposure metric. Generally, in time series studies, ambient monitor measurements are used as a surrogate for personal exposure to outdoor air pollution. This results in exposure misclassification which varies from pollutant to pollutant, depending on the degree of correlation between ambient and personal exposure. This is further complicated by the multi-component nature of air pollution. Identifying whether or not the specific component of concern alone is responsible for the health effects or is acting as a surrogate for or interacting in combination with other correlated pollutants, is challenging and dependent on arbitrary adjustments using very complex multi-pollutant models. These factors have led to the conclusion that especially for pollutants for which the ambient and personal exposure are not highly correlated, it is not possible to describe concentration response functions with accuracy, and the ability to determine “thresholds” for individual pollutants alone is obscured (Brauer et al., 2002).

Only a few studies are available which have evaluated the degree and direction of exposure misclassification in time series studies (Zeger, 2000, Sheppard, 2005). These studies focused

on the single pollutant PM, and on one health effect, mortality. Thus, as stated by the authors of these studies, the results of these studies cannot be extrapolated to other pollutants, such as gaseous pollutants, and to other health indicators, such as morbidity. Further, as stated by the authors of these studies, the issue of exposure misclassification for multiple pollutants is very complex and has not been addressed. Thus, overall, the degree and influence of exposure misclassification in time-series studies of air pollution remains unknown.

Another challenge and potential limitation of time-series studies is their dependence on complex statistical models to control for meteorological and time varying factors which can confound the effects of air pollution. For pollutants for which meteorology drives the airborne levels (e.g. ozone), developing an accurate statistical model that captures these complexities is additionally challenging.

A thorough review of this topic focusing on particulate matter was conducted by the Health Effects Institute (HEI, 2003). In the studies where temporal effects were extensively examined, some of the estimates of the effects of air pollution were more sensitive to the degree of smoothing of temporal effects than either the convergence criteria or the method used to account for temporal effects. In some studies the original effect estimates were largely insensitive to either the method or degree of smoothing whereas in other studies, the changes were substantial enough to result in meaningful changes in the study conclusions. The reasons for these differences remain unknown. The HEI concluded “the appropriate degree of control for time, nor the appropriate specification of the effects of weather, has not been determined in time-series analysis.” Further, the HEI panel concluded that in the absence of adequate biological understanding of the time course of PM and weather effects, and their interactions, the sensitivity of future time series studies to a wider range of alternate degrees of smoothing and to alternate specifications of weather and other variables needed to be explored further.

Since this time, a number of studies have further evaluated the impact of various alternate time-series statistical model specifications on effect estimates in these studies. A number of these studies are included in the review of PM summarized below. In these studies, various changes in model parameters (e.g. degrees of freedom for smoothing time varying factors, segmentation by season, and selection of lag times) significantly influenced the PM risk estimates. Thus, the selection of the appropriate time series model to control for meteorological and time varying factors has still not been determined and remains a source of variability and limitation for use of these studies.

As mentioned above, one strength of time series studies is the ability to provide information on population risks. On the other hand, this could also be considered a limitation as these studies are unable to provide information on the actual risks posed to individual's, which is generally considered stronger information for purposes of determining causal relationships.

The large majority of panel studies of air pollution also depend on use of ambient monitor measurements and therefore have the same challenges and limitations as the time-series studies described above. Since these studies focus on health effects in individuals, the potential exposure misclassification at the individual level is potentially greater than for a community average exposure. An advantage of these studies is the ability to assess effects at the individual level. However, the smaller number of individual's included in these studies and associated statistical considerations limits the extrapolation of the results of these studies to the general population.

6.2 Cardiovascular and Systemic Effects

6.2.1 Heart Rate and Heart Rate Variability (pages 6-8 to 6-22)

In this review of the data on heart rate variability (HRV) and heart rate, as in their previous review, EPA continues to overstate the evidence that: 1) exposure to ambient PM produces changes in HRV and heart rate; and 2) very small acute and transient changes in HRV and heart rate, equivalent to those occurring in humans every day, have any biological significance; 3) the evidence is specific to PM_{2.5} and not other correlated air pollutants which EPA also asserts cause changes in HRV and heart rate. We note that the overall body of data from human clinical studies does not provide convincing evidence that PM produces meaningful changes in HRV and heart rate. In our view, for clinically oriented endpoints, EPA should place higher weight on the results of controlled human clinical studies in an overall causal determination. EPA places far too much emphasis on the results of highly confounded and model dependent observational epidemiology studies, for which in our view, the data are also inconsistent.

As we reported in our comments on the Criteria Document in the previous PM review, small transient changes in HRV occur as a result of homeostatic mechanisms in response to normal everyday stresses. In the table below, we present a few of the myriad of life events that produce such changes. It is clear that the small changes alleged to be produced by air pollution as reported in certain observational epidemiology studies (but are not confirmed in human clinical studies) fall within the range of normal and should not be viewed as the mechanism by which PM produces mortality or morbidity. EPA continues to confuse the use of HRV in the management of heart disease, which is based on long term changes in HRV and various cardiac outcomes, with acute and transient changes in HRV.

Studies show that HRV varies widely in healthy volunteers: 157+/-45 for men, 138+/-29 for women (Ramaekers et al, 1998). According to the report by the European expert Committee, to increase the risk of arrhythmia after an MI, the observed cutoff values of 24-hour measure of HRV, i.e., SDNN of <50ms and triangular index of <20 for moderately depressed HRV are broadly applicable. In the Pope (1999) study, average reduction of SDNN was about 2 ms for the commonly used index of a 10 ug/m³ increase in PM₁₀. This value falls well below that deemed to be clinically significant. This view was shared by the cardiologists at a Workshop on Cardiac Effects Associated with Particulate Matter. For example, Dr. Verrier with Boston Deconnis Hospital stated emphatically that the small changes in HRV reported in the air pollution studies are similar to those observed with many normal day-to-day activities and, in his view, did not explain the mortality reported in acute time-series studies.

Many physiological and psychological factors affect HRV. Some of the more important factors are discussed below. These data indicate that the minor HRV changes reported in semi-ecological studies are well within the range of normal and therefore do not explain the mortality associations.

1) Heart rate: Heart rate is inversely related to all measures of HRV. In one study, heart rate accounted for 12.5 to 22.6% of the variance in 2h SDNN and low and high frequency power.

2) Age: Old age is related to low HRV. In one study, decrease HRV due to aging resulted in 18% (SDNN), 10% (rMSSD), and 15% (pRR50) of subjects over age 60 falling below published cut-off points for risk of mortality (Umetani et al, 1997).

3) Gender: The data are conflicting on this issue. Since women live longer and develop cardiovascular disease at a later age than men, it has been postulated that healthy women would have greater heart rate variability than healthy men. However, recent studies suggests heart rate variability indices, denoting vegal activity, are not significantly different between men and women, whereas the spectral indices of low frequency power and low/high frequency ration were significantly higher in men (Ramaekers et al, 1998). Liao et al. (1995) observed lower LF and higher HF/LF ratio in women than men. Some studies suggest otherwise. Cowan et al. found significantly lower HRV in women than in men, for all indices except HF power. Huikuir et al reported increased high frequency heart rate variability among middle aged women than men. Umetani et al. (1997) Reported that influence of gender on HRV is age related and disappears by age 50. Stein et al found no difference in HRV among older subjects but found significantly higher LF/HF ratios in men, which is similar to the findings of Ramaekers et al. However, they found significantly higher HRV amount younger men (age 33+/-4) than women.

4) Race: Liao et al. (1995) found that blacks have lower LF, higher HF, and higher HF/LF ratio than whites.

5) Systemic illness: Congestive heart failure and diabetes mellitus (Burger et al, 1997) are known to be associated with low HRV.

6) Other: Many normal day-to-day activities and physiological events cause minor fluctuations in HRV. Some of these are listed below. In many cases, the magnitude of the change is comparable to that reported in air pollution studies.

Activity	HRV Measure	Direction of Change	Reference
Smoking	Multiple	Decrease	Tsuji , 1996
Use of beta adrenergic blocking agent	Multiple	Decrease	Tsuji , 1996
Use of diuretic	Multiple	Decrease	Tsuji , 1996
Diastolic blood pressure >/+ 90 mmHg	Multiple	Decrease	Tsuji , 1996
Consumption of 3 or more cups of coffee/day	Multiple	Decrease	Tsuji , 1996
Shift work impact on sleep	SDNN	Decrease	Van Amelsvoort , 2000
Simple mental and verbal activities e.g. reading or stressful book	SDNN	Decrease	Bernardi, 2000
Systolic blood pressure >= 90 mmHg	Multiple	Increase	Tsuji, 1996
Time of measuring HRV in the morning	Multiple	Increase	
Low job demand, high control High job demand, high control High job demand, low control	SDNN	Increase	Van Amelsvortt, 2000
Higher noise level on job	SDNN	Increase	Van Amelsvort, 2000
Post exercise	?	Increase	Seals, 1989

Concerning heart rate, studies show that a rise in heart rate is associated with increase risk of cardiovascular mortality. The Framingham study (Kanel et al., 1987) showed that odds ratio and 95% CI for each increment in HR of 40 beats/min. adjusted for age and systolic BP level were for all cause mortality 2.18 (1.68, 2.83) in men and 2.13 (1.59, 2.88) in women. For cardiovascular mortality, the values were 1.68 (1.19, 2.37) in men and 1.70 (1.08, 2.67) in women (Gillman et al, 1993). In order to produce this level of change, PM levels of thousands

of $\mu\text{g}/\text{m}^3$, or well over 10 times currently present in ambient air, would be required. We conclude that the heart rate changes reported in the air pollution studies, which are on the order of a fraction of a beat/minute, are not clinically significant and not likely to explain the acute mortality reported in air pollution epidemiology studies.

6.2.8 Blood Coagulation

In this section, EPA summarizes the new studies on air pollution and changes in various measures of blood coagulation. However, EPA provides no reference to judge the biological significance of these changes, which are inconsistent, and small in magnitude. Some information on this topic is presented below. We recommend that EPA should include this kind of information in the ISA.

Plasma fibrinogen has been shown to be an independent risk factor for cardiovascular events (Wilhelmsen et al, 1984; Thompson et al, 1995; Violi et al, 1996). There are many factors which may be associated with high plasma fibrinogen levels. These include genetic factors, old age, smoking, job stress, high basal body mass index, high blood pressure, plasma glucose, presence of diabetes, and glycated hemoglobin, and triglyceride levels (Ishizak et al, 1996); Ko et al, 1997; van der born et al., 1998; Yarnell et al, 2000; Margaglione et al, 1998). Association of high plasma fibrinogen and coronary events is also very complex. Some studies show that an increase in fibrinogen might not always increase risk in coronary events. Van der Born et al. compared fibrinogen levels in a group of Coronary Artery Disease patients with a group of people with genetic disorders, which are known to cause high fibrinogen levels. Results showed that increases in plasma fibrinogen due to genetic factors were not associated with increased cardiovascular events. On the other hand, in the CAD group, a rise in fibrinogen of 1 g/l was associated with a 45% increased risk of MI (odds ratio 1.45; CI 1.12-1.88). Violi et al. showed that a difference of 1.05 g/l in the level of fibrinogen has an odds ratio of 1.16 for cardiovascular events.

From the data by Violi and van der Born, about 1 gm/l in plasma fibrinogen was associated with higher incidence of coronary events, 16% and 45 %, respectively. In contrast, in a study of air pollution, Pekkanen et al (2000) reported an increase of 1.5% fibrinogen associated with an exposure of $61.7 \mu\text{g}/\text{m}^3$ NO_2 , a marker for traffic air pollution. This corresponds to an increase of 0.036 gm/l fibrinogen. This value is almost 30 times lower than the fibrinogen level viewed as being clinically significant, or 1 gm/l. In the new studies reported in the ISA, Ruckeri (2007) report in increase of 0.6% fibrinogen per $13 \mu\text{g}/\text{m}^3$ of PM and Liao et al (2007) report a decrease of 0.006 gm/dL albumin. Thus, the air pollution studies citing possible findings for PM on various measures of coagulability changes that are not viewed as clinically significant. Therefore, we conclude that the available data do no support a coagulability-related mechanism of PM toxicity for acute events

6.2.1.0 Hospital Admissions and ED Visits and 6.1.11.3 $\text{PM}_{2.5}$ Causal Determination

On pages 6-71, EPA briefly describes the design of the key MCAPS study by Dominici. et al. (2006). Then, on page 6-82, EPA has one sentence capturing the results, which they characterize as 0.4% (CI 0.0-0.8) increase in ischemic heart disease per $10 \mu\text{g}/\text{m}^3$ of $\text{PM}_{2.5}$. This single result also appears in table 6.3. Then, on page 6-98, under the heading of Causal Determination for $\text{PM}_{2.5}$, on pages 6-98-6-99, EPA references the results of the MCAPS study as the largest U.S.-based multi-city study, but now indicates an excess risk of 0.7%.

In our view, EPA's presentation of the results of the MCAPS study are highly misleading. First, presenting the findings as a single precise central national risk estimate fails to recognize the high degree of unexplained heterogeneity and variability in the results as presented in the publication of this key study. As noted by the authors of the study, for cardiovascular diseases, all estimates in the Midwest, northeastern, and southern regions were positive, while estimates in northwest, west, central, and southeast, were close to zero. This heterogeneity renders the national central estimate as presented in the draft ISA as meaningless, and weakens the firm conclusion that EPA reaches that the association between ambient PM and increased hospital admissions for cardiac disease is causal. This heterogeneity was also apparent, but to a lesser degree, in the data for respiratory related diseases. While the authors speculate that these dramatic regional difference might be due to difference in the chemistry of PM, the actual reasons for this high degree of heterogeneity are unknown. The 204 counties evaluated in the study were urban. Thus, one could reasonable expect substantial differences in lifestyles over a relatively small area. Medicare data key to the conduct of this study do not provide socioeconomic/behavioral information which may differ. Such factors are related to the health outcomes under study and could al be related to the placement of monitors. Also, the quality of Medicare data also varies from region to region. We recommend removing the notation of any central estimate from the ISA and recommend a conclusion that the association between PM and hospital admissions for both cardiovascular and respiratory causes is at most suggestive of a causal association.

6.5 Mortality Associated with Short-Term Exposure

Change in PM_{2.5} mortality associations in NMMAPS data, 1987 to 2000 (pgs 6-203 to 6-205)

In this section, EPA summarizes and interprets the results of an update of the key National Morbidity and Mortality Study (Dominici et al. 2007). In the past, EPA characterized this study as the highest quality time series study available given the multi-city nature and consistent methodology employed. The update was intended to provide an answer to the question on whether or not the risks per unit of PM₁₀ have changed from when they were last evaluated in 1987 to 1994 (data analysis available at the time of the last NAAQS review of PM), to 1995 to 2000 (data analysis available for the current NAAQS review). Dominici et al (2007) conclude that the national risk estimates for PM₁₀ have *declined* during the period 1987 to 2000. Specifically, they report a 14% *decline* in risk estimates for all cause mortality and a 23% *decline* in risk estimates for cardio-respiratory mortality. However, since this conclusion works against EPA's pre-conceived notion that the PM risks must be increasing and therefore the PM standards must be lowered, EPA has added a series of unjustified and poorly supported qualifying statements intended to discount the author's conclusions and the status of the study.

First, EPA states that the study by Dominici et al. (2007) is not actually an intervention study since the reduction of PM₁₀ was "very gradual" and not "sudden" as per other invention studies. EPA references studies in Hong Kong (Hedley, 2002) and Dublin (Clancy et al. 2002) where specific changes in fuels use led to more abrupt change in PM levels. Second, EPA also states that a "flaw" in the analysis performed by Dominici et al. (2007) is that the changes may reflect changes in PM composition during the duration of the study.

We note that the EPA's application of these qualifying statements is highly selective. EPA describes the update of the Harvard Six Cities Study by Laden et al. 2006 as an intervention study even though there was no specific regulatory intervention in the six cities during the timeframe of this study, the reduction of pollution was equally gradual and not "sudden", and the

PM composition also could have changed during the course of the study. Similarly, all of the various updates of the ACS study are construed as intervention oriented studies, without any of the qualifying statements above.

This inconsistent application of criteria, i.e., assigning a study as an intervention study, which receives higher weight in the EPA causality scheme, when the study reports a positive result, and manufacturing reasons to discount a study and lower its status when the study reports a negative results or results that does not supports EPA's policy objectives, is a serious data quality issue that EPA must address.

EPA also states on page 6-206 that there remains a PM₁₀ mortality association in the 1995-2000 time period. We disagree with this conclusion. All of the nine risk estimates for all cause, cardiopulmonary, and other cause mortality in the east, west, and national, were reduced to levels that were *not statistically significant*. Therefore, the correct conclusion is there is now no clear association between ambient PM₁₀ and mortality in the time period 1996-2000 in the key NMMAPS data. Since PM levels of all sizes including PM₁₀ have continued to decline during this period, EPA should conclude that the reductions have been sufficient to lower the risk of mortality to a level that is no longer significant today.

Finally, EPA fails to note that Dominici et al. reported rather dramatic reduction in PM_{2.5} risks with increased degrees of freedom for smoothing of time. This finding has been reported in other recent studies (e.g. Ostro et al. 2006, see below) and provides further proof to dispel the EPA notion that the risk estimates in time series studies are "robust" to the statistical model selected.

PM_{2.5} Mortality associations in 27 U.S Cities, Variable between 1997 and 2002 (pages 6-218 to 220 and conclusions on page)

In this section, EPA summarizes and interprets the study by Franklin et al. (2007) who evaluated the association between PM_{2.5} and all cause and various specific causes of mortality in 27 communities using a case cross-over design followed by a meta analysis to estimate summary effects over all 27 communities. In this study, effect modification of age and gender were examined using the case crossover model while effect modification by geographic location, annual PM_{2.5} concentration above and below 15 ug/m³ and use of central air conditioning were estimated using meta-regression. They reported a 1.21%, 1.78%, and 1.03% increase in all cause mortality, respiratory mortality, and stroke, respectively per 10 ug/m³ of PM_{2.5}. Franklin et al. described these values as triple those reported for PM₁₀ suggesting to the authors that combustion and traffic related particles are more toxic than larger sized particles. Franklin et al. reported that statistically significant effect modification occurred by age and geography: the effects of exposure were greater in subjects ≥ 75 years of age, in Eastern cities and for those without central air condition, particularly in cities where peak PM exposures are in the summer. Evidence also suggested that women may be more susceptible than men to effects of PM_{2.5}, although that lacked statistical significance. Franklin et al. also reported slightly higher, but not statistically different, risk estimates in cities below 15 ug/m³, versus cities above 15 ug/m³ which the authors claim provides evidence for "health effects below the current NAAQS standard³". EPA repeats this assertion in their summary.

Inspection of the city specific risk estimates for the 27 communities presented in the Figure 1 of this study suggests there is no clear evidence for an association between PM_{2.5} and all cause mortality. In three of 27 communities, the risk estimates were negative and statistically significant, in three other communities, the risk estimates were negative but not statistically

significant, and in six communities, the risk estimates were near zero (null). Furthermore, negative or null associations were observed in many of the cities with the highest daily average concentrations of PM. For example, in four of the five cities with the highest daily average PM_{2.5} values (Riverside 28 ug/m³, Los Angeles 22.3 ug/m³, Birmingham, 20.4 ug/m³, Cincinnati, 18.6 ug/m³), negative associations were reported.

The summary estimates for all cause, respiratory, and cardiovascular mortality presented in Table 2 are based on the meta-regression analysis. Since the test for heterogeneity for these mortality indicators was highly statistically significant (Q-statistic $P \leq 0.001$), the risk estimates presented in Table 2, which are unadjusted for factors that account for heterogeneity, are of questionable value, i.e., the Q statistic indicates the data should not have been pooled. This likely explains the numerous risk estimates in Table 2 that are not statistically significant. Six of twelve estimates across three lag times (0, 1, and 0-1) and 4 different indicators of mortality are not statistically significant. Franklin et al provide no biological basis for selecting the results at lag 1 as definitive. The results at lag 1 are much higher than at lag 0 and lag 0-1 but the size the risk should not guide the choice of the preferred results, an approach equivalent to lag-mining.

Franklin et al use the results at lag 1 to support their conclusion that the risk estimates are "triple" those recently reported for PM₁₀ by Dominici et al. (2005) and Zeka et al. (2005). This conclusion is unfounded. First, again, it is not clear why the results at lag 0 and 0-1, which are in the same range as those reported by Dominici et al. and Zeka et al. (2005), were summarily discounted by the authors. Second, Dominici et al. and Zeka et al. evaluated a different spectrum of cities than those evaluated by Franklin et al. Given the very high degree of heterogeneity observed across risk estimates in cities in all of these studies, it is not possible to make comparisons of one form of PM versus another unless the same cities are evaluated. Third, Zeka et al. also report positive associations for both respiratory and cardiovascular mortality for PM₁₀ whereas Franklin et al did not report a positive association for cardiovascular mortality. This further complicates the comparison, and appears to indicate that the toxicity of PM₁₀ rather than PM_{2.5} may be more complex. We also note that the association for respiratory mortality but not cardiovascular mortality is not coherent with the results of chronic studies which demonstrate the opposite finding, i.e., cardiovascular mortality but not respiratory mortality.

Franklin et al. 2007 report that use of air conditioning resulted in a marked reduction of PM_{2.5} risk, accounting for 60% of the variation in respiratory deaths, and 35% of all-cause deaths. They suggest this protective effect modification was due to filters on air conditioners that reduce exposure to fine PM. While we agree that this hypothesis could explain some of the variability, we note that Franklin et al. did not investigate any co-pollutants such as ozone and nitrogen dioxide that they, and others, have concluded cause acute mortality. It is well known that fine PM penetrates indoors whereas the highly reactive gaseous pollutant ozone readily reacts after penetrating into indoor environments and dissipates rapidly after those reactions. This is why the correlation between ambient monitor measurements and personal monitor measurements are much higher for fine PM than they are for ozone and in fact for other gaseous pollutants. Therefore, the hypothesis suggested by Franklin et al is not confirmed. The heterogeneity and variability observed in their data could be explained by the effects of other pollutants such as ozone that they chose not to evaluate, particularly since personal exposure to these pollutants are much more influenced by use of air conditioning than is fine PM. Franklin et al stated that the reason they chose not to consider confounding by other pollutants was based on the fact that gaseous pollutants had little impact on associations for PM₁₀ in other studies such as Dominici et al (2005). We note that there is a much higher correlation between ozone and fine

PM than ozone and PM₁₀. Therefore, the hypothesis that there will be no impact on risk estimates for PM_{2.5} has not been confirmed.

Franklin et al. report that since larger risk estimates were observed in cities with average ambient PM_{2.5} values below 15 ug/m³ than for cities with average values above 15 ug/m³, this “pattern” suggests that “health effects may still be seen below the NAAQS standard.” EPA repeats this conclusion in their summary of the study. This conclusion is unfounded. First, all of the six risk estimates (three causes of death for cities above and below 15 ug/m³) are not statistically significant. In our view, no firm conclusions can be drawn from a table of non-statistically significant results. Second, as noted above, the risk estimates for cities with some of the highest levels of average daily PM_{2.5} concentrations are not statistically significant. This “reverse” dose response defies biological explanation. Third, the overall estimates across all cities are so highly heterogeneous that no trend can be discerned. For example, the authors fail to explain why the risk estimates for Dallas and Milwaukee -- two cities with nearly identical daily average daily PM_{2.5} value of 13 ug/m³ and 14 ug/m³ respectively, and both below 15 ug/m³ -- have all-cause mortality associations of -5% and +10%, respectively.

Furthermore, the authors assumed linearity in their modeling exercises. This assumption virtually guarantees a steeper slope and higher coefficients versus a linear non-threshold model which is more appropriate for the outcomes under study which were largely non-cancer causes of death.

Overall, this study should not be used to draw the conclusion that mortality is occurring below 15 ug/m³ or any level of PM_{2.5}.

PM_{2.5} – Mortality Associations in Nine California counties, 1999-2002 (pages 6-220 to 6-221)

In this section, EPA summarizes and interprets the study of Ostro et al. 2006 who examined the association between daily PM_{2.5} concentrations and all cause and other specific cause mortality in nine California counties using a time series approach. Ostro et al. examined the associations in several subpopulations, including those > 65 years old, males, females, non-high school graduates, blacks, whites, and Hispanics. They used Poisson regression models incorporating natural or penalized splines to control for covariates that could affect daily counts of mortality, including time, seasonality temperature, humidity and day of the week. They used meta-analysis using random-effects models to pool the observations in all nine counties. Ostro et al. reported a 0.6% increase in all-cause mortality, with similar or greater effect estimates for several other subpopulations and mortality subcategories, including respiratory disease, cardiovascular disease, diabetes, age > years, females, death out of hospital and non-high school graduates. Ostro et al. reported that the results were generally insensitive to the model specification and type of spline model used.

The study by Ostro et al. is well reported. The main strength of the study is the effort to examine the influence of several key model specification changes on the risk estimates. However, careful examination of the results presented by Ostro et al. reveal several key differences in the conclusions versus those offered by the authors and repeated by EPA. First, the authors did not test for heterogeneity before combining the county specific risk estimates. From inspection of the data, there was a very high degree of heterogeneity in the individual county estimates. This brings into question whether or not these data should be combined at all for purposes of meta-analysis. The high degree of heterogeneity is very troubling and raises the question whether the associations in the combined analysis are meaningful at all. The heterogeneity in

this study cannot be explained by seasonal, geographic, or PM air pollution differences, as has been hypothesized in others time series studies where a high degree of heterogeneity has been observed (e.g. NMMAPS, Dominici et al.). The confidence limits in the combined analysis are relatively wide and many of the results are not statistically significant, raising the question of whether or not the associations can be used to assert causality.

The author's claim that the results were generally insensitive to the model specification and type of spline used is not supported by inspection of the data. Simply changing the degrees of freedom/year from 4 to 8 in the natural spline model significantly reduced the estimated increase in mortality, from a 40% reduction for all-cause mortality (from 0.5% to 0.3%) to 100% reduction (i.e. from 0.4% change to 0% change, per 10 ug/m³ PM_{2.5}) for cardiovascular mortality. The authors do not provide any information concerning model fit which would aid in determining selection of the optimal degrees of freedom. The mortality estimates are derived using non-parametric regression (penalized spline model) were consistently higher than those reported using parametric regression (natural splines). It is not clear why Ostro et al chose to focus on the results using penalized splines, other than the results are higher for this spline model.

The authors state that when co-pollutants highly correlated with PM_{2.5}, nitrogen dioxide and carbon monoxide, were included in the model, they tended to "*attenuate the magnitude and significance of the PM_{2.5} coefficient*". It is not clear why Ostro et al do not present the actual results for the two-pollutant models. We speculate the attenuation was rather high, given the fact that the results shifted from statistically significant to not statistically significant. Given these findings, we cannot see how Ostro et al. claim that the increased risks they choose to report as preferred, i.e. 0.6% increase in mortality per 10/m³ PM_{2.5}, are attributed to solely to PM_{2.5}, or why the results using natural splines at higher degrees of freedom are not equally valid.

Some other findings in Ostro et al. are difficult to explain, and the author's hypothesis and attempts to explain some these nuances in the data are internally inconsistent. For example, since higher risk estimates were reported for those with non-high school graduates (0.9 per 10 ug/m³) than for high school graduates (0.4 ug/m³), the authors hypothesize that this difference is attributed to factors associated with education such as nutritional status, access to health care, occupation, and psychosocial stress. However, this differential is not supported by trend observed when the findings are segmented by race, another surrogate measure for some of the above factors, as the increased mortality estimates for blacks and white were 0.1% and 0.8%, respectively per 10 ug/m³. Also, in contrast to most other studies, the increased risks for cardiovascular mortality were lower than those for respiratory mortality. Ostro et al offer several explanations why cardiovascular deaths are often higher in studies of air pollution, including higher power to detect an effect given the greater prevalence of circulatory disease, and preferential assignment of deaths as cardiac versus respiratory. However, they offer no meaningful explanation as to why the increase in mortality in their study does not follow this pattern, but is rather 3-fold higher for respiratory versus cardiovascular mortality.

None of the concerns above are recognized by EPA in their summary. Rather, EPA chooses to focus on the "consistency" between the results of Ostro et al. 2006 and Franklin et al. 2007 in cities that were evaluated in both studies. It is obvious that the results within Ostro et al (2006) were not internally consistent as many model selection parameters markedly influenced the results. Thus, EPA conclusion should be given the model driven heterogeneity in the results of Ostro et al. (2006) and lack of criteria to determine which of the various risk estimates are

“correct”, it is not meaningful to compare the results of this study to those of Franklin et al. (2007).

Summary of PM_{2.5} Risk Estimates (page 6-222)

EPA concludes that the risk estimates for all cause mortality for all ages range from 0.29% (Dominici et al, 2007) to 1.21% (Franklin et al 2007) per 10 ug/m³. In our view, this conclusion fails to capture the high degree of heterogeneity in the results of the new time series mortality studies which, if consistent, would result in the true risk range extending to no risk or null. Based on our critical review of the new data on the association between PM_{2.5} and mortality described above, we describe below the key uncertainties and sources of heterogeneity in the PM_{2.5} mortality estimate. We recommend that EPA include these points in their conclusions.

- Since the last review of the NAAQS standard for PM, only four new studies are available that evaluated the association between PM_{2.5} and mortality. Only one of these studies (Ostro et al. 2006) considered potential confounding by gaseous pollutants that EPA also considers to cause mortality. In this study, Ostro reported that inclusion of gaseous pollutants attenuated the mortality association for PM_{2.5}. Thus, since the last review, the relative importance of gaseous pollutants on mortality risk estimates for PM_{2.5} has not been resolved.
- Since the last review, two studies (Ostro et al. 2006 and Dominici et al.) critically evaluated how various model specification factors potentially influencing the PM_{2.5} mortality association. The results indicated that the acute PM_{2.5} mortality risk estimates are *not* robust to changes in various time series model specifications. The factors influencing the PM_{2.5} mortality association included the degrees of freedom used for smoothing time, the spline model selected for the regression, segmentation by season, and the lag times used.
- Since the last review, two studies (Dominici et al. 2007, and Ostro et al. 2006) evaluated the regional heterogeneity in the results of the PM_{2.5} mortality association. Both studies reported significant and unexplained regional (Dominici et al. 2007) and county-wide (Ostro et al. 2006) heterogeneity in PM_{2.5} mortality risk rendering any central or national estimate of acute mortality risk as unreliable and misleading.
- Since the last review, no new studies are available that evaluated the potential change in the PM_{2.5} mortality association with reduced ambient levels of PM_{2.5}. However, one higher quality study is available that evaluated the change in PM₁₀ levels (Dominic et al., 2007). The results of this study indicated that the risks attributed to a unit PM₁₀ exposure *decreased* from the period 1994-1997 to 1995-2000, and were not statistically significant at the latter time period.
- Since the last review, there have been no new studies on the key issue of the influence of measurement error on PM_{2.5} risk estimates.
- Since the last review, and number of epidemiology, human clinical, and toxicology studies have evaluated the chemistry of PM most important for health effects including acute mortality. The results of these studies are variable and conflicting and do not permit a conclusion as to the components of PM most responsible for acute health effects.

Chapter 7: Integrated Health Effect of Long Term Exposure

7.2 Cardiovascular and Systemic Effects

In this section, EPA presents a summary and their interpretation of several publications from the Multi-Ethnic Study of Atherosclerosis (MESA) (Diez Roux et al, 2008 and Allen et al. in press). In contrast to the author's findings, which indicate a lack of association between PM_{2.5} and various measures of atherosclerosis, EPA re-interprets the findings to arrive at a different conclusion, namely, that this collection of studies demonstrates that various measures of ambient PM are contributing to the progression of atherosclerosis.

On page 7-4, EPA has re-interpreted the results of the study by Diez Roux et al. (2008). In both the study abstract and the body of the paper, Diez Roux et al. clearly indicate they did not observe an association between ambient PM and PM-CAC. This was due to the fact that the associations were *weak* and all the confidence intervals included the null value. In contrast, EPA continues their approach of interpreting a very weak (i.e. very small) and statistically non significant risk coefficient as positive evidence for adverse health effects. We recommend that EPA correctly summarize the results of this study as reported in the peer-reviewed publication.

Allen et al. (in press) report a non statistically significant increase of 6% (95 CI -4 to 16) excess risk of abdominal aortic calcium per 10 ug/m³ of PM_{2.5} in a subset of individuals in the MESA study. They describe this association as "not persuasive." However, EPA considers the finding as positive evidence for adverse health effects (page 7-5).

On page 7-21, EPA concludes that the collection of studies provide *consistent evidence* that exposure to ambient PM_{2.5} is associated with subclinical measures of atherosclerosis. We disagree. EPA's conclusion is not supported by the results of these studies, as reported in the publications. Rather, EPA's conclusion is based on the unsubstantiated reliance of very weak and non statistically risk coefficients for PM-CAC and PM-AAC, and null findings for the other indicators of atherosclerosis, ankle brachial thickness (ABT) coronary intimal medial thickness (CIT). We strongly recommend that revise their conclusion that new publications from MESA fail to confirm an association between long term exposure to ambient PM_{2.5} and atherosclerosis.

7.4 Reproductive, Developmental, Prenatal and Neonatal Outcomes

7.4.1 Epidemiologic Studies

Low Birth Weight

On pages 7-53 to 7-56, EPA summarizes the epidemiology studies on low birth weight and on page 7-57 to 7-58, presents a section entitled "Issues in interpreting results of low birth weight studies." In general, the EPA approach can be characterized as accentuating the positive and eliminate the negative. For example, on page 7-54, EPA notes that Maisonet et al (2001) reported no association between ambient PM₁₀ and low birth weight at term whereas Bell et al. (2007b) reported an association between a exposure to PM₁₀ during pregnancy and the third trimester and low birth weight and exposure to PM_{2.5} during the second and third trimester and low birth weight. EPA states that a positive result was observed in the study by Bell et al. due to a larger sample size, which was able to detect a small risk, and measured concentrations were available in the study by Bell et al. but not in the study by Maisonet et al. Similarly, EPA explains the positive results by Parker et al. (2005) in light of the "reduced exposure misclassification by

including only women living within 5 miles of a monitoring station, and including only births at 40 weeks gestation.” EPA states that “reducing exposure misclassification should lead to a stronger association, if the association is causal.” This theme is repeated in the section on interpreting studies on low birth weight, where EPA concludes “*studies with negative results must be interpreted with caution when the comparison groups have significant exposure.*”

The entire discussion demonstrates EPA bias towards emphasizing the results of any study that reports a positive association, and attempting to explain away or qualify any study not reporting a positive association. All of the above studies are ecologic in nature and therefore have the same basic deficiency in that they all rely on ambient monitor measurements as a surrogate for personal exposure. The fact that Bell et al. 2007 included ambient measurements for PM_{2.5} whereas Maisonet et al. 2001 did not is a poor excuse to emphasize the results of Bell et al. over those of Maisonet et al. 2001. As noted below, the study by Bell et al. was not able to assess the effects of correlated gaseous pollutants, reported high heterogeneity in the ambient PM measurements. The sample size is rather irrelevant when the exposure is misclassified for everyone in the study.

Concerning the study by Bell et al. 2007b, which EPA places great reliance on, the exposure measurement scheme used did not account for individual mobility, either in daily living or in residential relocation into the area before the birth. Bell et al. make light of this as they claim this would have caused non-differential misclassification which would have biased the results towards the null. However, that assumption may not be correct. Far too much data are lacking to make firm statements regarding the potential for confounding. Several factors associated with relocation are also associated with low birth weight, for example, preferential selection of lower cost housing may have occurred, "leading to the potential for more exposure misclassification for mothers of lower weight infants". Also, the authors discuss the documented heterogeneity in air pollution concentrations that can exist at sub-county scales. "This variation can affect health effects estimates".

The effect modification by race indicated that potential uncontrolled confounding remained for differences in baseline health status, health care access, or occupational exposure as well as proximity to heavily traveled roadways. The latter could have steepened the exposure-response slope via "compositional clustering". This phenomenon occurs when the highest exposed individuals have the same attributes that make them more susceptible to the adverse effects of environmental exposures. In this study, the geographic areas having unhealthy behaviors and other SES deficits--shown in this study and in others to be associated with low birth weight--also have the highest presumed levels of exposure. That end of a linear model greatly influences the slope of that model, particularly since average concentrations for all pollutants in this study were below the NAAQS.

The authors concede that measuring the mothers' educational status as they did may not fully address the potential confounding by SES. They also admit that their binary measure of tobacco use (simply yes, no, or unknown) during pregnancy was probably less than adequate. Also, the alcohol and tobacco data were self reported. Also self-reported was LMP which showed a "terminal digit bias" that generated significantly more than expected frequencies every 5th day (i.e., 10th, 15th, 20th) than expected. This would have a material effect on the trimester analyses which would be affected most by this bias vs analyses based on the full duration of the pregnancy.

With the statistical modeling approach, maternal alcohol use was excluded from the analyses. The preferred method for covariate selection is via univariate analyses for each potential risk

factor, using birth weight as the response and each of the potential risk factors as explanatory variables. Each covariate exhibiting a statistically significant relationship with birth weight should be incorporated into models investigating air pollution exposure. Also, the binary measure of alcohol used during pregnancy could not fully capture its effect.

Two pollutant linear models were applied for gestational exposures. However, PM₁₀, PM_{2.5} and NO₂ were highly correlated. The pairs of these three pollutants were not included in the models. So this study can not distinguish the effects of these three pollutants. There was no information in the study about model fit, thus a complete assessment of the findings is not feasible.

County-level pollutant concentrations were used. As the authors said, "heterogeneity in air pollution concentrations can exist at sub-country scales and that this variation can affect health effect estimates." Ambient monitoring was not a good surrogate for personal exposure. Furthermore, there were missing data imputed using a weighted average of the concentrations on days with data available. This imputed data may also affect health effects estimates.

The authors can only speculate on a physiological mechanism, and they offer several possibilities. However, the evidence is quite thin on all of them at this point. So, all we are left with are weak statistical associations that are practically obscured by the enormity of the measurement errors that studies such as these inevitably make. Rather, this is an exceptional study on which to generate hypotheses, not test them.

Overall, the evidence that exposure to ambient PM_{2.5} cause low birth weight should be considered inconclusive.

Growth Restriction

On pages 7-62, EPA summarizes the studies on growth restriction. Again, the approach EPA take is to systematic make arguments to discount the results of negative studies while emphasizing the strengths of positive studies.

For example, in contrasting the results of Parker et al. (2005) who reports a positive association for PM_{2.5} and growth restriction, and Salam et al (2005) who reports no association, EPA harkens that the difference is due to less exposure misclassification in the study by Parker et al. (2005), since the monitors were located closer to the subjects residences, Parker et al. considered PM_{2.5} in addition to PM₁₀, and Parker et al employed a more stringent definition of intrauterine growth (IAG) . Again, both studies relied on ambient measurements as a surrogate for personal exposure and neither study considered the potential confounding effects of gaseous pollutants. EPA provides no scientific justification for why the stringency of IAG would influence the precision or direction of the risk estimates.

Infant Mortality

In this section, EPA continues their approach of relying of the results of observational studies that report very weak and non statistically significant risk coefficients to make conclusions on causality. For example, EPA places high importance of the results of two studies by Woodruff et al. (2008, 2005) even though the studies report non statistically significant results (page 7-67). EPA states that since many of the new studies on infant mortality demonstrate a higher risk for those in lower socioeconomic groups, these groups are higher risk. EPA ignores the alternate conclusion, which is that the new studies fail to adequately adjust for socioeconomic

status, which includes a host of well known risk factors for infant mortality. Overall, the evidence that exposure to ambient PM_{2.5} causes infant mortality should be considered inconclusive.

7.7 Mortality Associated with Long-Term Exposure

In this section, EPA summarizes and interprets the results of the new studies on long term exposure to PM. In our view, this section is very poorly written, presents a very biased interpretation of the new literature, and reads as an advocacy oriented text developed for the purpose of supporting a revised PM NAAQS. Due to the seriousness of these concerns, which we highlight below, we recommend that EPA consider identifying a new team to author this section of the ISA.

Harvard Six Cities Study (pages 7-104)

In this section, EPA summarizes the update of the Harvard Six Cities (H6C) Study published by Laden et al. (2006). EPA describes the study as a follow-up to the original H6C study using updated air pollution and mortality data during the period 1990-1998 vs. the original study period of 1874-1989. EPA reports that statistically significant associations are reported between long-term exposure to PM_{2.5} and mortality for data for the two periods (RR = 1.16 [CI:1.07-1.26], per 10 ug/m³). EPA further notes that since a statistically significant *reduction* in mortality risk was reported with reduced long-term fine particle concentrations, the results of Laden et al. strongly suggest that a reduction in fine PM pollution yields positive health benefits.

The summary by EPA fails to provide any critical analysis of the methodologies and assumptions used by Laden et al., for which as described below, there are serious scientific concerns. A few of these concerns were pointed out in the commentary on this study published by Gamble and Nicolich (2006). It is noteworthy that the author of this section of the ISA chose not to site this commentary, since it was critical of some of the conclusions made by Laden et al. However, in other section of the ISA where the original publication presents findings that raise concerns with methodologies used in chronic air pollution studies (e.g. see the section on the study by Janes et al.), the author of the ISA readily cites such commentaries, which are intended to discount any methodology concerns with these studies. This demonstrates a clear bias for including or excluding data for presentation in the ISA. This is a serious information quality concern.

The authors of the ISA also fails to note any information on the limitations of the underlying data from the H6C study upon which the analysis by Laden et al. is based, even those limitations appear in the published reanalysis of this study by Krewski et al. (2000). Again this illustrates the biased nature of how the data are presented in this section of the draft ISA, and is another serious information quality concern.

Concerns With the Methods to Estimate Exposure

The PM_{2.5} exposure data used in the update by Laden et al. are very uncertain and lacks consistency with the data from the first period of study. Whereas the data in period 1 were based on actual ambient measurements from the EPA AIRS database, the exposure estimates in period 2 by Laden et al. can be best described as best guess estimates. Since the Six Cities PM_{2.5} air monitoring stations were shut down long ago, PM_{2.5} concentrations in period 2 were based on PM₁₀ measurements and city-specific extinction coefficients which in turn were derived from humidity-corrected visibility data from local airports. Laden et al provide no validation for this approach. Further, the reductions in PM_{2.5} that Laden et al report are based

on the assumption that they will be equal to the reductions for PM₁₀. a different NAAQS pollutant with different control strategies. Again, Laden et al provide no validation that this assumption is correct. Even though the Pearson correlation between the estimated and observed annual mean PM_{2.5} from the Six Cities monitors during the years when both were available (1985-1987) was 0.93, the estimated daily PM_{2.5} concentrations after the shutdown of the Six Cities monitoring (1985-1998) were not good surrogates for observed values. The Pearson correlation may change over time.

Daily ambient PM_{2.5} concentrations were measured at a centrally located air-monitoring station in each community. Exposure assessment studies have shown that data from monitors at central sites do not adequately represent personal exposure (Lioy et al., 1990; Janssen et al., 1997, 1998; Ozkaynak et al., 1996; Haran et al., 2002; Dominici et al., 2003). Using data from monitors as a surrogate for personal exposure can potentially cause bias in estimates of the health effects of air pollution (Thomas et al., 1993). Furthermore the distance of cities to the monitor, rather than the number of monitors, a significant potential source for exposure misclassification, would seem to be of concern, but is not mentioned by Laden et al. or EPA. The distance for Portage, the city with the lowest risk, is the highest distance, i.e., 200km, whereas the city with the highest risk, Steubenville, is the lowest distance, 6 km, from the nearest monitor.

The other pollutants included in the original report of the H6C study by Dockery et al. demonstrated positive associations with mortality of equal magnitude to that of fine PM. For example, the relative risks for all cause mortality for the high to low range of PM_{2.5}, sulfates, TSP, SO₂, and NO₂ were 1.28, 1.28, 1.26, 1.26, and 1.25, respectively. Similar comparable results were observed for cardiopulmonary mortality. The single exception was ozone, for which a statistically significant beneficial effect was reported for cardiopulmonary mortality (relative risk 0.78, 0.64-.95). PM_{2.5} levels were also highly correlated with SO₂ and NO₂ levels; 0.85 and 0.78, respectively. However, the potential confounding of the PM association with other pollutants, or vice versa, was not evaluated by the original investigators, nor was this issue evaluated by Krewski et al. (2000) in their reanalysis of this study. The reason given by Krewski et al for not performing this critical analysis was "there was only one monitor in each of the only 6 cities."

From the time of the initial H6C study by Dockery et al. through the period of the update by Laden et al., the concentrations of other pollutants besides PM_{2.5} have also been steadily declining, reflecting the overall dramatic decrease in NAAQS pollutants across the U.S. To illustrate this point, we present below the reductions for sulfates, SO₂, NO₂, and acidity in the city of Steubenville, Ohio. While the percent reduction for PM_{2.5} was on the order of 24-31%, the percent reductions for other pollutants were even greater, i.e. 46%, 84%, and 46% for sulfate PM, SO₂, and NO₂, respectively.

Air Quality Comparison Data for Steubenville Ohio				
Pollutant	Period 1 1974-1989 Dockery et al. ^a	Period 2 1990-1998 Sarat et al ^b 2000	Change	Percent Change
PM _{2.5}	29	19-20 ug/m ³ 22 ^c ug/mg ³	7 ug/m ³	24-31%
Sulfate	13	6-7 ug/m ³	7 ug/m ³	46%
SO ₂	25 ppb	3-5 ppb	21 ppb	84%
NO ₂	22 ppb	9-11 ppb	12 ppb	46%
Acidity	12.8	NA	NA	NA but must have been dramatic given the large change in SO ₂ levels

^aDockery et al. (1993). *Reduction in fine particulate air pollution and mortality. NEJM* 329, 1753-1759.

^bSarnat et al. (2006). *Factors affecting the association between ambient concentrations and personal exposures to particles and gases. Env. Hlth Persp.* 14(5), 649-654.

^cLaden et al. (2006). *Extended follow-up of the Harvard Six Cities Study. Am. J. Respir Crit Care Med* 173, 667-672.

Laden et al do not even present any data on the levels or risks for other co-pollutants. We note that EPA has determined that these other pollutants also cause morbidity and mortality. Therefore, it is unclear how Laden et al., who chose not to consider any other pollutants beside PM_{2.5} in their update of this study, or, Dockery et al., who presented single pollutant risk estimates for these pollutants, but did not consider multi-pollutant effects or confounding in the original report of the H6C study, can make the firm conclusion that fine PM, alone, is the pollutant responsible for the mortality in period 1 or period 2, respectively, or in the case of Laden et al, for the reduction in mortality from period 1 to period 2.

It is noteworthy that in the other chronic cohort air pollution study that EPA places great reliance on, the American Cancer Society, inclusion of SO₂, considerably *reduced* the mortality estimates for fine PM (Krewski et al. 2000).

We conclude that the firm conclusion by Laden et al., repeated in the draft ISA, that PM_{2.5} is solely responsible for the mortality is unsupported. At best, the risk estimates by Laden et al. are for a marker for a complex mix of pollutants. Of note, even if additional pollution data had been available to Laden et al, the sample size of this study (N = 6) would have hindered an analysis of joint effects.

Concerns With the Analytic Methods to Assess Risk

Laden et al. used baseline individual data for behavioral risk factors and place of residence. They use a cox proportional hazards models below were adjusted for sex, smoking, education, and body mass from those earlier data.

The effect of each 10 $\mu\text{g}/\text{m}^3$ increase in average annual $\text{PM}_{2.5}$ was comparable between Period 1 (RR = 1.17; 95% CI: 1.08-1.26) and Period 2 (RR = 1.13; 1.01-1.27). Controlling for Period 1 exposures, each 10 $\mu\text{g}/\text{m}^3$ reduction in Period 2 mean $\text{PM}_{2.5}$ concentration was associated with a reduction in risk (RR = 0.73; 0.57-0.95).

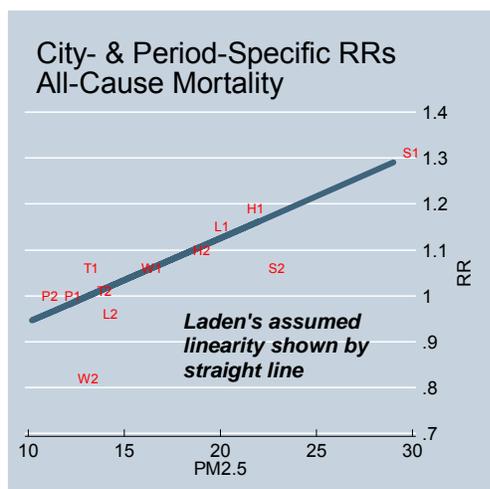
The table below summarizes the data analysis.

Adjusted proportional hazard mortality RRs for a 10 $\mu\text{g}/\text{m}^3$ increase in avg ambient $\text{PM}_{2.5}$ over the entire follow-up, and the RRs for avg $\text{PM}_{2.5}$ in Period 1 & the decrease in levels between the 2 periods				
	Model 1		Model 2	
	Cases	Entire f/u avg $\text{PM}_{2.5}$	Period 1 avg $\text{PM}_{2.5}$	Decrease in avg $\text{PM}_{2.5}$
Total mortality	2,732	1.16 (1.07-1.26)	1.18 (1.09-1.27)	0.73 (0.57-0.95)
Cardiovascular	1,196	1.28 (1.13-1.44)	1.28 (1.14-1.43)	0.69 (0.46-1.01)
Respiratory	195	1.08 (0.79-1.49)	1.21 (0.89-1.66)	0.43 (0.16-1.13)
Lung Cancer	226	1.27 (0.96-1.69)	1.20 (0.91-1.58)	1.06 (0.43-2.62)
Other	1,115	1.02 (0.90-1.17)	1.05 (0.93-1.19)	0.85 (0.56-1.27)

The authors infer a linear C-R function that looks similar to the C-R function from the original SCS study (Dockery, 1993). See graph below.

Orderings for exposure and outcomes have good correlation, but other explanations are just as plausible. For instance, the orderings for the percentage of people with sedentary lifestyles were also correlated with mortality RRs in 5 of the 6 cities (Lipfert, 2000; Moolgavkar, 2005). Compared to other air pollution epidemiology studies the RRs are higher than expected, perhaps implausibly so since those estimates are equivalent to increasing smoking among smokers by 25 pack-years (Moolgavkar, 2005). Such differences become glaringly apparent

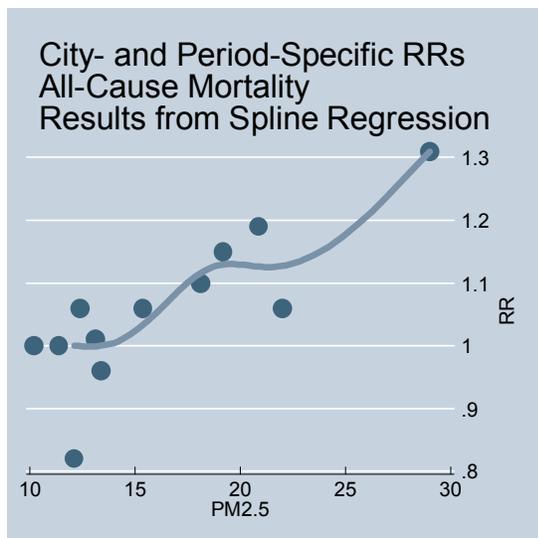
when the results are converted back to $\log(\text{RR})$, which is the actual statistical output from log-linear models prior to exponentiating the effect estimate.



Laden's conclusions regarding PM and mortality-- particularly the alleged linear concentration-response (C-R) function (see 1st graph)--are inaccurate and should be reconsidered. Graphing both the pooled all-cause using spline regression indicates the C-R functions are actually nonlinear for the complete follow-up and for Period 2 (see 2nd graph). Both scatter plots clearly show statistical flatness around RR = 1.0 for total mortality. The sole data point with a RR

significantly different from 1.0 was Steubenville in Period 1 (S1, first graph). Both of Harriman's RRs (H1 & H2, first graph) were just below statistical significance. So, the bulk of the data points were at statistical unity. A meaningful, albeit small, effect was not seen until exposure concentrations exceeded $18 \mu\text{g}/\text{m}^3$, i.e., a threshold effect which can be seen mostly clearly in the 2nd graph.

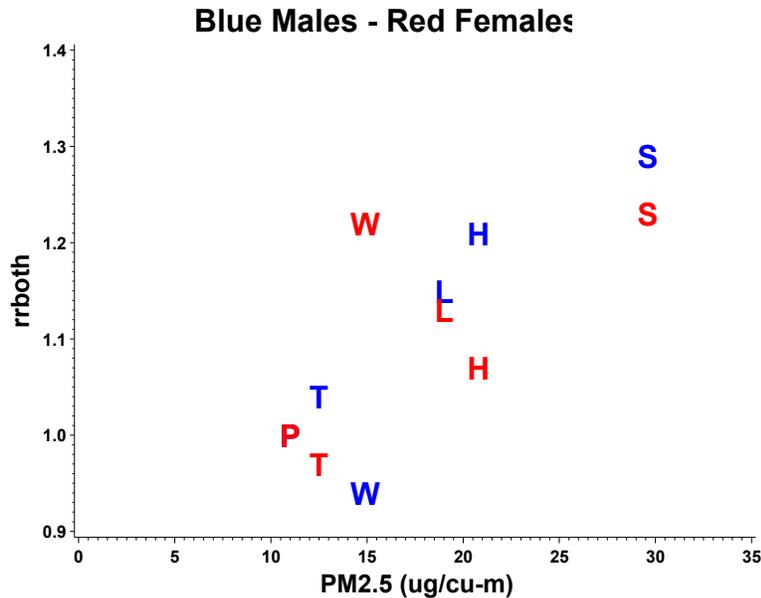
The authors' claim of linearity was limited to the range of data in the study. Thus, the model does not extend down to the 0,0 data point. Others may however be tempted to continue the line down to that point, particularly since this study included a cancer outcome (lung cancer).



For the purposes of understanding the implications of the linear no-threshold model, one can dismiss the actual RR values which are all relative to one city (Portage) and just consider them as general quantitative response values. A proper interpretation of a linear model is that those responses above zero/null -- usually in units of absolute risk, not relative risk -- are above the background rate of disease not attributable to PM. It would be inappropriate to label Portage as truly unexposed to PM (likewise for Watertown which had a lower mortality risk than Portage). So, the Portage RRs would not be a valid proxy for the PM-unassociated fraction of disease from which a linear threshold model might originate.

C-R functions in Periods 1 and 2 are not comparable as the data are approximately linear in Period 1 and nonlinear in Period 2. This becomes apparent by mentally connecting the data points ending in the number 2 (for Period 2) in the first graph. The heterogeneity between Periods 1 and 2 is so great that it appears statistically improper to combine these results. The effect estimate of RR = 1.13 in Period 2 is incorrect because it is based on a linear model and over-estimates risk for all cities except Harriman. There appears to be no association with PM in Period 2 as none of the cities have significantly elevated RRs. Three cities have effect estimates at or below 1.0 (one result <1.0 is nearly significant), and the most polluted cities show decreasing risk as concentration increases.

Cohort studies other than Six Cities studies tend to support nonlinear relationships or no association. Non-linear C-R functions were considered likely in the reanalysis of Six Cities and American Cancer Society (ACS) Study, and the accuracy of the linear models was questionable. Nonparametric analyses suggested thresholds greater than $30 \mu\text{g}/\text{m}^3$ for all-cause and lung cancer and about $20 \mu\text{g}/\text{m}^3$ for cardiopulmonary mortality in the updated ACS study. Threshold models for all-cause and nonmalignant respiratory mortality were statistically significant, while the linear models were not significant in the Adventist Health Study.



In the original Six Studies study (Dockery, et al, 1993), RRs by sex were not displayed in a graph. Those results are shown here in the chart to the left. There is a large difference for males and females in Watertown with females having a higher RR; females have a lower RR in the other cities. The relationship for males is strong ($r=0.90$), while the relationship for females is relatively weak ($r=0.67$). The relationship for the combined data is very strong ($r=0.99$).

In the Sensitivity section of the original Six Cities Study (Dockery, 1993) the authors state the following:

“However, positive associations between mortality and air-pollution levels were observed in all subgroups defined by occupational exposure and sex, and differences among the subgroups were not statistically significant.”

And partially based on this result, we have not seen sex considered in the air pollution models, other than to ‘account for it’ in the regression models. The simple ‘sex’ term is not significant, maybe because a simple term only accounts for the intercept of the regression, while a sex-by-PM interaction term (likely not used in the model) would account for differences in slope. A term is needed that determines if the male slope is different from the female, or maybe a term that assesses the significance of each slope separately.

If one just examines the data for the update by Laden et al, the findings for period 2 alone are not very convincing. None of the relative risks are statistically significant, 3 of the 6 cities have relative risks of less than 1.0 and below that for Portage, the comparison city, which has a relative risk of 1.01. The relative risk for Steubenville, with the highest PM levels, is lower than that for Harriman, which is counter-intuitive. As mentioned above, from plotting the data, one could visualize an apparent threshold for mortality at around the level of 18 ug/m³, or one could visualize there is no meaningful association.

Laden et al make the key assumption that the Cox proportional hazards model is valid. However, Laden et al. did not report the proportionality test for this model. In fact, the limited analysis of this test performed by Krewski et al in their partial reanalysis of the H6C study, discussed below, raises serious questions that the proportionality assumption.

Baseline data were used for the confounders used in the models even though several decades have since passed since their documentation. The authors admit that this could have lead to misclassification of confounders. Yet, they seem confident that such bias was minimal given the findings of other studies in which some of the same confounders were deemed not to be

significant. Aside from the time-varying aspects to confounding, the field of potential confounders was small. Anything that is associated with mortality and varies from city to city is a candidate for confounding. The opportunity for an unmeasured and influential confounder to have influenced the RRs is considerable in the SCS. Such residual confounding could be as strong as the effect estimates which are weak, as is typical for air pollution epidemiological studies. Also, individual level covariates were not available in the second period of follow-up. However, personal characteristics (age, race, gender, smoking history, occupational exposures, body mass) were typical confounders in air pollution cohort studies (Dominici et al., 2003).

Geographically-based studies such as SCS and ACS fail to take "compositional clustering" into account. This phenomenon occurs when people having more or less the same attributes make them more/less susceptible to the adverse effects of environmental exposures. In this extended SCS study, Steubenville is a classic example of this: that area of the Ohio Valley have low educational attainment, a factor related to both the risk for disease in general (mediated by behavioral factors) and the risk of exposure. "S1" is the most influential data point in the linear model presented above, so we can assume that such clustering biased the results upward.

Laden et al report that the relative risks (RR) for mortality decreased slightly between period 1 and period 2 from 1.17 (CI 1.08-1.26) to 1.13 (CI 1.01-1.27). For the entire period of follow-up they report a RR of 1.16 (1.07-1.26), which was similar to the RR of 1.14 (1.07-1.22) associated with annual mean PM_{2.5} in the year of death. The RR in the original H6C study was 1.13 compared to 1.17 in period 1, both per 10 ug/m³. The only apparent difference in methods was that in the original H6C study (Dockery et al 1993) the "ending date" of the study for each city was March or June of 1991. The end of follow-up for period 1 analysis was through 1989. Laden et al. noted that Dockery et al included "several months of follow-up, which we have assigned to Period 2." There appears to be nearly 1 ½ years less follow-up in period 1 than in the original follow-up by Dockery et al, which seems to us to be more than "several years." Laden et al indicate the months put into Period 2 rather than Period 1 occurred in 1990, when it appears the several months occurred in both 1990 and 1991. Figure A shows the difference in mean RR is due to increased RRs in Topeka, Harriman, and Steubenville in Period 1 of the update. The effect of "several months" less update and 66 less deaths (1430 vs. 1364) seem larger than expected. The difference in deaths by the cities showing difference in RRs were 7 less deaths in Topeka, 3 less deaths in Harriman, and 12 less deaths in Steubenville in Laden that Dockery. Are these differences in RR and E-R relationships an indicator of variability/uncertainty in estimates of health effect despite small differences in the database? Interestingly, if one assumes the starting point for comparison is the RR in the original H6C study (1.13), then there is no difference in change in risk through period 2, which also reported a RR of 1.13. This indicates the tenuous nature of the assertion by Laden et al., and repeated by EPA in the ISA, that this study provides *strong evidence* that reducing PM_{2.5} concentrations reduce mortality i.e. yield health benefits.

Laden et al also state that the reduction in risk observed was specifically for deaths due to cardiovascular and respiratory disease and not from lung cancer, a disease with a longer latency period and less reversibility, another conclusion repeated by EPA in the draft ISA. Laden et al. that the mortality effects of long term air pollution may be at least partially reversible over periods of a decade. Both of these conclusions or explanations are not supported by the data in the study.

The data reported in Table 3 show that the reduction in risks in Period 2 applies to all specific causes of death, namely total, CVD, respiratory, lung cancer, and other causes and are not specific to cardiovascular and respiratory deaths. In fact, lung cancer and other causes of death

had the largest reductions in risk, namely 1.06 and 0.85 versus 0.69 and 0.43 for cardiovascular and respiratory deaths, respectively. A finding of a large reduction in Other deaths does not support the authors hypothesis that reductions in PM_{2.5} produced health benefits since: 1) other deaths is not related to PM_{2.5} concentrations; 2) Other deaths shows a large reduction in deaths between period 1 and 2 than Total, CVD, and Respiratory mortality and slightly less reduction than lung cancer; 3) Total, CVD, respiratory and lung cancer mortality are hypothesized to be associated with PM_{2.5}; 4) Other deaths is the comparison no-effect group and the finding of a reduction must be due to some factor independent of PM_{2.5}; 5) the independent effect causing reduction in Other deaths may also be causing deaths attributed to PM_{2.5} reduction; and 6) it is not possible to adequately confirm this contradiction because concentration-response data for respiratory, lung cancer, and Other mortality was not included in the published report.

Overall, the study by Laden et al. falls well short of establishing causality due to its inherent limitations due to a quasi-ecological design which virtually guarantees bias due to confounding of varying types discussed above. The statistical models used in this study are questionable without observed covariates in the second period of follow-up and daily PM_{2.5} concentrations after the shutdown of the Six Cities monitoring (1985-1998). Furthermore, there were no tests for the key assumption of proportional hazards for the Cox models. The model did not differentiate acute from chronic effects, and it was difficult to differentiate historical from recent exposure. Likewise, it was difficult to assess adequate control for spatial confounding (Dominici et al., 2003). The fact that the results from this study are qualitatively consistent with most previous studies on PM_{2.5} should not necessarily infer validity. The biases operating in this study are likely to occur in other studies that use the same basic methodology.

Concerns With the Underlying Data in the H6C Study

The ISA also fails to note many of the limitations and concerns with the underlying data from the H6C itself. Many of these are detailed in the partial HEI sponsored reanalysis of this study conducted by Krewski et al. (2000). Some examples are presented below.

The original 6-cities study protocol was not found in the archives and could not be provided by the original investigators. The coding protocol allowed cigar and pipe smokers to be classified as "non smokers". The calculation of pack-years of smoking cigarettes was inconsistent as the rules for calculating this variable were not followed closely, especially in the earliest period of the study when a different form was used and a "correction factor" was applied then later dropped. Krewski et al estimated the change resulted in an underestimate of smoking pack-years of about 3% in cities where the old form was used. These cities are Watertown, Harriman, and St. Louis.

An error in a computer program resulted in some data for some subjects not being updated in all of the 6 cities. The percentage of subjects where early censorship occurred ranged from 0% in Watertown up to 11 and 12% in Portage and Topeka. The error rate for the education variable on the earliest form used was 18%. In Steubenville, code 1 meant grade school not completed whereas another from used in Harriman, Watertown, and St. Louis, it meant grade school not completed. Sometimes, the interviewers crossed out the "not", other times they did not, so the actual rate of error could have been higher. The overall error rates for recording occupational exposure were on the order of 5-6%.

The original data files for air pollutant exposures were not available. None of the reconstructed data files for fine PM provided by the original investigators could produce the exact PM air

pollution concentration averages reported in the 1993 NEJM article by Dockery et al. The levels of gaseous pollutants were not audited by Krewski et al.

Because the 6 cities study has involved, at most, 5 df for incorporating additional ecologic covariates, the rather thorough sensitivity analysis conducted by Krewski et al was mostly limited to the ACS study, i.e., could not be performed for the H6C study. Nonetheless, using the S-plus computing software provided by Grambsch and Therneau (1994). Krewski et al was able to use a flexible modeling approach to examine the validity of the Cox proportional hazards model assumptions that the hazard ratio for each covariate remained constant over time in the H6C study, and effect of each predictor of mortality is linear. Using the default 5 df regression with a spline model, they discovered in fact marginally significant time-dependent effects for both fine and sulfate PM. Also interestingly, the time dependent effects were highly dependent on the df used to model the effects. Whereas 4 and 5 df provided evidence from departure from the cox model, such departures were not observed for 3 df or less, and the latter, i.e., 3 df or less, fit the model considerably *less* well! The hazard ratio for this analysis was non-monotone, essentially decreasing to zero risk at about 5 years follow up, then increasing again up to 12 years, then falling off again. There was no discussion concerning how to determine the correct or most appropriate df.

The education level gradient in risk observed in the ACS study was also observed in the H6C study; the relative risk for all cause mortality for those with more than high school education was 0.98 (0.72-1.36) versus 1.45 (1.13-1.85) for those with less than high school education. Whether this is due to uncontrolled socioeconomic confounding, or some other factor, has never been explored. Rather, the assumption (unproven hypothesis) that this indicates those of lower socioeconomic status are more "susceptible" to the effects of PM has over time has gained acceptance as the preferred explanation.

When Steubenville was removed from the analysis, the relative risks for the high to low comparison were no longer statistically significant. This highlights the magnification approach of using the very highest to very lowest cities, alone, in the preferred analysis, as the range of fine PM exposures was reduced from delta 18 to 9 ug/m³ when this city was removed. Also, Steubenville was the only individual city for which the relative risks were statistically significant.

We note that in the update of this study by Laden et al., we are still relying on the same socioeconomic data collected many years ago. So, potential key items such as exploring further the reason for the education gradient, which may be related to difference in smoking cessation rates across the 6 cities, has never been explored. The data and potential confounding by the other pollutants is not even mentioned in the paper by Laden.

California Cancer Prevention Study (page 7-105)

In the summary of the study by Enstrom (2005), which fails to report a mortality association attributed to PM_{2.5} in the most polluted counties of California, the author of the ISA adds the following qualifying statement: "However, the use of average values for California counties as exposure surrogates likely leads to significant *exposure error*, as many California counties are large and quite topographically variable." This again illustrates the bias of the author of the ISA. No such qualifiers are placed on any of the other observational studies reporting positive associations, including many locations with locations that are "large and quite topographically variable", including the AHSMOG study which was also conducted in California, and the ACS study, which included cities across the U.S. and thus qualifies as "topographically variable."

AHSMOG

In the summary of this study, the ISA repeats and therefore accepts the speculative statement from the study that the reason increased risks were only reported in females and not males is that females may be more sensitive to air pollution, due to differences in dosimetry and exposure. The statement itself mixes two different issues. If females indeed have more exposure, then this alone could account for the difference in response in lieu of any "sensitivity" difference between males and females. However, the hypothesis that females are more exposed to ambient PM is speculative and counter-intuitive based on time activity patterns. The author of the ISA chooses to ignore the major methodology concern for this study, which is more likely to account for any differences in risk, namely, that smoking history is self reported, and therefore likely under-reported, especially given the nature of the study population (Seventh Day Adventists).

Women's Health Initiative (WHI) Study (page 7-106)

The publication of WHI study by Miller et al. (2007b) is well written. The study was performed in careful manner and the methods used were carefully documented. However, methodological issues along with biologically implausible findings limit the utility of the study. These shortcomings are not captured in the EPA summary. Also, the EPA summary conveniently fails to mention two letters to the publisher of this paper, New England Journal, that were very critical of some of the findings by Miller et al. 2007. This exclusion is another information quality concern.

Subjects in this study were postmenopausal women from the Women's Health Initiative (WHI) between the ages 50 & 79, so this group may not be fully generalizable to the general population. As the authors concede, this homogeneity could potentially magnify an effect, thus constituting a "sensitive" population. The age-related underlying risk of CVD would make them more susceptible to the effects of PM based on the current state of mechanistic knowledge in this area. The relative degree of susceptibility in this group is seen in the age range upon enrollment, 56-71; subjects aged to 62-77 during the run of the study. A group of women in, say, their twenties and thirties would likely have shown few or none of these effects from PM exposure. Statistical adjustment for BMI and other risk factors was done, but these adjustments cannot completely remove the biological effect of a unique and particularly homogeneous population with significant prevalence of factors associated with the outcome of interest.

The authors tout that this study, unlike previous studies, started out with subjects without previous CVD. A more accurate description would be "without clinically diagnosed CVD". In this group of postmenopausal women, one can assume that a significant percentage had sub-clinical underlying disease upon enrollment. Without subclinical disease before the beginning of the study, far fewer of these events would have been recorded since the natural history of the chronic disease process exceeds a mere 6-year period. Also, the subjects were not medically screened for, and therefore not excluded from the study, due to hypertension, another likely highly prevalent underlying condition in this study population. This potentially enhanced the sensitivity of this group to PM exposure and therefore to the outcomes under study.

This is a prospective observational cohort study but the exposure to various pollutants was determined ecologically, albeit possibly more precisely than most previous studies have done. The major limitation in the exposure classification is that a single year (2000) was used, leading

to exposure misclassification. As a means of comparison, other cohort studies of air pollution used multiple years of exposures. Exposure classification based on a single year may not reflect the true total local exposure over a lifetime, and the rate of decrease in PM_{2.5} concentration is different in different cities. (Janes et al, 2006). Using a single local monitor as an indicator for exposure, rather than the citywide exposure, may provide a surrogate for various neighborhood conditions rather than the PM_{2.5} exposure suggested by the authors. Therefore, the risk is not for PM_{2.5} exposure, but rather the risk for neighborhood conditions.

Cohort members were followed for a median time of 6 years. However, the cardiovascular outcomes of interest in this paper have a reasonably long latency between critical exposure time and occurrence (likely on the order of 10 years). Therefore, exposures 10 years in the past are the critical exposures that should have been tested for an exposure response.

The assumptions for the Cox Proportional Hazards model were not met; therefore, the risk estimates are biased. The Cox model is based on the assumptions that (1) the exposure estimates are accurate, and (2) the response is linear—that is the response only depends on the magnitude of the change, *not what the exposures are* (i.e., going from an exposure of 5 µg/m³ to 15 µg/m³ has the same response as going from 25 µg/m³ to 35 µg/m³). For example, in Table 3 for *Any First Time Cardiovascular Event* the HR and 95% CI are: 1.24 (1.09-1.41) for OVERALL; 1.15 (0.99-1.32) for BETWEEN CITIES; and 1.64 (1.24-2.18) for WITHIN CITIES. The authors state that the “within-city estimates tended to be larger than between-city estimates, but the differences were not statistically significant (p=0.07). However, on the model's original log scale, the coefficients are as follows:

Exposures	Coefficient x10 ⁻³	se x10 ⁻³
Overall	21.5	6.4
Between-cities	14.0	7.2
Within-cities	49.5	14.2

The table shows that the within-city estimate is about 3.5 times the between-city estimate. This ratio holds for the subcategories of First Events. (The ratio is smaller for the death category, about 1.5 for any death and ranging from less than 1 to 2.5 for the subcategories. The coefficients for death are more variable because of the smaller number of events.) They may not be statistically significant but they are definitely practically significant! Again, if the two model assumptions had been true, it should not have mattered if the comparison group was in the same city or in a different city.

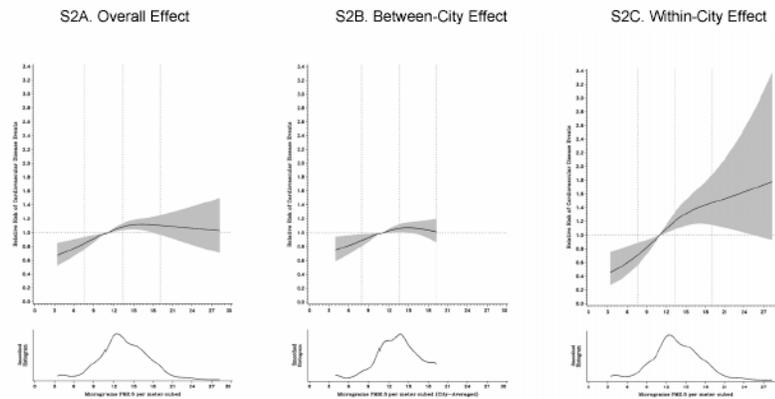
Several study findings seemed biologically implausible and do not fit the PM hypothesis. According to the *coherence* guideline for causality, the less severe (e.g. morbidity) outcomes should show stronger associations than more severe effects (e.g., mortality). These data show the reverse as associations with cardiovascular mortality are stronger than for first cardiovascular events (which include both fatal and nonfatal events). The authors attributed the higher risks to reduced misclassification of exposure (greater diagnostic certainty) and outcome and exclusion of men and women with previous CVD, and this is a plausible explanation. However, this is still conjecture.

The risks are considerably higher than those from other cohorts so the consistency guideline for causality may not be met. For example, for *Cardiovascular Death* (table below):

WHI (Miller et al. 2007)	Six Cities (Krewski et al. 2000)	ACS (Krewski et al. 2000)
1.76 (1.25-2.47) (overall)	1.19 (1.06-1.34)	1.13 (1.08-1.18)
1.63 (1.10-2.4) (between cities)		
2.28 (1.10-4.75) (within cities)		

The RRs above are implausibly high for low exposure to fine particles, as they are comparable to much higher exposures to tobacco combustion particle exposures. PM_{2.5} ambient exposure is about 10% that of a light smoker but, in this study, the risk of ambient PM_{2.5} is over 3 times greater than that estimated for the highest average ambient exposures. This difference in toxicity of ambient PM_{2.5} and cigarette smoke fine particulate is not plausible. Another seemingly implausible finding is that never having lived with a smoker (RR = 1.9) increased the risk of cardiovascular events compared to those formerly or currently living with smokers (RRs = 1.59 and 1.65, respectively); the trend is significant (p = 0.02). This is unexpected, as unmeasured PM exposure should be lower in the not-living-with-smoker group.

Figure S2. Fine particulate Matter Air Pollution and the Risk of Cardiovascular Events in Women.



The graphs demonstrate the observed relationship between the risk of cardiovascular events and PM_{2.5} concentration, including myocardial infarction, coronary revascularization, stroke, both definite and possible deaths from coronary heart disease, and death from cerebrovascular disease. Panel 2A shows the overall relationship between PM_{2.5} and risk of cardiovascular events, Panel 2B the effects between metropolitan areas, and Panel 2C demonstrates effects within metropolitan areas, with an indicator variable to adjust for each city. Risk is depicted in comparison with a reference value of 11 µg per cubic meter. The histogram in lower portion of each panel illustrates the density of exposure distribution for air pollution. All estimates adjusted for age, ethnicity, education, household income, smoking status, year of smoking, cigarettes per day, diabetes, hypertension, systolic blood pressure, BMI, and hypercholesterolemia.

The relationship between cardiovascular events and PM_{2.5} exposure is not linear for the overall effect and the between-city analysis (Supplemental figure 2 (S2) reproduced below). These graphs indicate the slope of the estimated RR is increasing at concentrations below about 15 ug/m³ and becomes flat thereafter. This response pattern is counter-intuitive to usual biological gradients.

In a letter to editor on this paper, Jerrett and Burnett (2007) conclude that Miller et al. have overstated the risk of death from cardiovascular disease associated with PM_{2.5} exposure. They argue that the increment change of PM_{2.5} used in the study, 10 ug/m³, is not available for most American cities. For example, they mention that the within city of New York, with 62 PM_{2.5} monitors covering 16 metropolitan areas around New York city, the within city 10th to 90th percentile exposure increment is only 3.3 ug/m³. Jerrett and Burnett also reflect on the within city and between city difference in risk coefficients reported by Miller et al.. They note that the difference across cities in the U.S. is likely due to difference in levels of sulfate PM whereas the variation within a city is driven by difference in PM_{2.5} coming mostly from traffic. Since these two sources of PM have different toxicities, the exposure increment Miller et al used to interpret their hazard ratio should reflect this difference. When this difference is accounted for, the hazard ratio for New York City for death from cardiovascular disease decreases from 2.28 to 1.31, which Jerrett and Burnett note is consistent with prior research.

We suspect the reason the author of the ISA fails to note the key point above is this would counter the argument that the study by Miller et al. reports *higher risks than previously reported*, a conclusion that would support lowering the NAAQS. Since the author of the ISA has gone on the public record criticizing EPA for not lowering the NAAQS further in the last review, keeping the letter to editor by Jerrett and Burnett out of the ISA is important to future advocacy efforts.

Medicare Cohort Study (pages 7-106 to 7-107)

In this section, EPA provides a summary and interpretation of the study by Eftim et al. (2008). The objective of this retrospective cohort study was to, in effect, conduct a sensitivity analysis of the findings from the Six Cities and ACS cohort studies by using a different cohort from the same geographic areas constructed from Medicare. The 2 new cohorts are labeled as Med-ACS and Med-SCS, respectively. The table below compares characteristics of these 2 studies with those earlier cohort studies.

Comparison of characteristics of the Medicare study vs. the ACS Study and SCS		
Characteristic	Medicare	ACS and SCS
Study design	Dynamic (open to enrollment)	Closed to enrollment
Study period	2000-2002	1982-1998; 1974-1998, respectively
Exposure period	2000-2002	ACS: 1979-83, 1999-2000; SCS: 1979-88, 1990-98
PM _{2.5} (ug/m ³), mean	Med-ACS: 13.6; Med-SCS: 14.1	ACS: 17.7; SCS: 16.4
Geographical areas	Counties	Metropolitan statistical areas
Population age	≥65 years	>25 years
Mortality outcomes	Total mortality	Total mortality
Exposure	Measured PM _{2.5} -- ecological	Measured PM _{2.5} & estimated PM _{2.5} from PM ₁₀ -- ecological
Time scale of exposure	Concurrent with study period	Before and during the study period
Individual-level risk factors	Age, sex	Age, sex, race, education, smoking, and more
Statistical model	Log-linear regression	Cox proportional hazards regression

The Medicare cohort consisted of about 40 million enrollees, but the analysis was restricted to those who did not change their address during the study period. Mortality rates were calculated on the same geographical locations as were the original studies. Using county of residence, the authors linked the Medicare participants to air pollution monitoring data from the EPA's Air Quality System on PM_{2.5}. No information was given about any proximity requirements for the exposure classification.

Results from scatter plots are shown below in their Figure 3. The authors state that "this shows that mortality rates tend to be higher in counties with higher average PM_{2.5} values, although the 2 counties with the highest values among Med-ACS counties have relatively low mortality rates". This will be assessed below.

The table provides the main results on the percent increase in mortality rate, again comparing the Medicare cohort results against the previous cohort studies' results. The authors modeled sociodemographic characteristics in various ways to control for confounding; however, only same-model comparisons will be drawn below.

Comparison of Results Across Studies: Estimated Percent Increase in Mortality Rate per 10 µg/m³ Increase in PM_{2.5}				
Study	Primary source	Model adjustment factors	Exposure period	% change in mortality rate (95% CI)
SCS	Krewski et al.	Individual-level age and sex	1979-1988	16.6 (7.3-26.1)
Med-SCS	[this study]	Individual-level age and sex	2000-2002	20.8 (14.8-27.1)
ACS	Pope et al.	Age, sex, race, smoking, BMI, education, ETOH, marital status, diet, occ exposure	1979-83, 1999-2000	6.2 (1.6- 11.0)
Med-ACS	[this study]	Individual level age & sex; incl 50 original SMAs from ACS, aggregated from 110 ACS locations	2000-2002	6.3 (3.8-8.9)

In the Med-ACS, three additional models (not shown here) were run, each of them producing percent increases substantially larger than the one reported above. The most influential model parameter in Med-ACS was the SMA area aggregation, shown above (the 6.3% increase). A competing model adjusting for individual-level age and sex, and for area-level education, income, poverty and employment (i.e., without the area aggregation) produced percent increases of 10.9%, so the area aggregation more closely approximated the methods of the original ACS study. Sensitivity analysis for what appeared to be the entire Medicare cohort showed that the effect estimates were insensitive to model specification, so the estimates appear to be robust.

Regarding smoking information, no data were available on the Medicare cohort. But the researchers used a workaround method devised by Zeger (2008) in which area-level COPD- and lung cancer-attributable standardized mortality ratios (SMRs) were calculated and characterized as a surrogate indicator of the long-term smoking pattern of individuals.

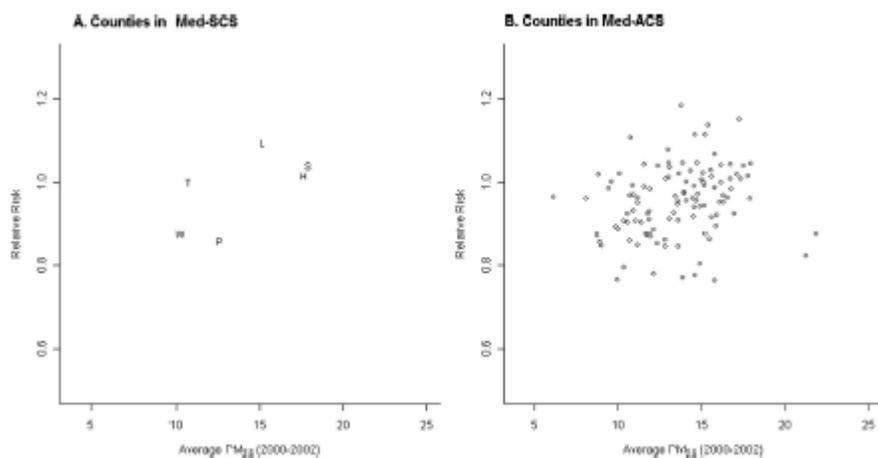


FIGURE 3. Adjusted mortality relative risk estimates plotted against average PM_{2.5} for (A) the 6 Med-SCS and (B) the 110 Med-ACS counties. T denotes Topeka, KS (the reference city for all plots); W Watertown, MA; L St. Louis, MO; S Steubenville, OH; H Harriman, Tennessee; P Portage, Wisconsin.

This was a semi-ecological study, as individual measures of PM exposure were lacking as is usually the case with such studies. Exposure misclassification biases are inherent to this study design, but the extent and direction of such

cannot be determined. No information is available about the linkage between static air quality monitors and residents' addresses (e.g., required to live within a set distance from one?). Individualized measures of sociodemographic factors (e.g., age, sex, income, and smoking) were also not available from the Medicare cohort. The surrogate strategy for modeling smoking was reasonable however. Given the area measures on exposure and SES, residual confounding was undoubtedly present to some degree.

While the authors' take on the scatter plots inferred a positive statistical relation between PM_{2.5} and the relative risk of mortality, a closer inspection suggests otherwise. The first panel (Med-SCS) gives a hint of positive linearity, but that is far from certain with only 6 data points. A linear regression re-creation on those scant data indicates a statistically insignificant results (95% CI: -0.01, 0.04) as well as a loose model fit ($R^2 = 0.42$, i.e., ~60% of the variability in risk is unaccounted for in the model). The Med-ACS scatter plot defies a linear interpretation, as its textbook scattershot.

The main point of this particular study is to gauge the sensitivity of the prior two cohort studies to population differences. The percent increases seen in the two original studies were replicated to a large extent in the Medicare cohort. This could easily be the result of using similar study methods with the same biases and limitations as the original studies. However, the characteristics of the studies differed on some potentially important factors that should have systematically produced some differing results unless the underlying methodology likewise imparts a systematic counterweight that biases towards the positive. The most obvious and biologically relevant difference was in the average age of the cohorts. Also, the estimates for the Medicare cohorts were not adjusted for confounding at the same individual level as in the original studies. And, the exposure concentrations were significantly lower in the Medicare cohort, yet the effect estimates were comparable. Perhaps the sensitivity of the older Medicare cohort produced the same statistical effects as did the original study with its younger cohort. If so, a sequel to the original studies would have produced significantly lower effects in a study population representative of the overall population.

The Medicare-based cohorts had only a 3-year window for death vs. a much longer (and biologically coherent) period in the original cohorts. The Medicare cohort's exposure and outcomes measure were constrained to the same short follow-up duration. Without prior information on past addresses and exposure levels, a meaningful exposure history--one that is in concert with the natural progression of chronicity--cannot be constructed. Quite simply, to attribute present-day deaths to present-day exposures to PM_{2.5} is biologically erroneous.

Potential confounding or effect modification resulting from between-region behavioral differences related to educational differences could not be assessed since Medicare data do not include individual-level information on education. Effect modification by educational level was noted in the Harvard Six Cities Study reanalysis. Age is a notable confounder in observational studies, but (as the author's state) the potential for such may be low because of the older age of the Medicare cohorts. The rationale for that statement is that the strength of several mortality risk factors, e.g., smoking and obesity, decreases in older ages (citations given).

This study unwittingly contributed to the body of science by quantifying the extent of geographic aggregation bias exhibited by the Med-ACS analysis when they used the larger SMAs as the unit of analysis instead of the smaller and more numerous counties. The percent increase estimated by the model adjusted for individual-level age and sex, and for county-level education, income, poverty, and employment was 10.9%. That estimate declined to 8.9% when the 110 locations were aggregated into 50 SMAs, with the same adjustments as the previous one. Dropping the county-level factors from the first model--leaving just the individual-level sex and age adjustments--reduced the estimate even further, to 6.3%. Since counties and SMAs would tend to be collinear to some extent, the degree of bias imposed by aggregation is best estimated by the two extreme estimates, i.e., the percent increase dropped from 10.9% to 6.3%. So, aggregating the analysis produced a downward bias of 42%. Of course, this suggests that the original ACS and Six Cities studies underestimated the effect of PM_{2.5} by that much since they used SMAs. Adding that bias to the effect estimates for those studies would produce absurdly high percent increases in the vicinity of 25% per 10 µg/m³ that pushes beyond the realm of credibility.

Besides the more sensitive older population and the likely aggregation bias, the higher effect estimates in the Medicare cohorts may also have been the result of a nonlinear causal relation between PM_{2.5} and mortality according to the authors. They base this on their claim that larger effect estimates were observed at lower exposure levels. However true the nonlinearity may be, the data from the Med-ACS are too scattered to be taken for anything but randomness (see scatter plot).

The lack of control for spatially correlated unmeasured confounders is major statistical limitation for epidemiologic studies that compare adjusted mortality rates with long-term air pollution exposures across different locations. The ACS re-analysis by HEI confirmed that while the original results held true, the CIs were larger after adjusting for spatial correlation. This validated the existence of the problem, and the authors acknowledge that this study could also be affected by the problem.

This was a well-designed and executed study that did all it can do with the available data and known statistical methods. Its uncanny resemblance to the original studies is comforting (and confirming) to the authors but, as mentioned previously, there should have been differences in the measured effect. Aside from the quantification of the aggregation bias, this study added nothing to the understanding of the true relation between PM_{2.5} and mortality.

Janes et al. (2007) pages 7-107 and 7-108

The ISA summarizes the findings by Janes et al. (2007a) who decomposed the associations between PM_{2.5} and mortality into two components, one for the local or county scale and another at the national level. They found that the effect estimates were different at these two spatiotemporal scales, raising the concern for confounding and bias in long term air pollution

epidemiology studies. Janes et al. concluded that when the association between PM_{2.5} and mortality at the national scale, which is more likely to be confounded than the association at the local or county scale, was set aside, there was little evidence of an association between exposure to PM_{2.5} and mortality. This led Janes et al to the conclusion that we should be very carefully about interpreting long-term trends of improving air quality and trends in reduced mortality as causal. The findings by Janes et al. are very important in the broader sense, since they raise serious general questions about the validity of the chronic air pollution studies which rely on spatial and city to city differences in long term trends in air pollution.

Since the key conclusions by James et al. are counter to the ISA author's pre-conceived views, he elects to inject a qualifying statement from Pope (2007), who commented on and was critical of the paper by Janes et al. 2007a. We note that Arden Pope is close colleague of ISA author and in fact has co-authored with the ISA author the key chronic air pollution studies that the Janes paper is critical of. In this manner, the author of the ISA is both injecting a criticism of the findings by Janes et al. and defending his own turf as an author and well known advocate for these types of studies. This sort of response has no place in the PM ISA, which is supposed to present an un-biased compilation of the existing information.

Furthermore, we find the comments by Pope (2007) on the paper by Janes et al. to be more political than factual. They state that the long term time trends as the primary source of exposure variability has been avoided in most air pollution epidemiology studies. This is clearly not the case. The studies by Laden et al (2006) and Pope et al. (2008) are examples of studies relying on long time trends as the source of exposure variability. Pope (2007) also criticized the analysis by Janes et al. on the basis that they eliminated the long-term average and spatial variability resulting in little exposure variability to exploit. In a rebuttal to Pope (2007), Janes et al. (2007b) successfully counter this criticism, noting that the standard error in the PM_{2.5} model they used was only twice the standard error in the PM_{2.5} coefficient in the ACS study. Therefore, there was considerable information left for estimating the long term effects of air pollution once the national estimate was set aside. Excluding the published response by Janes (2007b) from the ISA constitutes a significant data quality error that must be corrected.

7.6.5. Composition and Source-Oriented Analysis of PM

Netherlands Study (pages 7-10 to 7-11)

In this section, EPA summarizes and interprets the results of the update of the Netherlands Cohort Study on Diet and Cancer (NLCS)-Air study on traffic-related air pollution by Beelen et al. (2008). This summary provides a classic example of the biased approach taken by EPA to present data in the draft ISA.

The study by Beelen et al. was originally designed as a case control (CC) study and the authors report results using the CC approach in this publication. However, the draft ISA does not even mention the results of the CC analysis. The basic reason is the CC results report there is no statistically significant effect between any form of PM and mortality. In this manner, the draft ISA selectively focuses on another analysis termed the full cohort (CC) approach. Focusing on one selected set of results while excluding the other constitutes a serious data quality issue.

Beelen et al. is a semi-ecological study of associations between mortality and long-term exposure to specific air pollutants (black smoke (BS), NO₂, SO₂, PM_{2.5}) and traffic-related air pollution in a Dutch cohort. The non-pollution data were from NLCS that began in 1986 with 120,852 subjects aged 55-69 and living in 204 cities in the Netherlands. Mortality follow-up was

1987 to 1997. Participants completed an 11-page questionnaire on demographic characteristics, dietary habits and other risk factors for cancer.

Two sets of environmental pollutants were assessed and estimated at the 1986 home address. One set of exposure variables was mean outdoor air pollution concentrations of NO₂, SO₂ and BS for the years 1976-1985 and 1987-1996. PM_{2.5} concentrations were extrapolated from PM₁₀ measurements collected 1992-1996. The second environmental exposures were related to traffic. Relative risks (RRs) were based on differences between 5th – 95th percentiles. For PM_{2.5} that interval covered 10 µg/m³.

There were two analyses of these data: a full cohort (FC) analysis and a case-cohort (CC) analysis. As mentioned above, the study was originally designed as a CC study. Data from all 11 pages of the questionnaire were available on all cases in both FC and CC analyses, but referent populations were different. In the FC analyses the referent group consisted of non-cases in the FC, for which only the first page of the questionnaire was available. In the CC analysis the referent group consisted of non-cases in a randomly selected sub-cohort of 4,971 individuals selected from the total study population. All 11 pages of questionnaire data were available from the controls in the CC study. Because of the 10 additional pages of questionnaire, the CC analysis was able to include the larger number of adjustments for potentially confounding variables (n=16) that were unavailable for the full cohort analysis (n=4). The variables adjusted for in each analysis are listed in Table 1. Because of missing data on confounding variables, the number of cases and controls was less in the CC analyses.

Secondary analyses on BS found that the age-sex adjusted results between the CC and FS were comparable, but the fully adjusted results (available only from CC) were not. This they relate to the 40% of subjects in the CC that had missing values on one or more of the confounders. They also found that within the CC there was little difference between the limited and fully adjusted models when the analysis was restricted to subjects without missing values. As a result of these analyses, the authors judged that residual confounding in the FC was unlikely to be substantial. So, their analyses in the paper are mainly on the FC which produced higher RRs than did the CC.

A major question in interpreting results involves assessing whether to base conclusions on the FC analyses or the CC analyses. The authors' focus their conclusions more on the FC analyses. CC results are also presented, but are de-emphasized because, according to the authors, potential biases identified in the use of the smaller set of controls. This assumed bias is due to missing data on additional confounders and because of "random error in a downward direction, probably related to the small fraction of high exposed subjects and the skewness of the exposure distribution of the traffic variables." Supplemental Material provides limited data for validation of the bias assumption.

Adjustments Made for Potentially Confounding Variables in Full Cohort and Case-Cohort Analyses	
Adjustments in Full Cohort Analysis	Adjustments in Case-cohort Analysis
<i>N</i> Total mortality (natural causes): 15,287	<i>N</i> Total mortality (natural causes): 10,094
Age	Age
Sex	Sex
Smoking status	Smoking status
	Passive smoking
	BMI (Body Mass Index)
	Education
	Occupational exposure
	Marital status
	Alcohol use
Area level indicators of SES	Diet: including vegetable intake, fruit intake, energy intake, fatty acids, folate intake, fish consumption
Spatial Clustering (separate analysis)	Area level indicators of SES
	-

Mortality and Air Pollution Results

Full Cohort (FC) analysis PM_{2.5} exposures:

All natural causes	RR = 1.06 (0.97-1.16)
Respiratory mortality	RR = 1.07 (0.75-1.52)
Cardiovascular mortality	RR = 1.04 (0.90-1.21)

Case-Cohort (CC) analysis PM_{2.5} exposures:

All natural causes	RR = 0.86 (0.66-1.15)
Respiratory mortality	RR = 1.02 (0.56-1.88)
Cardiovascular mortality	RR = 0.83 (0.60-1.15)

There was no statistically significant association between PM_{2.5} and mortality in either of the analyses. The authors focused on the FC analyses, supposedly based on the 40% loss of subjects in the adjusted CC analyses. They indicated “residual confounding...is unlikely to be substantial” in the FC because they were similar to the results of the CC analyses when the analysis is restricted to subjects without missing values.

Cardiopulmonary mortality in the FC was singled out for additional analysis which adjusted for age, sex, smoking status, and area-level indicators of socioeconomic status. The highest effect models produced RRs in the 1.05-1.07 range (not statistically significant, NS) while the former models had RRs of 1.10 (NS).

Effect modification by cigarette smoking (FC only) and SES/education (CC only) was examined on. While the RRs were all low and NS for cardiovascular disease, never smokers had a higher RR (≈ 1.15) than did the ex-smokers (≈ 0.95) or current smokers (≈ 1.05). While current smoking had a statistically significant effect on the relation with respiratory disease (RR ≈ 1.50), never

smokers showed marginally significant effect modification for lung cancer ($RR \approx 1.50$). All-cause mortality also showed a marginally significant positive modification from never smokers.

Full Cohort Analysis: All mortality outcomes were generally elevated for three traffic variables. All but the relation of respiratory mortality and traffic intensity within a 100-m buffer [$RR = 1.21$ (1.02-1.44)] were statistically insignificant. There was no elevated RR associated with the fourth variable of 'traffic intensity on the nearest major road and distance to this road,' but no data were shown for this variable.

Case-cohort Analysis: There were no statistically significant associations between mortality and any of the 4 traffic variables in the CC analyses and RRs were generally lower than those from the FC analyses except for respiratory mortality. While the effect estimate was nearly the same as in the full cohort ($RR = 1.23$), this association was NS (95% CI: 0.89-1.68) due to the smaller size of that sample. The exceptions were for traffic intensity in a 100-m buffer where RRs for Respiratory and Lung Cancer mortality were essentially the same in both CC and FC analyses.

The FC results are, as was claimed, more precise because of a larger number of subjects. Clearly this is true, but it is not a substantive reason for focusing on a FC analysis unless the CC data were biased or unreliable, especially since the study was actually designed as a case-cohort study (p 196).

Key questions regarding perplexing issues were not addressed adequately, casting doubt on the execution of this study.

- *Since this was designed as a case-cohort study, why is 40% of the sub-cohort control sample data missing?* This casts doubts on the execution of the study.
- *Why is only part of the data supporting the bias arguments reported in the supplementary material?* It is difficult to verify that the arguments apply to all comparisons because only parts of the relevant data are presented.

Neither moderately strong associations nor statistically significant weak associations between $PM_{2.5}$ and mortality were typically found in this study. The lone statistically significant finding in the traffic intensity in a 100-m buffer for respiratory mortality could easily be attributed to chance. Out of 20 FC comparisons there was one significant finding, which is expected on the basis of random variability. Furthermore, the analysis of effect modification produced some biologically incoherent findings regarding smoking status (e.g., never smokers having a higher lung cancer and CVD RR than current smokers). The authors had no explanation for this.

The authors concluded incorrectly there was an association with traffic variables. The reason for the reduced RRs in the CC analysis was, in their opinion, not due to reduction of confounding, but because of sensitivity to sampling variation. Circumstantial evidence comes down on both sides of this argument, so such bias is inconclusive. Even if present, the statistical associations would generally remain weak and NS so this is probably a trivial pursuit.

The authors remark that their results (the first in Europe using air monitor data) are comparable to the ACS study (Pope, 2002). However, that ACS study found the increased risk for cardiovascular disease mortality, not respiratory-related mortality as in this study. Like the ACS study, the exposure metrics were based on area/geography, not individuals, leaving considerable opportunity for residual confounding. Likewise, traffic exposure metrics were ecological.

Because moving is a time-dependent variable (inconsistent with the modeling strategy), the CC analysis excluded subjects who moved during the follow-up period. Again, using BS as a $PM_{2.5}$

exposure surrogate, effects were higher for those stationary subjects (only the CC had this data). The authors suggest that the increased RRs were due to more accurate exposure assessment. They concede bias in this approach, but do not elaborate on it.

Background PM_{2.5} concentrations decreased over time (Beelen et al. 2007) but the correlations between the two time periods 1976-85 and 1987-96 were highly correlated (>0.9). So, the authors could not evaluate which of those two time periods was most influential on the RR. This also made it hard to isolate the effects of specific pollutants because they act as indicators of a mixture of air pollutants coming from the same sources (Kjellstrom et al. 2002).

On the whole, this study fails to match the authors' assertions that they found evidence of health effects from long-term exposure. It was characterized by weak and NS effects along with lingering questions over execution and emphasis one set of weak and NS results (FC) over another set of negative results (CC). EPA should be characterizing this study as one that essentially reports no association between ambient exposure to various forms of PM and mortality. Instead, they characterize it as a clearly positive study, consistent with other previous studies.

Ozkaynak and Thurston (1987)

The draft ISA states that given the dearth of published source-oriented studies of the mortality impacts of long-term PM exposure, and given the recent Medicare Cohort study now indicates that such ecological cross-section studies can be useful for evaluating time trends and/or comparing across pollution components, it may well be that examining past cross section studies comparing source-oriented components of PM may be informative. The ISA then launches into a description of the study by Ozkaynak and Thurston (1987).

We do not agree with the EPA rationale for including this study. First, as described above, the Medicare Cohort study by Eftim et al. has a number of methodological concerns that limit its' usefulness. Second, the Medicare Cohort study did not provide useful information for comparing across pollution components. The study by Eftim et al. provided no information whatsoever concerning PM source related information, which is the focus of the paper by Ozkaynak and Thurston (1987). Therefore, it is not clear why the 20 plus year old study by Ozkaynak and Thurston (1987) provides any value or is included in the ISA. Perhaps the inclusion of this paper is related to the fact that the author of the draft ISA is a co-author on this paper, a poor reason to include an old study in a compilation that is supposed to capture the new information on PM. We recommend that the paper by Ozkaynak and Thurston be deleted from the ISA.

7.6.6. Within-City Effects of PM Exposure

ACS, Los Angeles (pages 7-111 to 7-112)

In this section, the ISA provides a summary and interpretation of the study by Jerrett et al. (2005b). Again, the summary and interpretation presented illustrates the bias of the author of this part of the ISA. The further illustrates another concern namely having someone who is a co-author of a key study be responsible for summarizing and interpreting their own data in a scientific support document. This constitutes a serious conflict of interest and information quality concern. The obvious short-comings of the study and rather high degree of evidence of confounding, which we describe below, are completely over-looked in the summary.

This is a study of the Los Angeles subset of the 1982-2000 ACS cohort follow-up study by Pope et al (2002) and comprises 5,856 deaths. The original studies compared risk between cities using an average PM_{2.5} exposure so each city had one exposure value. Jerrett et al suggest that giving all residents of the city the same PM exposure may have produced effect modification by education level and exposure misclassification, the latter through the dependence on central monitors which may have reduced risk estimates 2- to 3-fold in Six Cities. The range of exposures within LA was 20 ug/m³ compared to the range of 16 ug/m³ for the mean values of all the cities in the full ACS cohort.

The objective of the study was to test the hypothesis of a within city spatial gradient of air pollution by estimating individual-level exposures by using several methods to spatially interpolate concentrations (based on data from the year 2000) from 25 PM_{2.5} monitors to residences of cohort members to 267 zip code areas in L.A. The authors also assessed impact of traffic by proximity to freeways (within 500 or 1000 meters). The set of 44 confounding variables used in previous analyses were included to promote comparability. This study also added 8 ecologic variables relating to the neighborhood (e.g., poverty, crime rate, racial composition, education, unemployment) that had not been analyzed previously. The range of exposures within LA was 20 µg/m³, which is larger than the range of 16 µg/m³ for the mean values of all the cities in the full ACS cohort.

RRs for cardiopulmonary mortality were similar in total ACS cohort and LA subset RRs are nearly 3 times those in the updated ACS cohort for total and lung cancer mortality (Pope, 2002) and about 1½-2 times greater controlling for neighborhood confounders. Inclusion of increasing numbers of confounding terms reduces RRs to statistical insignificance.

Relative Risks (RR) per 10 ug/m ³ increase of PM _{2.5} in Los Angeles based on 267 zip codes (this study) and updated ACS cohort based on mean concentrations of 51 cities (Pope, 2002)			
COD/Study	Model Adjustments		
	Age, sex, race	44 individual confounders	Neighborhood ecologic confounders
Total mortality:			
This study	1.24 (1.11-1.37)	1.17 (1.05-1.30)	1.11 (0.99-1.25)
Pope (2002)	-	1.06 (1.02-1.11)	-
Cardiopulmonary:			
This study	1.20 (1.04-1.39)	1.12 (0.97-1.30)	1.07 (0.91-1.26)
Pope (2002)	-	1.09 (1.03-1.16)	-
Lung Cancer:			
This study	1.60 (1.09-2.33)	1.44 (0.98-2.11)	1.20 (0.79-1.82)
Pope (2002)	-	1.14 (1.04-1.23)	-

Jerrett et al provide RRs for other causes of death (COD) that are not reported in other analyses of ACS or other cohort studies. Except for IHD (ischemic heart disease), these other COD may be considered negative controls (i.e., no association is expected) as these COD are not hypothesized to be causally associated with PM_{2.5} exposure, and were not being specifically tested in this study (see Figure below).

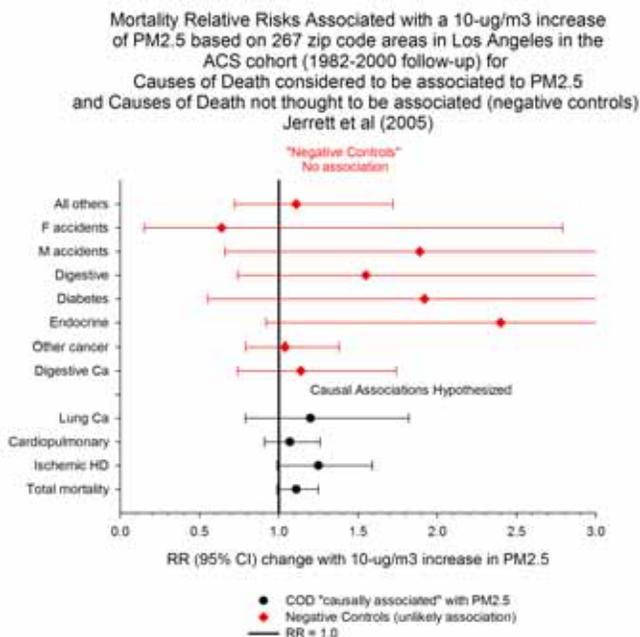
This study included two new traffic-related variables: intersection with freeways within both 500 and 1000 meters. Results (not shown) indicated that the effect of traffic was particularly elevated for lung cancer, endocrine and digestive mortality, but those estimates were imprecise and none of the statistical effects were statistically significant.

Jerrett et al concludes that effects of PM_{2.5} on mortality “may be even larger than previously reported” and results are more specific, and those associations are stronger for IHD than for cardiopulmonary and total mortality. EPA repeats this theme in their summary of this study in the ISA.

This study includes several features that are an improvement over previous analyses of the ACS cohort. The use of individual-level (vs group-level) exposure estimates are based on

residence in a zip code and interpolation from the closest air pollution monitors. More causes of death were reported, including many which can be considered “negative controls” because no association with air pollution is hypothesized. These include deaths from accidents, endocrine disorders, diabetes, and digestive disorders. Additional confounding ecologic variables were included, air conditioning being one of the more important of these.

Inexplicably, the summary in the ISA fails to note that the RRs of the negative controls show a wider range of values compared to those causes thought to be associated with PM_{2.5}. Even more striking is the fact that half of RRs for the control (non hypothesized) causes of deaths are **greater** than the RRs for COD where an association with PM_{2.5} is expected,



e.g., IHD, cardio-pulmonary and lung cancer. The same situation occurred in the traffic variables. It seems unlikely that all “unexpected” COD would be subject to the same kind or degree of confounding, which suggests bias or random variability as possible explanations for results. Model results in the graph include 44 confounding terms plus neighborhood confounders (e.g., air conditioning).

Inclusion of increasing numbers of confounding variables reduces RRs and makes them no longer statistically significant as the lower confidence intervals include zero. Inexplicably, even though the study claims to provide an improvement since it considered all of these confounding variables, the authors as well as EPA fail to note that the results were not statistically significant when these factors were included in their analysis. For instance, the model only including the PM_{2.5} term for cardiopulmonary mortality produced a RR = 1.20 (95% CI: 1.04-1.39). This decline prevailed across all COD except other cancers. The addition of the 44 potential ecological confounders lowered the RR to 1.07 (0.91-1.26), shown in the graph. Yet, there is evidence that significant confounding remained.

The authors, but not EPA in their summary of the ISA, comment on a least one of these unexplained increases, namely that the association with diabetes may adversely affect diabetics more than the general population, or it “may indicate some uncontrolled confounding because we expect type 2 diabetics to live in neighborhoods with poorer social environments.” Certainly no plausible biological explanation exists. Given the RR = 2.38 for diabetes--*which is 3 times that of IHD which has the strongest “expected” association*--the apparent uncontrolled confounding is substantial. If confounding is indeed producing these increased risks, then all the so-called “causative associations” in the ACS cohort study are strongly biased. Also, none of the models included co-pollutants (SO₂, NO₂, CO) other than ozone.

The authors state that PM-mortality associations remained robust to the freeway buffer, and risk estimates were unchanged when this variable was included in the model. Those associations are not robust. The fact that RRs are virtually unchanged when included in the model suggests they are not associated with any effect. The authors also state that imprecision in freeway exposures may have biased RRs toward the null and that their results agree with “recent evidence” of associations between traffic and cardiopulmonary mortality and lung cancer. Despite this imprecision and alleged downward bias, the RR was 1.44 (0.94-2.21) for lung cancer and residence within 500 meters of a freeway.

Jerrett states that the results generally agree with recent evidence of associations between traffic and cardiopulmonary mortality (Hoeck et al., 2002) and lung cancer (Nyberg et al. 2000). These results are, however, not comparable as they do not show an association with traffic pollution whereas Hoek et al. reported a two-fold increased cardiopulmonary RR. It would be more accurate to state that the results of Jerrett et al. which report no effects from traffic are consistent with an accurate interpretation of the study of Beelen et al, the full study follow-up of Hoeck, which also indicate no effects from traffic air pollution. Jerrett et al found no increased risk with RRs of 0.92 and 0.98 for 500m and 1000 m, respectively. This indicates no increased risk, and the RR is slightly less protective when residence is further from the traffic, contrary to an exposure-response association. For lung cancer there is a non-significant positive association at 500 meters and a negative association at 1000 meters. The argument is unconvincing that these negative results are due to exposure misclassification. Although Hoek used 100 m from a freeway or 50 m from a major road instead of 500 m, Hoek also showed a 17% increase in total mortality associated with traffic. Jerrett et al showed a deficit for proximity to a freeway.

Exposure estimates are based on 2000 exposure data and are likely to bias the estimated risks and error estimates in an unknown direction. Yet, the mortality data are from subjects enrolled in 1982 and followed through 2000. Thus, the mortality events began approximately 20 years before the exposure measures when exposures were approximately 130% higher. These levels did not necessarily change uniformly over the Los Angeles region because of differential growth patterns. It is not known what the effect on the estimated risk this exposure estimation error had. In the Pope et al (2002) analyses of the same cohort there was consistently lower risk estimates using 1980 exposure data. Jarrett's suggestion that there is not much difference in the rate of change by considering the Pope et al (2002) results is not an accurate comparison. Jarrett considers spatial variation and Pope only considers the average exposure at the two time points. Also, the discussion of different error structure and bias toward the null are not appropriate because the likely changes in exposure from the 1980s to the 1990s are not spatially uniform.

The increased range of the PM exposure values compared to the full ACS cohort underscores the weakness of ecological exposure ascertainment. The mean of concentrations from air pollution monitors in a large metropolitan area results in extensive exposure misclassification, which has been an important criticism of these studies from the beginning. The authors suggest that using the mean produced an underestimate of risk, but the aforementioned uncontrolled confounding is likely to be the larger of the two biases.

Given the level of sophistication of the analyses it is unusual that the statistical testing for mis-estimation based on incorrect exposure was limited to a simple algebraic manipulation of coefficients and an 'eyeball' test for differences [see 4th paragraph of Discussion Section of Jarrett et al. The statistical modeling of the spatial distribution is very sophisticated, but the initial level is flawed. With the level of expertise demonstrated, the authors could have used time-varying exposure techniques with some assumptions about the rate of decline in the pollution levels.

In sum, these data fail to establish a causal association between PM_{2.5} and specified causes of death. None of the associations with PM_{2.5} are statistically significant in the models with the most complete control of confounding bias, a key point not mentioned in the summary of this study in the draft ISA. We recommend that this point is added to the summary. Also, there was substantial evidence of uncontrolled confounding. Again, this key point was not noted in the summary in the ISA and must be included. The authors' conclusions infer that health effects from PM_{2.5} may, as a result of this study, be larger than previously believed, a point repeated in the ISA. This conclusion too is unwarranted as it ignores the totality of results from their analysis and should be deleted from the ISA.

7.6.7. Effects of Different Long-term Exposure Windows

In this section, the ISA summarizes and interprets the results of Schwartz et al. (2008). The summary overlooks a number of key methodological concerns and limitations. First, the analysis is based on the data from the H6C study. As we have described previously under our review of Laden et al. (2006), the underlying data from this study have serious limitations. Second, the analysis by Schwartz et al. focused solely on the single pollutant PM_{2.5}. As described in our comments under the study of Laden et al., the reductions of other pollutants besides PM_{2.5} in the six cities were even greater than those for fine PM. EPA has concluded these pollutants also potentially cause mortality. Since these co-pollutants were not evaluated by Schwartz et al, any conclusions such as "the response curve for PM_{2.5} was linear, clearly continuing below the level of the current U.S. air standard of 15 ug/m³" are meaningless.

Schwartz et al. used the Cox proportional hazards regression models to estimate adjusted mortality rate ratios, treating air pollution as a time-varying covariate, unlike previous studies of this cohort. Various model specifications were used: penalized splines, linear splines, and distributed lags (up to 5 years before the death) within a Bayesian model averaging (BMA) framework. The statistical adjustments were for the same roster of risk factors/ confounders as in the original study. These data were collected on subjects at enrollment and were not updated. Survival times were calculated as the death date (or 31 Dec 98 if still alive) minus the enrollment date.

The validity of the Cox proportional hazard model was not tested by Schwartz et al. Abrahamowicz et al. (2003) provided the theory for a method to test this assumption and attempted to apply it to the data from the American Cancer Society cohort. Due to computer limitations, they were not able to apply the method to the whole data set. However, a subset analyses indicated that the assumption of a constant size of the effect of the exposure on the hazard was not met for the PM component of the study.

Unlike the original study (Dockery et al, 1993) that used average PM_{2.5} concentrations for each city, this study's measurement came from centrally located air monitors in each community through 1988. After that, PM_{2.5} was estimated through prediction models using EPA data for PM₁₀. In our critique of the study by Laden et al. we note the serious limitations of this non-validated approach.

Death ascertainment was non-uniform: before 1979, a certified nosologist determined the underlying cause of death (COD); after that, the COD was determined by what was listed by the National Death Index (NDI). This inconsistency almost certainly produced some period-specific misclassification in an unknown direction.

Air pollution from PM_{2.5} was constructed as a time-varying covariate, a major analytical upgrade from past studies which did not use a time-variant approach. However, potentially confounding factors were collected only at enrollment. Thus, substantial confounding may still be possible should any of those poorly-measured factors have a statistical relation with both mortality and the exposure concentration. The potential for their effect to overwhelm the weak effect of a pollutant is high.

Schwartz et al report a robust examination of the shape of the C-R function for mortality clearly indicated linearity, i.e., there was no evidence of non-linearity. RRs above about 12 µg/m³ were statistically significant. However, this conclusion is based on assumption that the highest probability models were for the same year exposure (72%) and for lag one year (26%). However, it is non-sensical that if these models are correct, the combined same and / lag 1 year model carries a probability of near zero.

In the distributed lag exploration, the authors claimed that the estimated effect of a 10µg/m³ increase in PM_{2.5} for lung cancer remains elevated up to 3 years after exposure, i.e., "up to 3 years preceding the death" (which is a counterintuitive way to portray that concept). However, a closer inspection of Figure 5 shows that these relative risks were not only weak (<1.2), they were also statistically insignificant. Lung cancer mortality RR estimates were about 10% higher than those for all-cause mortality, but they were far less precise. All said, the distributed lag analysis produced no result that was significantly different from a null effect.

7.6.8. Summary and Causal Conclusions (pages 7-115 to 7-117)

The summary is based on a biased and non critical review of the new studies. We strongly recommend that section is re-written to reflect the issues we have described in our preceding detailed comments on these studies. We present below key statements in the summary that must be revised.

On page 7-115, EPA states: “The recent evidence is largely consistent with past studies, further supporting the evidence of associations between long-term PM_{2.5} exposure and increased risk of human mortality in areas with mean concentration from 14 to 29 ug/m³ (Figure 7-8).”

Figure 7.8 provides a list of studies and effect estimates but provides no indication of the levels of PM_{2.5} present in the areas of study. Therefore, there is no support for the stated 14-29 ug/m range of PM_{2.5} levels. If for example the lower end of the range comes from the update of the Harvard Six Cities Studies, as per our comments, if one just examines the data for the update by Laden et al, the findings for period 2 alone are not very convincing. None of the relative risks are statistically significant, 3 of the 6 cities have relative risks of less than 1.0 and below that for Portage, the comparison city, which has a relative risk of 1.01. The relative risk for Steubenville, with the highest PM levels, is lower than that for Harriman, which is counter-intuitive. As mentioned above, from plotting the data, one could visualize an apparent threshold for mortality at around the level of 18 ug/m³, or one could visualize there is no meaningful association.

One page 7-115, EPA states: New evidence from the Six Cities cohort study shows a relatively large risk estimates for reduced mortality risk with decreases in PM_{2.5}.(Laden et al., 2006).”

comments

As noted in our detailed comments, there are serious concerns with both the methodologies and underlying data used in this study. The first methodological concern is the very poor and incomplete exposure assessment methods used.

The study claims to focus on reductions in levels of ambient PM_{2.5} yet there were no actual PM_{2.5} measurements employed during the period of study. Rather, estimates are based on levels of another NAAQS pollutant, PM₁₀. The exposure metric is purely based on compliance monitoring measurements, and lacks the improved exposure assessment methodologies employed in more modern studies such as the MESA-Air study. These include, in addition to area compliance monitoring, outdoor neighborhood, indoor, and personal measurements, distance to monitors, home infiltration rates, etc. The authors make very strong statements about the correlation between reductions of PM and mortality, yet ignore the fact that in the areas of study, there were even more dramatic reductions in other pollutants that are also presumed to cause morbidity and mortality. The authors use the Cox proportional hazard model without validation even though the model has been demonstrated to not be valid in a reanalysis of another key air pollution epidemiology study. The authors ignore an obvious non-linearity of the data. The numerous concerns with the underlying data for this study are described in our previous .

On page 7-115, EPA concludes: “The results of new analyses from the six cities cohort and ACS study in Los Angeles suggest that previous and current studies may have underestimated the magnitude of the associations (Jerrett et al, 2005b).

As described in our detailed comments on the study of Jerrett et al. 2005, EPA ignores the fact that when all of the many sophisticated covariates that the authors claim improve the methodology of the study are included, the risk estimates for PM_{2.5} become non statistically

significant. Therefore, a more accurate conclusion statement would be that the study by Jerret et al. indicates slightly higher but non statistically significant risk estimates for PM_{2.5}. EPA also fails to note that the associations that Jerrett et al. report for exposure to PM_{2.5} and non hypothesized causes such as accidental deaths and digestive disorders, used as negative controls, were in fact higher than the risks reported for the hypothesized causes, i.e., cardiovascular mortality. This clearly indicates there is serious uncontrolled confounding of factors other than air pollution, likely, a variety of socioeconomic factors that are magnified through the use of the intra-city analysis approach. Therefore, no firm conclusions can be drawn from this study until the confounding is eliminated.

On page 7-115, EPA concludes: “The recent WHI cohort study (Miller et al.) shows even higher cardiac risks per ug/mg³ than found in the ACS study. The WHI study also considered within vs. between city mortality, as well as morbidity co-associations with PM_{2.5} in the same populations. The first showed that the results are not due to between city confounding, and the morbidity analyses show the coherence of the mortality associations across health endpoints, supporting the biological plausibility of the air pollution-mortality associations found in these studies.”:

As described in our detailed analysis of this study, the EPA conclusions fail to note the serious concerns raised by Jerrett and Burnett (2007) who stated that that Miller et al. overstated the risk of death from cardiovascular disease associated with PM_{2.5} exposure. They argued that the increment change of PM_{2.5} used in the study, 10 ug/m³, is not available for most American cities. For example, they mention that the within city of New York, with 62 PM_{2.5} monitors covering 16 metropolitan areas around New York city, the within city 10th to 90th percentile exposure increment is only 3.3 ug/m³. Jerrett and Burnett also reflect on the within city and between city difference in risk coefficients reported by Miller et al.. They note that the difference across cities in the U.S. is likely due to difference in levels of sulfate PM whereas the variation within a city is driven by difference in PM_{2.5} coming mostly from traffic. Since these two sources of PM have different toxicities, the exposure increment Miller et al used to interpret their hazard ratio should reflect this difference. When this difference is accounted for, the hazard ratio for New York city for death from cardiovascular disease decreases from 2.28 to 1.31, which Jerrett and Burnett note is consistent with prior research.

Expert Elicitation

On pages 7-115 to 7-116, EPA presents a summary of the “Expert Elicitation” (EE) findings (Roman et al., 2008). We strongly recommend that this summary is deleted from the ISA. This effort, which was sponsored and administered by EPA, does not constitute “new scientific data.” Rather, this work presents the distillation of a collection of opinions from a group of scientists hand picked by the Agency for the sole purpose of inflating PM mortality risk estimates and increasing the estimated “benefits” of regulatory initiatives aimed at reducing PM air pollution. As described below, the process used to elicit these opinions was clearly biased. We disagree with the conclusion in the ISA that EE in it’s current state of development, or specifically the EE on PM can be used to make the firm conclusions on the key uncertainties on concentration response functions for PM mortality. Therefore, we request that the above statement is deleted from the draft report.

Our view on the current status of EE is consistent and supported by the recommendations of the National Research Council Committee on Improving Risk Assessment Approaches (CIRAA), a committee commissioned by EPA for the purpose of providing advice on improving risk assessment at EPA. In the chapter on uncertainty and variability, the CIRAA express serious concerns with both the methodology and use of EE. This discussion was provided in the

context of the specific EE report on PM which the Committee used as an example to express their concerns. This text appears on pages 93-95 of the report.

First, the CIRAA did not consider the information from the EPA PM EE report to be useful for weighing risk management options.

"Expert elicitation can provide interesting and potentially valuable information, but some critical issues remain to be addressed. It is unclear precisely how EPA can use this information in its risk assessments. For example, in its regulatory impact analysis of the National Ambient Air quality Standard of PM_{2.5}, EPA did not use the outputs of the expert elicitation to determine the confidence interval for the concentration-response function for uncertainty propagation but instead calculated alternate risk estimates corresponding to each individual expert's judgment with no weighting or comparing of judgments (EPA, 2006). It is unclear how that type of information can be used productively by a risk manager, inasmuch as it does not convey any sense of the likelihood of various values, although seeing the range of commonality of judgments of individual experts may be enlightening."

Second, the CIRAA also expressed concerns with the concept of combining or weighting the expert judgments to arrive at a central conclusion, given the various biases that exist in this kind of exercise.

"Formally, combining the judgments can obscure the degree of their heterogeneity, and there are important meteorological debates on the merits of weighing expert opinions on the basis of their performance on calibration exercises (Evans et al, 1994; Budnitz et al, 1998). Two other problems are the need to combine incompatible judgments or models and the technical issue of training of the phenomenon being estimated (for example the risk of a particular disease at an environmental dose). Although methods have been developed to address various biases in expert elicitation, mischaracterization is still expected (NRC, 1996, Cullen and Small 2004)."

The CIRAA expressed serious reservations concerning the underlying cognitive tendencies that influence expert judgment and which cannot be accounted for. In our view, many of these individual concerns apply directly to the case of the EPA PM EE effort, particularly those asterisked below.

"Some findings about judgment in the face of uncertainty that can apply to experts are provided in box 4.3"

availability: the tendency to assign greater probability to commonly encountered or frequently mentioned events*

anchoring and adjustment: the tendency to be over-influenced by the first information seen or provided in an initial problem formulation*

representativeness: the tendency to judge an event by reference to another that in the eye of the expert resembles it even in the absence of relevant information

disqualification: the tendency to ignore data or strongly discount evidence that contradicts strongly held convictions*

belief in law of small numbers: the tendency of scientists to believe small samples form a population to be more susceptible than is justified*

overconfidence: the tendency of experts to overestimate the probability that their answers are correct*

"Other practical issues are the cost and time required for expert elicitation, *management of conflict of interest*, and the need for a substantial evidence base on which experts can draw to make expert elicitation useful."

We note due to the makeup of the EPA PM EE panel, the issue raised by the CIACA, management of conflict of interest is a valid concern. Many of the EPA PM EE panel had taken public position on the issues they were opining on, were reviewing either their own studies, or studies conducted by colleagues with whom they have close academic relationships.

Given the concerns expressed by the NRC CIRAA on the methodology and use of EE in risk assessment, and our own concerns expressed in more detail below, we do not agree that EPA should use the results of the EE either qualitatively or quantitatively in the PM NAAQS review.

"Given all these limitations, there are few settings in which expert elicitation is likely to provide information necessary for discriminating among risk-management options. The Committee suggests that it be used only when necessary for decision-making and *when evidence to support its use is available*. The general concept of determining the level of sophistication in uncertainty analysis (which could include expert elicitation or complex QUA) based on decision-making needs is outline in more detail below."

Using the NRC CIRAA framework, we list below the very serious concerns that are specific for the EPA PM EE effort.

availability: EPA has placed high importance on chronic PM mortality in previous NAAQS reviews. The expert EE panel members were clearly aware of this fact, thereby introducing a bias to assign a higher probability to a commonly mentioned event.

anchoring and adjustment: In previous PM NAAQS reviews, EPA placed high importance on the results of the ACS study, a fact clearly known to the panel, especially since the panel included a number of co-authors of this study. EPA introduced further bias by emphasizing the ACS study in the background materials provided to the expert panel. EPA introduced the ultimate bias when they invited the lead author of the ACS study, Arden Pope, to make a presentation during the EE deliberations. The objective of the presentation was to address and dispel any limitations of the study that the experts may have had. All of these activities ensured that the ACS study would receive primary importance in the PM mortality risk estimates, thereby introducing serious anchoring and adjustment bias.

disqualification: EPA set up an expert selection process that was designed to maximize the number of experts on the panel engaged in the conduct of observational epidemiologists, with well know opinions on the key questions, which were: 1) are the association's causal (yes); 2) is there a threshold for the effects (no). This was achieved by basing the selection of the initial expert list on the number of publications. It is well known that is very easy to publish, for example, time-series observational air pollution studies. All one needs is access to publicly available air pollution and morbidity/mortality records and the standard programs to develop correlations between the two. Based on the pilot EE for which there was a different spectrum of

experts and results, i.e., a higher percentage of those engaged in human clinical or toxicology research, EPA excluded most of these experts, who are known to have a higher tendency to have opinions different than the "strongly held views". The few people who remained on the final panel with differing views were thereby marginalized, introducing a serious member disqualification bias. EPA then provided a list of studies that did not include those reporting no association between PM and mortality, or those suggesting that threshold for health effects may actually exist, depending on the methodology of analysis used. EPA thereby disqualified these studies from consideration.

belief in law of small numbers: Based on review of the various science documents EPA has recently produced for criteria pollutants, we conclude that EPA now assumes that there exists for all criteria pollutants no threshold below which at least some individual may be affected by exposure. We term this the EPA doctrine of "infinite population susceptibility." The new causality scheme EPA has adopted for NAAQS reviews places unqualified high emphasis on the results of observational epidemiology studies of air pollution. These studies report very small relative risks that are 2 to 3 orders of magnitude below those that would normally be required to support causality. EPA continues to confuse these small relative risks reported in the studies themselves from observational epidemiology studies, with larger potential population risks derived from their risk assessment process, which are based on exposure to the general population. Therefore, we conclude that EPA has a near unqualified belief in the law of small numbers.

overconfidence: In our view, many of the scientists EPA included on the PM EE effort fall in the category of those inclined to overstate the confidence in observational epidemiology data in general, and specifically, the results of the studies EPA selected to consider in this effort. First, the panel consisted of a high percentage of experts conducting observational epidemiology studies. These panel members have a vested economic and professional interest in promoting these types of studies. We note that many of the panel members have received EPA funding, and EPA continues to provide extensive funding for observational epidemiology research. Second, many of the key studies that EPA selected to focus on were authored by the panel members or colleagues, e.g. trained or worked at the same university. Therefore, these experts were in many cases opining on their own data, or the data of colleagues, introducing a significant bias towards being less critical of the findings, resulting in overstating the confidence in the results.

Chapter 8: Public Health Impacts

8.1.1. Mortality Associated with Short Term Exposure to PM

This section presents a biased summary of the literature citing only studies that support EPA's pre-conceived conclusion that there is no threshold for the health effects of PM. For example, in summarizing the studies conducted during the time of the last review, EPA does not reference the study by Smith et al. (2000) that report a threshold for PM_{2.5} mortality in Phoenix or the study by Nicolich and Gamble (1999) that report a threshold for TSP mortality in Philadelphia. We recommend that these data are referenced in the ISA.

The summary of the study by Samoli et al. (2005) in the draft ISA fails to mention that the authors report the curve for respiratory mortality that suggests a *threshold* model was reasonable. The summary also understates the high degree of city to city heterogeneity in the results. For example the concentration response functions for model cities London, Athens, and

Crakov (Figure 1) are completely different from one another, and there is an unexplained regional difference. We suggest adding these finding to the summary.

8.1.2. Mortality Associated with Long-Term Exposure to PM

This section fails to mention the study by Abrahamowitz et al. (2003) that examined the concentration response relationship in the key American Cancer Society data set using a flexible non-parametric modeling approach. Abrahamowitz et al. reported that levels of sulfate PM below 12 ug/m³ had little impact on morality suggesting a possible “*no effect threshold.*” We strongly recommend adding this key finding to this section.

8.1.3. Summary of Concentration-Response Relationship

EPA states that studies using various statistical methods “consistently” find that a no-threshold log-linear model adequately portrays the PM-mortality C-R relationship in multi-city studies. As mentioned, the data on this topic are limited and not consistent. We recommend that EPA revise their conclusion to consider the following key points

Short Term Mortality

- The large majority of studies examining the shape of the concentration response function for mortality have relied on AIC and model fit criteria that are were not developed to assess scientific theories of etiology and are therefore inappropriate for making firm conclusions on the nature of mortality concentration response function (HEI, 2003).
- The results of existing studies are inconsistent. Some studies report a log-linear relationship with no evidence for a threshold while other studies report the existence of thresholds. It is apparent that there are methodological differences that explain some of these differences in results.
- In studies of multi-cities, there is a high degree of unexplained city-to-city heterogeneity in the mortality concentration response functions that render any central conclusion of limited utility.

Long-Term Studies

We recommend that EPA clearly indicate that there are a very limited data from which to draw firm conclusions concerning the shape of the concentration response function for chronic PM mortality and thus, the assessment of risks for chronic PM mortality must reflect this degree of uncertainty and present alternate approaches with equal weight of consideration of estimates based on the assumption of log-linear with no threshold versus estimates based on the assumption of a threshold.

We recommend that EPA note that in an examination of the key ACS study, a *threshold* for mortality was reported for one form of fine PM (sulfates) but the reverse relationship, i.e., steeper slope at lower concentrations, was reported for another form of fine PM (PM_{2.5}) (Abrahamowicz et al. 2003).

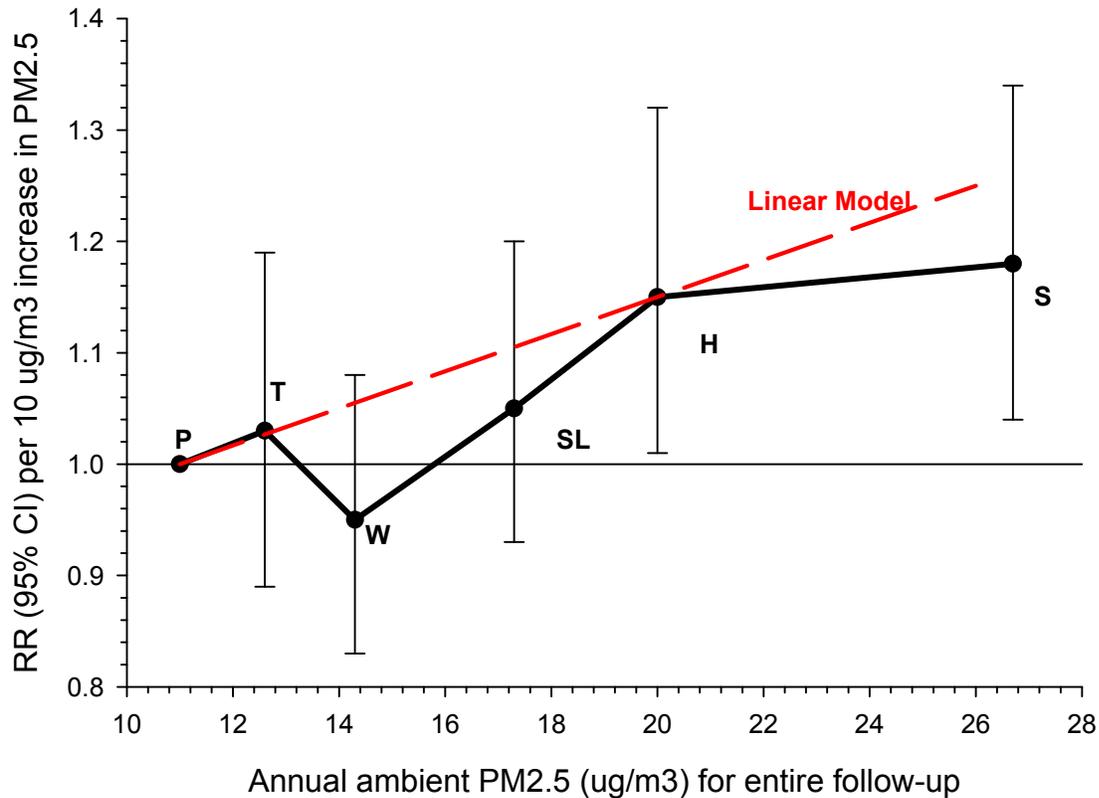
We strongly recommend deleting the reference to log-linear model recommendation from the EPA Expert Elicitation report (Roman et al, 2008). As mentioned in the detailed comments this report, there are many methodological problems and biases in this effort. The opinion of a narrow list of EPA selected scientists, most of whom were observational epidemiologists with well know views on the topic of the nature of the concentration response function for mortality, should not be construed to be scientific fact or data. The recent NRC report on improving risk assessment commissioned by EPA (NRC, 2008) clearly indicates that there is insufficient scientific data to support the use expert elicitation for decision making.

We recommend that EPA note that in the update of the Harvard Six Cities Study (Laden et al.), the authors' claim of linearity was limited to the range of data in the study. Thus, the model does not extend down to the low ambient levels. Further, we suggest the conclusions of Laden et al regarding PM and mortality are inaccurate and should be modified. Graphing the data indicate the concentration-response functions are nonlinear for the complete follow-up and for Period 2. Thus the effect estimates from linear models overestimate risk and graphical displays suggest a threshold or no association. In the following we consider total mortality, but similar findings for cardiovascular mortality can be confirmed by the reader graphing the data.

The effect estimate of 1.16 (1.07-1.26) for the total follow-up is based on a linear model. It is incorrect and overestimates risk as the relationship is not linear as readily observed when the data are plotted. Visual representation suggests a threshold below 20 $\mu\text{g}/\text{m}^3$ (Figure 1).

Concentration-response functions in Periods 1 and 2 are not comparable as the data are approximately linear in Period 1 and nonlinear in Period 2. This becomes apparent by connecting the data points in Figure 2 in Laden et al. The heterogeneity between Periods 1 and 2 is so great that it appears statistically improper to combine these results. The effect estimate of 1.13 in Period 2 is incorrect because it is based on a linear model and over-estimates risk for all cities except Harriman. There appears to be no association with PM in Period 2 as none of the cities have significantly elevated effect estimates, 3 cities have effect estimates at or below 1.0 (one nearly significant), and the most polluted cities show decreasing risk as concentration increases, see Table 2 in Laden et al.

FIGURE 1
Concentration-response Functions for total mortality
for the entire follow-up period (1974-1998)
based on estimated rate ratios for each city and linear model
(Laden et al, 2006)



● Total Mortality 1974-1998
 — RR = 1.0
 - - - Linear model

P = Portage
 T = Topeka
 W = Watertown
 SL = St Louis
 H = Harriman
 S = Steubenville

In conclusion, EPA should revise the conclusions of this section as follows:

A critical review of the update of the Harvard Six Cities Studies by Laden et al. (2006) indicates there is likely no mortality association present in the update period, or alternatively, a mortality threshold at 18 ug/m³. Other studies also tend to support nonlinear relationships or no association for chronic PM mortality. Non-linear concentration-response functions were

considered likely in the reanalysis of Six Cities and ACS and the accuracy of the linear models was questioned (Abrahamowicz et al. 2003). Nonparametric analyses suggested thresholds greater than 30 $\mu\text{g}/\text{m}^3$ for all-cause and lung cancer and about 20 $\mu\text{g}/\text{m}^3$ for cardiopulmonary mortality in the updated ACS study (Pope et al, 2002). Threshold models for all-cause and nonmalignant respiratory mortality were statistically significant while the linear models were not significant in the Adventist Health Study (Abbey et al., 1999) In a pilot study on traffic air pollution in the Netherlands, the effect estimates were generally higher and significant for higher exposure but lower and non-significant at lower background concentrations of black smoke (Hoek et al., 2002). In the full study, $\text{PM}_{2.5}$ risk estimates were negative in the *a priori* designed case control analysis, and small and non-statistically significant in the full cohort analysis (Beelen et al, 2008). Two additional cohorts showed no associations between mortality and $\text{PM}_{2.5}$, one of hypertensive males (Lipfert et al. 2000) and the other a newly published study of elderly Californians who were part of the first ACS cohort CPS (Emstrom et al. 2005)

References

Abbey et al. (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am. J. Resp. Crit. Care Med.* 159, 373-382.

Abrahamowicz et al. (2003). Flexible modeling of exposure-response relationship between long-term average levels of particulates air pollution and mortality in the American cancer Society study *J. Toxicol. and Epid. Part A* 66, 1625-1643.

AEA (2005). Methodology for the Cost-Benefit Analysis for CAFÉ. Volume 2: Health Impact Assessment. AEA Technology Environment. UK

Allen RW, et al (2009). Fine particulate air pollution, proximity to traffic, and aortic atherosclerosis: The multi-ethnic study of atherosclerosis. *Epidemiology* (in press).

Beelen, et al (2008). Long-term effects of traffic-related air pollution on mortality in a Cutch cohort (NLCS-Air study). *Environ Health Perspect* 116:196-202.

Bell ML, Ebisu K, et al (2007). Ambient air pollution and low birth weight in Connecticut and Massachusetts. *Environ Health Perspect* 115(7):1118-1124.

Burger AJ, Charlamb M, Weinrauch LA, D'Elia JA (1997). Short - and long-term reproducibility of heart rate, variability in patients with long-standing type 1 diabetes mellitus, *Am J Cardiol* 80(9):1198-202.

Brauer et al. (2002). Exposure misclassification and threshold concentration in time series analyses of air pollution health effects *Risk Analysis* 22, No. 6, 1183-1193.

Diez Roux AV, Auchincloss AH, et al (2008). Long-term exposure to ambient particulate matter and prevalence of subclinical atherosclerosis in the Multi-Ethnic Study of Atherosclerosis. *Am J Epid* 167(6):667-675.

Dominici F, et al (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA* 295(10):1127-1134.

Dominici FR, Peng D, et al (2007). Particulate air pollution and mortality in the United States: did the risks change from 1987 to 2000? *Am J Epid* 166(8):880-888.

Eftim SE, et al (2008). Fine particulate matter and mortality. A comparison of the 6-cities and American cancer society cohorts with a medicare cohort. *Epidemiology* 19 (2):209-216.

Enstrom et al. (2005). Fine particulate matter air pollution and total mortality among elderly Californians, 1973-2002. *Inhalation Toxicology* 17, 803-816.

EPA (2006a). Provisional Assessment of Recent Studies on Health Effects of Particulate Matter Exposure. EPA/600/R/063. July 2006.

Franklin M, Koutrakis P, et al (2008). PM_{2.5} composition & daily mortality in 25 U.S. communities. *Epidemiology*.

Franklin M, Zeka A, et al (2007). Association between PM_{2.5} and all-cause and specific-cause mortality in 27 US communities. *J Expo Sci Environ Epidemiol* 17(3):279-287.

Gamble J. and Nicolich M. (2006). Comments on the updated Harvard Six Cities Study. *Amer. J. Resp. Crit. Care Med.* 174, 722-724.

Gardner SY, Lehmann J, Costa D (2000). Oil fly ash-induced elevation of plasma fibrinogen levels in rats, *Toxicol Sci* 56:175-180.

Grant, L. (2002). Letter from EPA to Recipients of the Third Draft EPA PM Criteria Document, June 5, 2002)

Greenbaum, D. (2002). Letter from the Health Effects Institute to the Scientific Community, May 30, 2002.

HEI (2003O. Revised Analyses of Time-Series Studies. Health Effects Institute Special Report.

Hoek G. et al. (2002), Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. *The Lancet*

Huikuri HK, Pikkujamsa SM, Airaksinen J, Ikaheimo MJ, Rantala AO, Kauma H, Lilja M, Kesaniemi A (1996). Sex-related differences in autonomic modulation of heart rate in middle-aged subjects, *Circulation* 94:122-125.

Janes H, et al (2007). Trends in air pollution and mortality. An approach to the assessment of unmeasured confounding. *Epidemiology* 18(4):416-423.

Janes H. et al. (2007b). Partitioning evidence of association between air pollution and mortality. *Epidemiology* 18 (4), 427-428.

Jerrett M, et al (2005). Particulate air pollution, social confounders, and mortality in small areas of an industrial city. *Social Science and Medicine* 60:2845-2863.

Jerrett M, et al (2005). Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology* 16(6):1-10.

Jerrett. M. and Burnett, R. (2007). Letter to editor on the study of Miller et al. 2007. *N. Engl. J. Med* 356 (20), page 2105.

Jensen-Urstad K, Reichard P, Jensen-Urstad M (1999). Decreased heart rate variability in patients with type 1 diabetes mellitus is related to arterial wall stiffens, *J Int Med* 245:57-61.

Kannel WB, Kannel C, Paffenbarger RS, Cupples LA (1987). Heart rate and cardiovascular mortality: the Framingham Study, *Am Heart* 113:1489-94.

Ko GT, Yeung VT, Chan JC, Chow CC, Li JK, So WY, Tsang LW, Cockram CS (1997). Plasma fibrinogen concentration in a Chinese population, *Atherosclerosis* 131(2):211-7.

Krewski et al. (2000). Re-analysis of the American Cancer Society and Harvard Six Cities Studies - Part II - Sensitivity Analysis. Health Effects Institute.

Liao D, Barnes RW, Chambless LE, Simpson Jr RJ, Sorlie P, Heiss G (1995). Age, race and sex differences in autonomic cardiac function measured by spectral analysis of heart rate variability - The ARIC study, *Am J Cardiol* 76:906-912.

Liao D, Creason J, Shy C, Williams R, Watts R, Zweidinger R (1999). Daily variation of particulate air pollution and poor cardiac autonomic control in the elderly, *Env Health Perspectives* 107(7):521-525.

Laden F, et al (2006). Reduction in fine particulate air pollution and mortality. Extended follow-up of the Harvard six cities study. *Am J Resp Crit Care Med* 173:667-672.

Lipfert FW, Baty JD, et al (2006). PM2.5 constituents and related air quality variables as predictors of survival in a cohort of U.S. military veterans. *Inhal Toxicol* 18(9):645-657.

Lipfert FW (2005). Traffic density as a surrogate measure of environmental exposures in studies of air pollution health effects. Long-term mortality in a cohort of US veterans. *Atmosph Env* 40:154-169.

Lipfert FW, Perry HM, et al (2000). The Washington University EPRI veterans' cohort mortality study: Preliminary Results. *Inhal Toxicol* 12(11) Suppl 4:41-73.

Maisonet et al. (2001). Relation between ambient air pollution and low birth weight in the Northeastern United States. *Health Perspect* 109 Suppl 3:351-356.

Margaglione M, Cappucci G, Colaizzo D, Pirro L, Vecchione G, Grandone E, Di Minno G (1998). Fibrinogen plasma levels in an apparently healthy general population -- relation to environmental and genetic determinants, *Thromb Haemost* 80(5):805-10.

Miller K, et al (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med* 356(5):447-458.

NEEDS (2006). Final Report on the Monetary Valuation of Mortality and Morbidity Risks from Air Pollution. New Energy Externalities Developments for Sustainability. Project no: 502687. EU Sixth Framework Programme.

Nicolich and Gamble (2000). Evidence of a threshold effect for TSP in Philadelphia data set. *J. Environmental Medicine* 1, 279-290.

NRC (2008). Science and Decisions: Advancing Risk Assessment. National Research Council, National Academy Press, Washington, D.C.

Ostro B, et al (2006). Fine particulate air pollution and mortality in nine California counties. Results from CALFINE. *Environ Health Perspect* 114(1):29-33.

Parker JD and Woodruff TJ (2008). Preterm birth after the Utah valley steel mill closure: a natural experiment. *Epidemiology* 19:820-823.

Peng RD, Dominici F, et al (2005). Seasonal analyses of air pollution and mortality in 100 US cities. *Am J Epid* 161(6):585-594.

Pope CA, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW (1995). Particulate air pollution as a predictor of mortality in a prospective study of US adults, *Am J Respir Crit Care Med* 151:669-674.

Pope (1999). Heart rate variability associated with particulate air pollution. *Am Heart J.* 138,, 890-899.

Pope CA, Verrier RL, Lovett EG, Larson AC, Raizenne ME, Kanner RE, Schwatz J, Villegas M, Gold RD, Dockery DW (1999). Heart rate variability associated with particulate air pollution, *Am Heart J* 138(5):890-9.

Pope. C.A. et al. (2008). Fine-particulate air pollution and life expectancy in the United States. *N. Engl. J. Med* 360, 376-86.

Pope and Burnette (2007)). Confounding in air pollution epidemiology, the broader context. *Epidemiology* 18 No.4, 424-426.

Puett RC, Speizer JH, et al (2008). Chronic particulate exposure, mortality, and coronary heart disease in the nurses study. *Am J Epid* 10:1161-1168.

Rabl A. (2006), Analysis of air pollution mortality in terms of life expectancy changes: relation between time series, intervention, and cohort studies. *Env. Hlth a Global Access Science Source* 5.

Rabl A. (2003). Interpretation of air pollution mortality: number of deaths or years of life lost? *J. Air Waste Management Assn.* 53(1): 41-50.

Ramaekers D, Ector H, Aubert AE, Rubens A, Van de Wert F (1998). Heart rate variability and heart rate in healthy volunteers: is the female autonomic nervous system cardioprotective? *Euro Heart J* 19:1334-1341.

Roman et al. (2008). Expert judgment assessment of the mortality impact of changes in ambient fine particulate matter in the U.S. *Env. Sic Techol.* 42, 2268-2274

Samoli et al. (2005). Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. *Env. Health Perspect.* 113, 88-95.

Seals DR, Chase PB (1989). Influence of physical training on heart rate variability and garone flex circulatory control, *J. Appl. Physiol.* 66:1886-1895.

Seaton A, MacNee W, Donaldson K, Godden D (1995). Particulate air pollution and acute health effects, *Lancet* 345:176-78.

Seaton A, Soutar A, Crawford V, Elton R, McNerian S, Cherie J, Watt M, Agius R, Stout R (1999). Particulate air pollution and the blood, *Thorax* 54:1027-32.

Schwartz J, Zanobetti A (2000). Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. *Epidemiology* 11, 666-672/

Schwartz et al. (2008). The effect of dose and timing of dose on the association between airborne particles and survival. *Env. Health Perspect.* 116, 64-69.

Sheppard L et al (2005). Exposure and measurement contribution to estimates of acute air pollution effects. *J. Exp. Anal. Env. Epi* 15, 366-376.

Smith et al. (2000) Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. *J. Air and Waste Association* 50, 1367-1379.

Stein PK, Bosner MS, Kleiger RE, Conger BM (1993). Heart rate variability: a measure of cardiac autonomic tone, *Am Hart J* 127(5):1376-1381.

Stein PK, Kleiger RE, Tottman JN (1997). Differing effects of age on heart rate variability in men and women, *Am J Cardiol* 80:302-305.

Samoli E, Analitis A, et al (2005). Estimating the exposure-response relationships between particulate matter and mortality within the APHEA multicity project. *Environ Health Perspect* 113(1):88-95.

Schwartz J, et al (2008). The effect of dose and timing on the association between airborne particles and survival. *Environ Health Perspect* 116(1):64-69.

Shepherd et al. (2005). Exposure and measurement contributions to estimates of acute air pollution effects. *J. Exp. Anal. Env. Epid.* 15, 366-376.

Task Force for the European Society of Cardiology, and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use, *Circulation* 93:1043-1065.

Thompson SG, Kienast J, Pyke SD, Haverkate F, van de Loo JC (1995). Hemostatic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris - European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group, *N Engl J Med* 332(10):635-41.

Tsuji H, Venditte FJ, Manders ES, Evans JC, Larson MG, Feldman CL, Levy D (1996). Determinants of heart rate variability, *J Am Coll Cardiol* 28(6):1539-46.

Umetani K, Singer DH, McCraty R, Atkinson M (1997). Heart rate variability: Aging effects over nine decades (Abs), *J Am Coll Cardiol* 251A:980-(144).

Umetani K, Singer DH, McCraty R, Atkinson M (1997). Heart rate variability: Gender effects over nine decades (Abs), *J Am Coll Cardiol* 251A:980-(143).

Utell M (2000). Workshop on cardiovascular effects associated with air pollution, University of Rochester.

VanAmelsuort LG, Shouten EG, Mann AC, Sweene CA, Kok FJ (2001). Occupational determinants of heart rate variability, *Int. Arch. Occup Env. Health* 73:255-262.

Van der Bom JG, de Maat MP, Bots ML, Haverkate F, de Jong PT, Hofman A, Kluff C, Grobee DE (1998), *Arteriosclerosis Thromb Vasc Biol* 18(4):621-5.

Veglio M, Cametti G, Comp A, Lussiana L, Inadriggile G, Pecchio D (1999). The effects of exposure to moderate altitude on cardiovascular autonomic function in normal subjects, *Clin. Auton. Res.* 9, 123-127.

Violi F, Criqui M, Longoni A, Castiglioni C (1996). Relation between risk factors and cardiovascular complications in patients with peripheral vascular disease - results from the A.D.E.P. study, *Atherosclerosis* 120(1-2):25-35.

Watkinson PW, Campen MJ, Costa DL (1998). Cardiac arrhythmia induction after exposure to residual oil fly ash particles in a rodent model of pulmonary hypertension, *Toxicol Sci* 41, 209-216.

Woodhouse RR, Khaw K?T, Foley A, Meade TW (1994). Seasonal variations of plasma fibrinogen and factor VII activity in the elderly: winter infections and death from cardiovascular disease, *Lancet* 343:435-39.

Woodruff TJ, Darrow LA, et al (2008). Air pollution and postneonatal infant mortality in the United States, 1999-2002. *Environ Health Perspect* 116(1):110-115.

Woodruff TJ, Parker JD, et al (2006). Fine particulate matter (PM2.5) air pollution and selected causes of postneonatal infant mortality in California. *Environ Health Perspect* 114(5):786-790.

Zanobetti A and Schwartz J (2008). Particulate air pollution and survival in a COPD cohort. *Environmental Health* 7:48 (<http://www.ehjournal.net/content/7/1/48>).

Zeger et al. (2000). Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environmental Health Perspect.* 108, 419-426.

Zeka A, Zanobetti A, et al (2005). Short term effects of particulate matter on cause specific mortality: effects of lags and modification by city characteristics. *Occup Environ Med* 62(10):718-725.

Yarnell JW, Sweetnam PM, Rumley A, Lowe GD (2000). Lifestyle and hemostatic risk factors for ischemic heart disease: the Caerphilly Study, *Arterioscler Thromb Vasc Biol* 20(1):271-9.