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

Paul Switzer, Ph.D.
Department of Statistics
Stanford University
August 2003

The fourth draft of the Particulate Matter [PM] Criteria Document [DCD4] was issued by EPA in June 2003. DCD4 examines a number of statistical issues relevant to connections between ambient PM variations and corresponding variations in health and mortality indicators. The discussion centers on relating temporal variations through time series modeling, and on relating geographic variations using cohort studies. Although DCD4 comments critically on a number of issues, its conclusions seem to reflect an unwarranted leap over many of the criticisms that it, itself, has raised. In addition, there are several issues not adequately addressed in DCD4, and DCD4 is sometimes completely uncritical of the literature that it cites in support of its conclusions. In this review I address several statistical issues in DCD4 and indicate how they affect conclusions drawn about the health effects of PM. The issues addressed here are grouped into the following categories.

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

1. CONFOUNDING OF WEATHER AND PM EFFECTS

Because PM variations are correlated with variations in weather, special care is needed in separating PM effects from the much larger effects of weather. DCD4 cites HEI reanalysis studies that point to the sensitivity of PM effect estimates to the modeling of weather effects, see Health Effect Institute (2003). However, reasons for the PM-effect sensitivity to temperature modeling are not explored. Therefore, a distinct possibility remains that more careful modeling of PM and weather could show that the reported PM effects are substituting for some of the weather effects.

While DCD4 describes PM effect sensitivity to alternative modeling of weather, it does not at all address the key assumption that weather and PM effects are additive, an assumption that is built into all the PM effect estimates that are cited by DCD4. The additivity assumption is very strong and it presumes that the incremental effects of PM would be the same at any level of temperature and humidity. Thus, for example, the presumption is that incremental PM effects are the same at moderate temperature and humidity as they are at extreme temperature and humidity. If this assumption should fail, then additive modeling of PM effects, as relied on by DCD4, can lead to uninterpretable estimates of PM effects. This is especially true when the proportionality of PM effects, as a function of PM exposure, is in question. Without a systematic exploration of non-additivity, we cannot conclude that some part of the weather effects is mistakenly attributed to PM.

There are several strategies for incorporating non-additivity, i.e., allowing for differential PM effects at different levels of the confounding variables. One strategy is to allow for the estimation of a joint response surface that includes both PM and weather. In addition to providing better separation of effects, the joint response surface provides a fuller understanding of PM effects and a better guide to regulation and public policy. A joint response surface could be of the spline type or other parametric or semi-parametric form. Roberts [2003a] explored the mortality effects of PM and temperature, in combination, using a nonparametric response surface. His analyses were applied to time series data from Pittsburgh and Chicago and these analyses indicate that the additivity presumption may not be plausible.

A second strategy for incorporating non-additivity of PM effects is to stratify, effectively to get different PM effect estimates for different temperature strata, see Morris and Naumova (1998) who examined mortality and morbidity effects of CO for different temperature strata. A similar strategy has been used in studies that seek to separate seasonal effects from PM effects. For example, winter and summer PM effects could be separately estimated by including two PM effect parameters in the model. Where separate season-specific estimates of PM effects have been obtained, it is not uncommon for these effect estimates to be different; see Lumley and Sheppard (2000) and Smith (2000) for example. Roberts (2002), in a recent reanalysis of the Chicago mortality time-series data found positive PM mortality effects in summer and negative PM mortality effects in winter.

An advantage of the stratification approach to non-additivity is that it is simple to describe and apply in a uniform manner to data from multiple cities. This approach has been applied by Roberts (2003a) who obtained separate PM mortality effect estimates for three temperature strata using daily data from Chicago and Pittsburgh. His findings show that PM effects are not uniform across temperature strata. The warm temperature stratum showed a positive PM mortality effect, the cold temperature stratum showed a negative effect, and the moderate temperature stratum showed no effect.

Similar issues arise with regard to confounding of PM effects with those of other co-pollutants. Although DCD4 reports that the PM effect estimate is little changed when co-

pollutants are entered additively into the model for daily mortality, it goes on to say (sec 8.4.8.2) that multi-pollutant analyses could be misleading when the spatial heterogeneity of the co-pollutants is very different from that of PM. For example, for a co-pollutant that is spatially more heterogeneous than PM, the ambient concentration measurement error could be much larger than that for PM, which would affect the relative sizes of effect estimates. Also, co-pollutant studies have ignored PM and co-pollutant interactions, essentially relying on additivity of effects.

Finally, DCD4 (p. 8-212) suggests that a principal components analysis, in the context of multiple pollutant exposure, could provide useful new information, citing Mar *et al*. (2000). This point of view seems to derive from the fact that the new principal component variables are uncorrelated with one another, unlike the original pollutant variables. However, principal component coefficients (loadings) are themselves unstable if the pollutant variables have sizable cross-correlation. It would be inappropriate to rely on interpretations of these coefficients to draw conclusions about the relative importance of different pollutants.

2. CONFOUNDING OF TIME TRENDS AND PM EFFECTS

DCD4 relies heavily on multi-city time series analyses for its estimates of the health effect of PM, in particular the 90-city NMMAPS study, Samet et al. (2000), subsequently reanalyzed in HEI (2003). This study acknowledges the possibility of confounding of PM effects with slowly evolving time-varying influences on mortality, including seasonal effects, epidemics, trends in population characteristics, health care, etc. This is an inherent problem in time series studies of PM effects. To address the time confounding issue, the models used by NMMAPS incorporate an explicit time trend adjustment, represented by a richly parameterized function that is simultaneously estimated with the PM effects. Consequently, only short-term effects of PM are effectively estimated. [These short-term effects are the ones that are confounded by weather variations on the same time scales, as described above.]

However, the recent reanalysis of the 90-city study by HEI (2003) found that PM effect estimates are sensitive to the parameterization of the time trend and weather variables. DCD4 (p.8-197) notes this sensitivity as a possible cause for variability of PM effect estimates. As expected, a time trend with richer parameterization will capture shorter scale fluctuations in mortality and reduce the apparent PM effect. There is no obvious level at which to cut off time trend effects in favor of PM effects and this conundrum is not resolved. The lack of robustness of PM effect estimates remains an unresolved issue and may point to deeper confounding problems.

But the reanalysis cited above did not address the additivity assumption in which the PM effect is forced to be the same regardless of the level of the time trend adjustment. This is known to be an issue because studies of PM effects by season indicate that the PM effects can indeed be different in different seasons; see Smith (2000) for example. Thus

the issue of confounding of time trends and PM effects is largely unresolved although the reanalysis of the 90-city study provided important new information.

3. HETEROGENEITY OF PM EFFECTS AND EFFECT MODIFICATION

DCD4 (sec. 8.2.2.3) correctly emphasizes the importance of the 90-city study, citing Samet *et al*. (2000) and Dominici *et al* (2000a, 2002) where the same modeling strategy was used for all cities in the study. 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 could no longer be discerned, i.e. the formal test of the null hypothesis of homogeneity is not rejected at the conventional 5% level of statistical significance. However, as always, it is dangerous to draw conclusions from the failure to reject a null hypothesis without an adequate study of the power of the statistical test against specific heterogeneity alternative hypotheses. It is likely that the power of the homogeneity test is inadequate to detect inter-city effect differences that would prove important to an understanding of the data. Such a study of power could and should be undertaken.

While DCD4 (sec. 8.4.7.1) cautions that the issue of inter-city heterogeneity of PM effects is not completely resolved, it seems nevertheless to rely heavily on combined analyses for its conclusions. Combined multi-city analyses are cited not only in connection with PM effect estimates but also in connection with the DCD4 discussion of concentration-response modeling (sec. 8.4.6) and mortality displacement (sec. 8.4.9.1) both of which are discussed later in this review.

There are apparent differences among certain city groupings. The grouping of northeast US cities shows a combined effect twice as large as the overall grouping. While there are conjectured explanations such as differing chemical composition of PM, none of these conjectures has been examined with data. A test for PM-effect homogeneity between regional groupings has not been reported, although it seems plausible that the homogeneity hypothesis would indeed be rejected based on the grouping.

The approach used in NMMAPS both in their original analysis, Samet. et al. (2000) and in their reanalysis, HEI (2003) is to consider inter-city PM effect differences as genuine but unexplained *random* differences. Adopting this random effects approach implicitly introduces the notion of a population of cities for which the 90 study cities are treated like a random sample of cities. This approach also introduces the notion of a *population* mean PM effect. Under the adopted model, this population mean PM effect is reported as a single PM effect summary statistic, see DCD4 (Table 8-34) for example. However, the "population" mean is a model construct and its estimation, per se, should not be of great interest in the real regulatory world. Nevertheless, the random effects model *is* useful for refining the individual city-specific estimates of PM effects and their precision and it

does allow for the incorporation of a spatial model for regional-scale variation of PM effects.

This is not to say that PM effect estimates for different cities could not be combined in a somewhat more meaningful way. For example, one could obtain a population weighted combined PM effect estimate for the 90 cities together with an associated confidence interval. If, for the moment, we allow the correctness of a linear PM effect in every city, then this combined population weighted estimate can be interpreted in terms of the overall effect of a simultaneous fixed PM reduction in all 90 cities. One could also divide the pooled data into 5-year periods and examine the variability among combined PM effect estimates for each period. Such data splitting provides a simple check on robustness of the modeling exercise.

A third approach to differences among PM effect estimates is to relate such differences to characteristic differences among the 90 cities of the study. This is called "effect modification and is a potentially useful approach. DCD4 (sec. 8.1.3) carefully and wisely distinguishes between the issues of confounding and effect modification. Effect modification arises as an issue only when PM effects have been separately computed from different data sets using the same model and estimation procedure, as in the 90-city study. Effect modifiers are not time-series variables that could be confounded with time-varying PM. Rather, effect modifiers are exogenous variables whose values differ among cities.

DCD4 (p.8-33) discusses a variety of possible and sometimes plausible PM effect modifiers for the diversity of PM effect estimates. Some putative PM effect modifiers for which data are available are variable demographic characteristics, climate statistics, proximity to pollutant sources, or statistical summaries of pollutant concentrations akin to climate statistics, and the statistical precision of the effect estimates. Samet *et al*. (2000) could not identify PM effect modifiers among those that they examined. Other potential effect modifiers, such as differences in chemical composition of PM in different cities, have not been sufficiently examined, likely because relevant data are not readily available. While there are many proposed effect modifiers, few have been investigated, so we are still in the dark regarding inter-city differences.

The approach typically used to test putative effect modifiers treats the separately estimated PM effects for each of the 90 cities as dependent-variable data for a second-stage regression analysis as described in DCD4 (p. 8-33). A better approach, potentially with more statistical power, would be to simultaneously estimate PM effects at all cities using a parameterization of the PM effect or the baseline mortality rate that depends on the putative effect modifiers. An interesting possibility is to use the long-term PM *average* concentration as an effect modifier for the baseline mortality in the multi-city time-series studies, perhaps in conjunction with demographic descriptors. This modeling approach creates the possibility to simultaneously estimate both acute and long-term PM effects.

The fact that few if any convincing PM effect modifiers have been found to account for differences among PM effect estimates is disconcerting because the discrepant PM effect estimates remain unexplained. Unresolved discrepant PM effect estimates, derived from the same modeling approach, could also result from inadequacies in the modeling approach such as an incorrect treatment of confounding variables or an incorrect characterization of the exposure-response relationship. So failure to resolve the heterogeneity issue will cloud other analyses of the data that are based on multiple cities.

Finally, it is important to draw attention to regulatory implications of unresolved discrepancies among PM effect estimates for different cities. For example, in a number of cities the analysis suggests that PM health effects are absent. In such cities where PM effects are absent, would any health benefit be derived by reducing ambient PM?

4. HETEROGENEITY OF EXPOSURE

There is no disagreement that exposures to ambient PM will vary across the population on any given day for which a single ambient concentration is reported. DCD4 discusses certain issues associated with the relationship of ambient PM concentration and levels of human exposure to ambient PM which it terms "measurement error"; see DCD4 (sec.8.4.8), for example. There are two sources of population exposure variability for a given ambient concentration. One obvious source is the heterogeneity among individual microenvironmental trajectories, such as variations in time spent outdoors, variations in residential and workplace penetration and air exchange factors. A second source of exposure variation is the spatial heterogeneity of PM concentrations, which induces different exposures relative to the monitoring site(s) used to measure ambient PM.

Personal exposure studies such as the Toronto study by Clayton *et al*. (1999) show substantial heterogeneity of PM exposure among individuals within the same city. Likewise, multiple ambient PM monitoring sites within a single city show important spatial differences in ambient PM such as the Chicago and Los Angeles study by Ito *et al*. (1995) and the Chicago and Pittsburgh study by Roberts (2003b). In the typical time-series study for a city, such as Dominici *et al*. (2000a), the time-varying measure of exposure is obtained by averaging data from available ambient monitoring sites in that city. For ecological long-term PM effect studies, such as in Pope *et al.* (2002), the single exposure number for each city is obtained as both a time-averaged and spatially averaged concentration for all monitors in that city.

If the concentration-response relationship were exactly linear, and if the population average exposure to ambient PM was in constant proportion to the reported ambient PM, then it could be argued that the estimated effect per unit increase in *reported* PM is not affected by population variability in exposure. But it is important to distinguish between the unit effects of reported PM and the unit effects of PM exposure. The proportionality factor relating population exposures to reported PM is likely to be different in different cities. The estimated unit effects of PM would not then be comparable across cities

without an understanding of city-specific relationships between exposures and reported PM. Heterogeneity across cities in the relation between monitored PM concentrations and average population exposures will directly affect PM effect estimates. Combined PM effect estimates across cities, such as those reported by NMMAPS, implicitly and implausibly assume that the relation between monitored ambient PM and ambient PM exposure is the same across cities.

In cities with multiple PM monitors, estimated unit effects of PM from time-series studies can vary widely, depending on which monitor or combination of monitors is used as the ambient PM measure, see Ito *et al*. (1995) and Roberts (2003b). Roberts obtained these PM effect estimates using a Poisson regression adjustment model comparable to the model used in the 90-city study by Samet *et al*. (2000), but care was taken to use estimating procedures that are not affected by computational issues identified for the S+ implementation of GAM models. Both Ito *et al*. and Roberts 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.

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. Furthermore, when PM10 is averaged over all 12 sites, it is not as strongly associated with mortality as the four selected sites. It could be argued that, if one monitor records proportionally lower ambient PM than a second monitor, the first monitor will show a correspondingly larger unit PM effect because both monitors are used to explain the *same* community-wide time series of health effects. However, in the cited studies the disparity in effect estimates among different monitors is not closely associated with the magnitude of measured ambient PM. If the true PM exposure-response function were nonlinear, such as a response function with a threshold, then the relationship among effect estimates for different monitors could be quite complicated.

Another implication of within-city spatial variability goes to the issue of co-pollutant confounding. PM is reported to be spatially more homogeneous within a city than gaseous co-pollutants that are potential confounders, as seen for example in Ito *et al*. (1995). Then the city average for PM may be more representative of PM exposure than the city average for the gaseous pollutants, leading to the possibility that the PM effect estimate will carry effects properly attributable to gaseous pollutants, see DCD4 (sec.8.4.8.2).

Finally, there is the regulatory question that is posed by the heterogeneity of PM effect estimates. Averaging heterogeneous PM effect estimates does not make the heterogeneity go away. The regulatory question concerns the implied reduction in health effects that 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, even based on random-effects models. This could be due to differences in chemical composition of PM in such cities, or due to the

weak relation between ambient PM and PM exposure in such cities, or other specific attributes of such cities.

5. THE RELATION BETWEEN EXPOSURE AND RESPONSE

In the preceding section, I referred to implications for PM health effect estimates of the relation between ambient PM concentration and population PM exposure. But the implications for ambient PM health effect estimates cannot be fully understood without also considering the relation between PM exposure and health effects. Much of the work on the measurement error approach, exemplified by Dominici *et al*. (2000b) 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 PM level. Furthermore, one could double the health effect improvement by doubling the PM reduction, so there is no obvious regulatory threshold based on health effects under this assumption.

DCD4 discusses the issue of exposure-effect proportionality, vis-à-vis exposure thresholds, in several places (section 8.4.6, for example). When non-proportional effects are allowed in the effect estimation model, 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). In these studies, the response is modeled as 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. A better approach could use equally-spaced discretized levels of ambient PM, say L_1,L_2,.. with nested indicator variables I(PM<L_1), I(PM<L_2),.. . This kind of analysis would give direct estimates of the incremental ambient PM health response at each succeeding PM concentration level, together with an interpretable estimate of its uncertainty.

Many of the disparate separate city estimates of PM response functions, reported in DCD4, seem 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. However, in several multi-city studies PM response functions 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. 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 putative benefits of ambient PM reductions in any particular city cannot be deduced from the pooled response function. The conclusions of DCD4 rely strongly on questionable commonality and linearity of the PM-effect response function.

However, such pooling of response functions across cities ignores monitoring/exposure heterogeneity among cities, as described in the preceding section of this review.

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, DCD4 discounts the importance of studies that show response thresholds, see Smith et al. (2000) for example, in favor of pooled response functions that are difficult to interpret.

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*. (2000b). 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. Individual-level response modeling is not incorporated in the above-cited reference where it is implicitly assumed that a common linear response function applies to all individuals. There are various approaches to individual-level response modeling, the simplest being to use a parametric family of response functions with parameters treated as random effects distributed across the community. As an example, a parametric family of response functions could allow for the possibility of individually varying response thresholds.

To see the possible consequences of heterogeneous non-linear response at the individual level, consider the following illustrative example:

	Day 1	Day 2	Day 3
Individual Exposures:	1 2 3	2 3 4	3 4 5
Individual Responses:	0 0 0	0 0 0	1 1 1
Community Exposure:	2.00	3.00	4.00
Community Response:	0.00	0.00	1.00

The above table should be interpreted as follows: On day 1, personal exposures to a pollutant vary and are equally divided among concentration values 1, 2, 3 . Similarly, on day 2 and day 3, personal exposures to the pollutant vary and are equally divided among concentrations 2, 3, 4 and 3, 4, 5, respectively. There are no health responses on day 1 and day 2, but all individuals respond on day 3. The community-level exposure and response for these three days are obtained by averaging the individual exposures and responses. If the exposure-response function were assumed linear, then the linear regression fitted to the individual level exposure-response data would indicate an effect reduction of 0.25 per unit reduction of the pollutant. However, a linear regression fitted to the community-level data indicates an effect reduction that is twice as large. At the individual level, the empirical exposure-response function exhibits a threshold:

Exposure:	1	2	3	4	5
Response:	0	0	0.33	0.50	1

This example is meant only to demonstrate how imposed linearity of exposure-response can be misleading when there is heterogeneity of individual exposure.

Personal exposure studies are sometimes used to infer non-ambient personal PM exposure, via regression of daily total personal PM exposure on daily ambient PM. Non-ambient PM exposure can be seen as a co-pollutant, albeit presumably with the same exposure-response function. The question is whether non-ambient PM effects might be confounded with ambient PM effects. DCD4 reports that the two PM exposures are not correlated over time (Section 5.5.2) and therefore non-ambient PM exposure should be ruled out as a potential confounder of ambient PM exposure. However, the non-ambient PM is a computed residual from a regression, therefore one needs to account for the downward bias in the naïve correlation between the computed non-ambient PM and the ambient PM regressor variable. [The ratio of ambient PM exposure to monitored PM could depend on the monitored value, for example because daily and seasonal air exchange in indoor environments could be correlated with ambient PM.] The possibility of non-ambient PM exposure as a possible confounder has not been ruled out.

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.

6. LAG SELECTION AND DISTRIBUTED-LAG MODELS

DCD4 (Section 8.4.5) discusses the issue of model selection in connection with choosing time-series lags that maximize PM effects. A common 1-day lag was chosen for the 90-city study (Dominici et al. 2002) 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 because it had the largest overall estimated effect among the three lags that were considered, and it was the only lag choice with clear overall statistical significance based on the HEI reanalysis. 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.

An alternative approach to lag selection is to use a distributed-lag model, where PM effects extend over several days and separate coefficients are 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 could include non-causal negative lags as a check on

the credibility of the distributed lag model, but none of the reported studies considered this possibility.

However, the distributed-lags literature cited in DCD4, such as the Schwartz (2003) 10-city study, has at least one 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. Summed lag effects, as reported in DCD4 (8-237) are typically several time larger than single lag effects, but they do not account for distributed-lag weather and co-pollutant effects.

7. MORTALITY DISPLACEMENT

Analyses of time-series PM mortality data, using distributed-lag models, have also been used to claim that reported PM mortality effects do not represent mortality displacement of frail individuals with short longevity, see Schwartz (2000). On the one hand the series of effect coefficients as a function of lag are characteristic of mortality displacement, and on the other hand, the sum of the effects coefficients are larger than that obtained by using a single effect coefficient which is used as an argument against mortality displacement. Neither of these claims is derived from an analysis in the context of an explicit mortality displacement model, so the conclusions are simply interpretations. This study used distributed lag model estimates that were pooled across 10 cities, further obscuring the clear differences among cities in characteristics of the lag coefficients.

Roberts (2003a) examined the behavior of distributed lag coefficients using a frail population model and concluded that neither the serial pattern of these coefficients nor their sum is necessarily indicative of mortality displacement. Thus further study is needed to understand what information is provided by distributed-lag models relevant to the question of mortality displacement. The simulation study reported by Zeger et al. (1999) is based on a decomposition of the mortality time series into shorter and longer time scales, but does not use an explicit frail population model to generate daily mortality. A characteristic of mortality displacement in frail populations is that the initial excess of deaths following a high pollution day is recouped gradually over an extended period, which was not reflected in their choice of simulation model. The most promising approach is that of Smith et al. (1999), Murray and Nelson (2000), and Roberts (2003a) who study PM effects explicitly in the context of frail population models. In the examples that they have studied, it appears that any excess PM mortality is indeed consistent with mortality displacement in frail populations with mean lifetime on the order of weeks.

DCD4 notes the inconsistency of mortality-displacement conclusions obtained on the one hand by interpretations of distributed lag models and frequency decomposition, and on the other hand by frail population models (8-273). But, somewhat incongruously, DCD4

is dismissive of the early suggestions of mortality displacement for PM. The policy implications are admittedly large if the acute effects of PM are indeed confined to a frail population.

DCD4 also cites a calculation of PM longevity effects (8-273) by Brunekreef (1997). The claim is that the life expectancy of 25 year olds is reduced by 1.11 years for each $10\mu g/m^3$ of time-averaged PM. The estimate is derived from the ecological study of long-term PM mortality by Pope *et al.* (1995). The Brunekreef model calculation was not checked against available data but it would be a good idea to do so, if only to see whether the implications of the PM mortality effect estimates are in accord with actuarial data. Checking would require a comparison of demographically adjusted mortality tables for different cities with different time-averaged PM, a plausible and necessary undertaking.

8. LONG-TERM PM-MORTALITY STUDIES

DCD4 refers to several long-term ecological cohort studies of PM health effects, of which Pope *et al.* (2002) is the latest and most comprehensive. In these long-term studies PM and mortality for each city are represented by single average numbers that do not vary over time. The ecological studies cited by DCD4 are cohort studies that are limited to enrolled individuals for whom individual covariate information is available such as demographic information and smoking habits. The individual covariate information is used to adjust crude mortality rates for the enrolled cohort so as to even out the mortality comparisons between cities. PM health effects are inferred by relating time-averaged adjusted mortality to time-averaged monitored PM across cities.

Both the Pope *et al.* (2002) cohort study of long-term PM effects and the Dominici *et al.* (2002) time-series study of short-term PM effects involve a comparable number of U.S. cities. However, geographic variation in the cohort studies takes the place of time variation in the time-series studies. City-specific effect modifiers in time-series studies, as discussed earlier, become confounding variables in the cohort studies. A putative confounding variable in a cohort study is one that shows geographic covariability with PM. Thus, demographic adjustments in the cohort studies are a way of accounting for potential confounding of PM effects by demographic variables. Similarly, between-city variations of co-pollutants and climate variables could be related to between-city variations of PM and thereby contribute to confounding of PM effects.

On the other hand, where it is possible to split the data record into time segments, then a separate cohort-based analysis of PM effects could be done for each time segment, as was done in a very limited way in Pope *et al.* (2002). For example, it would be straightforward to repeat the analysis using cohort deaths in 5-year intervals. Then the different time-segment PM effect estimates in the cohort study are analogous to different-city PM effect estimates in the time-series studies, and issues of PM effect heterogeneity would need to be addressed. A cohort study that looks at a single time period is, in this respect, analogous to a time-series study for a single city.

In some important ways, however, a multi-city cohort study suffers from disadvantages vis-à-vis a single-city time-series study. For example, in a time-series study the population at risk is the same each day while in the cohort study the population at risk in each city is different and models are needed to bring the separate at-risk populations into alignment. Also, the assigned PM concentration for a city needs to be related not to the average city-wide PM exposure but rather to the average PM exposure of the cohort assigned to that city. Further, it is reasonable to suppose that this exposure measurement error will be different for different cities.

With a fairly dense collection of cities, unmodeled geographic effect differences could be partially modeled as spatially autocorrelated residuals, as was done in this study. This is a useful addition that is akin to allowing for temporal autocorrelation of residuals in time-series studies to account for unmodeled variations in effects over time. Even modest autocorrelation can sometimes have important consequences, especially on the precision of effects estimates. This is analogous to incorporating smooth time trends as explanatory variables in time-series studies. This cohort study states that incorporating spatial autocorrelation and using a spatially smooth residual field takes care of unmodeled risk factors; however, this could only work if the unmodeled risk factors were themselves spatially smooth fields. Unmodeled risk factors can be geographically correlated with PM without having a locally smooth spatial structure.

A strength of this study is the stratification approach that was used to estimate PM effects separately for different age groups, sexes, education levels, and smoking status. The stratification approach bypasses some of the additivity assumptions that pervade the time-series studies. Figure 4 of Pope *et al.* (2002) shows clearly that PM effects can be different for different age groups, sexes etc. Strict additivity would have enforced a common PM effect at all levels of all control variables, as in the time-series studies. Of course, some additivity assumptions are necessary to avoid multi-way stratification of the data with severe loss of estimation precision. Generalized additive models were used presumably at the point of combining effects of control variables but it was not clear to me how this was done. DCD4 (8-106) reports that the convergence problem of the S-Plus estimation routine did not impact PM effect estimates in this study.

Although pollutants other than PM were considered in Pope *et al.* (2002), it does not appear that these co-pollutants were used to adjust the baseline mortality when the PM effect was estimated. Perhaps, more significant, is that there was no explicit adjustment for climate variables -- variables for which ample information is available for any time period. Climate effects would not be efficiently modeled by a nonspecific spatial trend. Omission of these potentially important confounders is a significant shortcoming of that study.

9. CONCLUSIONS

DCD4 appropriately emphasizes multi-city studies, in particular the 90-city study, because a common modeling approach was used. Thus 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 DCD4. This review describes some of these deficiencies and offers suggestions for strengthening the analysis. Because of the deficiencies in DCD4, we cannot draw comfortable conclusions regarding the circumstances and magnitudes of ambient PM health effects, or whether reported PM health effects are causative. Below I briefly summarize points made in this review.

1. **Sensitivity of PM estimates to model specification.** This issue was brought to light in the HEI reanalysis in the context of time and weather adjustments, and serves as a cautionary tale. The reported effects of PM are often difficult to discern and are inconsistent among cities, regions, seasons, and time lags. Such inconsistencies may be suggestive of modeling inadequacies, particularly in regard to unmodeled confounding and unexplained effect modifiers.

2. **Enforced additivity in the analysis model**. The analysis model assumes that the PM health effect is necessarily the same at any temperature, in every season, and at any level of the co-pollutants. Limited analyses show that this assumption is likely to be seriously violated. There are at least three approaches to mitigate the problem, depending on availability of data – joint response surface modeling of PM and its confounders, stratification of the analyses based on confounder categories, or making the PM response be a parametric function of covariates.

3. **Enforced linearity of exposure-response**. There is evidence that PM health effect reductions would be different at different PM levels, depending on geographic location. This has important implications for regulation. Enforced linearity conceals heterogeneity of response, and pooling of response functions to obtain linearity is not statistically justified and leads to regulatory dilemmas.

4. **Unexplained heterogeneity of PM health effect estimates.** There are significant differences among estimates of PM health effects for different cities and using different PM monitors within the same city. There is no reconciliation of these differences in DCD4, which makes it hard to argue from the epidemiologic data for a causative role for PM, and which casts doubt on the completeness of the model under which the data have been analyzed.

5. **Incomplete characterization of the relations between ambient PM exposure, individual PM exposure, individual PM susceptibility to health effects, and community level health effect measures**. While there has been progress in modeling and understanding relations between the time variations of individual exposure and ambient PM, the important link to individual response functions on the health effects side has not been made. The topic is important because it

provides modeling guidelines for community level studies and elucidates the anticipated benefits of PM reductions.

6. **Insufficient attention to the issue of mortality displacement.** Some studies suggest that acute PM mortality effects are consistent with mortality displacement in frail populations. This issue is important for public policy and needs to be studied more intensively. In the meantime DCD4 should not be so dismissive.

The fourth draft of the PM Criteria Document (DCD4) emphasizes the "coherence" of PM health effect estimates across a number of epidemiologic studies. In fact, it is difficult to see this coherence even within the multi-city studies that DCD4 appropriately emphasizes. 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.

That PM effect estimates are delicate is not surprising given that they are superimposed on much stronger effects due to concomitant weather variations, for example. These widely varying, sometimes negative, PM health effect estimates are symptomatic of probable model shortcomings. Given the difficulty of the task of estimating PM health effects, it is a matter of perspective whether one uses the term "coherent" to describe the diverse findings summarized in DCD4. The perspective of my comments is to point to likely model deficiencies and to suggest alternatives that might lead to PM health effect estimates that are truly more consistent and therefore more credible.

REFERENCES

Brunekreef, B. (1997) Air pollution and life expectancy: is there a relation? *Occup. Environ. Med.* 54: 781-784.

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.; Samet, J.; Zeger, S.L. (2000a) Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modeling strategy. *J. Royal Statist. Soc.* A 163: 263-302.

Dominici, F.; Zeger, S. L.; Samet, J. (2000b) 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.

Mar, T.F.; Norris, G.A.; Koenig, J.Q.; Larson, T.V. (2000) Associations between air pollution and mortality in Pheonix, 1995-1997. *Environ. Health Perspect.* 108: 347-353.

Morris, R. D.; Naumova, E. N.; (1998). Carbon monoxide and hospital admissions for congestive heart failure: evidence of increased effect at low temperatures. *Environ. Health Perspect.* 106: 649-653.

Pope, C.A.III; Burnett, R.T.; Thun, M.J.; Calle, *et al*. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Medical Assn.* 287: 1132-1141.

Pope, C.A.III; Thun, M.J.; Namboodiri, M.M.; *et al*. (1995) Particulate air pollution as a predictor of mortality in a prospective study of US adults. *Am. J. Respir. Crit. Care Med.* 151: 669-674.

Roberts, S. (2002, 2003a) Personal communications. Stanford University Department of Statistics.

Roberts, S. (2003b) 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. (2000) Harvesting and long term exposure effects in the relation between air pollution and mortality. *Am. J. Epidemiol*. 151: 440-448.

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

Schwartz, J. (2003) Airborne particles and daily deaths in 10 U.S. cities. Special reanalysis report, Health Effects Institute pp. 211-218.

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.

Zanobetti, A.; Wand, M. P.; Schwartz, J.; Ryan, L. M. (2000) Generalized additive distributed-lag models: quantifying mortality displacement. *Biostatistics* 1: 279-292.

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

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.