Key Scientific Observations and Conclusions:
Comments on the Epidemiological Conclusions Reached in the Revised Chapter 9 of the Draft PM CD

Allen S. Lefohn, Ph.D.
A.S.L. & Associates
Helena, MT 59601
asl@attglobal.net
www.asl-associates.com

September 16, 2004

1. Key Observations and Conclusions

1.1 Historical Concerns About the Statistical Methodology

In Chapter 9, the EPA has presented its assessment of the epidemiologic literature relating monitored PM concentrations to health effects. Although the Agency has expressed concern about the uncertainties associated with the extensive body of epidemiologic evidence of associations between health effects and PM (see Section 8.4), it concludes that there is a convincing case for a causative association even though as new information became available and new modeling approaches were tried, inconsistencies in PM effect estimates continued to accumulate.

The CD examines a number of statistical issues relevant to connections between ambient PM variations and corresponding variations in health and mortality indicators. In Chapters 8 and 9, the EPA has carefully discussed the assessment of confounding by co-pollutants and adjustments for meteorological variables, statistical issues in the use of multi-pollutant models, adjustments for meteorological variables, the question of lags, measurement error, concentration-response relationships for ambient PM, and the question of

heterogeneity of particulate matter effects estimates. The CD often contains important

caveats regarding modeling issues since effect estimates will be strongly dependent on

modeling. Unfortunately, many of the caveats mentioned are put aside by the Agency in

drawing its conclusions in both Chapters 8 and 9.

Prior to writing the PM CD, the EPA had expressed concern about the strengths and

limitations of the extensive body of epidemiologic evidence of associations between health

effects and air pollutants. In its Air Quality Criteria for Carbon Monoxide (EPA, 2000), the

EPA noted on pages 6-4 to 6-6 that there were important statistical issues that needed to be

considered in critically assessing the available epidemiologic database for CO and other air

pollutants. Many of these concerns, as noted in the Carbon Monoxide CD, were especially

pertinent to daily time-series studies, which were the majority of the epidemiologic studies

available for ambient CO and which form a large part of the epidemiologic database for other

criteria air pollutants as well. The authors of the CO CD noted that statistical uncertainties,

coupled with existing biological and epidemiologic uncertainties, could pose difficulties in

judging the quantitative accuracy of pollutant effect estimates themselves and, perhaps in

some cases, their qualitative validity (see page 6-4). On pages 6-4 through 6-6, the EPA noted

5 key statistical issues in the Carbon Monoxide CD.

> (1) Sensitivity of effects estimates to different choices of statistical models/model
> *specifications*. Effects estimates for CO and other air pollutants can vary,
> depending on different choices of statistical models or specifications for
> important model parameters. For example, as the following sections show, effects
> estimates for CO or other pollutants at times differ with different choices of
> metric for the same pollutant, with different choices of modeled covariates
> (independent variables), with different lag or moving-average structures for air
> pollutants and covariates, with different choices of time spans in nonparametric
> smoothing procedures, and with different choices of adjustment for
> autocorrelation and overdispersion in the data. Also, in parametric models,
> different choices of functional forms of modeled concentration-response
> relationships can lead to different interpretations of results. Furthermore, effects
> estimates for CO and other pollutants in single-pollutant models frequently differ

from those derived from multi-pollutant models. It is increasingly clear that ambient air pollution health effects may arise from exposure to multiple pollutants, and that single-pollutant models do not necessarily describe adequately the effects of ambient CO and other pollutants. Recent increased attention to multi-pollutant models is, therefore, highly appropriate. Even so, time series of ambient CO levels often are highly correlated with time series levels of other pollutants, such as PM. Thus, effects estimates for different pollutants remain subject to confounding in multi-pollutant models. Scientific consensus as to optimal modeling strategies for time series air pollution studies has not yet been achieved. Application of nonparametric techniques, which generate exposure-response surfaces for more than one pollutant at a time, may well prove useful in future analyses. Nevertheless, however, much progress has been made in evaluating issues related to uncertainties associated with model selection and alternative specifications for important model parameters, as illustrated by U.S. EPA (1996) evaluations of time-series analyses of ambient PM exposure effects. As mentioned above, small health effects estimates generally have been observed for ambient air pollutants, and small effects would be expected on biological and epidemiologic grounds. At the same time, because the effects estimates are small, they can be sensitive to different model specifications. These can calculate to substantial differences in estimated numbers of cases of illness or mortality attributable to ambient air pollution exposure. On balance, there remains uncertainty as to the proper choice of effects estimates to employ in estimating risks of exposure to ambient CO and other air pollutants in the human population.

(2) *Measurement error in ambient CO metrics*. In this document, Appendix 3-A and Sections 4.2.2 and 4.2.3 indicate that there is substantial spatial variability in ambient CO concentrations, and that fixed-site CO measurements may not adequately index widely varying actual population exposures to ambient CO. Current evidence also suggests that ambient CO is more spatially heterogeneous than other criteria pollutants assessed in epidemiologic studies to date (e.g., $PM_{2.5}$, PM_{10}, ozone [O_3]). In many instances, misclassification of exposure leads to downward bias in statistical effects estimates. Thus, effects estimates for ambient CO may be biased downward in available epidemiologic studies, and downward bias may be stronger for estimates of CO effects than for estimates of other pollutants' effects. However, this has not been confirmed. Further research is needed to quantify the degree to which fixed-site measurements of ambient CO and other pollutants overestimate or underestimate actual population exposures to these pollutants. This research will require characterization of relationships between fixed-site CO measurements and personal CO exposures over time. It will also require accurate apportionment of total CO exposure into CO exposures of ambient and nonambient origin. Further research also is needed to characterize influences of measurement error on estimates of air pollution effects in statistical models for CO and other ambient pollutants.

(3) *Potential confounding of air pollution and weather effects.* Meteorologic events and ambient air pollutant concentrations may be highly correlated on short time scales, even when longer time trends have been filtered. It is essential to model joint effects of air pollution and weather with great care. Such joint

modeling has been conducted only rarely in time series studies of ambient air pollution. To date, simple additive or proportional assumptions generally have been made in modeling health effects of air pollutants and weather. These assumptions are not necessarily fully adequate, largely because health effects estimates for air pollutants are small and subject to large proportional differences with different model specifications. One example of recent efforts to model better possible complex relationships involving combinations of meteorological factors (e.g., temperature, barometric pressure, etc.) and to assess their impacts on air pollutant health effects estimates was the modeling of "synoptic weather pattern" effects in conjunction with ambient PM exposures (as described in U.S. EPA 1996). Application of nonparametric techniques to jointly model CO and meteorologic effects also may prove useful in future CO analyses.

(4) *Insufficient reporting of statistical uncertainty.* In available studies, statistical uncertainty generally has been assessed rather superficially, without formal consideration of the model tuning performed by the investigators. For example, lag times and averaging times for air pollutant metrics are sometimes selected to maximize statistical effects estimates for pollutants. This may, at times, lead to inflated reported effects estimates and, perhaps, unduly narrow confidence intervals for these estimates. In future studies, uncertainty arising from model tuning should be assessed more carefully. In this effort, resampling or simulation procedures, which would recreate the entire model estimation process, should be considered as a means for evaluating the accuracy and robustness of reported effects estimates.

(5) *Health effects averaged over extended time periods*. None of the available time series studies of ambient CO are capable of assessing the incremental effect of pollutants over extended time periods. For example, current models cannot confidently predict whether reduction in pollution will decrease monthly rates of hospital admissions or mortality, even if they imply a reduction of admissions on days with low pollution. This public-health-related issue cannot be addressed by daily time series analysis, using only admission or mortality counts. In future studies, investigators also could consider time averaged health effects over, say, 1 or 3 mo, in relation to pollution exposure metrics for the corresponding periods. Consideration of extended time-averaged health effects would tend to allow for detection of more chronic impacts beyond any short-term "harvesting" that might be observed in daily analyses. (In time series studies of air pollution, *harvesting* is a short-term elevation in the frequency of a health outcome during or just after a short period of elevated ambient air pollution, followed quickly by a short-term reduction in frequency of the same outcome below baseline frequency, then by a return to baseline. It has been argued that presence of harvesting would suggest that elevated air pollution exposure hastens occurrence of the health outcome by only a short time, but brings about little or no net increase in occurrence of the outcome. It also has been argued that absence of harvesting would suggest that, without the elevated air pollution exposure, the outcome might have been delayed for a long time or might not have occurred at all.)

1.2 Current Concerns About the Statistical Methodology

A key question comes to mind when reviewing the statistical concerns described above: Do the concerns described in the CO CD, taken individually or collectively, affect the time-series epidemiologic analyses purporting to show associations between health effects and PM? EPA believes that the pattern of epidemiologic time-series results reinforces the conclusions of the previous PM CD that there is a strong relationship between health effects and PM 24-h average concentrations. On page 9-45, the EPA concludes that the epidemiological evidence continues to support likely causal associations between $PM_{2.5}$ and PM_{10} and both mortality and morbidity from cardiovascular and respiratory diseases, based on an assessment of strength, robustness, and consistency in results. In addition, the Agency concludes that epidemiologic studies suggest no strong evidence for a threshold in PM-mortality relationships. However, before one agrees with the Agency, it is important to review the statistical concerns described by the EPA in Chapter 8 and assess whether the pattern of inconsistent epidemiological results contradicts EPA's decision not to provide appropriate weighting to the concerns in making its final judgment on the existing evidence. Of interest, when reviewing Chapter 9, one discovers that the words "uncertainty" and "uncertainties" are used only 5 times in the section that focuses on the evidence for linking epidemiological results with human health effects. The following are the instances:

1. While this synthesis focuses on what has been learned since the last PM NAAQS review, it also highlights important remaining **uncertainties** that remain and recognizes the value of continuing PM research efforts in a number of key areas. Although detailed delineation of research needs is beyond the scope of this document, such recommendations are to be discussed in later PM research needs documents and/or research plans to be prepared by EPA. (page 9-1).

2. Still fewer studies have used $PM_{10-2.5}$ measurements. The effect estimates are almost all positive and similar in magnitude to those

reported for $PM_{2.5}$ and PM_{10}, but few reach statistical significance. Measurement error likely contributes to greater **uncertainty**, reflected by wider confidence intervals, in effect estimates for $PM_{10-2.5}$ than for $PM_{2.5}$ and $PM_{10.}$ (page 9-26)

3. Whereas the resulting analysis using all the daily data showed clearly statistically significant positive PM-mortality associations, the results for the analyses using 1-in-6 day data sets were quite inconsistent. Hence, the use of air quality data with many missing days adds **uncertainty** to results for PM-health outcome associations. (page 9-38 – 9-39)

4. Much progress has been made in sorting out contributions of ambient PM_{10} and its components to observed health effects relative to other co-pollutants. Despite continuing **uncertainties**, the evidence overall tends to support the above conclusions that ambient PM_{10} and $PM_{2.5}$ are most clearly associated with mortality/morbidity effects, acting either alone or in combination with other covarying gaseous pollutants, with more limited support being available with regard to $PM_{10-2.5}$. (page 9-44).

5. Hence, although at the time of the 1996 PM AQCD, the epidemiologic evidence was viewed as substantiating well PM_{10} or $PM_{2.5}$ associations with human mortality and morbidity, **uncertainties** remained with regard to (a) the contribution of specific PM constituents to PM toxicity and (b) the biological plausibility of the reported effects and/or the mechanisms of action underlying them. (page 9-46).

Clearly, even though considerable discussion has taken place both in Chapter 8 and at

CASAC deliberations, the authors of Chapter 9 have made the decision not to factor into its

conclusions the large uncertainties associated with the use of the epidemiologic methodology

and the resulting inconsistencies in PM effect estimates that continued to accumulate.

Without a clear statement in Chapter 9 that indicates the ramifications of these serious

deficiencies associated with the effects estimates, a selective set of data within the PM

Criteria Document will be cited in the EPA's Staff Paper and used in risk assessments

without acknowledgement of the large uncertainties associated with the inputted data.

Statistical uncertainty and risk analyses performed by the Staff Paper authors will not include

considerations associated with the highly uncertain risk estimates, which actually include values that are positive, negative, and no effect estimates. Pooled estimates will be used that will blur the actual uncertainty associated with the estimates.

Many of the statistical caveats raised throughout Chapters 8 and 9 indicate a pattern of inconsistent results that is troubling. Examples of the growing pattern of inconsistent and inconclusive findings include the following:

- Instability of PM mortality effect estimates resulting from different model specifications of weather effects and time trends.

- Instability of PM effect estimates resulting from different selections of monitoring sites within cities.

- Increased heterogeneity of PM effect estimates across cities.

- Greater diversity of findings among studies and across study areas.

- Contradictory results from mortality displacement studies.

- PM effect lags that are inconsistent across cities and across studies.

- Exposure-response relations that are inconsistent across cities and across studies.

- Inconsistencies between short-term and long-term effect studies, such as respiratory effects of fine particles.

- Contradictory findings among long-term studies.

For example, the NMMAPS reanalysis of the 88 largest U.S. cities plays a rather prominent role in current considerations of the PM CD for the effects of ambient PM on mortality and morbidity. The reanalysis of this study in HEI (2003) reduced PM mortality effect estimates by a factor of two and increased their associated standard error estimates. The combination of decreases in effect sizes and increases in standard errors for individual-city PM effects meant that genuine inter-city effect differences were more difficult to discern

statistically. The PM CD reports that even greater spatial heterogeneity appears to exist across newly reported study results (8-311, 9-37).

Figure 8-1 illustrates the estimated excess risks for PM mortality (1 day lag) for the 88 largest U.S. cities as shown in the revised NMMAPS analysis. The CD stresses in many places that there are more positive than negative estimates of relative risk. However, an interesting observation is that there are many negative estimates. The presence of so many negative estimates requires some explanation of the reasonableness of the estimation procedures used in the epidemiological studies. It is interesting to note that, except for three cities (Oakland, New York, and Little Rock,), all of the remaining cities have confidence limits that contain 0. The prevalence of negative risk estimates in the collection requires very careful consideration since if they are neglected, then the positive estimates should also be neglected. These inconsistencies are suggestive of model inadequacies, particularly in regard to confounders and effect modifiers.

The claim that consistent effect estimates exist (page 9-26) is clouded by the incompatible comparisons in Figure 9-4. In Chapter 9, the CD summarizes results from a number of PM mortality studies such as the comparative display in Figure 9-4 (page 9-24). This figure, and others like it, is misleading because some of the plotted risk confidence intervals are derived from single-city studies, while each multi-city study is represented by a single confidence interval that conceals evident differences among cities and does not incorporate city-specific risk uncertainty. For example, the study with the shortest confidence interval is actually an artificial composite of 90 separate and highly variable model estimates.
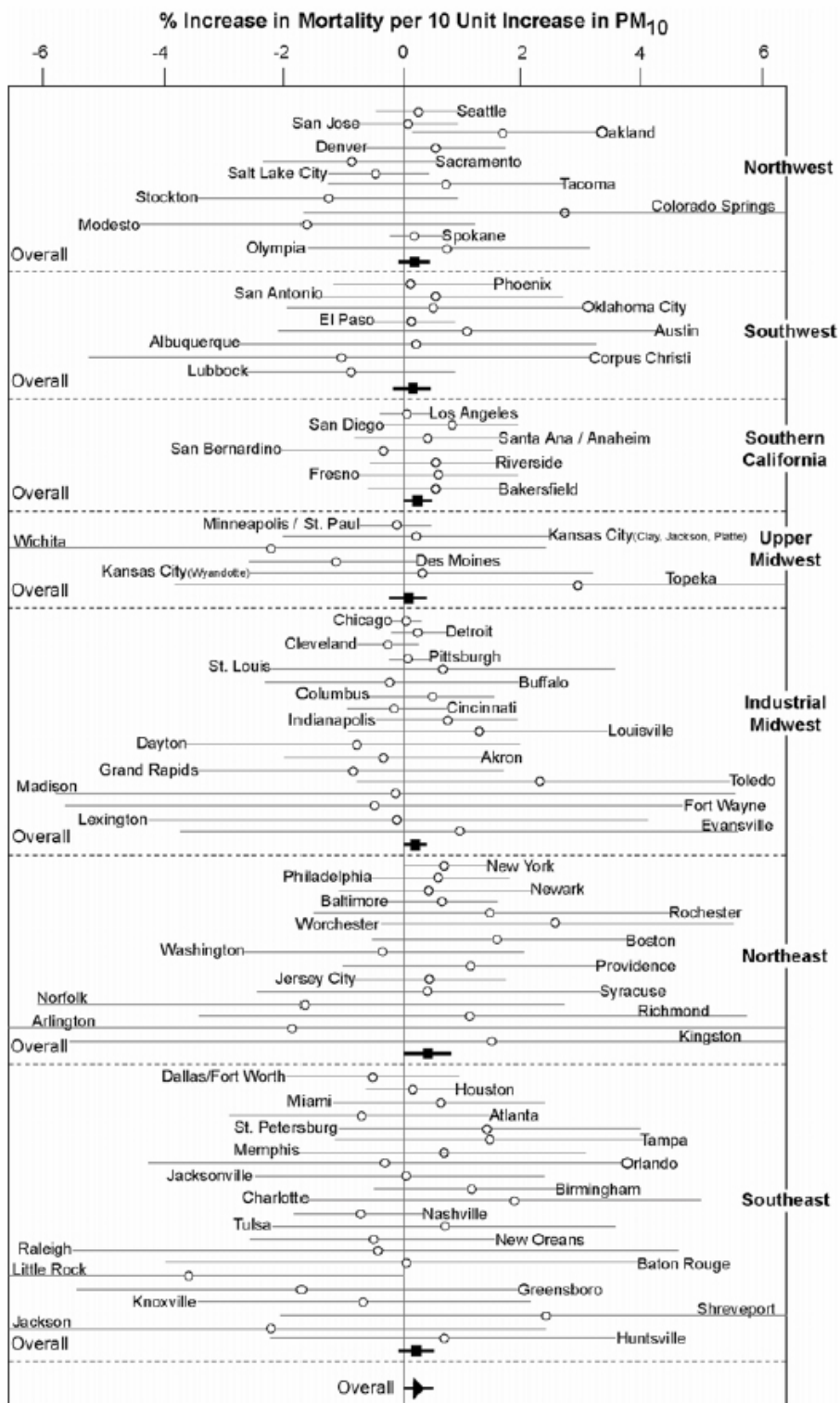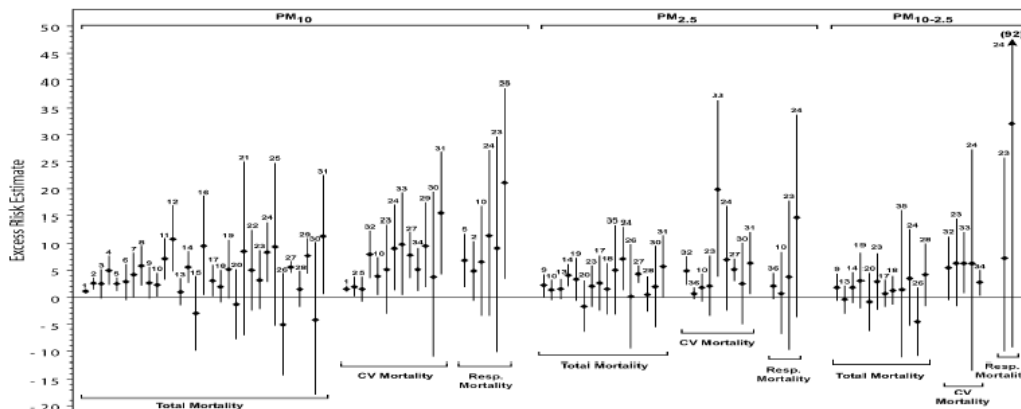
Figure 8-1. Estimated excess risks for PM mortality (1 day lag) for the 88 largest U.S. cities as shown in the revised NMMAPS analysis.

Figure 9-4. Excess risk estimates for total nonaccidental, cardiovascular, and respiratory mortality in single-pollutant models for U.S. and Canadian studies. PM increments: 50 μg/m3 for PM10 and 25 μg/m3 for PM2.5 and 10-2.5. Results presented from time-series studies that did not use GAM or were reanalyzed using GLM.

Beginning in January 2000, Drs. Switzer and Lefohn have commented on the various versions of the PM CD. Several of the concerns raised with the time-series analyses were similar to the important statistical issues that EPA believed needed to be considered in critically assessing the available epidemiologic database for CO and other air pollutants (EPA, 2000). Over the period 2000 - 2004, we have focused our attention on a number of statistical concerns, which are still not adequately addressed in either Chapters 8 or 9 of the PM CD. These concerns include the following:

1. Confounding of weather and time trends with PM effects
2. Heterogeneity of PM effects and effect modification
3. Heterogeneity of exposure within study areas
4. The relation between exposure and response
5. Lag selection and distributed-lag models
6. Mortality displacement
7. Long-term PM-mortality studies

In the previous drafts of Chapter 8 of the PM CD, the authors emphasized the "coherence" of PM health effect estimates across a number of epidemiologic studies.

10

However, as we have noted in our previous comments, in fact, it is difficult to see this coherence, even within the multi-city studies that the CD appropriately emphasizes. The latest version of Chapter 8 cites recent studies and re-analyses of earlier studies that, taken together, demonstrate the increasing unreliability and inconsistency of the model-based estimates of PM effects on mortality. Most importantly, the pattern of inconsistent results has apparently expanded with the accumulation of new studies and the reanalysis of earlier studies. Even among those analyses that do estimate health effect reductions from reduced PM, quantitative estimates differ by at least an order of magnitude. What is more striking are the inconsistencies and the sensitivity of PM effect estimates to modeling choices, including additivity assumptions, seasonal differences, regional grouping, spatial heterogeneity, lags and multiple lags, and treatment of gaseous pollutant confounders.

The CD has appropriately emphasized multi-city studies, in particular the 90-city study, because a common modeling approach was used. Thus, one would anticipate that the heterogeneity of PM effect estimates is less attributable to disparate model selection. However, a multiplicity of cities does not guarantee that there are not important model deficiencies in the common model and the statistical methods relied upon by the CD.

In reaching its most current conclusions concerning its assessment of epidemiologic evidence in its revised version of Chapter 9 of the PM Criteria Document, the EPA has considered information in relation to several criteria (see Page 9-21) noted at the outset of Section 9.2 of the CD: (1) the *strength* of reported associations, in terms of magnitude, statistical significance, and statistical power/precision of effects estimates; (2) the *robustness* of reported associations to the use of alternative model specifications, potential confounding by co-pollutants, and exposure misclassification related to measurement error; (3) the

consistency or general concordance of findings in multiple studies of adequate power, and in different persons, places, circumstances and times; (4) temporality, in terms of lag periods between exposure and observed effects; (5) the nature of concentration-response relationships; and (6) information from natural experiments or intervention studies as to the extent to which reductions in PM-related air pollution have been observed to be associated with improvements in health measures.

Dr. Switzer (2004c) has carefully evaluated the epidemiologic assessment criteria presented in Section 9.2.2. He has excluded the final speculative topic dealing with natural experiments because he believes that these experiments are uncontrolled and unadjusted observational studies, subject to strong publication bias. His conclusions are as follows concerning the 5 criteria:

Strength of reported associations. PM effect coefficients vary substantially from city to city and are not consistently positive or consistently statistically significant. The single-city effect estimates for multi-city studies should have been examined in assessing reported associations. When estimates from the 90-city mortality study are averaged, the net effect is very small.

Robustness of epidemiologic associations. PM effect estimates are sensitive to modeling of weather, long-term trends and seasonality, and selection of time lags. The way in which weather and co-pollutants are confounded with PM has still not been adequately explored because covariation models have not allowed for interaction. The recommended further investigation of weather effects [9-34] could even show that remaining PM effects are substituting for heretofore unmodeled weather effects.

Consistency of epidemiologic findings. PM effect estimates are not consistent from city to city, are not consistent within cities when different monitoring sites are compared, and are not consistent from season to season. Inconsistencies are suggestive of model inadequacies, particularly in regard to confounders and effect modifiers. There is no way to anticipate the consequences of PM reduction, given the diverse array of epidemiologic findings.

Temporality and the question of lags. Most studies assessed in Chapter 9 did not examine the pattern of effect estimates as a function of time lag, nor did they check non-causal negative lags. The extent of lag selection bias remains unresolved.

Statistical properties of PM effect estimates as a function of time lag are difficult to anticipate because of temporal autocorrelation of PM time series.

Concentration-response. The assumption that health effect responses are proportional to the ambient PM concentrations is central to many of the reported epidemiologic findings and conclusions. Departures from proportionality can have profound consequences both assessing PM health effects and for regulatory standards. There is a diversity of findings regarding the proportionality and the power of statistical tests for proportionality is weakly understood. Pooling of response functions across cities to obtain linearity is not statistically justified and leads to regulatory dilemmas. A separate 24-hour standard is superfluous under presumed linearity of concentration-response. The relation between individual-level response and the community-level response used in time series studies is not well understood. And finally, some studies suggest that acute PM mortality effects are consistent with mortality displacement in frail populations, an important regulatory issue not addressed in Chapter 9.

Dr. Switzer (2004c) concludes that the available epidemiologic information does not meet the criteria for a convincing case for a causative association, nor does this information provide a basis for anticipating the effects of PM reductions. As new information became available and new modeling approaches were tried, inconsistencies in PM effect estimates continued to accumulate. Although Chapter 9 does enumerate important caveats regarding the role of modeling choices on PM effect estimates, many of its own caveats were put aside in drawing conclusions. Without a clear understanding of the reasons for inconsistent effects estimates, one cannot rule out the possibility that PM effect estimates are model artifacts.

It is clear that the PM effect estimates are delicate; this is not surprising given that they are superimposed on much stronger effects due to, for example, concomitant weather variations. These widely varying, sometimes negative, PM health effect estimates are symptomatic of probable model shortcomings. Given the difficulty of estimating PM health effects, it is clear that consistent inconsistencies are demonstrated throughout the revised chapter. One simply cannot draw comfortable conclusions regarding the circumstances and

magnitudes of ambient PM health effects, or whether reported PM health effects are causative.

An alternative to the cause-and-effect explanation provided in Chapter 9 is that many of the results cited may be mostly associated with modelling artifacts. Ultimately the CASAC and the EPA must answer a serious question: If the time-series data are the most important information available for establishing Federal PM standards, are the data good enough to use in the decision-making process? Based on the evidence presented in Chapters 8 and 9 in the PM CD, one simply cannot draw comfortable conclusions regarding the circumstances and magnitudes of ambient PM health effects, or whether reported PM health effects are causative. Thus, one might conclude that there is still too much uncertainty remaining in the epidemiological time-series results. In addition, many of the concerns expressed by the EPA in the Carbon Monoxide CD (EPA, 2000) about the strengths and limitations of the extensive body of epidemiologic evidence of associations between health effects and air pollutants have not been adequately addressed in the most current version of the PM CD. The growing pattern of inconsistent and inconclusive findings is troublesome and presents both scientists and policymakers with a very difficult decision. Simply stated, the science is just not substantial enough at the moment to provide the foundation upon which a clear path can be built leading directly from the science to the policymaking decision arena.

2.   Heterogeneity of Exposure

2.1   Introduction

With the release of the August 2004 revised version of Chapter 9, when discussing spatial variability, there are serious inconsistencies throughout the chapter. With the exception of a few longitudinal panel studies, most epidemiologic studies of PM health effects rely on ambient community monitoring data providing 24-h average PM concentration measurements. The use of ambient concentration can lead to misclassification of individual exposures and to errors in the epidemiological analysis of pollution and health data, depending on the pollutant and on the mobility and lifestyles of the population studied. Any *gradient* that may exist between a fixed-site monitor and the outdoor microenvironments near where people live, work, and play, affects the concentration profile actually experienced by people as they go about their daily lives. If outdoor PM constituent concentration profiles are either spatially or temporally variable, it is likely that exposure misclassification errors could be introduced in the analysis of PM air pollution and health data. Variability in PM concentrations across study areas could influence epidemiologic study results. The authors of the CD note the importance of spatial variability on page 9-38, when they state, "*Greater spatial variability in PM levels would be expected to increase exposure error, potentially affecting epidemiologic study results in those areas*" (emphasis added).

Aspects of the spatial and temporal variability of the 24-hour average $PM_{2.5}$ concentrations for 27 U.S. metropolitan statistical areas (MSAs) for the period 1999, 2000, and 2001 were considered in the main body of Chapter 3 (Section 3.2.5) and in Appendix 3A. In Chapter 3, the CD points out that whereas high correlations of $PM_{2.5}$ provide an

indication of the spatial uniformity in temporal changes in directions (i.e., concentrations at both sites increasing or decreasing together) across urban areas, **the correlations do not imply uniformity in the $PM_{2.5}$ concentrations themselves** (page 3-46). It is important to note that in Chapter 3, the 90th percentile difference in concentrations (P_{90}) and the coefficient of divergence (COD) were used to provide an indication of the degree of spatial uniformity in $PM_{2.5}$ concentrations across urban areas. A COD of zero implies that both data sets are identical, and a COD of one indicates that the two data sets are completely different. Figure 3-18 (page 3-48), which is reproduced on the next page, illustrates examples of the varying degree of spatial variation in concentrations between pairs of sites that are highly correlated (i.e., r>0.9) in three MSAs in Columbia (SC), Chicago (IL), and Detroit (MI). Although the correlation coefficients were greater than 0.9 for all site pairs in Figure 3-18, there is a high degree of variation as indicated by both the P_{90} and the COD. Pairs of sites showing high correlations and CODs < 0.1 and P90's ≤ 5 µg/m3 (as in Columbia, SC) indicate spatial homogeneity in both $PM_{2.5}$ concentrations and in their temporal variations. Pairs of sites showing high correlations (r > 0.9) and CODs > 0.2 and P90's > 10 µg/m3 (as in Detroit, MI) indicate spatial heterogeneity in concentrations but homogeneity in their day-to-day changes. That is, although the absolute concentrations on a given day are different between site pairs, both sites tend to increase and decrease together from day to day. Although the three MSAs exhibited high correlations among their monitoring pairs, Chapter 3 stresses that in many cases, correlation coefficients ranged from high to low values among monitoring pairs within the various MSAs.

Columbia SC 1999 & 2000

45-079-0007 vs 45-079-0019
r = 0.97
COD = 0.06
$P_{95} = 2.7 \mu g/m^3$

Chicago IL 2000

17-031-2001 vs 17-031-4201
r = 0.94
COD = 0.14
$P_{95} = 5.5 \mu g/m^3$

Detroit MI 2000

26-099-0009 vs 26-163-0033
r = 0.93
COD = 0.22
$P_{95} = 12.7 \mu g/m^3$

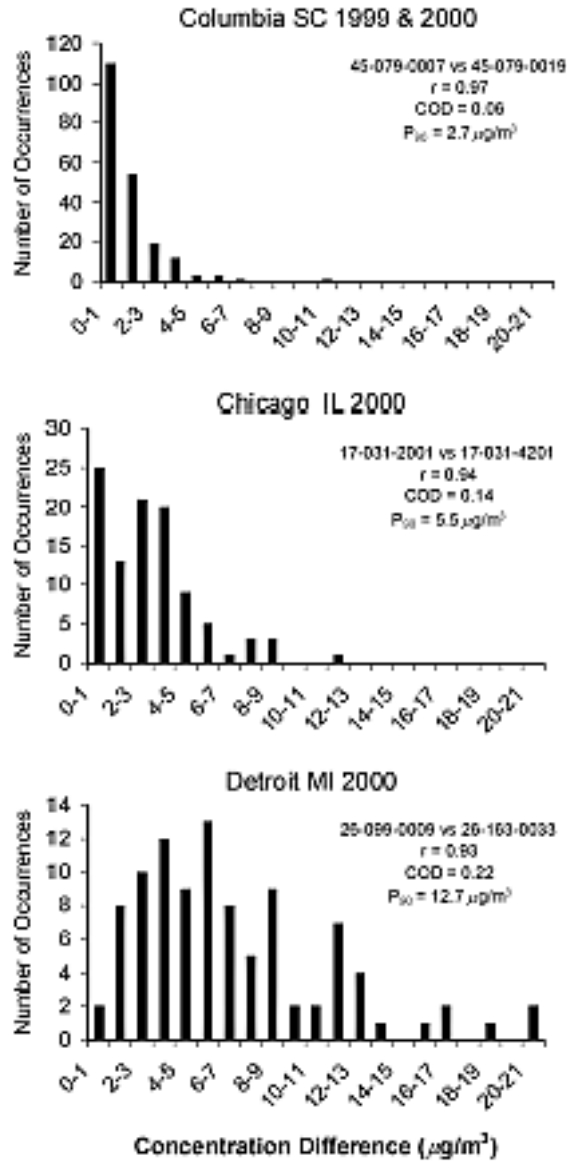Concentration Difference ($\mu g/m^3$)

Figure 3-18 in CD on Page 3-48.

17

Table 8-40 on page 8-281 illustrates statistics summarizing the spatial behavior of PM$_{2.5}$ concentrations, based on detailed analyses presented in Appendix 3A. Figures 3A-1 – 3A-27 (pages 3A-4 – 3A-30) in Appendix A summarize the information about the spatial and temporal variability of 24-hour average PM$_{2.5}$ concentrations. The summary table is presented in this review. The mean Pearson correlation coefficient for all site pairs considered in a given MSA, the average of the annual mean concentrations, the range of annual means at the sites considered, the average 90th percentile value (P90) of the absolute concentration difference, and the average coefficient of divergence (COD) are shown in Table 8-40 for MSAs satisfying data completeness criteria used for inclusion in Appendix 3A. Data in Table 8-40 show the ranges in the metrics (annual means) considered for all the MSAs included in the analyses. The CD points out that substantial concentration gradients exist on many days across some MSAs. The effects of outlying sites on the summary statistics were examined for the Atlanta, GA, Washington, DC, Seattle, WA and Los Angeles MSAs by removing them from the analyses. Their deletion either had no effect (as in Washington, DC) or a very large effect (as in Seattle, WA). In addition to outlying monitoring sites, located outside the main urban air shed, monitoring sites within the urban core can also enhance the spatial variability in MSAs, as shown for Detroit, MI. As discussed in Chapter 3, there are a number of factors that contribute to spatial variability in ambient PM$_{2.5}$ concentrations in urban areas.

The CD concludes on page 8-280 that "… concentration gradients can exist in MSAs whose monitoring sites are highly correlated and that use of correlation coefficients alone is not enough to characterize spatial variability". Because of incomplete data capture for some

18

TABLE 8-40. SUMMARY STATISTICS SHOWING MEAN SITE-PAIR PEARSON

CORRELATION COEFFICIENTS, ANNUAL MEAN PM$_{2.5}$CONCENTRATIONS (μg/m$_3$), THE RANGE IN ANNUAL MEAN CONCENTRATIONS (μg/m$_3$), MEAN OF 90$_{th}$PERCENTILE DIFFERENCES IN CONCENTRATIONS

BETWEEN ALL SITE PAIRS (μg/m$_3$), AND COEFFICIENTS OF DIVERGENCE

(COD) FOR MSAs MEETING SELECTION CRITERIA GIVEN IN APPENDIX 3A.

VALUE IN ( ) REFERS TO NUMBER OF SITES.

Mean Correlation	Annual Mean Concentration (μg/m$_3$)	Range in Annual Means (μg/m$_3$)	Mean P$_{90}$ (μg/m$_3$)	Mean COD
Eastern U.S.				
Philadelphia, PA-NJ (5)	0.89	15.3	2.3	5.1	0.12
Washington, DC (6)	0.84	14.6	3.4	7.1	0.17
Washington, DC* (5)	0.85	14.6	3.4	6.1	0.17
Norfolk, VA (5)	0.96	13.5	0.7	3.6	0.08
Columbia, SC (4)	0.92	15.6	1.8	4.3	0.09
Atlanta, GA (7)	0.71	20.2	4.5	10.3	0.18
Atlanta, GA* (6)	0.78	20.6	3.5	8.5	0.15
Birmingham, AL (5)	0.83	20.3	5.3	11.5	0.18
Tampa, FL (4)	0.74	12.4	1.6	4.3	0.12
Central U.S.				
Cleveland, OH (8)	0.9	17.1	6.2	8.8	0.17
Pittsburgh, PA (11)	0.81	17.9	8.2	11.8	0.16
Steubenville, OH-WV (5)	0.86	17.7	2.4	8.1	0.18
Detroit, MI (10)	0.89	16.7	6.4	9	0.17
Detroit, MI ** (9)	0.92	15.9	4.7	8.2	0.16
Grand Rapids, MI (4)	0.96	12.7	1.2	4.6	0.13
Milwaukee, WI (8)	0.9	13.7	1.3	4	0.12
Chicago, IL (11)	0.89	17.6	6.1	7.4	0.14
Gary, IN (4)	0.75	15.8	3.6	7.7	0.18
Louisville, KY (5)	0.9	17.1	2.7	5.1	0.12
St. Louis, MO-IL (11)	0.83	17.4	5.6	8.8	0.15
Baton Rouge, LA (3)	0.95	14.3	0.4	2.7	0.08
Kansas City, KS-MO (6)	0.91	12.6	2.4	4.2	0.13
Dallas, TX (7)	0.94	12.6	2.2	4.1	0.11
Western U.S.				
Boise, ID (4)	0.88	9.5	1.6	6.4	0.17
Salt Lake City, UT (6)	0.91	11.3	5.0	7.9	0.21
Seattle, WA (5)	0.62	8.9	6.1	10.8	0.3
Seattle, WA* (4)	0.85	10.4	3.0	6.1	0.17
Portland, OR (4)	0.83	7.7	2.8	5.3	0.19
Los Angeles, CA*** (5)	0.61	17.9	13.9	21.4	0.28
Los Angeles, CA (6)	0.77	21	5.4	12.2	0.18
Riverside, CA (5)	0.89	27.5	5.0	12.3	0.17
San Diego, CA (4)	0.79	15.8	2.4	8.7	0.17

* outlying site removed.

** interior site removed.

*** Results from analysis including site in Lancaster, CA (included in L.A. MSA, but located across mountains to east of downtown LA)..

individual monitors on a given day in a particular MSA, when large concentration gradients exist across that MSA, day-to-day differences in calculated area-wide 24-h average PM levels may not accurately reflect the day-to-day changes that would be obtained by the full set of monitors. On page 8-283 the CD states, "The above results provide clear evidence that fine particle concentrations may be less homogenous in at least some MSAs than has been previously assumed". Although not all major metropolitan areas were included in the spatial variability analysis, there is a sufficient number of major population centers included in the analysis summarized in Chapter 3 to indicate that the assumption of spatial homogeneity across all major metropolitan cities is inaccurate.

The CD notes on page 9-36 that new exposure studies indicate that fine particle measurements at central monitoring sites are good indicators of personal exposures to $PM_{2.5}$ in time-series studies. This claim rests largely on correlation coefficients calculated in a single study that was both geographically and temporally limited. What is really relevant is whether or not reductions in ambient monitor-site PM would produce predicted reductions in community mortality. Since different PM monitoring sites within the same community can lead to quite different PM effect estimates, the correlations with exposure cited in Chapter 9 appear somewhat beside the point. Similarly, Chapter 9 reports on page 9-36 that fine particle concentrations are well correlated between monitors in some cities, but the relevance of this observation is not clear. For example, Cook County is said to have well-correlated fine particle concentrations between monitor pairs but mortality effect estimates based on different monitors can be quite different.

In cities with multiple PM monitors, estimated unit effects of PM from time-series studies can vary widely (8-285), depending on which monitor or combination of monitors is used as the ambient PM measure; see Ito *et al*. (1995) and Roberts (2003). For example, of the twelve Chicago sites that monitored PM_{10} during 1987-1994, four sites showed consistent significant positive association with same day mortality, while the other eight sites showed negligible and non-significant mortality associations.

Weak cross-correlations of ambient PM measurements between monitors in the same city suggest that a PM-effect analysis, based on a composite ambient time series for that city, is likely to be misleading. In particular, weak cross-correlations, as seen for example in Salt Lake City, imply that human exposures to ambient PM will likewise be weakly correlated with the composite ambient PM measurement. Thus, reported health effects in cities with weak correlations between monitors should be viewed skeptically.

EPA correctly points out that high cross-correlations among monitor PM time series are a poor indicator of spatial homogeneity (8-280 bottom). In particular, even when correlations are high, there are commonly large differences in average levels or percentile values of PM among monitors as seen in Table 8-40. Correlation should not, therefore, be used as an indicator of spatial homogeneity.

If cross-correlations among monitors are indeed high, and average population exposure is indeed approximately proportional to the monitored PM values, and the exposure-response relation is indeed linear, then PM effect estimates should be about the same using any standardized combination of monitors to represent exposure. The fact that this has been empirically contradicted suggests that exposure-response may not be linear, that monitored PM poorly represents population exposure, and that effects of PM

reductions would be hard to anticipate. On page 9-36, the CD states "…there may still be

substantial variability in $PM_{2.5}$ concentrations across some urban areas, as discussed in

Chapter 3.

2.2   Inconsistencies Presented in Chapter 9 of the CD

In Chapter 9, on page 9-14, the authors state,

Much new information on the distribution of $PM_{2.5}$ and $PM_{10-2.5}$
concentrations across cities is available from the new monitoring networks
and is presented in detail in Chapter 3. In general, $PM_{2.5}$ is more evenly
distributed than $PM_{10-2.5}$ in terms of both daily/seasonal/yearly averages
and day-to-day correlations, although there are significant differences
among cities.

Using "correlation" as a measure of spatial variability is inconsistent with the

discussion in Chapter 3. Correlation is not a measure of spatial variability. As indicated in

Chapter 3 and parts of Chapter 8, the 90th percentile difference in concentrations (P_{90}) and

the coefficient of divergence (COD) were used to provide an indication of the degree of

spatial uniformity in $PM_{2.5}$ concentrations across urban areas.

On page 9-20, the CD states,

For example, new data reinforce our earlier understanding that ambient
concentrations of fine particles (measured as $PM_{2.5}$) are typically more
highly correlated and/or are more uniform across community monitors
within an urban area than are coarse particles (measured as $PM_{10-2.5}$),
although in some areas the differences are much less pronounced than in
others. More limited data and knowledge of the behavior of ultrafine
particles suggest that spatial distributions of their concentrations (which
decrease quickly from peak levels around major highways) are likely more
similar to those for coarse particles (which decrease quickly from peak
levels around primary sources) than for other (accumulation-mode) fine
particles. Further, new studies reinforce our earlier understanding that fine
particles generally infiltrate indoors much better than do either coarse or
ultrafine particles. **Thus, central site ambient concentration
measurements are a better surrogate for population exposure to**

accumulation-mode fine particles, measured as PM$_{2.5}$, than for either coarse or ultrafine particles, although there may be large differences in PM$_{2.5}$ concentrations across many urban areas (emphasis added).

What is really relevant is whether or not reductions in ambient monitor-site PM would produce predicted reductions in community mortality. Because different PM monitoring sites within the same community can lead to quite different PM effect estimates, the correlations with exposure cited by the CD seem somewhat beside the point. New exposure studies do not indicate that fine particles measurements at central monitoring sites are good indicators of personal exposures of PM$_{2.5}$ in time-series studies. As indicated earlier, if outdoor PM constituent concentration profiles are either spatially or temporally variable, it is likely that exposure misclassification errors could be introduced in the analysis of PM air pollution and health data. Thus, the use of central monitoring site data obtained by averaging spatial heterogeneous monitoring sites may introduce exposure misclassification errors that contribute to the greater heterogeneity of risk estimates derived from different locations in studies of both mortality and morbidity effects (see discussion on page 9-36).

On page 9-37, the CD notes,

Variability in PM concentrations across study areas could influence epidemiologic study results. For larger metropolitan areas, including monitors in outlying areas may bias the exposure estimate and reduce the correlation between the averaged concentration and the true population exposure. From among those U.S. cities in which epidemiological studies have been conducted, areas with more uniformity in PM$_{2.5}$ concentrations include Chicago and Detroit, whereas areas with more spatial variability include Seattle and Los Angeles. There are a number of factors that could influence spatial variability of PM concentrations, including topography, location of major PM sources, and weather patterns. Greater spatial variability in PM levels would be expected to increase exposure error, potentially affecting epidemiologic study results in those areas.

Upon reviewing Table 8-41 on page 8-282, one finds that both Chicago and Detroit are ranked "Intermediate' (applying the P_{90} ranking) in spatial variability and not ranked "Relatively Homogeneous". Seattle is ranked either as "Intermediate" or "Heterogeneous", depending upon whether an outlying site is removed. Los Angeles is ranked either as "Heterogeneous" or "Very Heterogeneous," depending upon whether an outlying site is removed. The point here is that neither Chicago, Detroit, Seattle, nor Los Angeles was ranked in Table 8-41 as "Relatively Homogeneous." Thus, all four sites experienced various degrees of spatial heterogeneity. Thus, as indicated by the authors of Chapter 9 on page 9-37, greater spatial variability in PM concentrations would be expected to increase exposure error and thus potentially affect epidemiologic study results in those areas.

As noted in Chapter 3, spatial variability differences may not be strictly related to the distance between monitors, especially where topography and sources of primary PM play a role. One key conclusion reached in the discussion of spatial variability in Chapter 3 is that whereas high correlations of $PM_{2.5}$ provide an indication of the spatial uniformity in temporal changes in directions (i.e., concentrations at both sites increase or decrease together) across urban areas, the correlations do not imply uniformity in the $PM_{2.5}$ concentrations themselves. *An important observation that has a direct impact on measurement error in the epidemiology analyses is that in many cases, as described in Chapter 3, correlation coefficients ranged from high to low values among monitor pairs within the various MSAs.* Based on this observation and the data provided in Appendix 3A, it can be concluded that the results in Chapter 3 provided clear evidence that fine particle concentrations are less homogeneous than had been previously assumed.

2.3   Ramifications of Spatial Variability

Only two (New York, Oakland) statistically significant positive associations between PM and mortality were found in the 90-city NMMAPS study and a statistically significant negative association was found in one (Little Rock). Thirty-seven percent of the 90 cities resulted in negative estimates of effect in city-specific analyses. In cities with multiple PM monitors, estimated unit effects of PM from time-series studies can vary widely [8-285], depending on which monitor or combination of monitors is used as the ambient PM measure; see Ito *et al.* (1995) and Roberts (2003).

The representativeness of the monitoring sites for population exposure of both the particle metrics and gaseous pollutants may be less than optimum. In reports submitted to the U.S. EPA that commented on epidemiology analyses, Dr. Paul Switzer (2003, 2004a, 2004b, 2004c) notes that spatial variability is an important consideration that tends to complicate many of the assumptions made in the various epidemiology analyses. As pointed out in Chapter 3, considerable spatial variability exists at many MSAs as measured by differences in absolute concentrations as well as moderate to low correlation coefficients.

It is difficult to extrapolate spatial variability observations from one city to another without first examining the data. As described in Chapter 3, a range of correlations of $PM_{2.5}$ concentrations was found between monitoring sites in the cities selected for the spatial variability analysis. Although some sites may be highly correlated with each other within an MSA, some monitors within a specific MSA may not be highly correlated with each other. Thus, within a specific MSA, concentration fields may not be

uniform. Considerable spatial variability exists at many MSAs as measured both by differences in absolute concentrations as well as moderate to low correlation coefficients.

Examples show that different PM monitors within the same city can provide quite different estimates of PM effects in time-series studies (e.g., Ito *et al*., 1995). Ito *et al*. (1995) show that it is not necessarily the case that straightforward monitor averaging provides more precise estimates of PM effects than other monitor combinations or even single monitors. As indicated in Ito *et al*. (1995) and the comments on the CD provided by Dr. Switzer (2003, 2004a, 2004b, 2004c), it is incorrect to assume, as has been done in parts of the CD, that site-to-site correlation is high within all cities. As emphasized by Dr. Switzer in his comments, although the latest revised draft of Chapter 8 of the CD critically describes a number of issues, the conclusions in Chapters 8 and the latest revision of Chapter 9 appear to reflect an unwarranted leap over many of the criticisms that it, itself, has raised. In addition, there were several issues not adequately addressed in the revised Chapter 8 of the CD and not adequately summarized in the revised Chapter 9. The CD is sometimes completely uncritical of the literature that it cites in support of its conclusions. The variability in site-to-site correlation within a geographic area can affect relative risk estimates. Data from the $PM_{2.5}$ monitoring network in 1999, 2000, and 2001 indicate that relatively high site-to-site correlations are not necessarily exhibited in most cities. As noted in Chapter 3 and stressed in parts of the revised Chapter 8, the correlation between sites is highly variable. As noted by Dr. Switzer (2003, 2004a, 2004b, 2004c) and mentioned earlier, weak cross-correlations of ambient PM measurements between monitors in the same city suggest that a PM-effect analysis, based on a composite ambient time series for that city, is likely to be misleading.

3. **The Importance of the Assumption of Linearity between Exposure and Response**

As discussed by Dr. Switzer (2004c), a linear (proportional) concentration-response relationship is key to many of the inferences and conclusions that Chapter 9 draws from the studies that it has reviewed, although only a single page in Chapter 9 is devoted to this central topic. Concentration-response linearity is central to many of the inferences and conclusions concerning PM effects drawn in Chapter 9, including discussions of exposure measurement error, confounding investigations, heterogeneity investigations, and fundamental modeling of time trends and weather. So central is linearity hypothesis that it is incorporated *a priori* in many of the studies that Chapter 9 has reviewed.

Below is a succinct discussion about the concentration-response (C-R) functions used in epidemiological studies as described by Abt Associates Inc. (2003) in their report to the U.S. EPA, Office of Air Quality Planning and Standards. The concentration-response (C-R) functions used in the risk assessment are empirically estimated relationships between average ambient concentrations of PM and the health endpoints of interest (e.g., short- and long-term exposure mortality or hospital admissions) reported by epidemiological studies for specific locales. Most of the epidemiological studies use a method referred to as "Poisson regression" to estimate exponential (or log-linear) C-R functions in which the natural logarithm of the health endpoint is a linear function of PM. The Poisson regression is essentially a linear regression of the natural logarithm of the dependent variable on the independent variable. The form of the exposure-response relationship is:

$$y = B\,e^{\beta x}, \qquad (1)$$

where x is the ambient PM level, y is the incidence of the health endpoint of interest of

PM level x, β is the coefficient of ambient x concentration, and B is the incidence at x=0,

i.e., when there is no ambient PM. The relationship between a fixed ambient PM level,

x_o, for example, and the incidence of a given health endpoint associated with that level

(denoted as y_o) is then

$$y_0 = B\,e^{\beta x_o}, \qquad (2)$$

C-R functions use daily average PM levels as input that relate these levels to the

daily incidence of the health endpoint. The difference in health effects incidence, $\Delta y = y$

$- y_0$, from y to the baseline incidence rate, y_0, corresponding to a given difference in

ambient PM levels, $\Delta x = x - x_0$, can be derived by subtracting equation (2) from equation

(1), which yields:

$$\Delta y = y_o[e^{\beta \Delta x} - 1]. \qquad (3)$$

Alternatively, the difference in health effects incidence can be calculated

indirectly using relative risk. Relative risk (RR) is a measure commonly used by

epidemiologists to characterize the comparative health effects associated with a particular

air quality comparison. The risk of mortality at ambient PM level x_0 relative to the risk of

mortality at ambient PM level x, for example, may be characterized by the ratio of the

two mortality rates: the mortality rate among individuals when the ambient PM level is x_0

and the mortality rate among (otherwise identical) individuals when the ambient PM level

is x. This is the RR for mortality associated with the difference between the two ambient

PM levels, x_0 and x. Given a C-R function of the form shown in equation (1) and a

particular difference in ambient PM levels, Δx, the RR associated with the difference in

ambient PM, denoted as $RR_{\Delta x}$, is equal to $e^{\beta \Delta x}$. The difference in health effects incidence, Δy, corresponding to a given difference in ambient PM levels, Δx, can then be calculated based on this RR:

$$\Delta y = y_o(RR_{\Delta x} - 1). \qquad (4)$$

Equations (3) and (4) are simply alternative ways of expressing the relationship between a given difference in ambient PM levels, Δx, and the corresponding difference in health effects incidence, Δy. These equations are the key equations that combine air quality information, C-R information, and baseline health effects incidence information to estimate ambient PM health risk. Note that Equations (3) and (4) are essentially independent of the initial absolute value of the ambient PM concentration.

Based on the Abt Associates Inc. discussion above, it is clear that much of the linear (i.e., proportional) exposure-response relationship is key to many of the inferences and conclusions that Chapter 8 draws from the studies that it has reviewed. As discussed by Switzer (2003, 2004a, 2004b, 2004c), much of the work on the measurement error approach, exemplified by Dominici *et al*. (2000) and Zeger *et al*. (2000), is solidly tied to an assumption of *proportionality*, i.e., the health effect reduction that follows from a fixed decrease in ambient PM is assumed to be the same regardless of the current ambient PM level.

The CD discussed the issue of exposure-effect proportionality, vis-à-vis exposure thresholds, in several places. When non-proportional effects are allowed in the effect estimation model, the estimated ambient PM-effect relationship often departs from proportionality, as can be seen for many cities in multi-city studies, such as Daniels et al (2000), Dominici *et al*. (2002), and Moolgovkar (2003). In these studies, the response is

29

modeled as a low-order parametric spline function of ambient PM. Application of the

spline response model to different cities yielded a variety of response shapes, often with

inadequate precision, suggesting that there are statistical difficulties distinguishing

between linearity and non-linear spline models. Formal tests for response-function

linearity will have low statistical power against plausible non-linear alternatives, not

necessarily alternative models with zero-effect thresholds. Indeed, EPA concluded that

the available information simply does not allow for a clear choice (8-308).

However, notwithstanding the heavily weighted preference for linear response

functions, non-linearity will have important consequences for the estimation of PM health

effects that cannot be summarized by a single coefficient of proportionality. Therefore,

direct estimates of the concentration-response relation should replace *a priori*

assumptions of proportionality, and statistical tests for distinguishing linear and non-

linear response functions should have adequate power.

Instead, Chapter 9 appears to reverse the logic by putting the burden on

disproving the proportionality assumption without regard to statistical power (Switzer,

2004c). As an example of this reversed logic, Chapter 9 states that available studies do

not provide strong evidence of a clear concentration threshold for health effects (9-42). It

is not evident that the strict proportionality assumption that Chapter 9 takes as a null

hypothesis is even biologically plausible (i.e., the health effect reduction that follows

from a fixed decrease in ambient PM is assumed to be the same regardless of the current

PM level).

 Switzer (2003, 2004a, 2004b, 2004c) points out that the conclusions of the CD

rely strongly on questionable commonality and linearity of the PM-effect response

function. Many of the disparate separate city estimates of PM response functions, reported in the CD, appear more like non-proportional response functions, and those that are more or less proportional have varying proportionality constants indicative of different PM effects in different cities. Switzer (2003, 2004a, 2004b, 2004c) notes that in several multi-city studies, PM response functions were pooled across cities, as in Schwartz and Zanobetti (2000) and Daniels et al (2000), even though city-to-city differences among PM-effect response functions are not obviously in the range of sampling variability. Such pooling across cities could create a pooled response function that is roughly linear, as pointed out in the two studies cited in this paragraph. However, a pooled response function is not readily interpretable, and the benefits of ambient PM reductions in any particular city cannot be deduced from the pooled response function.

Such pooling of response functions across cities ignores monitoring/exposure heterogeneity among cities, as described in the previous section. Furthermore, a pooled PM-effect response function has no concrete interpretation in the presence of heterogeneities of various kinds. *Given the inter-city heterogeneity of PM response functions, a combined PM response function that applies to no city, nor to the group of cities treated as single data set, provides little insight for standard-setting purposes.* Unfortunately, the CD discounts the importance of studies that show response thresholds (e.g., Smith *et al*. 2000), in favor of pooled response functions that are difficult to interpret.

The assumed linear concentration-response relationship also has implications regarding the need for separate 24-hour and annual standards. The proportionality of concentration-response implies that a reduction in the annual average will reduce both

31

long-term and short-term effects in the same proportion. A separate 24-hour standard is superfluous in the linear concentration-response context. Advocating for a separate 24-hour standard must imply non-linearity in concentration-response.

4.   Conclusion

In reaching its most current conclusions concerning its assessment of epidemiologic evidence in its revised version of Chapter 9 of the PM Criteria Document, the EPA has considered information in relation to several criteria (see Page 9-21) noted at the outset of Section 9.2 of the CD: (1) the *strength* of reported associations, in terms of magnitude, statistical significance, and statistical power/precision of effects estimates; (2) the *robustness* of reported associations to the use of alternative model specifications, potential confounding by co-pollutants, and exposure misclassification related to measurement error; (3) the *consistency* or general concordance of findings in multiple studies of adequate power, and in different persons, places, circumstances and times; (4) temporality, in terms of lag periods between exposure and observed effects; (5) the nature of concentration-response relationships; and (6) information from natural experiments or intervention studies as to the extent to which reductions in PM-related air pollution have been observed to be associated with improvements in health measures.

The available epidemiologic information does not meet the criteria for a convincing case for a causative association, nor does this information provide a basis for anticipating the effects of PM reductions. As new information became available and new modeling approaches were tried, inconsistencies in PM effect estimates continued to accumulate. Although Chapter 9 does enumerate important caveats regarding the role of

modeling choices on PM effect estimates, many of its own caveats were put aside in drawing conclusions.

The inconsistencies and the sensitivity of PM effect estimates to modeling choices are recognizable. Statistical concerns, which include weather and time trend covariation, additivity assumptions, seasonal differences, regional grouping, spatial heterogeneity both between cities and within cities, time lag selection, and treatment of gaseous pollutant confounders have been discussed many times. That PM effect estimates are delicate is not surprising given that they are superimposed on much stronger effects due, for example, to concomitant weather variations.

Clearly, even though considerable discussion has taken place both in Chapter 8 and at CASAC deliberations, the authors of Chapter 9 have made the decision not to factor into its conclusions the large uncertainties associated with the use of the epidemiologic methodology and the resulting inconsistencies in PM effect estimates that continued to accumulate. Without a clear statement in Chapter 9 that indicates the ramifications of these serious deficiencies associated with the effects estimates, a selective set of data within the PM Criteria Document will be cited in the EPA's Staff Paper and used in risk assessments without acknowledgement of the large uncertainties associated with the inputted data. Statistical uncertainty and risk analyses performed by the Staff Paper authors will not include considerations associated with the highly uncertain risk estimates, which actually include values that are positive, negative, and no effect estimates. Pooled estimates will be used that will blur the actual uncertainty associated with the estimates. *Without a clear understanding of the reasons for*

inconsistent effects estimates, one cannot rule out the possibility that the PM effect

estimates are model artifacts.

       An alternative to the cause-and-effect explanation provided in Chapter 9 is that the results we are citing may be mostly associated with these modelling artifacts. One is left with answering a serious question: If the time-series data are the most important information available for establishing Federal PM standards, are the data good enough to use in the decision-making process? Based on the evidence presented in Chapters 8 and 9 in the CD, one simply cannot draw comfortable conclusions regarding the circumstances and magnitudes of ambient PM health effects, or whether reported PM health effects are causative. Thus, one might conclude that there is still much uncertainty remaining in the epidemiological time-series results and many of the concerns expressed by the EPA in the Carbon Monoxide CD about the strengths and limitations of the extensive body of epidemiologic evidence of associations between health effects and air pollutants have not been adequately addressed in the most current version of the PM CD. The growing pattern of inconsistent and inconclusive findings is troublesome and presents both scientists and policymakers with a very difficult decision. Simply stated, the science based on epidemiological results is not substantial enough at the moment to provide the foundation upon which a clear path can be built that leads directly from the science to the policymaking decision arena.

## 5.	References

Abt Associates Inc. (2003). Particulate Matter Health Risk Assessment for Selected Urban Areas: Draft Report. Prepared for the Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Contract No. 68-D-03-002.

Daniels, M.; Dominici, F.; Samet, J. M.; Zeger, S. L. (2000) Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am. J. Epidemiol. 152: 397-406.

Dominici, F.; Zeger, S. L.; Samet, J. (2000) A measurement error model for time-series studies of air pollution and mortality. Biostatistics 1: 157-175.

Dominici, F.; Daniels, M.; Zeger, S.L.; Samet, J.M. (2002) Air pollution and mortality: estimating regional and national dose-response relationships. J. Am. Statist. Assn. 97: 100-111.

Health Effects Institute (2003) Revised analyses of the National Morbidity, Mortality, and Air Pollution Study (NMMAPS). Special report. Boston, MA.

Ito, K.; Kinney, P.L., Thurston, G.D. (1995) Variations in PM-10 concentrations within two metropolitan areas and their implications for health effects analyses. *Inhal. Toxic.* 7: 735-745.

Moolgovkar, S. H. (2003) Air pollution and daily mortality in two U.S. counties: season-specific analyses and exposure-response relationships. *Inhalation Toxicology*. 15:877:907.

Roberts, S. (2003) Combining data from multiple monitors in air pollution mortality time series studies. *Atmospheric Environment* 37: 3317-3322.

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.

Smith, R. L.; Davis, J.M.; Sacks, J.; Speckman, P.; Styer, P. (2000) Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama. Environmetrics 11: 719-743.

Switzer, P. (2003) A Review of Statistical Methods Used in Time-Series Epidemiologic Studies of Ambient Particulate Matter and Acute Health Effects Cited by the June 2003 EPA Draft PM Criteria Document. Report submitted to the U.S. EPA by the American Petroleum Institute.

Switzer, P. (2004a) A Review of Statistical Methods Used in Time-Series Epidemiologic Studies of Ambient Particulate Matter and Acute Health Effects Cited by the

December 2003 Revised Chapter 8 in the EPA Draft PM Criteria Document. Report submitted to the U.S. EPA by the American Petroleum Institute.

Switzer, P. (2004b) A Review of Statistical Methods Used in Epidemiologic Studies of Ambient Particulate Matter and Mortality Cited by the June 2004 Revised Chapter 8 and Chapter 9 in the EPA Draft PM Criteria Document. Report submitted to the U.S. EPA by the American Petroleum Institute.

Switzer, P. (2004c) The EPA Assessment of Epidemiologic Studies of Ambient Particulate Matter [August 2004 Draft of Chapter 9 of the PM Criteria Document] An Assessment of the EPA Assessment. Report submitted to the U.S. EPA by the American Petroleum Institute.

U.S. Environmental Protection Agency (2000) Air Quality Criteria for Carbon Monoxide. Office of Research and Development, Washington, D.C. EPA 600/P-99/001F. June 2000.

Zeger, S. L.; Thomas, D.; Dominici, F.; Samet, J. M.; Schwartz, J.; Dockery, D.; Cohen, A. (2000) Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ. Health Perspect*. 108: 419-426.