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

Paul Switzer, Ph.D.
Department of Statistics
Stanford University
September 17, 2004

The August 2004 draft of Chapter 9 of the EPA Particulate Matter Criteria Document [EPA9] offers an assessment of the epidemiologic evidence for PM health effects. Chapter 8 of the Particulate Matter [PM] Criteria Document, summarizing epidemiologic studies of PM and health effects, was presented by EPA in June 2004. The EPA9 assessment draws primarily from the Chapter 8 discussion [EPA8] and presents an evaluative summary of a diverse array of findings from a large number of studies. The majority of these epidemiologic studies examine concurrent temporal variations of PM and mortality or morbidity (mostly PM_{10}), together with a small number of cohort studies that use geographic variation rather than temporal variation. The epidemiologic assessment is presented in Section 9.2.2 of EPA9, which is organized into six parts:

1. Strength of reported associations
2. Robustness of reported association to alternative model specifications
3. Consistency of findings in multiple studies
4. Lags between exposure and effects
5. The nature of concentration-response relationships
6. Information from so-called natural experiments

My comments are organized to respond to each of the EPA9 epidemiologic assessment topics. Page references in this assessment refer to EPA9, with occasional reference to EPA8. I excluded the final speculative topic dealing with natural experiments because these experiments are uncontrolled and unadjusted observational studies, subject to strong publication bias.

1. Strength of Reported Associations

Estimates of PM mortality effects derived from many time-series studies vary from city to city, are often non-significant, and even negative. Estimates rarely exceed a fraction of a percent of daily mortality for the range of controllable PM concentrations.

EPA9 appropriately emphasizes the importance of multi-city studies, which it notes do not suffer from potential omission of negative findings [9-23, 9-27 bottom] and which use a common modeling framework. However, the multi-city studies are then grouped with other studies in Figure 9-4, where a 90-city study appears as a single point, for

example. EPA9 seems to lean heavily on Figure 9-4 in its conclusion that "almost all of the associations between PM_{10} and total mortality are positive and over half are statistically significant, including most all of those with more precise estimates". However, it does concede that the effect estimates are small [9-30].

There is some serious obfuscation in this conclusion drawn from Figure 9-4. In the first place, the 90-city study should not have been represented as a single point since the combined PM effect estimate does not refer to any specific location and there are unresolved issues regarding the heterogeneity of effect estimates among cities. The apparent precision of the combined effect estimate pertains to an artifactual model parameter rather than to a geographic location. EPA9 does itself point out that inter-city differences of PM health effect estimates are to be expected because of differences in demographics, climate and PM composition [9-36]. If effect estimates were plotted separately for each of the cities in multi-city studies, then one would see inconsistency in place of the apparent consistency of Figure 9-4, and a smaller fraction of positive and statistically significant findings. Indeed, the exercise of combining effect estimates from the largest multi-city study produces a PM effect that is typically a fraction of PM effect estimates in the studies selected for inclusion in Figure 9-4.

In any event, it is disingenuous to speak of the strength of reported associations in isolation of model selection issues and exposure errors, which are separately discussed in later sections. Nevertheless, the EPA9 summary conclusion [9-30] is that the "epidemiologic evidence is *strong* for associations between PM_{10} and $PM_{2.5}$ and mortality" based on a presentation of data that conceals variation and does not anticipate the later discussion of model selection and sensitivity.

It is important to understand the sources of inter-city differences among PM effects. Without a clear understanding, we cannot rule out the possibility that PM effect estimates are model artifacts. There has been a determined but incomplete effort to relate inter-city effect differences to characteristic differences among cities such as demographics, climate, etc. This is called "effect modification" and is a potentially useful approach. However, Samet *et al.* (2000) could not identify any statistically significant PM effect modifiers among those that they examined in their 90-city study. Disparities among PM effect estimates could also arise because of model inadequacies, for example through incorrect treatment of confounding variables or an incorrect characterization of the concentration-response relationship.

2. Robustness of Epidemiologic Associations

Because the PM effects are so small compared with the range of daily Poisson mortality variation, great care must be taken to assure that estimates are not sensitive to modeling choices. EPA9 begins its assessment of robustness of PM health effect estimates by stating "Many epidemiologic studies have also included assessment of whether the associations were robust to such factors as model specification" [9-32]. It would have been informative for EPA9 to list which of the cited studies (in Figure 9-4 and Figure 9-5

for example) did indeed investigate model robustness, and which issues of model robustness were examined and *not* examined. The use of the term "many" may be inappropriate.

Indeed, the EPA9 assessment devotes but a single page to the critical issue of model robustness and the role of weather and time trend specification in PM effects modeling. Earlier statements by EPA that weather confounding had been adequately addressed proved subsequently to be incorrect. Even those studies that did examine robustness of PM effects to weather confounding have done so in a limited way, which, for example, does not include weather-PM interactions.

PM variations are substantially correlated with weather variations. Therefore, special care is needed in separating PM effects from the much larger effects of weather. EPA9 cites HEI reanalysis studies [HEI 2003] that point to the sensitivity of PM effect estimates to the modeling of weather and time trends [9-33]. The HEI re-analysis demonstrated that greater flexibility in modeling these confounders could substantially reduce the apparent PM effect estimate and alter study conclusions, but there is no agreement regarding how much detail should be incorporated into confounder adjustments. So it is astonishing that the EPA9 assessment claims that PM-mortality associations are robust to model specification. The recommended further investigation of weather effects [9-34] could even show that remaining PM effects are substituting for heretofore unmodeled weather effects.

While EPA9 does acknowledge PM effect sensitivity to alternative modeling of weather, it does not adequately address a critical modeling assumption -- *additivity* of weather and PM effects, an assumption that is built into all the PM effect estimates cited by EPA9. The additivity assumption is very strong and it presumes that the incremental effects of PM would be the same at any temperature and humidity level. Studies that allow PM effects to differ between seasons are a step in the right direction, and such studies typically demonstrate that PM effects, corrected for additive weather effects, are nevertheless very different in different seasons; see Lumley and Sheppard (2000), Smith et al. (2000), and Moolgovkar 2003, for example. The issue of model robustness, with respect to weather adjustment, will not go away until the issue of non-additive effects is addressed.

Similar issues arise with regard to confounding of PM effects with those of other co-pollutants. EPA9 concludes that PM effect estimates are robust to the inclusion of co-pollutants [9-35], although this has not been a consistent finding. Furthermore, in all cases where multi-pollutant models have been introduced, the effects of all pollutants are taken to be purely additive. Models that exclude pollutant interactions may not be biologically plausible so it becomes difficult to interpret the results of additive multi-pollutant models. Among other shortcomings of the multi-pollutant is the common practice of forcing the same lag structure on all pollutants or not allowing for distributed lag effects for the co-pollutants of PM.

EPA9 also includes a curious discussion of PM exposure error in its section on model robustness [Section 9.2.2.2.3]. Presumably the point being made is that PM effect estimates are robust to different surrogate measures of PM exposure, such as ambient concentrations at monitoring stations. This has not been demonstrated. On the contrary, studies that have looked at varying combinations of monitors within the same city have found, for example, that PM_{10} effect estimates are not robust to the choice of monitoring sites; see Ito *et al.* (1995) and Roberts (2003). The argument for fine particles is made only obliquely in EPA9 by reference to a few studies that show reasonable correlation between monitoring site concentrations and personal exposure, and even these studies do not show a consistent picture; see Clayton *et al.* (1999). However, even strong correlations are not evidence for robustness of PM effect estimates in situations where there is not a proportional relation between exposure and response, as will be discussed later.

3. Consistency of Epidemiologic Findings

Section 9.2.2.3 of EPA9 is an assessment of consistency of PM effect estimates across studies. This section clearly outlines why one might not expect consistency among effect estimates, for example due to differences in pollutant composition and differences in population characteristics among cities. So in the final analysis we are left with *inconsistent* findings, with conjectures to explain the inconsistencies. A reading of this section would leave the impression that consistency of PM effect estimates is *not* a property of the PM epidemiologic literature. The claim that the larger studies show greater consistency fails to recognize the multi-city nature of the larger studies whose effect estimates are composites of inconsistent city-specific effect estimates.

Formal statistical tests to detect overall heterogeneity of effect estimates among cities have provided conflicting results [9-39], but in any event are not informative because of the acknowledged low power of these tests. The lack of power was clearly pointed out in the HEI Special Panel review [8-228].

The evident inconsistency of PM effect estimates across cities is also clearly seen in the multi-city studies that used a common modeling framework. As noted earlier many city-specific estimates in the multi-city studies are not significant and even come up negative. See for example Figure 8-1, which is reproduced here. Inconsistencies are suggestive of model inadequacies, particularly in regard to confounders and effect modifiers. The fact that city-specific effect estimates have been combined into a statistically significant small overall effect estimate does not make the consistency issue go away.

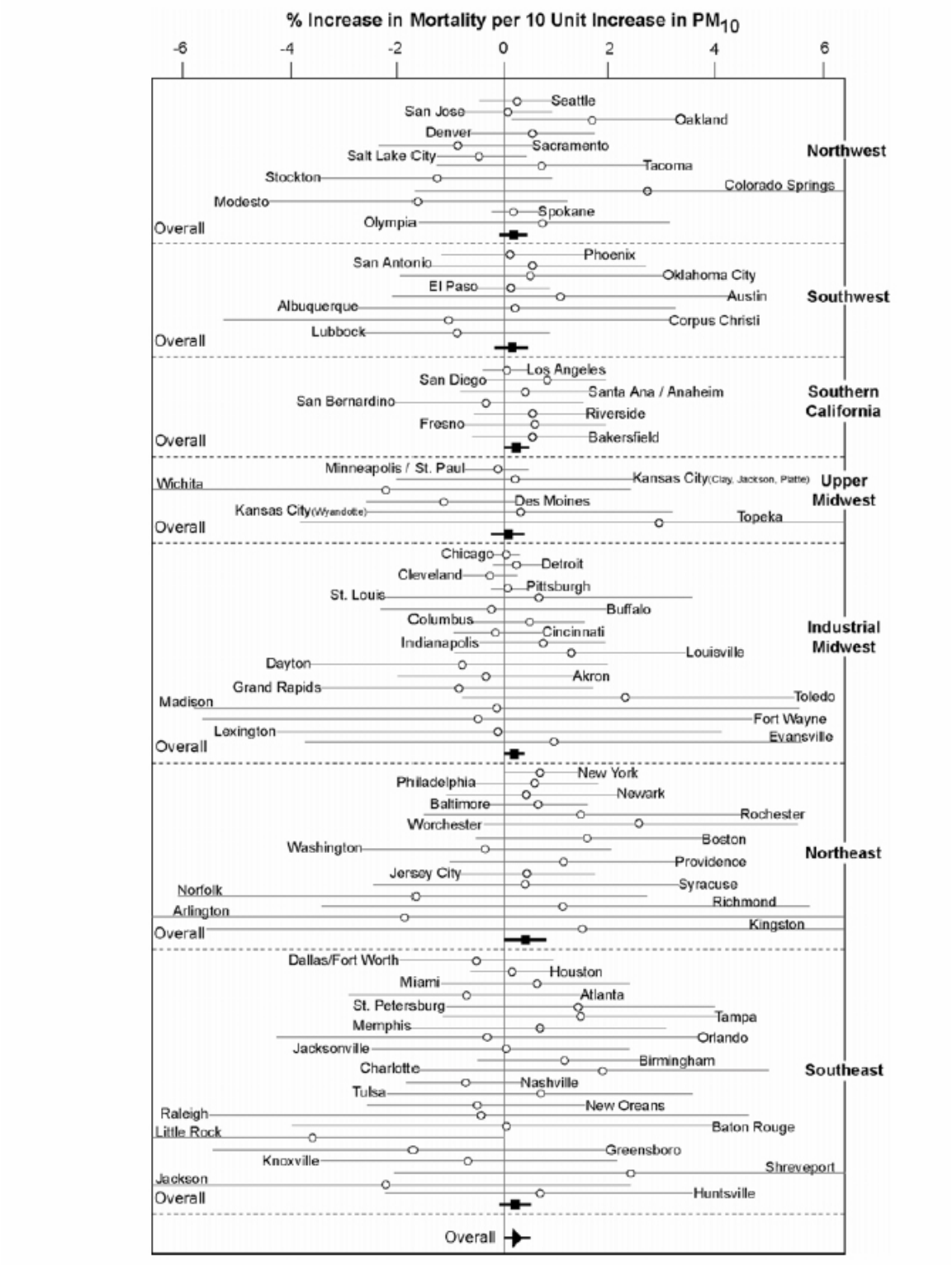

Figure 8-1. Estimated excess risks for PM mortality (1 day lag) for the 88 largest U.S. cities as shown in the revised NMMAPS analysis.

There are further inconsistencies in PM effect estimates that surface in studies that examine cities with multiple monitors. Using ambient PM concentration data from different monitors within the same city generates differing estimates of PM health effects; see Ito *et al.* (1995) and Roberts (2003). For example, of the twelve Chicago sites that monitored PM10 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.

EPA9 acknowledges that spatial variability of PM levels is an unresolved problem [9-38]. High temporal correlations between monitors within the same city do not imply that PM effect estimates will not be monitor-dependent. For example, if one monitor consistently records levels twice as high as another monitor, then the effect estimate using data from the first monitor will be half that for the second monitor, since both are being related to the same time series of health or mortality data. Even when PM time series are rescaled to a common standard deviation for all monitors, it is still the case that PM mortality effect estimates can vary substantially depending on which monitor or monitors are used to represent ambient PM; see Roberts (2003).

The patent inconsistencies across cities and studies raise an important regulatory question: what reduction in health effects could be expected from a specific regulatory standard? For example, based on results from the multi-city studies, it is reasonable to suppose that a reduction of ambient PM will produce no health benefit in some cities. Given the inter-city heterogeneity of PM response functions, a combined PM effect estimate that applies to no city provides little insight for standard-setting purposes.

4. Temporality and the Question of Lagged PM Effects

EPA9 discusses the issue of lag selection for time-series studies where the putative effect of a high-PM day could be delayed or spread over a number of subsequent days. EPA9 recognizes that maximized lagged effect estimates may be biased [9-39] and appropriately suggests that effect estimates should be derived for a series of lags. Simulation studies by Lumley and Sheppard (2000) have shown that lag selection bias can be of the same order as the estimated PM effect itself. However, EPA9 did not report which of the time series studies that were assessed may indeed be subject to lag selection bias.

For example, a common 1-day lag was chosen for all cities in the 90-city study (Dominici *et al.* 2002) so as to mitigate a strong model selection bias that would arise if the choice of lag was optimized separately for each city. However, there is still bias present because the 1-day lag was selected for the very reason that it gave PM effect estimates that were overall twice as large as the other lags considered [8-261], and it was the only lag choice with clear overall statistical significance based on the HEI reanalysis.

If PM effects are conjectured to be distributed over several days, then EPA9 claims that lag selection for greatest apparent effect would underestimate (rather than overestimate) the total multi-day effect. The circumstances under which this claim is true have not been explored and intuition may be tricky; temporal autocorrelation of ambient PM introduces issues that are difficult to intuit without systematic exploration.

In distributed-lag models, PM effects extend over several days and separate coefficients can be estimated for all lags included in the model, typically 5 to 30 days. This approach has some attractive possibilities and can potentially extract more information regarding short-term PM effects. As a salutary exercise, one should also include non-causal *negative* lags as a check on the credibility of the distributed lag model. However, "negative lag" checks are not discussed by EPA9 and appear to be absent from the cited distributed-lag model studies.

Naïve application of distributed lag models will have another serious shortcoming: if one is to allow PM effects to extend over several days then one should also allow effects of *confounding* variables, such as weather and co-pollutants, to extend over several days. Failure to allow for distributed-lags in confounding variables can lead to an exaggeration of the PM effects summed over lags. None of the cited distributed lag studies allowed for distributed lags for weather effects or co-pollutant effects.

Finally, EPA9 raises the issue of how to incorporate long-term trends in PM exposure in the cohort studies that span decades. It correctly concludes that further study is needed to evaluate the relation between health outcomes and long-term PM exposure where these exposures are changing over time [9-40]. But EPA9 does not say whether we should trust the findings of the long-term studies that did not adequately consider this issue.

5. The Relation Between Levels of Ambient PM Concentration and Health Effect Response Levels

A linear [proportional] concentration-response relationship is key to many of the inferences and conclusions that EPA9 draws from the studies that it has reviewed, although only a single page in EPA9 is devoted to this central topic. Concentration-response linearity is central to many of EPA9's inferences and conclusions concerning PM effects, including discussions of exposure measurement error, confounding investigations, heterogeneity investigations, and fundamental modeling of time trends and weather. Indeed, so central is the linearity hypothesis that it is incorporated *a priori* in many of the studies that EPA9 has reviewed.

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, EPA9 seems to reverse the logic by putting the burden on *disproving* the proportionality assumption without regard to statistical power. As an example of this reversed logic, EPA9 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 EPA9 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.

Furthermore, the proportionality assumption provides no basis for an acceptable regulatory threshold. For example, one could always double the health effect improvement by doubling the PM reduction, so there is no obvious regulatory threshold based on health effects under this assumption. Furthermore, the same reduction of effects could then be achieved by comparable absolute PM reductions in either a high-PM or a low-PM city.

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 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.

When non-proportional effects are allowed in the effect estimation model, then the estimated ambient PM-effect relation often departs from proportionality, as can be seen for many cities in multi-city studies, such as Daniels *et al* (2000) and Dominici *et al.* (2002), and in the Los Angeles – Chicago study by Moolgovkar (2003). In these studies, the response is 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, EPA8 concluded that the available information simply does not allow for a clear choice [8-308].

In the multi-city studies described in EPA9, PM concentration-response functions for different cities were pooled across cities, as in Schwartz, 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. However, a pooled response function, even if it appears linear, is not interpretable unless the same concentration-response relationship applies to every city. The putative benefits of ambient PM reductions in any particular city cannot be deduced from the pooled response function. Nevertheless, the EPA9 conclusion that accepts monolithic linearity of concentration-response draws principally from such pooled response functions [9-42].

Exposure measurement error will tend to flatten a non-linear concentration-response curve [Cakmak et al. 1999] making it harder to distinguish between linear and non-linear associations. Although their simulation study reports that specific threshold concentration-response models for a population could be distinguished even in the presence of exposure measurement error, we do not know to what extent their specific simulation findings can be generalized. More studies of this kind are needed that examine the statistical power to distinguish among competing concentration-response models.

EPA9 makes a reasonable point regarding the difficulty of interpreting "population" threshold models because it is reasonable to suppose that the response threshold would vary across individuals and even across time [9-41]. Better insights into the relationship between monitored ambient PM concentrations and anticipated community-level PM health effects could be obtained by modeling the relationship between monitored PM and *individual* PM exposure such as Dominici *et al.* (2000). However, individual-level *exposure* modeling should go hand-in-hand with the individual-level modeling of *response* to PM in order to build a model for community-level response to ambient PM. One can readily construct examples to illustrate that ignoring individual heterogeneity of concentration-response to PM can lead to misleading community-level concentration-response functions. The relation between individual level response and community level response remains unexplored.

Finally, if PM mortality effects were largely confined to a frail population with short longevity, referred to as 'mortality displacement', then there would be substantial public policy and regulatory implications, and this is recognized in EPA8 [8-316]. For example, it might be more effective to mitigate PM exposure of frail individuals through nursing home and hospital indoor PM requirements, as opposed to regulation of ambient PM. However, the EPA9 assessment is mute on this important topic, possibly because studies that examine mortality displacement are not in agreement.

6. Conclusions

EPA9 presented its assessment of the epidemiologic literature relating monitored PM concentrations to health effects using criteria that correspond to the section headings in this review. As detailed above, 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 becomes available and new modeling approaches are tried, inconsistencies in PM effect estimates continue to accumulate. EPA9 does enumerate important caveats regarding the role of modeling choices on PM effect estimates. However, many of its own caveats are put aside in drawing conclusions. Without a clear understanding of the reasons for inconsistent effects estimates, we cannot rule out the possibility that our PM effect estimates are model artifacts. Below I briefly summarize a few of the points made in this review. The body of the review should be consulted for details and a fuller discussion.

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 by EPA9 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 EPA9.

REFERENCES

Cakmak, S.; Burnett, R.T.; Krewski, D. (1999) Methods for detecting and estimating population threshold concentrations for air pollution-related mortality with exposure measurement error. Risk Anal. 19: 487-496.

Clayton, C.A., Pellizzari, E.D., Rodes, C.E., Mason, R.E., & Piper, L.L. (1999) Estimating distributions of long-term particulate matter and manganese exposures for residents of Toronto, Canada. *Atmospheric Environment*, 33: 2515-2526.

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.

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

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.

Samet, J. M.; Dominici, F.; Zeger, S. L.; Schwartz, J.; Dockery, D. W (2000a) National morbidity, mortality, and air pollution study. Cambridge, MA: Health Effects Institute; research report no. 94.

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.

Zeger, S. L.; Dominici, F.; Samet, J. (1999) Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology* 10: 171-175.