Comments on the PM Criteria Document

Peter Guttorp and Lianne Sheppard
National Research Center for Statistics and the Environment,
B211 Padelford Hall, Box 354323,
University of Washington, Seattle, WA 98195-4323

and

Richard L. Smith,
Department of Statistics,
University of North Carolina, NC 27599-3260

July 12, 2001

1 Introduction

This paper has been prepared as part of the public commentary process on the second external review draft of the Particulate Matter Criteria Document (EPA, 2001). The Criteria Document was made available for public review on April 11, 2001, and there is a 90-day period during which anyone can submit comments on it. The Criteria Document is an important part of the regulatory process which will lead to the establishment of a new particulate matter standard, and is supposed to represent an impartial review of the scientific literature. Most of our comments concern Chapter 6, "Epidemiology of Human Health Effects from Ambient Particulate Matter", though they are also relevant to Chapter 9,"Integrative Synthesis...", to the extent that the latter chapter draws on material in the former. Where possible, we have tried to refer to specific passages or pages in the Criteria Document, though our most serious reservations do not concern the way the document has treated individual papers that it has reviewed, so much as the overall approach that it has taken. Most of the published studies in this field of research have demonstrated a positive association between increased levels of particulate air pollution and adverse human health outcomes (mortality and morbidity). However, there are many questions of scientific interpretation that must be addressed before these studies can reasonably be said to justify tightened standards and increased regulation of ambient particulate matter. Many of these scientific interpretation issues are of a statistical nature.

In other words, they concern matters such as the design of a sampling scheme, the choice among different methods of statistical analysis, and the statistical interpretation of the results of an analysis. In general, we find that throughout Chapter 6, these statistical issues have been dealt with very poorly or ignored altogether. Nevertheless, there are by now a substantial number of papers in the published, refereed scientific literature that address statistical issues associated with particulate matter epidemiology. Some of these papers have been omitted entirely from the review, while others that are included in the Criteria Document have only been dealt with cursorily or in a manner that ignores their statistical content.

The remainder of this discussion is set out as follows. Sections 2 and 3 are primarily intended to fill in gaps in the CD. Section 2 reviews a special issue of the statistical journal *Environmetrics* on statistical analysis of particulate matter air pollution data. For some reason this special issue, the production of which was part of an EPA-funded effort to assess statistical aspects of the PM question, has been left out of the document. The very brief Section 3 aims to clarify some points concerning one of our papers which is described in the CD. These two sections are largely intended to fill in gaps in the current CD. Section 4 is a much broader review of one of the main methodological techniques used in the current PM literature, time series analyses of ambient PM exposure against either mortality or morbidity outcomes (the present review is primarily concerned with mortality studies). In this, we review a number of the methodogical issues raised by these analyses, including the combination of data from different cities (the main purpose of the NMMAPS study), publication bias, the effects of model selection, non-linear dose-response relationships and co-pollutants. Section 5 is a similar review for the other main form of research in this field, cohort studies of long-term effect. Finally, Section 6 summarizes our conclusions. Overall, we believe that for both the time series analyses and the cohort studies, the PM Criteria Document has failed to perform a remotely adequate job of summarizing the available literature in a manner that would allow true appreciation of the complex issues that these studies raise.

2 Environmetrics special issue

A special issue of the environmental statistics journal Environmetrics (vol. 11, Number 6, November/December 2000) was dedicated to statistical aspects of PM air pollution. Although substantial efforts were made to assure that these articles were made available to the PM CD staff in time for the original June deadline, and the issue appeared before the end of the year, these papers have not been taken into account in the CD. We describe the papers in the volume briefly below.

2.1 Sun et al. (2000)

The authors set forth a spatio-temporal model for daily PM_{10} in the Greater Vancouver area in British Columbia. The temporal structure is described by a single autoregressive model of order 1 for all stations. There is no evidence of leakage of correlation from space to time, i.e., the spatial correlation of the raw data and that of the residuals from the time series model are roughly the same. This is in contradistinction to the case of hourly data, where this type of leakage is serious. The spatial correlation by itself is, however, found to be heterogeneous. The resulting model is used for Bayesian prediction of the underlying PM_{10} field in a dense grid of points. The authors point out that the common assumption of spatial stationarity (or homogeneity) is violated in this case, as is quite common in environmental applications.

Interpolation of PM_{10} fields between monitoring stations is of potential importance in assessing the overall societal impact of new air pollution standards. In this paper, the proposed methodology performed well when evaluated using cross-validation, and this to some extent justified the rather complex approach taken (involving a heterogeneous model and hierarchical structure for the spatial dependence, in contrast to a simple geostatistical approach such as kriging). On the other hand, they noted some caveats in their approach, for example, that the interpolated spatial surfaces are very irregular (which complicates the interpretation) and that the model does not seem to do so well in predicting extreme levels of PM_{10}.

2.2 Dewanji and Moolgavkar (2000)

A point process model for recurrent events is applied to hospital admissions for chronic respiratory disease in King County, Washington, over the years 1990-1995. These data have also been analyzed by Moolgavkar *et al.* (2000). The analysis uses different temporal stratifications (varying from no stratification to half months), as well as pollution data on PM_{10}, CO and very fine particles (nephelometry). Temperature was taken into account either using a linear model or a cubic polynomial. All the pollutants are associated with hospital admissions. The effect of PM is stronger than that of CO in multi-pollutant models, in contrast to the previous analysis by Moolgavkar (2000). The effect of temporal stratification, even fairly coarse, is substantial, and decreases the effect estimates compared to those from the non-stratified model.

2.3 Lumley and Levy (2000)

The case-crossover design, which is commonly used in air pollution health effect studies, relies on a strong temporal stationarity assumption. This can be eliminated by using short enough time-frames. The standard analysis, namely conditional logistic regression models as in case-control studies, produces a persistent bias, which is due to a false analogy between the two designs. There are several reasons for this: in case-crossover studies the exposures are autocorrelated over time, while in a matched case-control study the exposures are independent. Two cases that occur on the same day will have the same (or very similar) exposure measures in case-crossover studies, while this constraint between strata does not occur in case-control studies. Finally, in a matched case-control study the stratification depends only on covariates and not on the response, while in a case-crossover study the stratification depends on the response. A simulation study indicated that for data similar to Seattle air pollution data, the degree of bias in this case was not much larger than the finite-sample bias. However, adjustment for meteorological factors or co-pollutants may introduce additional bias.

2.4 Lumley and Sheppard (2000)

The effect of selecting lags on the resulting model for particulate matter health effects is one of the main issues in model selection. Using simulated data with parameters similar to a Seattle $PM_{2.5}$ series, the bias resulting from the selection is shown to be similar in size to the relative risk estimates from the measured data. More precisely, the log relative risk from the measured Seattle data is about twice the mean bias in the simulated control data, and the published estimate of relative risk is only at the 90th percentile of the bias distribution in these control analyses. The selection rule used was to choose the lag (between 0 and 6) with the largest estimated relative risk. In comparisons to real data from Seattle for other years, and from Portland, OR, with similar weather patterns to Seattle, similar bias issues appeared.

2.5 Smith et al. (2000b)

Many ad hoc decisions go into model selection in air pollution health effects studies. The effect of some of these decisions on relative risk estimates for Birmingham, AL, PM_{10} data, previously analyzed by Schwartz (1993) and others, is illustrated. The response variable is non-accidental mortality. Specifically, the selection of meteorological variables, the selection of an exposure variable (as a weighted average of lagged PM values), and the possibility of nonlinear effects, such as threshold effects, are investigated. The results are sensitive to the inclusion of humidity in addition to temperature. This inclusion decreases the resulting PM_{10} coefficient. The model is highly sensitive to the definition of an exposure measure. For example, when lags 0-4 were averaged, there was no significant effect. In an attempt to account for a nonlinear PM-mortality effect, there appeared to be little effect of exposure below 80 $\mu g/m^3$, and a threshold analysis (as well as a generalized additive models approach) supported the conclusion that the main effect is at higher values of PM. Although this paper was based on an intensive analysis of a single data set (in contrast to other studies, such as the NMMAPS analysis discussed in Section 4.1 below, which combined data from many cities), it demonstrated the very wide range of interpretations that are possible using alternative, but statistically valid, analyses of the same data.

2.6 Clyde (2000)

A more systematic analysis of model choice is obtained using Bayesian Model Averaging. The same Birmingham, AL, data as analyzed by Smith *et al.* (2000b) were used. Several different calibrated information criterion priors were tried, in which models with large numbers of parameters are penalized to various degrees. After taking out a baseline trend (estimated using a GLM estimate with a 30-knot thin-plate smoothing spline), 7860 models were selected for use in model averaging. These included lags 0-3 of a daily monitor PM_{10}, an area-wide average PM_{10} value with the same lags, temperature (daily extremes and average) lagged 0-2 days, humidity (dew point, relative humidity min and max, average specific humidity) lagged 0-2 days, and atmospheric pressure, lagged 0-2 days. The model choice is sensitive to the specification of calibrated information criterion priors, in particular disagreeing as to whether different PM_{10} variables should be included or not. For example, some PM_{10} variable is included in all the top 25 AIC models, but only in about one third of the top BIC models. Both approaches give a relative risk estimate of about 1.05 (to be compared to Schwartz value of 1.11 for a 100 μg/m^3 increase), with credibility intervals of (0.94,1.17) for the AIC prior and (0.99,1.11) for the BIC prior. A validation study in which left out data were predicted using the different priors favored Bayesian model averaging with BIC prior over model selection (picking the best model) with BIC or any approach with AIC.

2.7 Remaining papers

The three remaining papers in this volume are Cox (2000), Phelan (2000) and Sheppard and Damian (2000). The paper by Cox is a summary of presentations and discussions at the PM workshop at the National Research Center for Statistics and the Environment at the University of Washington in Autumn of 1998. Phelan outlines a stochastic process approach to cost-benefit analysis of air pollution regulation, and Sheppard and Damian present a methodological approach to combining ecological and individual-level data in the analysis of air pollution

3 Two points of clarification about Phoenix

The CD refers at a number of places to two papers of ours based on data from Phoenix, AZ. We would like to clarify some issues related to the paper Smith *et al.* (2000).

The reference to this paper in Table 6-1 (page 6-23) notes the absence of a specific estimate for the fine and coarse particles effects. The estimates are as follows (for the same analyses as those actually reported in the paper). Translated into a relative risk for a 25 μg/m^3 increase of either fine (PM$_{2.5}$) or coarse (PM$_{10-2.5}$) particles, the RR for coarse particles is 1.046 (i.e. 4.6% increase in mortality) with 95% CI (1.019,1.074). The corresponding results for fine particles are 0.993 (0.885, 1.109). As noted in the paper, the results are for different populations, city mortality data being used for the fine particles analysis and region-wide data for coarse particles. These numbers may also make it possible to compare the results of the paper with others depicted in Fig. 6-4, page 6-53.

Table 6-2 on page 6-51 also notes the absence from Smith *et al.* (2000a) of statistics related to mean levels of PM$_{10}$ and PM$_{2.5}$. In the data set we used for Phoenix, over the time period for which deaths were available (2/1/95 to 12/31/97), the mean level of PM$_{2.5}$ was 13.2 μg/m^3, the mean ratio of PM$_{2.5}$ to PM$_{10}$ was 0.28, and the value of r, the correlation coefficient between PM$_{2.5}$ and PM$_{10-2.5}$, was 0.68. For three-day aggregate values, the mean and mean ratio are virtually the same, but the correlation coefficient increases to 0.74.

4 Time series analyses

In this section, we review several issues related to time series analyses of PM data, concentrating on those that take mortality as an endpoint.

4.1 The NMMAPS study

One of the most significant new developments in particulate matter research since the 1996 PM CD is the NMMAPS study (Samet *et al.* 2000a,b) which has combined evidence from initially 20, and in later parts of the report 90, of the largest cities in the U.S. This work has, naturally, been given considerable attention to the CD, though with very little attention to the actual methodology involved. Given that we believe this is very important to the interpretation of the results, the following comments are concerned primarily with the methodology rather than the results of the NMMAPS reports.

The simplest approach to combining regression estimates from different cities is a meta-analysis in which the results are weighted with weights inversely proportional to the variance of the individual city estimates. This approach can be criticized as not allowing for random effects between the cities — in effect, a simple meta-analysis assumes that the effect being measured is the same in all cities, whereas in fact, one would expect the effects to be different in different cities. (To give just one among many reasons why, it is obviously the case that the composition of particulate matter varies by city, and there is growing evidence that PM composition has an important influence on health effects, though the precise nature of that influence is far from being well understood. If PM effects are analyzed city by city, without explicitly taking PM composition into account, one has to expect that the results will differ to a greater extent that what would be explained by variability in the individual regressions.) The hierarchical model analysis introduced in the NMMAPS report and in Dominici *et al.* (2000) allows for random effects, but there are potentially many different ways of specifying such a model. An even more radical, but potentially very important, extension of the analysis is to allow for spatial dependence among the cities. However, after making a good start to the spatial analysis (Part I, Samet *et al.* 2000a, p. 68), there is actually very little discussion of the spatial model itself (for example, do spatial correlations based on the fitted model actually correspond to observed spatial correlations in the data?), and in Part II (Samet *et al.* 2000b), it is apparently abandoned in favor of a somewhat simpler, but possibly less realistic, regional analysis.

From a modern perspective, all these models may be estimated using Bayesian Hierarchical Models, but different specifications of the models (e.g. different prior distributions) do lead to different results, and there is still only an incomplete understanding on how the prior specification influences the properties of the resulting estimators. The methods of Dominici, Samet and Zeger are as good as anything anyone else has derived, but there is not a full understanding of their properties.

One example of the contrast produced among different prior distributions and modeling approaches is Table 4 of Part I (page 71), where the posterior probabilities that the overall effect is positive are notably lower when the spatial model is adopted than under either the univariate or bivariate non-spatial models. The most likely explanation of this is that if spatial dependence is really present but is ignored, as in the univariate and bivariate models, then the posterior variances of the parameter estimates are underestimated, resulting in too high a posterior probability (the same positive posterior mean, but a larger posterior variance, would lead to a smaller posterior probability of a positive effect, assuming no substantial change in the shape of the posterior distribution between the two analyses). In Table 4, even under the spatial model, all the posterior probabilities are still over 0.8, but the fact that there is this difference between the spatial and non-spatial models makes it more disturbing that the spatial model has not been pursued more vigorously in Part II of the report.

The weighted regression (or meta-analysis) approach is also mentioned at a number of places and specifically developed for the morbidity analysis in Part II (pages 32–35 for the basic methodology). This is a simpler form of the non-spatial hierarchical models analysis, and should lead to fairly similar results as the Bayesian approach, as the authors claim at several points. The potential disadvantages of the weighted regression approach are (i) the final estimates and standard errors do not fully allow for uncertainty in estimating the variance components, and (ii) the method of estimating the variance components, in column 2 of page 34, is less efficient than maximum likelihood or Bayes — note, in particular, the possibility that $\hat{\Omega} = 0$. From various comparisons made in the text, it appears that these issues do not affect the results too much, but they might if compared with a fully spatial analysis in Part II.

In conclusion, we believe that the hierarchical modeling approach is sound, and a major new contribution to the methodology of particulate matter research. Our main concern is whether the NMMAPS authors have really explored enough different versions of the model, and especially, that they might not have gone far enough in the spatial analysis. We do not feel that any of these issues are adequately dealt with in the CD.

4.2 Publication bias

In Section 6.4.4 (lines 26, 27 of p. 6-238), the CD explicitly claims that it is reasonable to select the most significant lag among a set of possible lags even though such a practice may bias the chance of finding a significant association. This statement is made in spite of evidence of model selection bias that results from this approach in the peer-reviewed literature (e.g. see the specific study examined by Lumley and Sheppard, 2000) and evidence to the contrary that can be compiled from studies reviewed in the CD alone. We now present an analysis of data from the CD that indicates the presence of selection bias in the published literature.

The NMMAPS study can be used as a gold standard against which to assess the presence of publication bias in other PM mortality effect analyses. Among the strengths of NMMAPS relevant for an analysis of publication bias, the data were consistently handled across all cities, city-specific models were specified using the same criteria in each city, and the cities to be included were not specifically selected based on outcome (size is a covariate, not an outcome).

We compare the compilation of city-specific results from NMMAPS (gleaned from Figure 6-1 on page 6-41) with estimates reported in 21 separate references in Table 6-1. To be eligible for this analysis, the paper had to report a total mortality effect estimate for a 50 μg/m^3 increment of PM$_{10}$, reside in the peer-reviewed literature and represent a distinct analysis or dataset. (Thus, for example, we excluded separate published analyses of the 10 cities analyzed by Schwartz (2000). We did not include Levy (1998) because in our opinion it was based on incomplete work.) All the estimates considered were city-specific with the exception of the Schwartz (2000) 10-city estimate and the Burnett (1998) 8-city estimate. Table 1 shows the included studies, cities, statistical significances (as indicated by the CI), and effect estimates gleaned from Table 6-1 of the CD.

We test the null hypothesis that there is no difference between the NMMAPS collection of results and the independently published set. We can test this hypothesis in two ways: by looking at statistical significance of the results and by considering positive point estimates of excess deaths. In both cases we use a two-sample test of proportions and rely on the asymptotic normality of this statistic. In NMMAPS, 11 out of 88 city-specific estimates were statistically significant (i.e. had confidence intervals that excluded 0) and 63 out of 88 gave positive point estimates for excess deaths. In contrast, out of the 24 separate confidence intervals reported in the 21 references, 19 of 24 were statistically significant.

First author and publication date	City	Statistical significance	Estimate for $50\mu g/m^3$ PM_{10}
Schwartz (2000a)	10 cities	Sig	3.4
Moolgavkar (2000a)	Cook County	-	.5-1
Moolgavkar (2000a)	Maricopa County	-	.25-1
Moolgavkar (2000a)	LA	-	.5
Ostro (1999a)	Cochella Valley	Sig	4.6
Ostro (1999a)	Cochella Valley	NS	2.0
Fairley (1999)	Santa Clara County	-	8
Pope (1999a)	Ogden	Sig	12
Pope (1999a)	Salt Lake City	Sig	2.3
Pope (1999a)	Provo	NS	1.9
Schwartz (2000)	Chicago	Sig	4.5
Lipmann (2000)	Detroit	NS	4.4
Gwynn (2000)	Buffalo	Sig	12
Mar (2000)	Phoenix	Sig	5.4
Tsai (2000)	Newark	Sig	5.7
Tsai (2000)	Camden	NS	11.1
Tsai (2000)	Elizabeth	NS	-4.9
Gamble (1998)	Dallas	NS	-3.6
Burnett (1998a)	8 Canadian Cities	Sig	3.5
Burnett (1998a)	Toronto	Sig	3.5
Wordley (1997)	Birmingham UK	Sig	5.6
Hoek (2000)	Netherlands	Sig	0.9
Ponka (1998)	Helsinki	Sig	18.8
Peters (1999a)	Czech Republic	Sig	4.8
Michelozzi (1998)	Rome	Sig	1.9
Wichmann (2000)	Frankfurt	Sig	6.6
Morgan (1998)	Sydney	Sig	4.7
Ostro (1998)	Bangkok	Sig	5.1

Table 1: Studies including PM_{10} mortality estimates in CD, Table 6.1.

This leads to a z-statistic of 7.29 and resoundingly rejects the null hypothesis of no difference. Similarly, of the 28 separate effect estimates reported, 26 were positive, leading to a z-statistic of 3.09 for this comparison. Again the null hypothesis of no difference is rejected. Thus by relying only on information summarized in the CD, it is reasonable to conclude that the statement on page 6-238 (lines 26-27) is inappropriate.

Since the study of air pollution health effects is no longer in its infancy, it is not appropriate for studies to continue to operate in a hypothesis-generating mode where a priori no single candidate model is preferred and the investigators report the model producing the results most consistent with the prevailing prior hypothesis. The standard of analysis in the epidemiologic health effects literature must shift away from hypothesis generation to hypothesis confirmation. Hypotheses ought to be stated a priori, then tested and reported. Only after this confirmatory analysis can more exploratory secondary analyses be done, analyses that may consider other possible models. Even then recognition of the potential bias due to model selection should be specifically acknowledged in the CD.

4.3 Model selection

From a statistical point of view, the common epidemiological practice of choosing variables (including lagged variables, co-pollutants, etc.) that maximize the resulting effect estimates is a dangerous approach to model selection, particularly when the effect estimates are close to 0 (i.e. RR close to 1). As has been demonstrated in Lumley and Sheppard (2000), the effect of choosing lags for PM_{10} in this fashion has a bias which is of the same order of magnitude as the relative risk being estimated. This, in particular, throws doubt over the results of Sheppard *et al.* (1999), which on the face of it yielded a convincing case for the effect of PM_{10} and/or CO on the rate of hospital admissions of asthmatic children, since the Lumley and Sheppard simulation study used parameter values corresponding to those in Sheppard *et al.* (1999). More importantly, it demonstrates through a specific study the magnitude and type of bias that may be operating in all air pollution epidemiologic studies that select the most significant lag after evaluating a set of lags.

Similar selection bias results were illustrated by Smith *et al.* (2000b). Thus, statistically speaking, doubts can be thrown over all studies which do not use an objective information criterion for selecting variables and/or lags. While it could be argued that, e.g., PM_{10} acute health effect generally appear to operate at lag 1 day) the analyses of Clyde (2000) and Clyde *et al.* (2000) find PM_{10} lag 1 as a strong predictor in most of the models ranked highest be either AIC or BIC, using a Bayesian model averaging procedure), the literature, perhaps due to the variable selection practice mentioned above, does not show substantial agreement as to which lag(s) to use.

While there are several model selection criteria (such as C_p, BIC, AIC, Bayes factors etc.), and no consensus within the statistical community regarding which criterion to use, there is agreement among statisticians that stepwise methods have serious drawbacks in terms of bias. In particular, when the estimated risk effects are very small, the epidemiological selection principle not only leads to bias in the estimates, but also to a false sense of scientific consensus, in that the estimates from models so selected will tend to be more similar than what is actually warranted by the data.

The advantage with the Bayesian model averaging procedure, as used by Clyde (2000) and Clyde *et al.* (2000), is that several models that are well supported by the data are considered simultaneously, rather than selecting a single *best* model. The standard error of relative risk estimates obtained in this fashion reflect the model selection procedure, while methods selecting a single model tend to ignore the selection, and calculate standard errors as if only the chosen model had been considered. Another feature of Bayesian model averaging is that it is straightforward to incorporate prior beliefs about important lags and variables in the analysis (although Clyde and co-workers have tended to weight all possible models equally). If the data disagree with the prior beliefs, this comes out of the analysis. A drawback, on the other hand, is that the methodology is somewhat sensitive to which criterion is used to rank the models (see Clyde, 2000).

4.4 Non-linear dose-response relations and thresholds

One of the critical questions associated with the translation of epidemiological studies into particulate matter standards is whether the dose-response relationship is linear and whether there is any evidence of a threshold, i.e. a critical level below which there is little or no effect of increasing air pollution on health.

Many of the early studies of PM-health relations were based on levels of PM much higher than those typically observed today. For example, in the notorious "London smog" of December 1952, in which there are estimated to have been 4,000 excess deaths as a result of air pollution, smoke levels reached as high as 3,000 μg/m^3. Schwartz and Marcus (1990) reported that mean smoke levels in London declined from about 500 μg/m^3 to about 60 μg/m^3 over the period 1958–1971, with corresponding decreases in the death rate. (The measure of air pollution used in these studies was "British smoke", which as an extremely rough guide is typically about twice the PM$_{10}$ level.) Schwartz and Marcus were possibly the first authors to claim that health effects actually persisted to the very lowest levels of PM, though their claims were quickly followed by a number of others — Pope (2000) has provided further historical perspective.

The issue, as it appears to us, is not whether very high levels of pollution are responsible for mortality effects — this seems to be established beyond reasonable doubt — but whether the effects really do persist to a level below that of the current PM$_{10}$ standard, which would justify a tighter standard. This requires critical examination of the shape of the dose-response curve across a wide range of dose levels, but particularly at those near to or below the current standard. It is not sufficient to argue that the relationship must be linear unless proved otherwise, though this reasoning is implicit in any hypothesis test that takes a linear relationship as the null hypothesis.

Our own attempts to examine this issue have produced confusing results. Smith *et al.* (2000a) examined both fine and coarse particle effects in Phoenix, concluding that there is a threshold (in the region of 20–25 μg/m^3) for fine particles, but not for coarse particles. (As a side comment on a statement made in the CD, page 6-247 remarks on the fact that the fitted nonlinear relationship for fine particles is roughly V-shaped, being a decreasing function of PM at low PM levels, and questions whether this is biologically plausible. We agree that it is probably not, but we did not claim that the negative slope is statistically significantly different from 0 — the confidence bands drawn in the paper show that it is not. On the other hand, we did quote p-values lower than 0.01 against the hypothesis of an overall linear effect.) Smith *et al.* (2000b) claimed evidence for an increasing slope in the dose-response relationship, and possibly for the existence of a PM$_{10}$ threshold at a level above 50 μg/m^3, in data from Birmingham. On the other hand, the same methods applied to data from Chicago (Smith *et al.* 1999) showed no evidence of a threshold and even a sharply increasing effect in the range 0–20 μg/m^3, a result which is also of questionable biological validity. More broadly, the CD openly acknowledges the difficulty in identifying thresholds, for example on page 6-9, and recognizes that even if thresholds do exist on an individual level, such effects may be masked when aggregated over the population (page 6-246).

Against this background, it is to be welcomed that there are some recent studies, notably Schwartz and Zanobetti (2000) and Daniels *et al.* (2000), that have sought to resolve the issue by combining data from a number of cities. However, we question whether the analyses have yet been taken far enough to establish anything conclusive.

For example, Daniels *et al.* (2000) took the same 20-cities data as in the NMMAPS study, and fitted a log-Poisson regression model including all the usual covariates (current day and 3-day averaged lagged days for temperature and dewpoint, both modeled via cubic splines, long-term trends also modeled by cubic splines, plus day of week and age-group effects), initially treating PM$_{10}$ (a) as a linear effect, but then modifying it to (b) modeling PM$_{10}$ nonlinearly using cubic splines, with fixed knots at 30 and 60 μg/m^3, (c) a threshold model. This analysis was initially conducted on a city-by-city basis, but then combined across cities by a hierarchical models analysis. The form of hierarchical model was to assume that the parameter estimates of interest for city c, $\hat{\phi}_c$ say, are distributed according to $\hat{\phi}_c \sim N[\phi_c, V_c]$ where ϕ_c are the random effects for city c and V_c is a covariance matrix for the estimates at city c (one presumes — in the paper, V_c is not actually defined), while the random effects ϕ_c are drawn independently from $N[\phi, D]$, ϕ and D having flat prior distributions. This defines a hierarchical structure from which one can draw posterior distributions by Gibbs sampling, though many other hierarchical structures are possible, if different assumptions are made for the inter-city effects. This scheme was used for the linear and spline models for the PM-mortality relationship; for the threshold model, noting the difficulty of estimating thresholds in individual cities, the authors did not attempt any hierarchical approach but simply combined the log likelihood across cities, implicitly assuming independence from city to city. Throughout the paper, beyond the direct comparison among approaches (a)–(c), there is no attempt to study the robustness of the conclusions against alternative model specifications, and the justification for the hierarchical model assumptions is not clearly made at all. Given the emphasis made on regional and spatial analyses in earlier analyses of the NMMAPS data, one would have expected some consideration of similar issues here.

Despite these criticisms, the analysis by Daniels *et al.* was a good first stab at the problem. However, as things currently stand the analysis is incomplete, and we anticipate that it will take a number of alternative analyses of the same or similar data before any definitive conclusions can be drawn. This is only to be expected, given the complexity of the issues involved and the number of alternative approaches that are potentially available for estimating non-linear dose-response effects simultaneously in a large number of cities. The analysis by Schwartz and Zanobetti (2000) also used nonlinear dose-response functions and combined data across cities via a meta-analysis approach, but this raises similar issues regarding the sensitivity of the analysis to alternative methods of statistical analysis, especially, alternative approaches to the meta-analysis. The narrowness of the confidence bands in Fig. 2 of Schwartz and Zanobetti, when compared with Fig. 3 of Daniels *et al.*, does lead us to question whether the Schwartz-Zanobetti approach is adequately allowing for inter-city variation.

The results of both papers imply that there is no strong evidence against a linear relationship, at least for all-cause, cardiovascular and respiratory mortality (Daniels *et al.* do suggest the existence of a threshold if cardiovascular and respiratory deaths are excluded), but we do not see these two studies as resolving these very complex issues. Our criticism of the CD (specifically, the section between pages 6-245 and 6-248) is that is has focussed exclusively on the results of these papers and has not paid any attention to the methodology of the analysis. However, without appreciating the methodology that was used, and its strengths and limitations, we do not think it is possible to form an overall scientific judgement of the results. At the very least, the CD should have highlighted the need for more work on these issues.

4.5 Co-Pollutants

The issue of whether co-pollutants need to be included in health effects analyses, or if the analysis becomes cleaner when only one pollutant at the time is included in the analysis, is subject to substantial, and in our view very confused, discussion in the Criteria Document. It appears that the authors are arguing that since many different co-pollutants tend to be correlated, the uncertainty of the health effects estimates tend to cloud the conclusions. This is only the case if one assumes a priori that particulate matter must have an effect on health independently of other pollutants. This, however, is what the analyses discussed in the document are attempting to investigate, and the conclusions are far from clear-cut.

In the NMMAPS report, the effects of PM_{10} are, generally, not much changed if the gaseous co-pollutants (O_3, SO_2, NO_2 and CO) are included as additional covariates in the models (see, in particular, Fig. 25 on page 27 of Part II). On the other hand, comparisons of PM_{10} for the primary pollutant, with each of the others as a primary pollutant, still does not show clear evidence that PM_{10} is the primary "culprit" as far as pollution-mortality effects are concerned. O_3 effects in summer, and each of the other gases overall, are statistically significant, or very nearly so, in at least one of the analyses reported (Figs. 26–29, pp. 27–28). Note that these analyses are based on the 20 cities, not the 90 cities. It would be of interest to see them repeated for the full 90 cities.

In many US cities ozone is only measured during the "ozone season", which generally does not include the winter (when particulate matter due to wood smoke is prevalent, especially in Western US). This adds substantially to the difficulty of separating out the effects of different pollutants.

The epidemiological evidence of the severity of fine particle health effects is simply not yet available: there is insufficient availability of $PM_{2.5}$ data to draw any firm conclusions. There are several studies in which the PM effects disappear when other pollutants are included in the model. There are also several studies with the opposite result. In our opinion, the most severe problem is that we do not yet have a firm grip on the composition of particulate matter in different parts of the United States. The criteria document authors seem to expect that health effects of particulate matter is a matter only of the size of the particles; not of the chemical composition of the particles. The variety of results with respect to co-pollutants can perhaps be caused by the variety of chemical compositions; this is certainly a likely explanation of the regional variability found in the 90-cities study.

5 Cohort studies

The claim that particulate matter causes long-term effects as well as short-term effects relies almost entirely on three prospective cohort studies, the Harvard Six-Cities Study (HSC — Dockery *et al.* (1993)), the American Cancer Society Study (ACS — Pope *et al.* (1995)) and the Adventist Health Smog Study (AHSMOG — Abbey *et al.* (1999)). The HSC and ACS studies were included in the 1996 PM Criteria Document and, as noted on pages 6-81 and 6-82 of the current draft CD, raised a number of questions — the four specifically listed there are (1) whether important confounding variables have been omitted, (2) the influence of other atmospheric pollutants besides PM, (3) the evaluation of time scales for long-term exposure effects, and (4) the existence of pollution thresholds.

The most significant new study published since the 1996 PMCD is a major re-analysis of the HSC and ACS studies sponsored by the Health Effects Institute (Krewski *et al.* (2000)). As a result of these re-analyses, the draft PMCD reports (p. 6-82) that "considerable progress has been made towards addressing further the above issues" and, while admitting that the results of the AHSMOG study have been less decisive, concludes that (p. 6-94) "there is evidence for an association between long-term exposure to PM (especially fine particles) and mortality". The further summary on chapter 9 (especially, Section 9.6.2.2, page 9-64) concludes that "One of the most important advances since the 1996 PMCD is the substantial verification and extension of the findings" (of the original HSC and ACS studies).

While we acknowledge that the HEI re-analysis was a very important study that added considerable depth and breadth to the original studies of Dockery *et al.* (1993) and Pope *et al.* (1995), we strongly dispute the implication, evident in the above quotes, that it has cleared up all the problems associated with the earlier studies. The re-analyses identified numerous methodological issues whose resolution is very far from clear at the present time.

We have no dispute with Part I of the HEI re-analysis, which was concerned with an audit of the data sources and verification that the original statistical analyses, as reported by the original authors, would indeed produce the results cited in the original papers. This part of the study was well executed and indeed helped to clarify a number of issues about exactly how the original authors performed their analyses. The comments below all refer to Part II of the re-analysis, which was called a "sensitivity analysis" but in reality went well beyond mere checking of the sensitivity of the results to certain assumptions in the original analyses, being a wholescale re-examination of the methodology that lay behind the study.

5.1 The ecological nature of the studies

The draft PMCD (page 6-2) cites Rothman and Greenland (1998) as classifying four common types of epidemiological study in order of increasing inferential strength, with "ecologic studies" as number 1 (lowest strength of inference), followed by 2. time series studies, 3. longitudinal panel and prospective cohort studies, and 4. case-control studies. The implication is that the cohort studies lie higher up the inferential food chain than the time series studies, and form a good basis for causal inference, though it is admitted that "the use of community-level or estimated exposure data may weaken this advantage, as in time-series studies".

In fact, we would argue that the three studies referred to are *primarily* ecologic studies. They would be convincing if they succeeded in correlating variations in mortality with variations in air pollution exposure *within* a community. But the comparisons they make are *between* communities. Taking HSC as an example, the original analysis employed a Cox proportional hazards analysis using various individual-level covariates (age, sex, smoking history, body-mass index and education level) to compute adjusted mortality rates for each city, and then (Dockery *et al.* (1993), page 1757 recomputed as Krewski *et al.* (2000), p. 76) plotted the resulting mortality rate ratios against mean levels of several air pollutants. For example, Portage, Wisconsin had the lowest adjusted death rate amongst the six cities and Steubenville, Ohio, had the highest; it was also the case that Portage had the lowest and Steubenville the highest of both fine and total particles (with several other pollutants showing a similar pattern).

A "pure" ecologic study would be one which compared the average mortality rates to the average pollutant levels without any adjustment for individual-level covariates. That would have the obvious flaw that differences observed among the six cities could be due to different distributions of those covariates (for example, more smokers in Steubenville than Portage) rather than air pollution effects. Certainly, the HSC study was better than that. But the assertion of a *causal* relationship between air pollution and long-term mortality rates amounts to the statement that there cannot be any other possible cause for these differences. This we would dispute. All the reported relative risks due to air pollution, derived from the HSC study, are based on regression on precisely these six data points.

The re-analysis tried to test these conclusions by including some other covariates in the analysis, such as occupational type and an indicator of population mobility. It was the case, for example, that among the six cities, Steubenville had the highest proportion of the population working in "dirty" occupations. Despite this, associations between total mortality and pollution still remained under the re-analyses, though they did find one curious fact, that the association does not seem to be present among the segment of the population with a post-high-school education (this fact is noted and highlighted in the draft PMCD).

Nevertheless, the fact that the relationship between standardized death rates and mortality was not destroyed by the inclusion of a small number of specific alternative covariates does not mean that the original conclusions have been proved correct. The weak inferential basis for making causal assertions in this study remains, however many alternative covariates are tested.

5.2 The ACS study

Although the ACS study was not as carefully carried out as the HSC study (for example, the participants were largely volunteers rather than selected by randomization), it involved many more participants (552,138 as against 8,111) and many more cities (a total of 154). (Just as a point for comparison, the AHSMOG study also involved a relatively small sample size, 6,338 subjects, and this fact may be responsible for the inconclusive results of that study.) Although the general points about the ecological nature of the study are just as true of ACS as they are of HSC, with the much larger number of cities, there are many more possibilities for alternative modeling of the inter-city data. Indeed, we would regard the innovations made by the re-analysis team in this respect as one of the major contributions of the entire study.

In the rest of this subsection, we comment on three of these which may all be thought of as addressing, in different ways, the issue of ecological effects, (i) ecologic covariates, (ii) random effects models, and (iii) spatial analyses.

5.2.1 Ecologic covariates

As an attempt to evaluate whether inter-city differences in mortality rates could be due to other city-level variates than particulate matter pollution, the re-analysis developed a suite of 30 "ecologic covariates" (20 of which were actually used in the analysis). These included demographic and socioeconomic variables (e.g. percentage of whites and blacks, poverty level, mean income), climate and physical environment variables (e.g. mean altitude, mean temperature) and health service indicators (number of physicians, number of hospital beds). They also included alternative air pollution indicators (CO, NO_2, O_3, SO_2). With the ecologic covariates introduced one at a time into the analysis, only two had a substantial impact on the coefficient due to sulfate particles in the total-mortality analysis. These were population mobility and SO_2 (Table 34, p. 180, of Krewski *et al.* (2000)). Moreover, when SO_2 was treated as the primary covariate, the relative risk due to SO_2 was higher than that due to sulfate particles, and unaffected if sulfate particles were also included in the analysis. Other results largely confirmed the same pattern. The re-analysis team also conducted rather limited analyses using multiple ecologic covariates.

The idea of incorporating ecological covariates is evidently a controversial part of the work. The original authors of the ACS study, commenting at the end of the Krewski *et al.* report (page 275), remarked "From the very beginning of the reamalysis, we were opposed to the idea of taking a myriad of ecologic variables and including them as covariates in the models...[In the ACS study] we considered the original work to be a straightforward, clean, elegant way to generate and test a specific well-defined hypothesis".

Of course, it is perfectly true that introducing a very large number of irrelevant covariates has the potential to weaken a genuine effect which is present in the data. But the passage just quoted seems to be denying the possibility of ecological effects. Although imperfect, we believe the ecological covariates analysis was an important part of the re-analysis and, to some extent, was successful in demonstrating that a number of plausible ecological covariates could not in fact explain the differences in mortality rates. We feel that they could have tried multiple regression analyses to a greater extent than they did, recognizing that the differences in mortality could be due to combinations of ecological factors rather than any one factor operating on its own. Another possible extension of the analysis would be to allow interactions among different ecological covariates. The interpretation of the two variables that were significant remains open to dispute. Population mobility may be related to educational status, and it was observed earlier that the PM effect does not seem to be present among those if high educational attainment. The findings about SO_2 relate to the whole issue of co-pollutants, to which we return later.

5.2.2 Random effects models

Even if it were correct that the differences in air pollution were the major factor explaining differences among mortality rates in the different cities, it would be scarcely credible that air pollution could be the only effect. Even after adjusting for air pollution, one would expect to see differences among the cities that go beyond individual-level variation. One of the methodological contributions of the re-analysis was an analysis that allowed for random effects in cities to explain other sources of variation. The results (Table 50, p. 213) showed no great sensitivity in the point estimates and confidence intervals when the random effects model was included. However, it was noted by the Review Panel that the estimated values of τ, the standard deviation of the city random effect, was comparable with the uncertainty in the estimated PM effect, and this in turn could complicate the interpretation of the PM effect. "If a large component of the variance is unexplained in the data, a model including sufficient variables to identify this residual variation might produce different regression coefficients for the variable of interest" (Krewski *et al.* (2000), page 259).

5.2.3 Spatial analyses

Going beyond a simple random effects model, a major finding of the HEI re-analysis was that there seems to be substantial spatial correlation among both air pollution and adjusted mortality rates. Although this could be a separate issue from that of "ecological bias", they are connected in that spatial correlation implies there are other sources of inter-city variation than pure variation in the level of air pollution. In a sense, the spatial correlation that remains after known covariates are taken into account can be regarded as representing additional variability due to unknown covariates.

Spatial correlation was detected by drawing maps, by formal tests of spatial correlation (Moran's I and G tests), and by performing regression analyses that adjusted for spatial correlation. Since our main concern in this commentary is the possible effect of spatial correlation on the conclusions about particulates and mortality, we concentrate on the third of these issues. Within the framework of spatially adjusted regression analyses, three kinds of analysis were carried out that tried to allow for spatial correlation in the regression. The first of these was based on a simple regional classification with random effects due to region. This analysis, while certainly a good first start on the problem, cannot be expected to adjust for all the effects of spatial correlation. The second analysis was based on first passing the data through a spatial filter designed to achieve approximate uncorrelatedness, and then regressing the filtered data. Although it seems promising, this method has uncertain properties —for example, the definition of the spatial filter is *ad hoc* and, even if it were derived from a specific spatial autocorrelation function, no allowance is made for the effects of estimating the autocorrelation structure.

The third analysis that attempts to adjust the regression for the effects of spatial correlation is one based on a specific spatial model. The model selected by the authors of the study was the SAR (simultaneous autoregressive) model, in which the map of the US was covered by so-called Thiessen polygons, one city in each polygon, and two cities considered to be "neighbors" if their polygons touched each other. The dependence between neighboring cities is represented by a correlation parameter ρ. There are various reasons why this model is not especially suitable for the kind of spatial dependence being studied. The tiling of the map does not correspond to any physical model of the spatial variation, and has some counterintuitive properties, e.g. if new cities were added to the study the Thiessen polygons and hence the assumed correlations would change, but it seems implausible that the correlation between two cities in the study would change according to which additional cities were also included. The model is also oversimplified in that a single parameter ρ is assumed to characterize the spatial dependence across the entire country. In our view, it would be more appropriate to use a continuous random field model of the kind common in geostatistics and environmetrics, and the authors might also have explored the possibility of nonstationarity in the spatial dependence structure.

In spite of the incomplete nature of the spatial analysis, it did have a significant impact on the results. For example, in an analysis including both sulfate particles and SO_2 (Krewski *et al.* (2000), pp. 210–211), the RR for sulfate dropped from 1.20 to 1.08 (95% CI: 0.91 to 1.28) though that for SO_2 was less affected (RR from 1.35 to 1.31; CI 1.12 to 1.50). If such a substantial change is possible through only a one-parameter addition to the model, it can only be speculated what would happen with more realistic spatial models.

5.3 Threshold effects

As noted in our introduction to this section, one of the issues identified in the 1996 CD as an issue needing clarification (in connection with cohort studies) was that of threshold effects. Unfortunately, the re-analysis shed little light on that issue.

Although there may not be evidence for a strict "threshold" in PM-mortality studies, there may well be an nonlinear dose-response effect, with the incremental effect due to PM changes being higher at high levels of PM than at low levels. The re-analysis investigated this issue at two different points.

First, Fig. 6 of the report (Krewski *et al.* (2000), page 162) shows standardized residuals of mortality against either sulfate concentration or fine particles, for all-cause mortality, for cardiopulmonary mortality, and for lung cancer mortality. All six figures are of similar shape. A widely dispersed scatterplot has been smoothed using cubic splines, and the resulting smoothed curve superimposed on the scatterplot with confidence bands. The shapes of the fitted curves vary, and e.g. for the dependence of all-cause mortality and cardiopulmonary mortality on fine particles actually show a higher slope at low PM levels (10–15 μg/m^3) than higher. However, in all six plots the width of the confidence bands, relative to the total variation in mortality rate, makes it hard to give any definitive conclusion about whether the overall relationship is linear or not.

In contrast, Figs. 10 and 11 on page 175, in which the ordinate is a log hazard ratio but otherwise supposedly conveying the same information as Fig. 6, gives a very different impression — a much more definitive shape to the curve which, in the case of $PM_{2.5}$, shows a statistically significant decrease between about 16 and 21 μg/m^3. We are extremely puzzled about this.

5.4 Co-Pollutants

As noted in our discussion of ecological covariates, SO_2 showed up as a significant variable (though not other atmospheric pollutants that were also considered). Consistently throughout the re-analysis study, when SO_2 and particulates were treated on equal footing, SO_2 came out showing a stronger effect than particles. There is some controversy about the interpretation of this (page 6-86 of the draft CD) because the sulfate measurements were complicated by an artifactual component which did appear to influence the results. SO_2 is generally a precursor to sulfate particles and it is quite possible that while it is the particles that have the health effect, it is the SO_2 that is more easily detected and measured, thus creating an apparently stronger effect for SO_2 than for sulfate. However, this is hypothetical: the exclusive focus on particulate matter as a pollutant of interest does not seem to us to be justified by the current epidemiological analyses.

5.5 Are they really measuring long-term effects anyway?

Throughout the discussion of time series and cohort studies, the impression created is "time series studies prove there are short-term effects and cohort studies prove that there are long-term effects". Evidently, "acute effects" deaths are also being included in the cohort studies, and there is no direct way to separate the two.

The draft CD (page 9-61) cites the 1996 CD that "PM effect size estimates for total mortality indicate that a substantial portion of the deaths reflected cumulative PM impacts above and beyond those exerted by acute exposure events", and goes on to report the "substantial verification and extension" of these findings by the re-analysis. In other words, measure the RR for acute effects using time-series studies, and that for acute and chronic effects combined using cohort studies, and if a significant difference exists, it must be due to chronic effects.

However, even when considering the time series analyses and the cohort analyses separately, there exist substantial difference from one analysis to another due to model selection, and moreover, the estimated RRs, even if statistically significantly greater than 1, always have wide confidence intervals associated with them. To read much interpretation into differences in RR levels from completely different types of analysis does not seem justified.

The re-analysis addressed part of this issue, in the case of the HSC study, by including PM as a time-dependent covariate. For ACS, the study was based on a single PM measurement (from 1982) for each city, and the difficulty with interpreting the actual numerical value of a RR is that the 1982 level may be completely unrepresentative of historical levels of PM, beyond some loose expectation that the most polluted cities in 1982 were probably also the most polluted cities in earlier years. This difficulty has been acknowledged both by the original authors and in the reanalysis. For HSC, some examination of this issue is possible because the study authors did have available historical records of PM (though not as far back as one would like to perform a genuine "lifetime exposures" analysis). When this is included in the model, the effect is considerable: comparing model 5 (treating PM as constant) with model 6 (time dependent) in Table 14 of Krewski *et al.* (2000), the estimated RR drops from 1.31 (95% CI: 1.13-1.52) to 1.16 (1.02-1.32), in effect, a halving of the estimated effect. This kind of sensitivity, to how the historical PM variable is treated, underlines the extreme difficulty of separating short-term and long-term effects in this kind of analysis.

5.6 Summary of cohort re-analysis

It is not our purpose here to criticize the re-analysis itself, which accomplished an enormous amount under intense time pressure. Virtually all the points mentioned here were brought up either by the re-analysis team themselves, or in the HEI Review Panel commentary. The draft CD seems to have concluded that the HEI re-analysis ended up confirming all the major claims that were made in the original HSC and ACS analyses. However, careful reading of the re-analysis shows that there are in fact numerous very important issues of methodology and interpretation, to which the re-analysis certainly made significant contributions, but which cannot be considered resolved at the present time. They may never be.

6 Conclusions

One of our concerns about the PM Criteria Document is its failure to cover all relevant literature, and in Section 2 of the present discussion, we have attempted to fill in one of the omissions, concerning the special issue of *Environmetrics* on statistical analysis of particulate matter air pollution. More broadly, however, we feel that the CD has not succeeded in adequately conveying the statistical and scientific interpretation issues that are involved in drawing conclusions from such complex issues from large epidemiological data sets. The bulk of this review has concentrated on two such kinds of studies: the time series analyses of PM exposure and mortality in Section 4, and the cohort studies in Section 5. Both kinds of studies raise a number of common issues: how to combine data from a large number of cities, especially when there is spatial dependence; the possibility of threshold effects or, more generally, a nonlinear dose-response curve; and the issue of co-pollutants. There are also some specific issues for each kind of study. The time series studies appear to have been substantially affected by publication bias, at least on the basis of our comparison of results from the NMMAPS study (which we presume to be free of publication bias) with other published analyses in the literature. There is also the question of model selection bias, which has not received anywhere near adequate treatment in the epidemiological literature. Bayesian model averaging is a relatively new technique developed by statisticians, and while not free of potentially problematic assumptions of its own, does offer a possible route to deriving results without this kind of bias. On the side of the cohort studies, we feel that the very broad issues raised by the ecological nature of the studies still needs further discussion, though we recognize that the HEI-sponsored re-analysis introduced a number of important new methodological developments.

7 References

Abbey, D.E., Nishino, N., McDonnell, W.F., Burchette, R.J., Knutsen, S.F., Beeson, W.L. and Yang, J.X. (1999), Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am. J. Respir. Crit. Care Med.* **159**, 373–382.

Cox, L.H. (2000), Statistical issues in the study of pollution involving airborne particulate matter. *Environmetrics* **11**, 611–626.

Clyde, M. (2000), Model uncertainty and health effect studies for particulate matter. *Environmetrics* **11**, 745–763.

Clyde, M., Guttorp, P. and Sullivan, E. (2000), Effects of ambient fine and coarse particles on mortality in Phoenix, Arizona. *J. Exposure Anal. Environ.*, submitted.

Daniels, M.J., Dominici, F., Samet, J.M. and 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. Epidemiology* **152**, 397–406.

Pope, C.A. (2000), Invited commentary: Particulate matter-mortality exposure-response relations and threshold. *Am. J. Epidemiology* **152**, 407–412.

Dewanji, A. and Moolgavkar, S.H. (2000), A Poisson process model for recurrent event data with environmental covariates. *Environmetrics* **11**, 665–673.

Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G. and Speizer, F.E. (1993), An association between air pollution and mortality in six U.S. cities. *N. Engl. J. Med.* **329**, 1753–1759.

Dominici, F., Samet, J.M and Zeger, S.L. (2000), Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy (with discussion). *J.R. Statist. Soc. A* **163**, 263–302.

Environmental Protection Agency (2001), *Air Quality Criteria for Particulate Matter, Vols. I and II.* Second external review draft, Office of Research and Development, United States Environmental Protection Agency, Washington, D.C.

Krewski, D., Burnett, R.T., Goldberg, M.S., Hoover, K., Siemiatycki, J., Jerrett, M., Abrahamowicz, M. and White, W.H. (2000), *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality.* A Special Report of the Institute's Particulate Epidmiology Reanalysis Project. Health Effects Institute, Cambridge, MA.

Lumley, T. and Levy, D. (2000), Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics* **11**, 689–704.

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

Moolgavkar, S.H., Hazelton, W.D., Luebeck, E.G., Levy, D. and Sheppard, L. (2000), Air pollution, pollens and respiratory admiossions for chronic obstructive pulmonary disease in King County. *Inhalation Toxicology* **12** (suppl. 1), 157–171.

Phelan, M.J. (2000), Timing and scope of emissions reductions for airborne particulate matter: a simplified model. *Environmetrics* **11**, 627–649.

Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E. and Heath, C.W. (1995), Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am. J. Respir. Crit. Care Med.* **151**, 669–674.

Rothman, K.J. and Greenland, S., eds. (1998), *Modern Epidemiology.* Second edition. Lippincott-Raven, Philadelphia.

Samet, J.M., Dominici, F., Zeger, S.L., Schwartz, J. and Dockery, D.W. (2000a), National morbidity, mortality and air pollution study. Part I: methods and methodologic issues. Research Report 94, Health Effects Institute, Cambridge, MA.

Samet, J.M., Zeger, S.L., Dominici, F., Curriero, F., Coursac, I, Dockery, D.W., Schwartz, J. and Zanobetti, A. (2000b), National morbidity, mortality and air pollution study. Part II: morbidity, mortality and air pollution in the United States. Research Report 94, Health Effects Institute, Cambridge, MA.

Schwartz, J. (1993), Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology* **137**, 1136–1147.

Schwartz, J. and Marcus, A. (1990), Mortality and air pollution in London: a time series analysis. *Am. J. Epidemiology* **131**, 185–194.

Schwartz, J. and 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.

Sheppard, L. and Damian, D. (2000), Estimating short-term PM effects accounting for surrogate exposure measurements. *Environmetrics* **11**, 675–687.

Sheppard, L., Levy, D., Norris, G., Larson, T.V. and Koenig, J.Q. (1999), Effects of ambient air pollution on nonelederly asthma hospital admissions in Seattle, Washington, 1987–1994. *Epidemiiology* **10**, 23–30.

Smith, R.L., Davis, J.M. and Speckman, P. (1999), Human health effects of environmental pollution in the atmosphere. Chapter 6 of *Statistics in the Environment 4: Statistical Aspects of Health and the Environment*, edited by V. Barnett, A. Stein and F. Turkman. John Wiley, Chichester, 91–115.

Smith, R.L., Spitzner, D., Kim, Y. and Fuentes, M. (2000a), Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. *Journal of the Air and Waste Management Association* **50**, 1367–1379.

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

Sun, L., Zidek, J.V., Le, N.D. and Özkaynak, H. (2000), Interpolating Vancouver's daily ambient PM_{10} field. *Environmetrics* **11**, 651–663.