HUMAN HEALTH EFFECTS OF ENVIRONMENTAL POLLUTION IN THE ATMOSPHERE

by

Richard L. Smith
Jerry M. Davis
Paul Speckman

DRAFT: JANUARY 2 1998

To appear in *Statistics for the Environment 4: Health Effects*, edited by V. Barnett, F. Turkman and A. Stein (1998). John Wiley: Chichester.

HUMAN HEALTH EFFECTS OF ENVIRONMENTAL POLLUTION IN THE ATMOSPHERE

Richard L. Smith [1], **Jerry M. Davis** [2] **and Paul Speckman** [3]

Summary. In recent years, much attention has been given to the human health effects of atmospheric pollutants, especially particulate matter. This has been the focus of particularly heated debate in the USA, as new regulations introduced in 1997 by the United States Environmental Protection Agency (USEPA) have considerably tightened the existing standard. Similar regulations are also being considered by several countries of the European Union. Much of the debate revolves around claims that particulate matter in the atmosphere directly influences mortality, hospital admissions with respiratory diseases, and so on. In this chapter, we take a critical look at one of these issues, the influence of PM_{10} (particulate matter of aerodynamic diameter no more than 10 μm) on deaths in the elderly population. Two data sets are considered, one from Birmingham, Alabama, and the other from Chicago. In both cases we find a significant PM_{10}–mortality relationship in some of the models fitted, but not in others. Other issues considered include the existence of a threshold below which PM_{10} has no discernable influence, the interaction with other pollutants, and the mortality displacement or harvesting effect (the theory that the direct effect of PM_{10} is limited to a very small subset of the population who are already critically ill and whose death is only advanced by a few hours or days as a result of air pollution). For the latter phenomenon, a compartment-type model is introduced and analyzed using a Markov chain Monte Carlo procedure. The results show that even when all these alternative effects are considered, there remains a considerable amount of unexplained association between particulates and mortality, but there appear to be too many uncertain issues to allow us to make definitive statements about a causal relationship.

Keywords: Bayesian inference, Generalized additive model, Harvesting phenomenon, Influence of other atmospheric pollutants, Linear regression, Markov chain Monte Carlo algorithms, Mortality displacement, Non-linear relationships, PM_{10}, Poisson regression, Variable selection.

---

[1] Address for correspondence: Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, USA. Email: rs@stat.unc.edu. This research was primarily conducted at the (US) National Institute of Statistical Sciences, Research Triangle Park, NC, supported in part by the United States Environmental Protection Agency (USEPA) under Cooperative Agreement #CR819638-01-0. The work has not been internally reviewed by the USEPA and no endorsement by that agency should be inferred. We thank Jerome Sacks for numerous conversations and insights. This paper is a revised and extended version of one presented at the 1997 Joint Statistical Meetings of the American Statistical Association (Smith *et al.* 1997a).

[2] Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695-8208, USA.

[3] Department of Statistics, University of Missouri, Columbia, MO 65211, USA.

1 INTRODUCTION

In recent years there has been increasing concern over the effects of air pollution on human health. The debate has been particularly intense in the USA, where in July 1997 the United States Environmental Protection Agency (USEPA) introduced new standards for ozone and particulate matter, citing extensive published research on human health effects as the reason why a tightened standard was necessary. In the case of ozone, a previous standard based on the maximum hourly value for each day was replaced by one based on the eight-hour average ozone level, but with a sharply reduced threshold (the old standard was 120 ppb for hourly ozone; the new standard is 80 ppb for the eight-hour average). In the case of particulate matter, the previous standard was based on PM_{10} (particulate matter of no more than 10 μm in aerodynamic diameter) and established an acceptable annual average level of 50 $\mu g/m^3$ and a daily maximum of 150 $\mu g/m^3$. The new standard is based on $PM_{2.5}$ (particulate matter of no more than 2.5 μm in diameter) and permits an annual average level of 15 $\mu g/m^3$ and a daily maximum of 65 $\mu g/m^3$. The annual average standard has been particularly contentious because many US cities are currently above that level and industry groups have argued that it would be extremely costly to implement the new regulations. Health care groups, however, have argued that epidemiological studies, based largely on current EPA monitoring of PM_{10}, show a strong influence of particulate matter on a variety of adverse health outcomes, including deaths in the elderly population, increased hospital admissions due to asthma and other respiratory complaints, and increased infant mortality. Moreover, it is believed that the very small particles are primarily responsible for these effects, hence the switch to a standard based on $PM_{2.5}$. At the present time, the new regulations are in force but penalties for non-compliance will be delayed for five years to allow time for additional scientific research. The issue has also come under close scrutiny in Great Britain, The Netherlands and several other European countries as similar regulation is considered there.

In this chapter, one of the key issues in this debate will be reviewed, the influence of PM_{10} on mortality in the elderly population. A number of authors have analyzed data on this, and have found a statistically significant relationship. Based on data from Birmingham (Alabama) and Chicago, we find that the relationships are highly model-dependent. We also raise questions over the the existence of a threshold below which PM_{10} has no discernable influence, the interaction with other pollutants, and the mortality displacement or harvesting effect (the theory that the direct effect of PM_{10} is limited to a very small subset of the population who are already critically ill and whose death is only advanced by a few hours or days as a result of air pollution). The latter question is examined by trying to reconstruct the size of the critically ill population through a Bayesian analysis. Overall, we believe that there are too many uncertain issues to allow us to make definitive statements about a causal relationship between PM_{10} and mortality.

2 METHODOLOGY

One of the first papers to consider the particulates-mortality relationship carefully was Schwartz and Marcus (1990), who were concerned primarily with data from London

in the 1960s. In that paper, the authors addressed a number of the difficulties involved in inferring a causal relationship from the available data. Among these are

- the effect of autocorrelation,
- the influence of long-term trends,
- the possible existence of a threshold level of particulates, below which there is no observable effect,
- whether the particulates effect is confounded with the weather,
- whether the effect due to particulates can be separated from that due to other forms of air pollution.

In the case of the London data, there was a tenfold drop in particulate levels over the period of the study, and there seems little doubt that this had direct benefits on human health, but the interpretation in the case of modern US cities is much less clear cut.

The main results in the present paper are based on classical linear regression models of the form

$$y_t = \sum \beta_j x_{jt} + \epsilon_t \tag{1}$$

where y_t is some transformation (logarithm or square root) of the daily death count on day t, $\{x_{jt}\}$ are covariates representing the long-term trend, meteorology and air pollution, and $\{\epsilon_t\}$ are treated as independent $N(0, \sigma^2)$ variables with unknown σ^2. In other studies, either maximum likelihood Poisson regression or variants of the generalized estimating equations technique due to Liang and Zeger (1986) have been adopted, but the alternative estimation methods make little difference to the kinds of models fitted or the conclusions obtained from the studies. Questions of overdispersion (with respect to a Poisson model) and autocorrelation among the residuals have also been considered, but in the data sets analyzed here these do not appear to be important features, provided one adequately models the long-term trend. Much more critical issues, in our view, concern how to define the covariates in (1).

First, we consider the long-term trend. All the data sets exhibit significant trends well beyond anything that can be explained in terms of meteorology and air pollution. For example, Fig. 1 shows weekly death counts for six years in Chicago, together with a trend fitted both by a simple `loess` smoother in S-PLUS (solid curve) and by the B-spline method to be described below (dashed curve). It is clear that there is a strong seasonal variation, but that it is irregular — the peak occurs at a different time each year and is of much greater height some years than others. This has been modeled by treating the trend as a single continuous curve represented as a linear combination of B-spline basis functions (Green and Silverman, 1994). In the analysis of Chicago which follows, the eventual analysis was based on 56 months' data and this was modeled by a B-spline

representation with 55 degrees of freedom. For Birmingham, we used 20 degrees of freedom for 41 months of data.

The next issue to consider is meteorology. It is accepted that extreme meteorological conditions may be correlated with air pollution and therefore act as confounders in the analysis. In most analyses, temperature and humidity are treated as the main meteorological confounders, the latter represented by either specific humidity or dewpoint. It matters little which of these two is adopted, but in the present analysis we use specific humidity. It is also necessary to consider some lagged variables, and to allow for nonlinear effects. The most important nonlinear relationship is that between deaths and temperature, which is decreasing over most of the range of temperatures but increasing at high temperatures. In the present analysis this is allowed for by introducing two regressors, T and $(T - T_0)_+$ ($x_+ = \max\{x, 0\}$) where T is either daily mean or daily maximum temperature and T_0 is a changepoint. The fitted curve is thus a broken straight line with a change of slope at T_0. In Birmingham, T is daily max temperature and T_0 set equal to 30^oC; in Chicago, T is daily mean temperature and $T_0 = 22^o$C. In both cases the value of T_0 was chosen on the basis of initial plots of the data. One indicator of the success of this strategy was that it eliminated any need for seasonal interactions in the model. In some earlier analysis, temperature has been modeled as a linear term only, but then an interaction between the temperature coefficient and season has been found, the coefficient being positive in summer and negative in winter. It seems likely that this is an artifact created by the fact that the true temperature-mortality relationship is non-linear. In the current analysis, seasonal interactions were tested, but not found to be statistically significant.

In both Birmingham and Chicago, the actual variables included were selected through standard variable selection techniques. For example, in Birmingham this led to the regressors T and $(T - 30)_+$ where T is maximum temperature lagged four days, as well as minimum temperature lagged three days, current day's specific humidity, square of current day's specific humidity, and specific humidity lagged two days. However a number of different meteorological models were tried, and the subsequent conclusions regarding the PM_{10} effect are not overly sensitive to the choice of variables, though it does appear to be important to include both temperature and humidity (or dewpoint) among the variables considered.

Finally, we consider which variable or variables best represents the particulates. This has been the source of considerable confusion in the literature. For example, Schwartz and Dockery (1992) and the subsequent "HEI" reports of Samet *et al.* (1995, 1997) have used two-day averages of total suspended particulates (TSP) in their studies of Philadelphia. Schwartz (1993) used three-day averages of PM_{10} excluding the current day (average of lags 1, 2 and 3) in his analysis of Birmingham, Styer *et al.* (1995) used three-day averages of PM_{10} including the current day (lags 0, 1, 2) in Chicago, Pope *et al.* (1992) used five-day averages of PM_{10} (lags 0, 1, 2, 3, 4) in Utah Valley, and so on. In each case the final exposure measure was decided based on some kind of statistical criterion of best-fitting model, but previous discussions of this question have not (with one exception) drawn attention to the sensitivity of the results to which measure is selected.

3 LINEAR MODELING RESULTS

The data from Birmingham, Alabama, are elderly (aged 65 and over) nonaccidental deaths for the period August 1985–December 1988, together with daily meteorology and PM_{10} readings. The data set is approximately the same as that of Schwartz (1993). The analysis which follows used square root of daily death counts as the dependent variable y_t. The square root transformation is a natural choice because this is the variance stabilizing transformation for the Poisson distribution. Maximum likelihood Poisson regression, including the standard logarithmic link function, has also been applied and produces results very similar to the following.

In the case of Chicago, the data are the same as in Styer *et al.* (1995), but restricted to the period from April 15, 1986 to December 31, 1990. Styer *et al.* used data from January 1, 1985 to December 31, 1990. The reason for omitting the first 15 months' of data is that within this period the PM_{10} readings are available only every six days, and this seems rather unsatisfactory for a comparison of different exposure measures. The remaining data have approximately daily PM_{10} values, though still with about 15% missing. For this data set, the most satisfactory regression was found to be a simple linear regression taking y_t as logarithm of daily death count.

To ensure a uniform scale for presentation of the results, all regression coefficients for PM_{10} are expressed as 10,000 times the increase in log deaths associated with a 1 $\mu g/m^3$ rise in PM_{10}. Thus if the regression coefficient is 8 (a fairly typical value), then according to the model, a 10 $\mu g/m^3$ rise in PM_{10} would produce an increase of 0.008 in log deaths, or a *relative risk* of $e^{.008} \approx 1.008$. Results from square root regression were converted to this scale using a Taylor expansion. The reference level of a 10 $\mu g/m^3$ change in PM_{10} is chosen because this is a reasonable guess of how much PM_{10} values might actually be reduced if standards were tightened. Some other authors have used a 100 $\mu g/m^3$ change in PM_{10} as the basis for relative risk calculations, but this seems misleading since in all US examples, the current level of PM_{10} is less than 100 $\mu g/m^3$, so the possibility of a reduction by this amount does not exist.

Daily values of PM_{10} will be denoted by pm with the appropriate lag — pm_0 for today's value, pm_1 for yesterday's, and so on. We also use $pmmean$ with appropriate lag for three-day average. Thus $pmmean_0$ is the average of pm_0, pm_1 and pm_2, while $pmmean_1$ is the average of pm_1, pm_2 and pm_3.

With these conventions, the regression of Birmingham deaths on $pmmean_1$, including regressors for the log-term trend and for meteorology as described in section 2, produces a coefficient of 9.6, standard error 4.9, t value 1.98. This is borderline statistically significant and is close to the result obtained by Schwartz (1993), who used a different meteorological model but the same measure of PM_{10}. However, other measures of PM_{10} produce quite different results. For example, taking $pmmean_0$ as the regressor produces a coefficient of 5.5, standard error 4.8, not significant. If we include all five daily values $pm_0, ..., pm_4$ in the regression, the coefficients are –8.7, 11.4, –5.2, 6.7 and 4.2, each with standard error about

5. In other words, we get negative coefficients for two of the five days. If we introduce pm_0, pm_1 and pm_2 into the model one at a time, the coefficients are –2.3 (standard error 3.7) for pm_0 on its own, 6.4 (standard error 3.8) for pm_1 on its own, and 5.2 (standard error 3.8) for pm_2 on its own. None of these is statistically significant but again the coefficient for pm_0 is negative.

These results are in line with the independent analyses of Roth and Li (1996). They also demonstrated great sensitivity to the lags of PM_{10} included in the model, including a negative coefficient for pm_0, even though the data set they used (1988 to 1993) only partially overlapped the one used here.

It seems unlikely that the current day's PM_{10} has a protective effect, but the results demonstrate the sensitivity of the estimates to the exposure measure. On balance, it does not appear that one can claim a significant PM_{10}–mortality relationship on the basis of this data set.

The results for Chicago are more complex. Chicago is a much bigger city than Birmingham, as reflected in the mean number of elderly nonaccidental deaths per day (83 for Chicago, 15 for Birmingham), and this should make it easier to detect statistically significant results. In this analysis, trend was modelled via a B-spline with 56 knots, one for each month of data. For the meteorological variables, attention was restricted to daily mean temperature $mntp$ and daily mean humidity $mnsh$, each for the current day and for lags of 1, 2, 3 and 4 days. For example, in the case of temperature these five values are denoted $mntp_0$, $mntp_1$, ... ,$mntp_4$ where the suffix denotes the lag. In addition, a new variable $tg22 = (mntp - 22)_+$ was created to allow for possible changes in the temperature effect above 22^oC, together with its lagged values $tg22_1$,...,$tg22_4$. We also considered squared terms in all of these variables. After routine variable selection a model was constructed including the variables $mntp_0$, $mntp_1$, $mntp_3$, $mntp_4$, $tg22_0$, $tg22_0^2$, $tg22_1$, $tg22_1^2$, $mnsh_0$, $mnsh_1$,$mnsh_3$ and $mnsh_3^2$ (Smith *et al.* 1998). Taking this as the basic model and adding different terms to represent particulate matter, the coefficients for $pmmean_0$ and $pmmean_1$ added separately were 7.3 and 3.9 respectively, each with standard error 2.3. In other words, the result was significant when based on $pmmean_0$ but not when based on $pmmean_1$. When $pm_0, ..., pm_4$ are inserted together, the estimates are 4.8, 3.9, 0.4, 0.8 and 2.2, each with standard error about 2.0. When pm_0, pm_1 and pm_2 are inserted one at a time, the estimates are 4.7, 5.1 and 0.05, each with standard error 1.7. Thus we do get a statistically significant result when based on either pm_0 or pm_1, but somewhat smaller in magnitude than that based on $pmmean_0$.

The overall evidence here is that PM_{10} does affect mortality in Chicago, but it is still surprising how sensitive the result is to different lags. An earlier analysis (Styer *et al.* 1995) discussed seasonal variation of the PM_{10} coefficient and suggested that it is significant in the spring and the fall, but not in the summer or winter. In the present analysis, we have also considered the possibility of a season $\times$ PM_{10} interaction and have concluded that this is not statistically significant. However this may be yet another manifestation of the sensitivity of all these kinds of results to model specification — in this case, the

different strategies for meteorological and seasonal adjustment used by Styer *et al.* and in the present research.

4 NONLINEAR RELATIONSHIPS BETWEEN PM_{10} AND MORTALITY

One of the original questions raised by Schwartz and Marcus (1990) was "is there a threshold"? In the context of fixing standards, this is a critical question, because an important issue is whether particulates at levels within the current standards have a measurable adverse effect.

One way to pursue this is to model the PM_{10} component of the model nonlinearly. In Fig. 2 a nonlinear relationship is plotted, using a B-spline representation, together with pointwise 95% confidence bands. The PM_{10} measure in each case is taken to be the one which in section 3 produced the most significant result, $pmmean_1$ for Birmingham and $pmmean_0$ for Chicago. The discussion in this and subsequent sections does not take into account the uncertainties associated with the selection of these particular exposure measures. Also, the rest of the model (trend and meteorology effects) is treated exactly as in the models of section 3. The effect is calculated in terms of relative risk (RR) with respect to the median level of the PM_{10} variable (45 for Birmingham, 37.5 for Chicago). Thus the RR is 1.0 and the confidence band of width 0 at this point. Elsewhere, the confidence band gives a measure of the uncertainty of the fitted curve.

In both cases the estimate and confidence bands indicate an increasing effect at higher levels, above 80 $\mu g/m^3$ for Birmingham and above 100 $\mu g/m^3$ for Chicago. These are both above the USEPA standard for the mean, though within that for daily maxima. The effect at lower levels is less clear. In Birmingham, the point estimate shows a steady decrease in effect as $pmmean_1$ decreases to 0, but the confidence bands cast severe doubt on the statistical significance of this. Schwartz (1993) gave a very similar plot of the estimated effect, but without any confidence bands. In the Chicago plot, there is a sharp drop in the estimated risk below about 20 $\mu g/m^3$, and the confidence bands suggest this is statistically significant. This would imply that there is no threshold in this case, though it is a little hard to explain this precise shape of curve. It is possible that it is an artifact of the B-spline representation used to fit the curve, though some exploration has been carried out with alternative representations, producing similar shapes of plot. We return to this point in section 5.

An alternative approach is to look for a threshold directly. This can be done by fitting the PM_{10} dependence through a relation of the form $\beta_1(P - P_0)_+$ where P is our measure of PM_{10} — $pmmean_1$ for Birmingham, $pmmean_0$ for Chicago — and P_0 is a threshold value. The relevant parameters in this case are therefore the threshold P_0 and the regression coefficient β_1. In Fig. 3, the profile log likelihood for P_0 is plotted for each of the two data sets. Once again the rest of the model is as in section 3. The calculation is conducted by fixing P_0, then estimating the rest of the model by least squares, then computing the residual sum of squares and hence the maximized likelihood. In each case

the profile log likelihood is normalized to have maximum value 0. Also shown on the plot is a horizontal line at –1.92, this being the level that determines the 95% confidence limit according to the standard $\frac{1}{2}\chi^2_1$ limiting distribution of the likelihood ratio statistic. In fact for this problem, being of the form of changepoint estimation, the standard asymptotic theory does not apply but nevertheless the $\frac{1}{2}\chi^2_1$ bound is shown as a reference point.

In the case of Birmingham, there is a formal "maximum likelihood estimate" at $P_0 = 68$. It seems unlikely that this value is of much meaning in itself, but a realistic interpretation from Fig. 3 is that the data provide no grounds to discriminate between any two values of P_0 below about 80. It is true that a null hypothesis $P_0 = 0$ would not be rejected, but neither would any other null value below 80, including 50 (the current US standard for the long-term mean).

For Chicago, again, the interpretation is more complex. The "maximum likelihood estimate" is at $P_0 = 0$ and the profile log likelihood drops away sharply as P_0 increases. There is a secondary peak at $P_0 = 105$ but this is almost certainly spurious. A formal test for P_0 would not reject any null hypothesis in the range $[0, 50]$, but the balance of evidence is that the value of P_0 is likely to be close to 0 if not actually equal to 0.

5 ALTERNATIVE APPROACH VIA GENERALIZED ADDITIVE MODELS

An alternative approach to the analyses of sections 3 and 4 is provided by generalized additive models (Hastie and Tishirani 1990), which we have implemented using the S-PLUS routines described in detail by Chambers and Hastie (1993). This approach allows us to combine the variable selection features of section 3 with the nonlinear modelling approach of section 4. Our description here concentrates on Chicago since the corresponding analysis for Birmingham has been described elsewhere (Smith *et al.* 1997b, 1998).

The basic additive model is defined by the equation

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_j \tag{2}$$

where x_{jt} is the value of the j'th covariate on the t'th day. As in ordinary least squares, $\mathrm{E}\{\epsilon_j\} = 0$ and $\mathrm{Var}\{\epsilon_j\} = \sigma^2$. The f_j terms are arbitrary univariate functions with an f_j modeled for each covariate. In the linear case $f_j(x_{jt})$ reduces to $\beta_j x_{jt}$ as in (1).

The nonlinear fits so far have been based on a B-spline representation for an arbitrary nonlinear function, but for the meteorological and particulate matter effects in the present analysis the `loess` procedure is used (Chambers and Hastie, 1993). This is a smoother based on fitting a linear model to subsets of the data. The fitted value at each $x_{jt} = X_j$ say is obtained from a weighted least squares fit in a small neighbourhood of X_j, where the weights are constructed to be large in a neighbourhood of X_j, but to decrease to zero outside of this neighbourhood.

To avoid possible difficulties for the algorithms created by missing data, it was decided to fill in missing data in the daily PM_{10} values by linear interpolation between the nearest available data points. This is in contrast to the analyses in sections 3 and 4, where days with missing data were simply omitted from the analysis. In the case of Chicago there are 232 missing days' out of the 1721 days of the analysis, and the maximum number of consecutive missing days is five.

The initial analysis attempted to build a model to predict deaths from meteorology alone; particulate matter was only added at a later stage. This is the same strategy as employed for the modelling of section 3. An initial list of variables consisted of $mntp_k$ and $mnsh_k$ for lags $k = 0, 1, ..., 4$, as in section 3. Backwards and stepwise selection were used together with the AIC criterion to select variables. For each variable x_j selected, a separate decision has to be made whether the corresponding function $f_j(x_j)$ should be linear or non-linear. In the non-linear case it was modeled by the S-PLUS `loess` function with default bandwidth $f = 0.5$, and the decision between linear and non-linear functions was made using the approximate nonparametric F test described in Hastie and Tibshirani (1990) and Chambers and Hastie (1993). Throughout the analysis, the trend continued to be modelled as a linear combination of 56 B-spline basis functions, while the response variable was taken as logarithm of daily elderly deaths, as in section 3.

This analysis produced a model including linear terms in the variables $mntp_0$, $mntp_3$, $mntp_4$, $mnsh_0$ and $mnsh_2$, and non-linear terms in the variables $mntp_1$ and $mnsh_1$.

Having thus produced a model for the meteorology and trend effects, variables representing particulate matter were added as in section 3. In particular, when three-day averages of PM_{10} lagged either 0 or 1 day ($pmmean_0$, $pmmean_1$) were added one at a time, we obtained regression coefficients 7.4 for $pmmean_0$ and 4.5 for $pmmean_1$, each with standard error about 2.3. Note that these results (both estimates and standard errors) are very similar to the ones produced in the earlier analysis of section 3, the main difference being a slight increase in the $pmmean_1$ coefficient (previously 3.9).

This analysis was extended to consider the effect of individual days' PM_{10}, as in section 3. This was achieved by adding the variables pm_0, pm_1, $...pm_4$ both one at a time and all together. When added one at a time, the two variables producing significant results were pm_1 (coefficient 4.6, standard error 1.7, t-value 2.8) and pm_0 (4.5, 1.6, 2.7). When all five variables were tested in a backwards selection process, pm_3, pm_2 and pm_4 were dropped in that order, leaving a model with pm_0 and pm_1 as the two significant variables (coefficients 3.56, 3.60, standard errors 1.73, 1.75). Thus we find, as in the analysis of section 3, that the most significant results are those due to the current day and one-day lagged value of PM_{10}, with the other days barely having any effect. Comparison of several different models based on an approximate C_p criterion showed two models — the one with $pmmean_0$ alone and the one with pm_0 and pm_1 — virtually tied for the best model overall.

The substantive questions raised by these analyses are why the pattern of one-day values is so different from Birmingham, and why the current day's value is so important

— intuitively, one would expect there to be at least some delay in the response to a high particulate level. The question raised by this is whether it is possible that PM_{10} is itself a by-product of some chemical reaction involving other pollutants and that the other pollutants are the real cause of death. Such explanation is plausible in view of studies showing that PM_{10} and ozone are produced by similar chemical processes involving the same precursors (Meng *et al.* 1997).

Study was also made of whether the effect due to PM_{10} was nonlinear. For this part of the analysis, PM_{10} was again represented by the variable $pmmean_0$ which was found earlier to be the most significant single measure of the effect. In this case, a nonparametric F test for nonlinearity found no evidence to reject a linear relationship, the P-value being .88. Nevertheless, for comparison with the results of section 4, a nonlinear component in $pmmean_0$ was fitted, with appropriate confidence bands.

Plots of the nine f_j functions from equation (2), together with pointwise 95% confidence intervals, are shown in Fig. 4. The $pmmean_0$ effect (bottom panel) is to be compared with the right-hand plot in Fig. 2 — the sharp kink near 0 in the latter plot is not so obvious in Fig. 4, but there is still a comparatively rapid increase in the dependence on $pmmean_0$ over the range [0,20]. Once again the nonlinear curve suggests that the PM_{10} effect persists all the way down to 0, but the width of the confidence bands do not allow us to make any definitive statement about the presence or absence of a threshold below which PM_{10} has no effect.

The GAM analyses reinforce the conclusions from sections 3 and 4 in a number of ways. They confirm that most of the meteorological effects except for temperature are linear, and provide an alternative approach to those terms which are nonlinear. The variable selection procedure leads to somewhat different variables being selected, but what is important, the different meteorological adjustments lead to similar conclusions regarding the significance of particulate matter, which is the real concern for us. Finally, the analyses confirm that the most significant PM_{10} variable is $pmmean_0$, with the biggest individual contributions being from the current day and the one-day lagged value. The bottom line is that although both the linear regression and GAM analyses involve a number of individual decisions which are somewhat arbitrary, they lead to very similar conclusions, so providing some reassurance that those arbitrary decisions are not exerting a great influence on the final results.

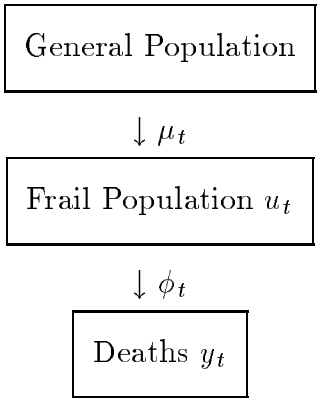
6 MORTALITY DISPLACEMENT

Mortality displacement (also known as harvesting) refers to the possibility that the observed PM_{10}–mortality relationship may be due to a very small subpopulation of critically ill individuals whose deaths are advanced only by a very short time period as a result of a high air pollution episode. At present, very little is known about this, though a number of authors have identified it as an important problem; for example, Samet *et al.* (1995) highlighted it as one of the major unresolved issues in current studies.

In this section, we outline a possible approach to the problem based on identification of a latent variable representing the size of a hypothetical "frail population", followed by a Bayesian MCMC approach to the estimation of this variable. The idea of using this kind of model has been suggested by other authors — for example, in unpublished notes relating to "phase II" of the HEI study, J. Samet and S. Zeger have proposed a model very similar to the one outlined here. However, to the best of our knowledge, the present study is the first attempt to work out the detailed consequences of this. As we shall see, the approach is not entirely successful, since although the analysis appears to identify a definite mortality displacement effect, similar conclusions are found in simulated data generated from the usual linear model.

6.1 A model for mortality displacement

The model may be represented as follows:



The top box represents the general population, assumed to be effectively of infinite size. The middle box represents the frail population, whose size on day t is u_t. On day t, a number of individuals z_t, assumed to be mutually independent of each other and of the past history of the process, and Poisson distributed with mean μ_t, transfer from the general population to the frail population. Then, among the $u_t + z_t$ individuals within the frail population, it is assumed that y_t of them die on day t. The distribution of y_t, conditionally on $u_t + z_t$, is assumed to be Binomial with individual probability of death ϕ_t. Thus $\{u_t,\ t = 1, 2, ..., \}$ satisfy the random difference equation

$$u_t = u_{t-1} + z_{t-1} - y_{t-1}. \tag{3}$$

The key to statistical inference within this model is given by the following:

Proposition 1. Suppose u_1 has a Poisson distribution with mean λ_1. Then for each $t > 1$, u_t has a Poisson distribution with mean λ_t, given by the difference equation

$$\lambda_t = (\lambda_{t-1} + \mu_{t-1})(1 - \phi_{t-1}). \tag{4}$$

Furthermore, $\{y_t,\ t \geq 1\}$ are mutually independent random variables, where y_t has a Poisson distribution with mean ν_t, given by

$$\nu_t = (\lambda_t + \mu_t)\phi_t. \tag{5}$$

Proof. Suppose, as an inductive hypothesis, that the statements about u_t are true for all $t \leq T$, and the statements about y_t are true for $t \leq T-1$. We shall argue that they must then be true for $t \leq T+1$ in the case of $\{u_t\}$, and $t \leq T$ in the case of $\{y_t\}$. This will suffice to prove the result for all $T \geq 1$, since the result is true by assumption for u_1.

The inductive hypothesis means that u_T has a Poisson distribution with mean λ_T. The migration z_T has an independent Poisson distribution with mean μ_T. Therefore, $u_T + z_T$ has a Poisson distribution with mean $\lambda_T + \mu_T$. Conditionally on $u_T + z_T$, the number of deaths y_T has a Binomial$(u_T + z_T,\ \phi_T)$ distribution. It is then readily verified (from moment generating functions, or directly) that y_T and $u_T + z_T - y_T$ are *independent* and each Poisson distributed, with means $(\lambda_T+\mu_T)\phi_T$ and $(\lambda_T+\mu_T)(1-\phi_T)$ respectively. This immediately gives the Poisson distributions of y_T and u_{T+1}, while the mutual independence of $\{y_t\}$ follows from the fact that all $\{y_t,\ t > T\}$ depend on the history of the process only through u_{T+1}, and we have shown that this is independent of y_T and hence by induction of all $\{y_t,\ t \leq T\}$. This completes the proof of the Proposition.

The beauty of this result is that the mutual independence of the y_t's means that we can write down directly the likelihood function of the observed death counts, as a function of the mean initial size λ_1 of the frail population, the migration rates $\{\mu_t\}$, and the death probabilities $\{\phi_t\}$. It remains to specify models for the latter two quantities, for which we adopt standard log-linear or logistic-linear relations

$$\begin{aligned} \log \mu_t &= \beta_{1,0} + \sum_{j=1}^{p} \beta_{1,j} x_{jt}, \\ \log\left(\frac{\phi_t}{1-\phi_t}\right) &= \beta_{2,0} + \sum_{j=1}^{p} \beta_{2,j} x_{jt}, \end{aligned} \tag{6}$$

in terms of known covariates $\{x_{jt},\ j = 1, ..., p\}$ available for day t, and parameters $\{\beta_{1,j},\ \beta_{2,j},\ j = 0, ..., p\}$. Note that constant terms $\beta_{1,0}$ and $\beta_{2,0}$ are assumed to be present, and are treated separately from the other covariates; this is because of certain technical differences in the way these parameters are treated within the Monte Carlo simulation, to be described.

The model allows for the possibility that an individual may migrate to the frail population and die the same day, thus in effect incorporating the possibility that some individuals by-pass the frail population. It also includes two limiting cases in which the distinction between the frail population and the general population disappears. One is when $\lambda_t \equiv 0$ and $\phi_t \equiv 1$: in this case the number of deaths on day t is Poisson with mean μ_t and the

model is effectively the same as one of the standard models based on Poisson regression with no mortality displacement. The second limiting case is when $\lambda_1 \to \infty$: in this case, μ_t and the daily fluctuations in u_t have no influence on the observed deaths over any finite period, and the model is effectively one of independent Poisson deaths with Poisson rate proportional to ϕ_t. Cases intermediate to these represent the ones we are really interested in. A final comment is that it may not be the case that the migration from the general population to the frail population is all in one direction; individuals in the frail population may recover, but it seems unlikely in this case that the migration rates in the two directions would be separately identifiable. Under this interpretation, μ_t represents a net migration rate from the general to the frail population.

6.2 Bayesian inference

The model in (3)–(6) is specified by the mean of the initial frail population size λ_1 and the regression parameters $\{\beta_{i,j},\ i = 1, 2,\ j = 0, ..., p\}$. To aid the specification of suitable prior distributions and to improve the stability of Monte Carlo estimation procedures, some reparametrization is desirable. First, we model λ_1 on a logarithmic scale:

$$\lambda_1 = e^{\beta_0^*} \tag{7}$$

where the prior distribution of β_0^* is flat. In other words, subsequent Monte Carlo evaluation of the posterior density uses β_0^*, rather than λ_1, as the parameter which determines the expected initial size of the frail population.

We also reparametrize $\beta_{1,0}$ and $\beta_{2,0}$ in terms of alternative parameters $\beta_{1,0}^*$ and $\beta_{2,0}^*$, defined as follows. Let D be the mean number of deaths per day. (In the example developed in section 6.3, which is based on elderly nonaccidental deaths in Chicago, $D \approx 83$.) Then write

$$e^{\beta_{1,0}} = De^{\beta_{1,0}^*}, \tag{8}$$

$$\phi^* = \frac{D}{\beta_{1,0} + \lambda_1}, \tag{9}$$

$$\beta_{2,0} = \log\left(\frac{\phi^*}{1-\phi^*}\right) + \beta_{2,0}^*. \tag{10}$$

The rationale behind these transformations is as follows. Long-term stability of the frail population means that the long-run migration rate must roughly equal the daily death rate D. Hence in (8), we expect $\beta_{1,0}^* \approx 0$. Taking $\beta_{1,0} + \lambda_1$ to be the approximate mean size of the frail population after daily migration, we then have ϕ^* in (9) to be the approximate mean death probability. This in turn means that in (10), we expect $\beta_{2,0}^* \approx 0$. Defining $\beta_{1,0}^*$ and $\beta_{2,0}^*$ in this way means that they are approximately independent of λ_1, which improves the performance of the Monte Carlo algorithm and also makes it more reasonable to assume that they have prior distributions independent of λ_1. In fact we do assume this, the prior distributions being improper uniform priors on $(-\infty, \infty)$. It does not matter that the arguments which have been used to justify (8)—(10) are deliberately

rather rough and heuristic, because any discrepancies from these will be picked up by the eventual posterior distributions of $\beta^*_{1,0}$ and $\beta^*_{2,0}$.

In the case of the remaining regression parameters $\{\beta_{i,j},\ i = 1,2,\ j = 1,...,p\}$, it is desirable to ensure that the uncertainty about these parameters is on roughly the same scale, but for the example to be discussed in section 6.3, this does not seem to be a problem. The reason for this assertion is that in an initial Poisson regression analysis without taking account of harvesting, the standard errors of all the parameters were within an order of magnitude of one another. In fact a change of scale was made in the definition of the PM_{10} coefficient to ensure this. With that proviso, however, no change was made in the original specification of these parameters. However, we still have to specify a suitable prior distribution for these parameters, and for this we proceed as follows.

Assume $\beta_{1,j},\ j = 1,...,p$ follow the hierarchical model

$$\begin{aligned} \beta_{1,j}|\mu_1,\tau_1 &\sim N(\mu_1,\tau_1^{-1}), \\ \mu_1|\tau_1 &\sim N(0,(m\tau_1)^{-1}), \\ \tau_1 &\sim G(a,b), \end{aligned} \tag{11}$$

where N and G denote the normal and gamma distributions respectively. We assume the same, but independent, model for $\beta_{2,j},\ j = 1,...,p$, in terms of hierarchical parameters μ_2 and τ_2. The model parameters m, a and b are assumed to be the same for both sets of regression parameters, and we arbitrarily set $m = a = b = 0.01$ to ensure diffuse but proper prior distributions.

Within the model (11), if we write down the joint density of $\beta_{1,1},...,\beta_{1,p},\mu_1,\tau_1$ and integrate out μ_1, we find that the remaining parameters have joint density proportional to

$$\tau_1^{p/2-1+a}\exp\left[-\tau_1\left\{b+\frac{1}{2}\sum_j\beta_{1,j}^2-\frac{(\sum_j\beta_{1,j})^2}{2(m+p)}\right\}\right]. \tag{12}$$

This suggests the following Gibbs-Hastings-Metropolis sampling scheme. Conditionally on $\{\beta_{1,j},\ j = 1,...,p\}$, the posterior density of τ_1 is $G(a',b')$, where

$$\begin{aligned} a' &= a+\frac{p}{2}, \\ b' &= b+\frac{1}{2}\sum_j\beta_{1,j}^2-\frac{(\sum_j\beta_{1,j})^2}{2(m+p)}. \end{aligned} \tag{13}$$

Therefore, we resample τ_1 according to the conditional distribution determined by (13). Then, conditionally on τ_1, the conditional joint density of $\{\beta_{1,j}\}$ is given by (12) multiplied by the Poisson likelihood for $\{y_t\}$. Thus we may resample the individual $\beta_{1,j}$ parameters according to a Hastings-Metropolis algorithm. This has been implemented via a random

walk Metropolis algorithm in which the initial trial distribution for $\beta_{1,j}$ is uniform of half-width $s_{1,j}$ about the current value. There is no clear-cut guideline to the choice of $s_{1,j}$ but the value that has actually been adopted is half the standard error of $\hat{\beta}_{1,j}$ according to the Poisson regression model without mortality displacement. This typically results in Metropolis acceptance rates of between 60% and 80% which is somewhat higher than the usually recommended optimum rates of between 15% and 50%, but alternative ways of specifying the $s_{1,j}$ parameters have not led to noticably better performance of the algorithm.

A similar procedure is, of course, used for resampling $\{\beta_{2,j},\ j = 1, ..., p\}$ along with the scale parameter τ_2. For $\beta_0^*, \beta_{1,0}^*, \beta_{2,0}^*$, a simple Metropolis updating procedure has been used with step lengths 0.5, 0.2, 0.2 respectively. These specifications are no better than rough guesswork but are given here so as to make explicit the procedure which has been adopted.

A number of parallel Monte Carlo runs were carried out with run lengths between 5,000 and 50,000 iterations and a variety of starting values, particularly for the parameter λ_1 which represents the initial size of the frail population. It was found that if λ_1 was taken too large, then the value drifted off to ∞ and the MCMC never appeared to reach convergence, but for moderate values of λ_1, the results apparently converged to a posterior distribution which did not depend on λ_1. We say *apparently* converged because with the high dimensionality of the problem and the slow running time of the algorithm, there does not appear to be any way to *ensure* convergence. For the final results reported in section 6.3, a run length of 50,000 iterations was used, with model parameters recorded every tenth iteration.

6.3 Results

The method outlined in the section 6.2 is now applied to the data on elderly deaths in Chicago. The covariates $\{x_{jt}\}$ are assumed to be those derived at the end of section 3, in which there are a total of 68 terms, i.e. 55 for the representation of trend as a linear combination of B-splines, 12 meteorological terms and one PM_{10} term — in this case, the variable $pmmean_0$ (three day mean including current day's value) was taken as a single variable representing the "best" effect due to particulate matter. The same 68 regressors were used in both parts of the model (6).

The main results are represented by Fig. 5. This shows the posterior density for four key parameters, (a) the mean size of the frail population, (b) the PM_{10} coefficient in μ_t, (c) the PM_{10} coefficient in ϕ_t, and (d) the mean number of days of life that would be lost, for each individual in the frail population, under the assumed model, if the level of PM_{10} were to rise by 10 $\mu g/m^3$. These results are based on a single long run of 50,000 iterations after some initial runs of shorter lengths, as described at the end of section 6.2. Table 1 shows the posterior means and standard deviations of the four key parameters.

Parameter	Posterior Mean	Posterior SD
Mean frail population size	823	221
PM_{10} in μ_t	5.23	7.04
PM_{10} in ϕ_t	8.17	2.50
Mean days lost	.081	.033

Table 1: Posterior means and standard deviations for the four key parameters graphed in Fig. 5

These results may be summarized as follows:

1. The estimated mean size of the frail population is very small. Consequently the mean time to death within this population is also small — given a mean of 823 individuals in the frail population, and a mean number of deaths per day of 83, the mean number of days each person lives in this population is only about 9.9. This interpretation is only believable if the frail population is confined to very sick individuals, though as we shall see in section 6.4, there is plenty of reason to doubt such a literal interpretation of the results.

2. Within the frail population, the posterior mean and standard deviation of the PM_{10} coefficient are very similar to the estimate and standard error computed from the least squares analysis.

3. The mean number of days of life lost, associated with a rise in PM_{10} of 10 $\mu g/m^3$, is very small — less than 0.1 days. It should be pointed out, however, that this conclusion is almost inevitable given the small size of the frail population: we have argued that each individual in the frail population survives for an average of 9.9 days, and under the fitted model, the effect of a 10 $\mu g/m^3$ rise in PM_{10} is to increase the death rate by about 0.82%. This leads at once to an expected $9.9 \times .0082 = .081$ days decrease in life length.

4. It may well be that the really critical parameter is the PM_{10} coefficient in μ. If this were positive, that would indicate that entry into the frail population is affected by PM_{10}, and this would lead to the conclusion that the PM_{10}–mortality relationship is not explained away by mortality displacement. In the present analysis, this parameter has a posterior mean of 5.23 and a posterior standard deviation of 7.04. Reinterpreted as a point estimate and standard error, this implies that it is not significantly different from zero, or in other words, that the observed PM_{10}–mortality relationship *can* be explained as entirely the result of mortality displacement.

6.4 A simulation study

The results of section 6.3 cannot be taken as definitively establishing the existence of a harvesting effect. We have no means of knowing for sure that the MCMC algorithm has converged to the true posterior density. Even if we knew that it had, the frequentist properties of a Bayesian procedure when applied to such a complicated high-dimensional problem are completely unknown.

To gain a greater insight into these aspects, a simulation study was conducted in which the observed log deaths $\{y_t\}$ were replaced by simulated values generated from the model (1). The covariates $\{x_{jt}\}$ were kept the same as in the real data, and the parameters $\{\beta_j\}$ and σ^2 were also assumed to be the same as those estimated when (1) was fitted to the data by ordinary least squares. However, independent random variables $\{\epsilon_t\}$ were generated, and hence new values of $\{y_t\}$. The process was repeated 20 times to create 20 simulated data sets, which were then analyzed using the algorithm of section 6.2.

Because of the enormous computation required, it was not feasible to repeat each of these simulations with 50,000 iterations of the MCMC algorithm, as in section 6.3. Instead, 5,000 iterations were used. To maintain comparability, just 5,000 iterations were used from the real data study. The starting values for the unknown parameters and the step lengths of the Metropolis updating scheme were also kept the same for all the data sets. Thus the results are strictly comparable with each other, irrespective of whether the MCMC algorithm has actually converged.

The results are summarized in Table 2. In this table, the Bayes posterior means and standard deviations are given for each of the four principal parameters, for each of 19 simulated data sets plus the real data (run no. 0). For the 20'th simulated data set, no estimates were obtained as in this case (only) the algorithm palpably failed to converge, the estimated λ_1 diverging to ∞. In all other cases, however, the MCMC algorithm *appeared* to converge, creating the impression of a finite frail population size.

In run no. 2, the estimated frail population size is a ridiculously small 388, with similarly meaningless values for the other parameters, e.g. a negative dependence on PM_{10} in the current daily death rate. It may be that if this simulation were continued, we would eventually find $\lambda_1 \to 0$, which as noted in section 6.1, is one of the limiting cases which reduces to a standard Poisson regression.

All the other simulations produced estimated frail population sizes which are larger than the one estimated for the real data. This may be interpreted as mild evidence that a harvesting effect actually exists: if we exclude the two simulations that produced obviously nonsensical results, then among the 19 data sets (18 simulated and one real) remaining, the real data set produced the strongest evidence of a finite frail population size as judged by the posterior density estimates. Nevertheless it is clear from this simulation that the actual size of the frail population, as estimated by the approach, should not be taken literally.

Run No.	Frail Pop Size		PM_{10} in μ_t		PM_{10} in ϕ_t		Days lost	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	799	160	8.4	8.0	8.1	2.8	.077	.028
1	1455	484	17.3	8.9	7.3	2.4	.130	.063
2	388	656	9.2	6.8	–4.9	7.4	.005	.036
3	1135	344	18.4	9.0	6.6	2.2	.091	.042
4	1022	257	6.7	6.5	4.2	2.3	.052	.031
5	829	217	6.7	4.9	4.6	2.8	.047	.034
6	1419	451	16.1	9.0	10.2	2.5	.180	.079
7	922	262	11.0	5.4	9.1	2.7	.104	.045
8	996	477	12.0	8.1	7.9	2.6	.099	.064
9	868	315	14.1	8.0	1.6	2.4	.019	.027
10	895	844	13.0	8.0	7.8	3.7	.084	.086
11	1025	648	6.7	5.4	6.3	2.8	.078	.062
12	1446	535	15.7	6.2	7.5	2.3	.137	.075
13	2036	543	.5	8.3	10.3	2.1	.259	.093
14	1184	304	7.3	7.1	10.8	2.4	.157	.057
15	912	382	13.3	9.3	10.6	2.7	.116	.056
16	1373	325	–16.1	8.5	9.5	2.4	.159	.051
17	1140	440	4.0	9.2	9.7	2.3	.136	.068
18	867	208	–2.5	6.1	5.0	2.4	.054	.031
19	1379	467	18.3	6.6	3.9	2.3	.066	.045

Table 2: Posterior means and standard deviations for the four key parameters from 5,000 MCMC iterations for the real data set (run no. 0) and for 19 simulations based on the linear model (1)

It may be that the difficulties highlighted in this section result from inadequate MCMC sampling, and that if we continued all the MCMC runs until it became clear whether the algorithm was converging or not, more of the simulations would show a diverging frail population size. Our own suspicion is that this is not the correct explanation, and that the simulation study shows an inherent limitation in this kind of Bayesian analysis. It should be emphasized that the direct evidence for a harvesting effect is very weak. For example, correlations between residuals from the linear model and lagged PM_{10}, which might be expected to be negative if the harvesting effect were real, do not in fact show any different behaviour from that which would be expected under a linear model with independent errors. It would be of interest to repeat the entire analysis using a larger data set, possibly one formed by combining data from different cities, to see whether the failure to obtain definitive conclusions in the current analysis is primarily a consequence of limitations in the available data.

6.5 Conclusions

The results of this section must be taken a highly tentative. Section 6.3 taken at its face value implies that the frail population size is finite, that the posterior mean and

standard deviation of the PM_{10} coefficient in ϕ_t are similar to the point estimate and standard error obtained in a linear regression analysis, and that no definitive conclusions can be made about the arguably more interesting PM_{10} coefficient in μ_t. On the other hand, section 6.4 shows that similar results were obtained in 18 out of 20 simulated data sets in which there was no harvesting, though with estimated frail population sizes larger than that obtained in the real data set based on the same number of MCMC iterations. Taken together, these results provide mild evidence that a harvesting effect exists, but allow very little to be said about the magnitude of the effect.

The consequences for the broader issue, of how to interpret studies showing a relationship between PM_{10} and mortality, are essentially to add further confusion to the whole debate. We would really like to know the extent to which life expectancy is reduced by exposure to high levels of particulate matter, but it seems impossible to obtain such information directly without detailed data on health histories of individual patients. It may be that a similar analysis to the one presented here, but for a much larger data set, would produce more definitive conclusions. Until further studies are available, we cannot dismiss the possibility that observed particulate-mortality associations are due primarily to mortality displacement.

7 OTHER ISSUES

In this section we mention, more briefly, two other issues that have been raised in connection with the particulates-mortality relationship.

One issue is the influence of other pollutants. This was a particular focus of Samet *et al.* (1997), who analyzed data from Philadelphia with respect to five "criteria pollutants": TSP (as a substitute for PM_{10}), O_3, SO_2, NO_2 and a lagged value of CO. When all five were included in the model together, they all produced statistically significant results, but the coefficient of TSP was not the one producing the greatest statistical significance, since both O_3 and lagged CO had larger t values. Also, the coefficient for NO_2 was negative, which the authors suggested was most likely a spurious result caused by linear correlations among the variables.

For Chicago, a more limited analysis has been conducted in which PM_{10}, represented by $pmmean_0$, was included in the model along with SO_2 and O_3. The exposure measure for SO_2 was taken to be the average of three days on lags 1, 2 and 3, while that for O_3 was the average of two days on lags 1 and 2. These were selected by a similar process to that which led to $pmmean_0$ being selected as the exposure measure for PM_{10}. When all three variables were included in the model, the PM_{10} coefficient was 7.1, standard error 3.0, and the t statistics for all three variables 2.4 for PM_{10}, 1.9 for O_3 and –1.9 for SO_2. The comparison with the earlier result for PM_{10} alone is confused by a large number of missing days for O_3, but if we re-fit the model with PM_{10} to those days in which O_3 and SO_2 readings are available, we get the same point estimate 7.3 as in our original analysis of Chicago using $pmmean_0$ (section 3). However the standard error in this case is 2.8. Thus

of the three pollutants, PM_{10} appears to be the most significant and the point estimate is little changed compared with the case when PM_{10} is fitted on its own. On the other hand the situation is similar to Philadephia in that one of the pollutants, SO_2, has a negative coefficient which is probably spurious but which does suggest that the estimates are influenced by correlations among the variables. When either O_3 or SO_2 is fitted on its own, we get a positive estimate for the effect, the t values being 2.1 and 1.3 respectively. Thus there is stronger evidence for PM_{10} being the most important pollutant than Samet *et al.* (1997) found for TSP in the case of Philadelphia, but there must remain doubts about whether this is a causal effect associated specifically with PM_{10}.

Yet another issue is whether one can find significant interactions among the effects of the different pollutants. Preliminary investigations of this point have suggested that the PM_{10} effect is greatest when O_3 is high and SO_2 is low; if correct, this would provide a possible explanation of why the SO_2 coefficient is negative in the preceding analyses. This issue needs to be investigated further.

The other main question of concern is that of errors in variables. All the variables are measured with uncertainty, but this is especially true of PM_{10} because the daily measurement is usually at only one site in a city and essentially nothing is known about variations in individual exposure. Particular concern has been expressed about the differences between indoor and outdoor exposure, and about the possibility that measurement error may be helping to mask confounding between PM_{10} and other variables. In an earlier Spruce volume, Zidek (1997) provided an excellent discussion of the role of measurement error in this kind of analysis, but as with the discussion of section 6, the task of quantifying the effect seems formidable. It remains a major issue in this debate.

8 CONCLUSIONS

This review has discussed several aspects of the problem which belie a simple interpretation of a causal relationship of particulates on mortality. The eventual outcome of the debate remains unclear, in our view. For the Birmingham data, the estimated effect is highly dependent on the particular choice of exposure measure, and even with that choice, suggests no meaningful relationship below about 80 $\mu g/m^3$. For Chicago, there is again considerable sensitivity to the choice of exposure measure, but the results are more robust than those for Birmingham and do suggest a significant result. Moreover, the nonlinear analysis suggests that this effect persists to low threshold values and possibly to $P_0 = 0$. On the other hand, the analyses of the mortality displacement effect, and of the interaction between PM_{10} and other pollutants in Chicago, raise more complicated issues which have not been resolved. The errors in variables problem has only been discussed very briefly but is another issue which complicates the interpretation of the regression analyses.

In the broader context of environmental regulation, it is often argued that where there is reasonable evidence of an adverse health effect, government should act without waiting for the evidence to become certain. In the present context, some commentators have

argued that the evidence in existing data is perfectly clear-cut and that it would simply be wasting time and lives to delay regulation. Our own approach is more cautious than this. While acknowledging that there is cause for concern, we cannot accept that the current studies provide definitive evidence of an effect specifically due to particulate matter. A number of critical scientific issues remain to be resolved.

REFERENCES

Chambers, J.M. and Hastie, T.J. (eds.) (1993), *Statistical Models in S*. Chapman and Hall, London.

Green, P.J. and Silverman, B.J. (1994), *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman and Hall, London.

Hastie, T.J. and Tibshirani, R.J. (1990), *Generalized Additive Models*. Chapman and Hall, London.

Liang, K.Y. and Zeger, S.L. (1986), Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Meng, Z., Dabdub, D. and Seinfeld, J.H. (1997), Chemical coupling between atmospheric ozone and particulate matter. *Science* **277**, 116–119 (July 4 1997).

Pope, C.A., Schwartz, J. and Ranson, M. (1992), Daily mortality and PM10 pollution in Utah Valley. *Arch. Environ. Health* **42**, 211–217.

Roth, H.D. and Li, Y. (1996), Analysis of the association between air pollutants with mortality and hospital admissions in Birmingham, Alabama, 1986–1990. Technical Report, Roth Associates, 6115 Executive Blvd, Rockville, MD 20852.

Samet, J.M., Zeger, S.L. and Berhane, K. (1995), The Association of Mortality and Particulate Air Pollution. In *Particulate Air Pollution and Daily Mortality: Replication and Validation of Selected Studies. The Phase I Report of the Particle Epidemiology Evaluation Project*. Health Effects Institute, Cambridge MA, pp. 1–104.

Samet, J.M., Zeger, S.L., Kelsall, J.E., Xu, J. and Kalkstein, L.S. (1997), Air Pollution, Weather and Mortality in Philadelphia, 1973–1988. In *Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants. The Phase IB Report of the Particle Epidemiology Evaluation Project*. Health Effects Institute, Cambridge MA, pp. 1–29.

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

Schwartz, J. and Dockery, D.W. (1992), Increased mortality in Philadelphia associated wuth daily air pollution concentrations. *Am. Rev. Respir. Dis.* **145**, 600-604.

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

Smith, R.L., Davis, J.M. and Speckman, P. (1997a), Assessing the human health risk of atmospheric particles. *Proceedings of the Section on Statistics in the Environment, American Statistical Association*, to appear.

Smith, R.L., Davis, J.M., Sacks, J., Speckman, P. and Styer, P. (1997b), Air pollution and daily mortality in Birmingham, Alabama: A Reappraisal. Technical Report (in

preparation), National Institute of Statistical Sciences, P.O. Box 14162, Research Triangle Park, N.C. 27709.

Smith, R.L., Davis, J.M. and Speckman, P. (1998), Airborne particles and mortality. To appear as Chapter 6 of *Case Studies in Environmental Statistics*, edited by L.H. Cox, D. Nychka and W.W. Piegorsch, Lecture Notes in Statistics, Springer Verlag, New York.

Styer, P., McMillan, N., Gao, F., Davis, J. and Sacks, J. (1995), The effect of outdoor airborne particulate matter on daily death counts. *Environmental Health Perspectives* **103**, 490–497.

Zidek, J.V. (1997), Interpolating air pollution for health impact assessment. In *Statistics for the Environment 3: Pollution Assessment and Control*, edited by V. Barnett and K.F. Turkman. Wiley, Chichester, pp. 251–268.

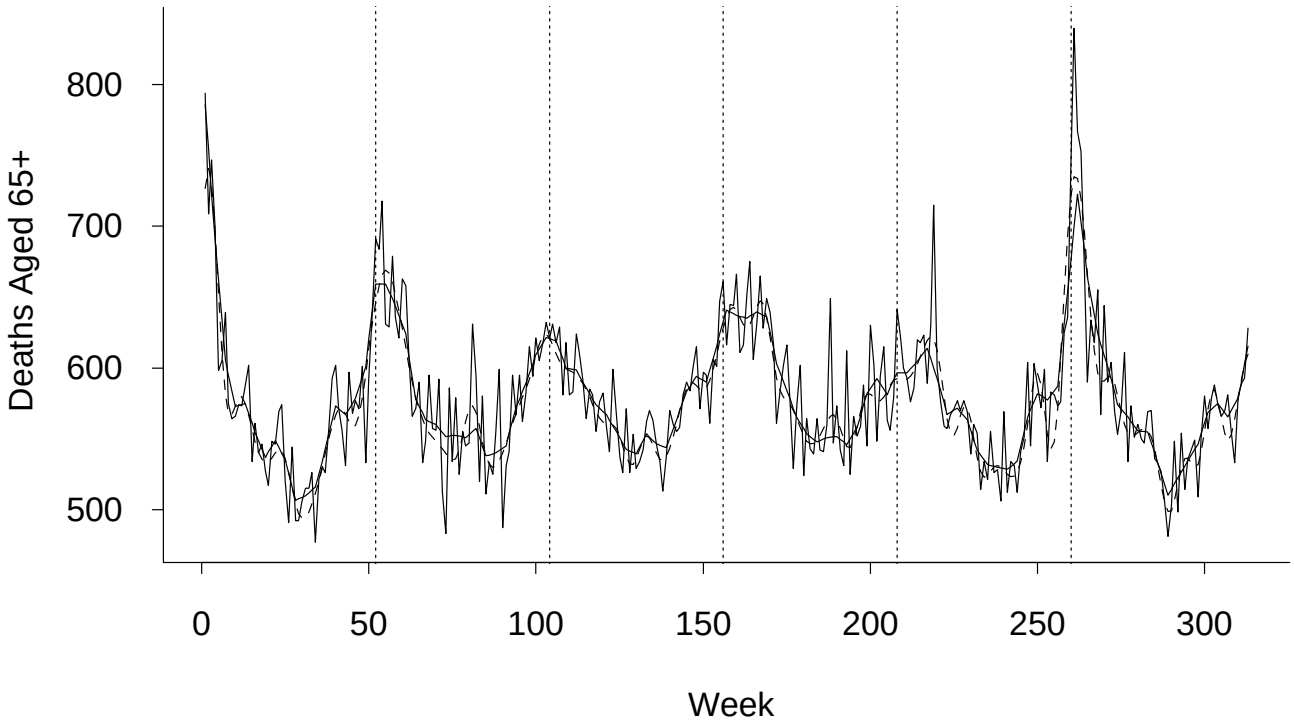


Fig. 1. Plot of weekly deaths in Chicago, 1985–1990, together with smoothed trends fitted by loess (solid curve) and B-splines (dashed curve)

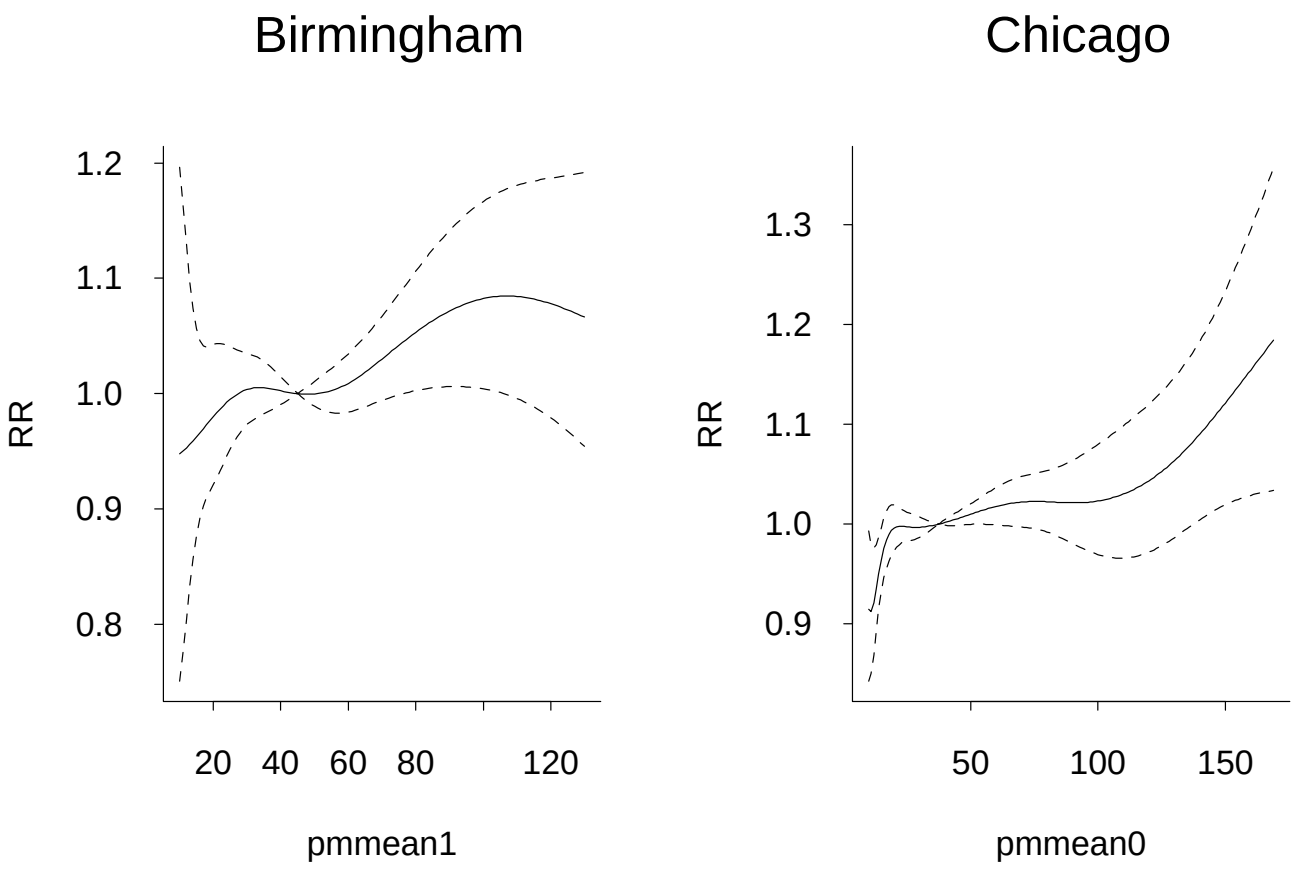


Fig. 2. Nonlinear estimates of relative risk, calculated with respect to the median PM_{10} level, with pointwise 95% confidence bands

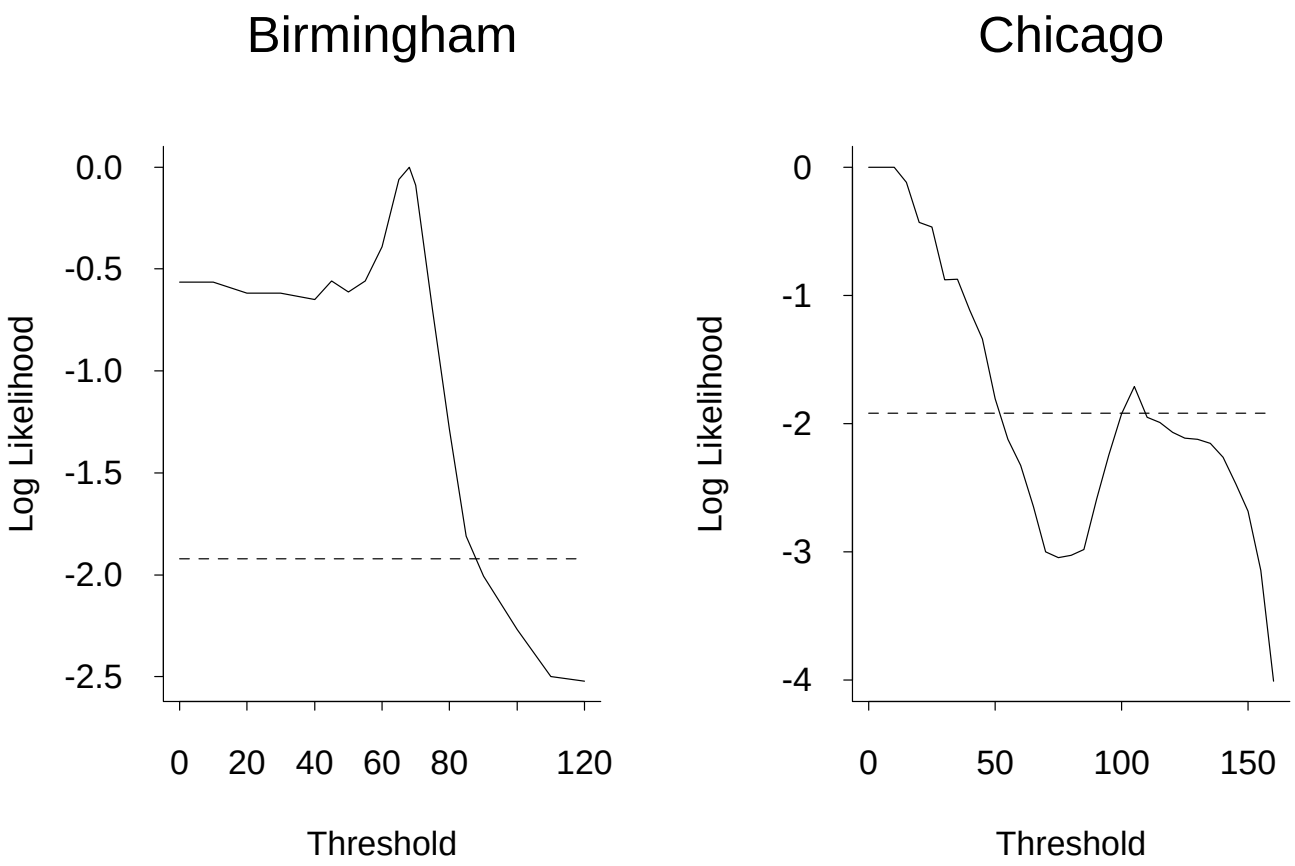


Fig. 3. Profile log likelihoods for the threshold parameter

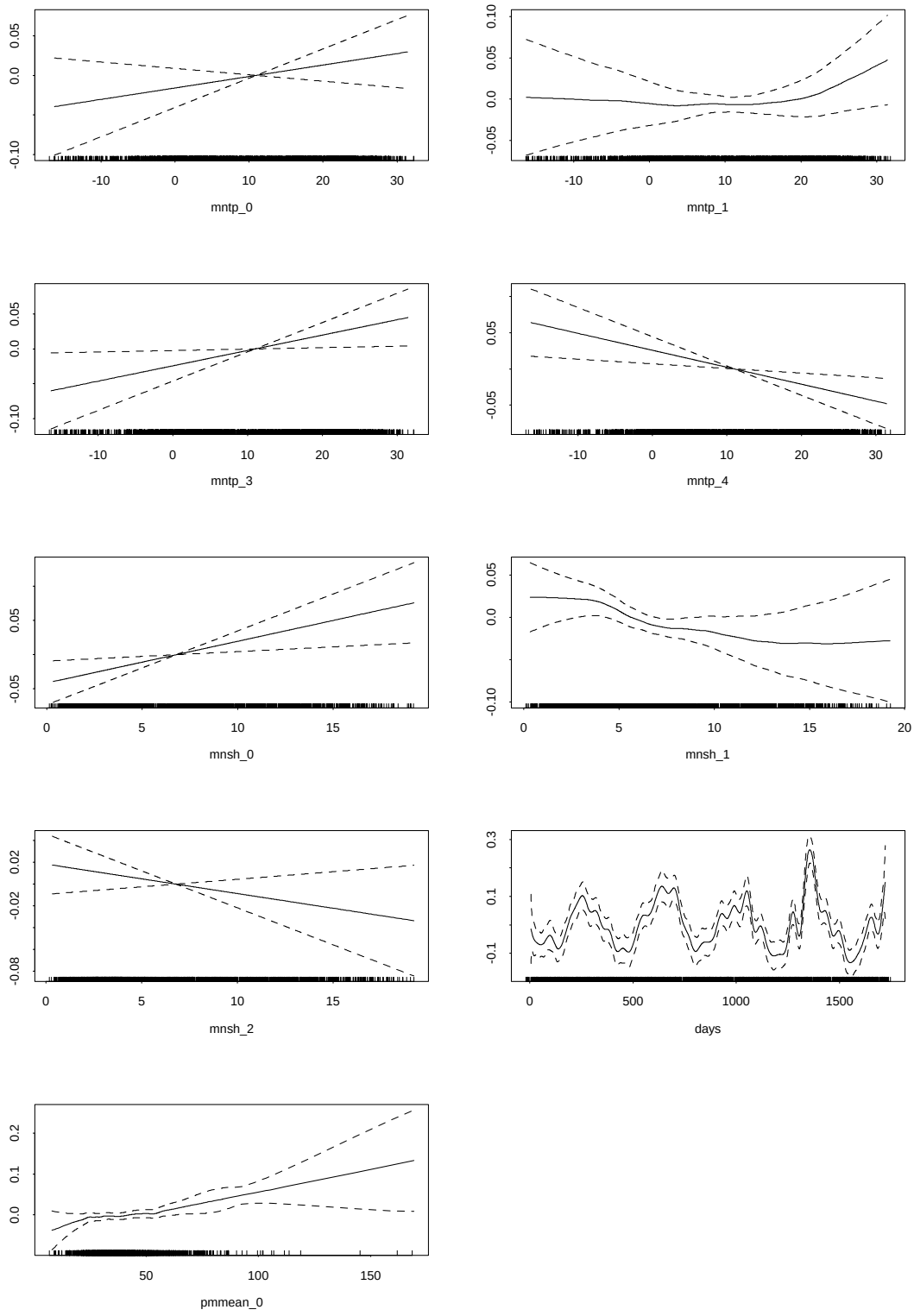


Fig. 4. Plot of 9 fitted functions produced by the GAM modelling scheme, with pointwise 95% confidence bands

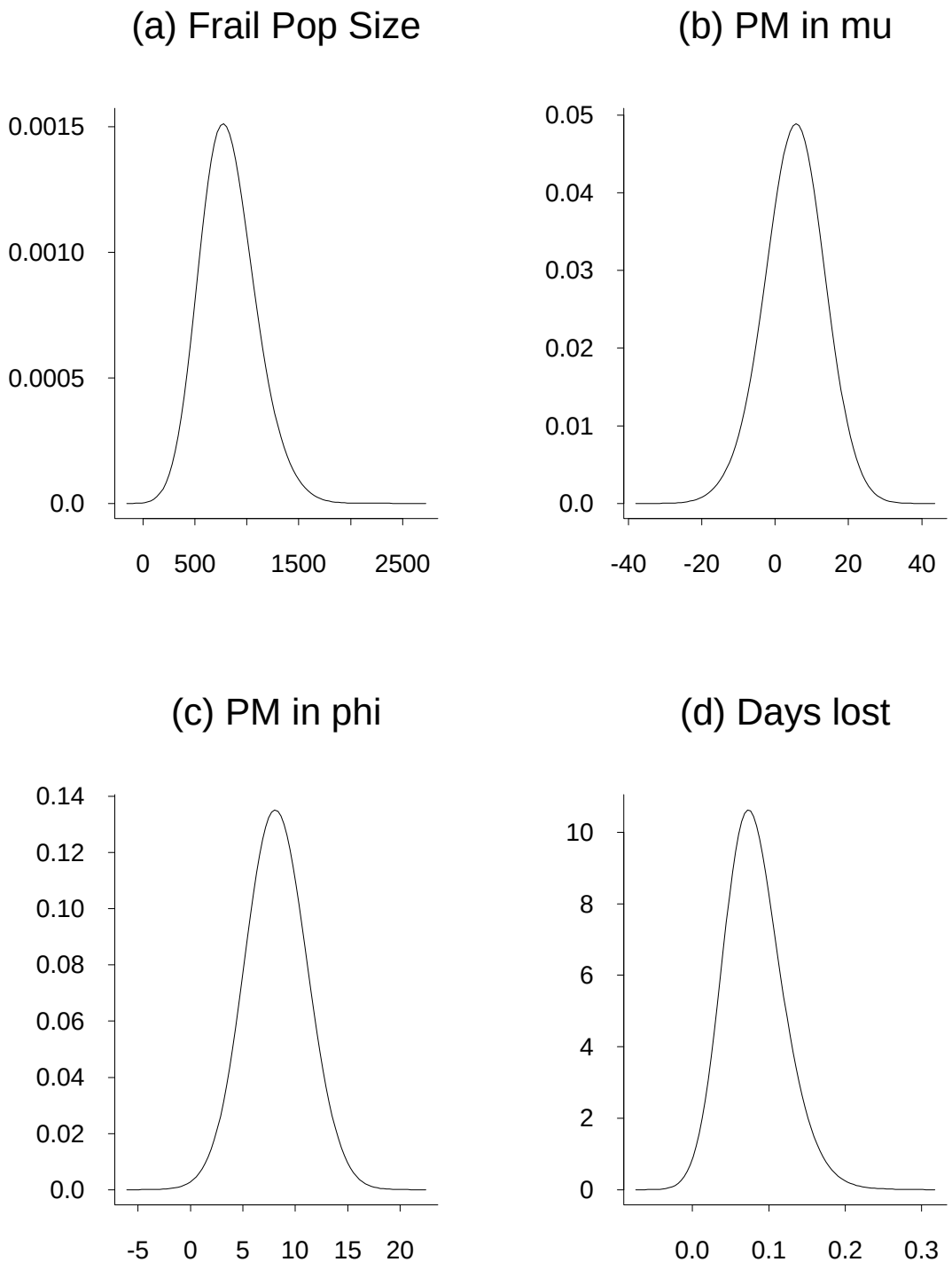


Fig. 5. Posterior density plots for four key parameters in mortality displacement analysis.