A Review of Statistical Methods Used in Time-Series Epidemiologic Studies of Ambient Particulate Matter and Acute Health Effects Cited by the April 2002 EPA Draft PM Criteria Document

Paul Switzer, Ph.D.
Department of Statistics
Stanford University
July 8, 2002

1. CONSISTENCIES AND INCONSISTENCIES

The third draft of the PM Criteria Document (DCD3) concludes that there is agreement and consistency of PM health effect estimates across a number of epidemiologic studies. In fact, it is difficult to see this consistency even within the multi-city studies that DCD3 appropriately emphasizes. Even among those analyses that do estimate health effect reductions from reduced PM, the 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 'consistent' to describe the diverse findings summarized in DCD3. 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.

2. CONFOUNDING AND ADDITIVITY

Because PM variations are correlated with variations in weather, seasons, and other air pollutants, great care is needed in articulating PM effects separate from the effects of these co-varying series. This is the confounding issue. Special care is needed in separating PM effects from those covariates that putatively have much larger effects than PM, such as weather and season. The emphasis in DCD3, on the contrary, has been on co-pollutants whose effects are likely of comparable magnitude to those of PM.

With few exceptions, the literature cited by DCD3 deals with co-varying confounders through an additivity assumption. The additivity assumption is very strong and it presupposes that the incremental effects of PM would be the same at any level of the

confounding variable. 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 the common and naïve additive modeling of PM effects can lead to biased and meaningless 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. An example of this approach is found in Morris and Naumova [1995] who found, for example, that CO effects were different at high temperatures than they were at low temperatures.

A second strategy for incorporating non-additivity of PM effects is to stratify, effectively to do separate analyses at different levels of the confounding variable. This strategy has been used in those studies that seek to separate seasonal effects from PM effects. An example of this approach is described in DCD3 [8-30], which refers to Samet *et al*. [2000]. 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. The advantage of the stratification approach to non-additivity is that its structural constraints are simple to describe. This approach could be extended to weather variables by creating weather 'bins' and by allowing for separate PM effect parameters for each weather bin. A recent reanalysis by Roberts [2002] of the Chicago mortality time-series also used three maximum- temperature bins with the result that the warm temperature bin showed a positive PM mortality effect, the cold temperature bin showed a negative effect, and the moderate temperature bin showed no effect.

No discussion of potential confounding is adequate without an investigation of the *joint* effects of PM and its putative confounders. In this respect DCD3 falls short because it relies on the results of studies where additive assumptions were built into the model from the start. I strongly recommend that the multi-city studies be reanalyzed in a way that allows for non-additive PM effects, especially in relation to weather but also in relation to co-pollutants. Stratification is an important diagnostic tool that could be applied to all 90 cities; inconsistencies revealed by stratification should be regarded as symptoms of unmodeled effects. In any event, it seems likely that either seasonal or weather stratification will reveal shortcomings of the strong additivity assumptions in DCD3.

Finally, DCD3 [8-217,218] 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 for to rely on interpretations of these coefficients to draw conclusions about the relative importance of different pollutants.

3. HETEROGENEITY OF PM EFFECTS AND EFFECT MODIFICATION

DCD3 [8-26] 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. Thus the patent diversity or heterogeneity of PM effect estimates cannot be attributed to differences in modeling strategies. There are different statistical approaches to the analysis of heterogeneous effect estimates. It is reasonable to suppose that city-to-city variations in PM effect estimates, including many *negative* PM effects, are not due statistical variability arising from data limitations. This appears to be ruled out although DCD3 does not cite a P-value for the rejection of this hypothesis.

The approach to heterogeneity favored by DCD3 [8-26-29], 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 estimated, with apparently high precision [8-29]. However, the population mean is purely a phantom model parameter so its estimation should not be of great interest in the real regulatory world. However, 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” in DCD3 [8-4 – 8-9] and is plausibly the most rational and most useful

approach. DCD3 carefully and wisely distinguishes between the issues of confounding and effect modification [8-4]. 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.

DCD3 [8-29] 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. 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.

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 [8-271]. 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 patent differences among PM effect estimates is disconcerting because the discrepant PM effect estimates remain unexplained. *DCD3 should be open to the strong possibility that unresolved discrepant PM effect estimates, derived from the same modeling approach, point to probable inadequacies in the modeling approach such as an incorrect treatment of confounding variables or an incorrect characterization of the exposure-response relationship.*

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 it appears that PM health effects are absent, for example in Cleveland. In such cities where PM effects are absent, would any health benefit be derived by reducing ambient PM?

4. HETEROGENEITY OF EXPOSURE

Epidemiologic city-based studies of PM health effects, such as those cited and relied on by DCD3, have assumed a common or homogeneous PM exposure for all individuals in a city. In fact, 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 studies by Ito *et al*. [1995]. 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.

Examples show that different PM monitors within the same city can provide quite different estimates of PM effects in time-series studies; see Ito *et al*. [1995]. The same might also be true for ecological long-term effect studies, but I don't know if this has been explored. The PM effect estimate variation in the time-series studies is statistically meaningful. Furthermore, 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. In any event, the within-city disparities of PM effect estimates begs for some explanation, since unexplained heterogeneity casts some suspicion on the statistical models used to relate PM changes to health effect changes.

Ito *et al*.'s [1995] examples of within-city heterogeneity of PM effect estimates have been recently updated by Roberts [2002] for Chicago, using additional monitoring stations and a longer study period. Of the twelve 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. PM10 averaged over all 12 sites was not as strongly associated with mortality as the big four. One site with daily PM10 showed smaller but significant mortality association. The PM effect estimates were obtained 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.

The issue of spatial variability within cities has other implications. A hypothetical example would be a situation in which some monitors typically record lower PM than others. Suppose that the low-PM monitors are roughly proportional to the high-PM monitors. With a linear exposure-response, the effect estimates for the low-PM monitors will be *larger* than those for the high-PM monitors because they are using the same community health effects data. Averaging across monitors will conceal the problem but not deal with it. Instead, one needs a somewhat sophisticated spatial modeling approach to properly combine information from multiple monitors. If the PM exposure-response

function has a threshold, then the estimated PM effects at low-PM monitors could be smaller than those at high-PM monitors.

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 [8-264].

A number of questions arise with regard to using ambient monitoring data as a measure of population PM exposure. DCD3 [8-252] addresses the ambient-to-exposure question principally as a problem of measurement error. Under a series of assumptions, using ambient PM as a surrogate for PM exposure would deflate the PM effect estimator in a model where effects are proportional to PM exposure, as described in Dominici *et al*. [2000b]. The deflation factor that connects ambient concentration with exposure could well be different for different cities and even for different monitors within the same city. For example, DCD3 [5-53] refers to personal monitoring studies that show ambient-to-exposure calibrations that vary by location.

Thus the spatial variability of the ambient-to-exposure calibration bears directly on the question of observed heterogeneity of estimates of PM effects across cities, as well as the heterogeneity of PM effect estimates when using different monitors within the same city. There is also 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. 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 PM exposure is the same at high and low PM exposures. However, there is a regulatory issue associated with presumed proportionality of exposure and effect. For a fixed PM reduction, the same effect reduction would be achieved in cities whose current PM levels are either high or low.

Furthermore, one could double the effect reduction in any city by doubling the PM reduction, so there is no obvious regulatory threshold based on health effects.

DCD3 discusses the issue of exposure-effect proportionality vis-à-vis exposure thresholds in several places [5-98, 8-246]. 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 might have been to 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 DCD3, 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. DCD3 goes further and reports a single overall proportional PM-effect function [8-247] by combining the disparate response functions for different cities. However, such an overall PM-effect function has no concrete interpretation or useful application because the city-to-city differences among PM-effect functions cannot reasonably be ascribed to sampling variability. Given the patent 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.

Better insights into the relationship between monitored ambient PM concentrations and anticipated community-level PM health effects can 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 simple 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 at the community level can be misleading when there is heterogeneity of individual exposure.

DCD3 cites a number of longitudinal personal exposure studies [5-22, 23], although most are limited in size and scope. These studies compare time series of total personal PM exposure with monitored PM on a daily basis, for selected individuals. Linear calibrations are used to infer personal exposure to ambient PM, without explicit accounting for spatial variability of ambient PM. The calibration between personal and ambient PM might be used to adjust PM effect coefficients derived from PM monitoring data as described in DCD3 [8-252], assuming throughout that the imposed linearity of exposure-response is correct. This raises the question of study design because the ambient-personal calibration coefficient for each individual should be seen as a random effect across the population with spatial structure. Therefore, it is important to design the exposure survey in a way that allows one to reasonably estimate a distribution of the calibration coefficients, and to infer the implications of calibration variability for modification of PM effect estimates. The use of pooled non-longitudinal data for recalibration of PM effects, such as PTEAM data, raises other issues that are not resolved in the DCD discussion [8-252].

Personal exposure studies are sometimes used also to infer non-ambient personal PM exposure, via subtraction of daily ambient PM from the daily total personal PM exposure. Non-ambient PM exposure can be seen as a co-pollutant, albeit with the same exposure-

response function presumably. The question is whether non-ambient PM effects might be confounded with ambient PM effects. DCD3 concludes that the two PM exposures are not correlated over time [8-252] 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 and therefore there is a downward bias in the naïve correlation between the computed non-ambient PM and the ambient PM regressor variable.

6. LAG SELECTION, DISTRIBUTED-LAG MODELS, TIME AVERAGING, AND MORTALITY DISPLACEMENT

DCD3 [8-237] discusses the issue of model selection in connection with choosing time-series lags that maximize PM effects. Even though a common 1-day lag was chosen for the 90-city study [8-27] to mitigate model selection bias, there is still bias present because other candidate lags were considered as part of the modeling process, see Samet *et al*. [2000]. Simulation studies by Lumley and Sheppard [2000] have shown that such model selection bias can be of the same order as the estimated PM effect itself. Although model selection bias is acknowledged in DCD3 [8-243], it seems not to have affected the conclusions of DCD3.

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, it is a good idea to see what effect estimates might be produced by distributed-lag models that include non-causal negative lags.

However, the distributed-lags literature cited in DCD3, such as the Schwartz [2000] 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 DCD3 [8-241], are typically several time larger than single lag effects, but they do not account for distributed-lag weather and co-pollutant effects. Finally, 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. But this claim is not derived from an analysis in the context of an explicit mortality displacement model, so the conclusions are simply interpretations.

Time-series data have been used also in an attempt to extract possible PM effects at longer time scales by using time-averaged health effects in place of daily data, such as Zeger *et al*. [1999], Zanobetti *et al*. [2000], and Schwartz [2000] as cited in DCD3 [8-244]. The proposal is to decompose the health effects time-series into a sum of component time series that represent increasing levels of temporal smoothing. It is

reported that higher levels of temporal smoothing correspond to larger PM effects estimates, with the conclusion that still larger PM effects extend beyond short lags. This kind of analysis has some problems. First, the smoothed time-series incorporate health effects that precede pollution, which is not in accord with causative models. Second, even modest time-series autocorrelation, either in the PM series or in the health effects series, could affect the derived sequence of time-averaged PM effect estimates in ways that are not obvious.

The foregoing studies are cited by DCD3 also as evidence that estimated PM mortality effects do not represent mortality displacement of frail individuals with short longevity. If this were the case, we should have seen *reduced* PM effects estimates for mortality time series with more smoothing, contrary to what has been reported. However, the model under which the sequence of PM effect estimates is obtained is not an explicit mortality displacement model so its conclusions are at best informal. What is sorely needed is a model analysis that explicitly accounts for and quantifies mortality displacement.

DCD3 also cites a related calculation of PM longevity effects [8-231] 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 PM mortality by Pope *et al.* [1995]. This 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.

7. LONG-TERM PM-MORTALITY STUDIES

DCD3 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 DCD3 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 how this was done or how the convergence problem of the S-Plus estimation routine impacted the 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.

8. CONCLUSIONS

DCD3 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 DCD3. This review describes some of these deficiencies and offers suggestions for strengthening the analysis. Because of the deficiencies in DCD3, 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. **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.

2. **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.

3. **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 DCD3, 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.

4. **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.

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.

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.; Munasinghe, R. L. 1995. Ambient air pollution and hospitalization for congestive heart failure among elderly people in seven large US cities. *Am. J. Public Health.* **85**: 1361-1365.

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) Personal communication. Stanford University Department of Statistics

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.

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.