

REGRESSION MODELS FOR AIR POLLUTION AND DAILY MORTALITY: ANALYSIS OF DATA FROM BIRMINGHAM, ALABAMA ¹

by

Richard L. Smith ², Jerry M. Davis ³, Jerome Sacks ⁴
Paul Speckman ⁵ and Patricia Styer ⁶

February 28, 2000

Abstract

In recent years, a very large literature has built up on the human health effects of air pollution. Many studies have been based on time series analyses in which daily mortality counts, or some other measure such as hospital admissions, have been decomposed through regression analysis into contributions based on long-term trend and seasonality, meteorological effects, and air pollution. There has been a particular focus on particulate air pollution represented by PM₁₀ (particulate matter of aerodynamic diameter 10 μm or less), though in recent years more attention has been given to very small particles of diameter 2.5 μm or less. Most of the existing data studies, however, are based on PM₁₀ because of the wide availability of monitoring data for this variable. The persistence of the resulting effects across many different studies is widely cited as evidence that this is not mere statistical association but indeed establishes a causal relationship. These studies have been cited by the United States Environmental Protection Agency (USEPA) as justification for a tightening on particulate matter standards in the 1997 revision of the

¹ The bulk of this work was carried out as part of various projects at the National Institute of Statistical Sciences, P.O. Box 14162, Research Triangle Park, NC 27709-4162. In particular, the research was supported by U.S. Environmental Protection Agency under Cooperative Agreement #CR819638-01-0 and by National Science Foundation Grant DMS-9208758. We thank the referee for constructive comments on an earlier version of the paper. This paper has not been subjected to the Environmental Protection Agency's internal peer review system and no endorsement by the Agency should be implied or inferred.

² Corresponding author: Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260; email rls@email.unc.edu

³ Permanent address: Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695-8208

⁴ National Institute of Statistical Sciences

⁵ Department of Statistics, 222 Math Sciences Building, University of Missouri, Columbia, MO 65211

⁶ 664 Windemar Dr., Ashland, OR 97520

National Ambient Air Quality Standard (NAAQS), which is the basis for air pollution regulation in the United States.

The purpose of the present paper is to propose a systematic approach to the regression analyses that are central to this kind of research. We argue that the results may depend on a number of *ad hoc* features of the analysis, including which meteorological variables to adjust for, and the manner in which different lagged values of particulate matter are combined into a single “exposure measure”. We also examine the question of whether the effects are linear or nonlinear, with particular attention to the possibility of a “threshold effect”, i.e. that significant effects occur only above some threshold.

These points are illustrated with a data set from Birmingham, Alabama, first cited by Schwartz (1993) and since extensively re-analyzed. For this data set, we find that the results are sensitive to whether humidity is included along with temperature as a meteorological variable, and to the definition of the exposure measure. We also find evidence of a threshold effect, with the greatest increase in mortality occurring above $50 \mu\text{g}/\text{m}^3$, which is the long-term average level permitted by the current NAAQS. Thus, on the basis of this data set, the need for a tighter NAAQS is not established.

Although this particular analysis is focussed just on one data set, the issues it raises are typical in this area of research. We do not dispute that there is a reasonable level of evidence linking atmospheric particulate matter with adverse health outcomes even within the levels permitted by current regulations. However, the impression has been created by some of the published literature that such associations are overwhelmingly supported by epidemiological research. Our viewpoint is that the statistical analyses allow different interpretations, and that the case for tighter regulations cannot be based solely on studies of this nature.

Keywords. Generalized additive modeling, linear regression, model selection, PM_{10} , Poisson regression, threshold effects, cubic splines.

1 BACKGROUND

A major focus of air pollution research in recent years has been the health effects of small particles in the atmosphere, PM_{10} or $\text{PM}_{2.5}$. The epidemiological evidence comes from two kinds of studies: time series studies and prospective studies. Prospective studies, the best known of which are the Harvard six-cities study (Dockery *et al.* 1993), the American Cancer Society study (Pope *et al.* 1995b) and the Adventist Health Study (Abbey *et al.* 1999), follow a fixed group of individuals over a long time span, monitoring health indicators (primarily, date of death) as a function of lifestyle factors such as smoking, and exposure to air pollution as derived from ambient monitoring stations in each of the cities covered by the study. Time series studies, in contrast, are usually concentrated within a single city, focus on a particular health outcome (e.g. daily deaths from nonaccidental causes among the population aged 65 and over), and are based on identifying associations

between short-term fluctuations in the health outcome with corresponding short-term fluctuations in the pollutant of interest. Because the data are typically derived from public data sources, there have been many more studies of the time series type than of the prospective type, and an impression has been built up in some of the literature that the weight of evidence is overwhelmingly in favor of causal relationships. For example, Pope *et al.* (1995a) reviewed around 80 published papers, most published within the decade 1985–1995, most concerned with levels of pollution which are common in modern cities in developed countries, and which showed an association between particulate air pollution and a very wide range of health outcomes. For the present paper, we consider only time series studies; the prospective studies which have been published raise numerous questions of their own, but these are quite different from the questions considered here.

The statistical methods used in time series studies typically involve linear or Poisson regression, with covariates based on long-term trend and seasonality, meteorology, and particulate and other ambient air pollutants. The methods are described in more detail in section 3. The overall flavor of most of the results is that after allowing for possible confounders as represented by the trend, seasonal and meteorological variables, there remains a clear statistically significant signal based on particulate air pollution. Nevertheless, some authors, including Styer *et al.* (1995), have drawn attention to possible inconsistencies in these results.

The purpose of the present paper is to examine these kinds of methods in relation to one particular data set, from Birmingham, Alabama. This city was studied for mortality by Schwartz (1993) and for hospital admissions by Schwartz (1994). These studies were based on Poisson regression models with covariates allowing for seasonal trends, meteorology and PM_{10} , and found a statistically significant effect due to PM_{10} . The mortality study was repeated by Samet *et al.* (1995), who developed new estimators and computational algorithms for fitting the models, but who adopted the same data and models as Schwartz. They essentially confirmed the numerical correctness of Schwartz’s results, but they did not consider the effect of alternative modeling strategies.

Another study of Birmingham data, over a largely different time period, was by Roth and Li (1996). They collected data on mortality and morbidity, PM_{10} and other pollutants (O_3 , SO_2 , CO). They fitted a wide variety of models based on different ways of handling the trend/seasonal and meteorological factors, different exposure measures for PM_{10} (e.g. one-day readings over various lags, three-day averages, etc.) and different combinations of pollutants. With mortality as response, they fitted 2,400 different models of which only seven showed statistically significant PM_{10} coefficients, and three of these were negative. It is not clear exactly what this kind of comparison means since many of the models being compared are very similar, so the results will be highly correlated. Moreover, Roth and Li did not make any attempt to identify a “best” model in any sense. Nevertheless, the result underlines the sensitivity of claims of statistical significance to model selection. They also found substantial variation of the PM_{10} coefficients from year to year. Putting all their results together, Roth and Li claimed that there was no evidence of any relationship between particulates and mortality in Birmingham.

The paper by Roth and Li helped draw our attention to the importance of different exposure measures of PM_{10} , a theme which is developed at some length later in sections 3 and 5. Apart from that, the results of the present paper have been derived entirely independently of Roth and Li.

We should also draw attention to the very recent study by Clyde (2000), which used Bayesian model averaging techniques to form an independent view of some of the issues discussed in the present paper.

For the present study, we have reconstructed the data from their original sources, and have developed independent methods of analysis. We focus particularly on the effects of model selection, different exposure measures, and on possible nonlinear relationships between particulates and mortality.

2 BIRMINGHAM DATA

To construct a data set for this problem, three sources have to be combined: mortality, meteorology and PM_{10} . The limitations of the present data set are mainly dictated by the availability of PM_{10} data.

2.1. PM_{10} data

The sampling period is January 1, 1985 through December 31, 1988, the same as in Schwartz (1993). During this period, PM_{10} data are available from the USEPA's aerometric data base for 13 monitors in or near the city of Birmingham. The 13 monitors do not necessarily represent different locations: when the type of monitor or the method of measurement changes, this is treated as if it were a different monitor.

For the first seven months of 1985, data from three monitors are available, but they are only collected every six days (the same sampling days for each monitor). Data of this nature are of limited use for studying daily mortality effects. However, from August 1985 onwards, there was usually at least one monitor collecting daily data, though not without some missing values.

A number of exposure measures have been used in previous studies. By an exposure measure we mean some function of present and previous days' pollution levels which is taken as an independent variable in the regression analyses. For example Schwartz and Dockery (1992a) used the average of current day's and previous day's TSP (total suspended particulates, the older measure of particles before PM_{10} became standard); Schwartz and Dockery (1992b) used just the previous day's TSP; Pope *et al.* (1992) used five-day averages of PM_{10} , and Styer *et al.* (1995) used three-day averages *including* the current day. For Birmingham, Schwartz (1993) used three-day averages of PM_{10} *excluding* the current day, i.e. today's exposure measure consists of the average PM_{10} for yesterday, the day before yesterday, and the day before that. As will be seen, the results are highly

sensitive to the choice of exposure measure, and in section 3 we consider the implications of this for our overall scientific assessment of the results.

One point, that came to light well after our initial work on the Birmingham data set was completed, is that not all of the PM₁₀ monitors are equally relevant to the study of mortality in Birmingham. Clyde (2000) drew attention to this feature and concentrated most of her analysis on data based on the monitor in the city of Birmingham, ignoring other nearby monitors which had been included in earlier studies. The present study uses, as alternatives, the average over all available monitors or the data from Birmingham alone, but it appears that the strongest associations are obtained using the monitoring data from Birmingham alone, so the main results are expressed in terms of that monitor.

2.2. Mortality data

Daily mortality data were available from the National Center for Health Statistics for 1985–1988. The data were classified in four ways: by gender (male/female), by race (black/non-black), by age (under 65/65+) and by cause of death (respiratory/cancer/circulatory/other disease/accidental). Although some of our work has been concerned with different age groups or with different causes of deaths, we have found that the strongest evidence of a PM₁₀-mortality association exists within the 65 and over age group, when all nonaccidental sources of mortality are combined into a single overall death count. From now on, therefore, we consider only that outcome variable.

2.3. Meteorological data

The Birmingham meteorological data for this study came from the U.S. National Climatic Data Center in Asheville, North Carolina. The data are publicly available through the ftp address ftp.ncdc.noaa.gov/pub/data/fsod/fsod_ascii.13876. This lists daily data at the BIRMINGHAM MUNI AP site, for which the WBAN number is 13876, and it is located at 33.57 °N and 86.75 °W at an elevation of 191m.

Although many meteorological variables are available on a daily basis, those actually adopted are those listed in Table 2.1. Further rationale will be provided in section 3 for the choice of these particular variables.

tmax:	daily maximum temperature (°C)
tmin:	daily minimum temperature (°C)
mntp:	mean of tmin and tmax
dptp:	average daily dew point temperature (°C)
mnsh:	average daily specific humidity (g/kg)
tg30:	larger of (tmax-30) and 0
mnshsq:	square of mnsh

Table 2.1: List of meteorological variables

2.4. Data summaries

Summary statistics for some of the main variables are in Table 2.2. Here “mort” is mortality in the 65+ age group excluding accidental causes. The variable “pm10” refers to daily readings and “pmmean” to three-day averages. Daily values were obtained either from the Birmingham monitor alone (these are denoted pm10^B or pmmean^B in the table) or from averages over all available monitors. Three-day averages were obtained by averaging the corresponding one-day values — if one or two of the one-day values was missing in any three-day period, the three-day average was based on the two days or one day remaining.

Variable	mean	SD	min	10%	25%	50%	75%	90%	max
tmax	23.39	8.61	-3.3	11.1	17.2	24.4	31.1	33.3	38.3
tmin	10.65	8.91	-12.2	-2.2	3.3	11.1	19.4	21.1	25.6
mns	9.31	4.89	0.8	3.1	4.9	8.7	14.0	16.2	18.3
pm10	47.23	23.77	8.0	21.0	29.0	44.0	59.3	79.0	163.0
pmmean	46.89	18.96	13.0	24.2	33.0	44.7	57.8	72.0	137.1
pm10 ^B	50.56	26.42	8.0	21.0	30.0	46.0	66.0	87.0	163.0
pmmean ^B	50.08	21.22	11.0	25.3	33.7	48.0	62.3	78.3	146.7
dptp	10.54	9.32	-19.1	-2.9	3.3	11.8	19.1	21.5	23.5
mort	15.06	4.25	3.0	10.0	12.0	15.0	18.0	21.0	32.0

Table 2.2: Selected summary statistics

Table 2.2 may be compared with Table 1 of Schwartz (1993), who gave a similar table for the data he used in his study. Apart from obvious transformations such as Fahrenheit and Celsius temperature scales, the meteorological data appear to be the same. The mean deaths reported in Schwartz’s study are a little lower than ours (e.g. Schwartz obtained 17.1 for mean daily nonaccidental deaths over all age groups; we obtained 19.8) which may have been due to his taking a slightly smaller geographical area. The deaths data used in the current study are for Franklin County, which includes Birmingham, whereas it appears that Schwartz used only data from the city of Birmingham. In the case of PM₁₀, there is a strange discrepancy in the number of days’ data: Schwartz reported 139, 332, 262 and 354 days’ data for respective years 1985, 1986, 1987, 1988, whereas our counts are 139, 332, 319 and 341 for the Birmingham monitor. We have no explanation for the extra 60 days’ data in 1987. So far as we can tell, Schwartz’s data set included other monitors besides Birmingham, though personal enquiries with Dr. Schwartz have failed to establish this definitively.

Despite these discrepancies in data sources, we do not believe that they are responsible for discrepancies in the results. For the most part, when models similar to those of Schwartz are also fitted to the present data, the results are very similar to those of Schwartz. The real points of difference are in the models being fitted, not in the data.

Plots of the data (see section 3) show that the largest part of the variability in the data is due to systematic seasonal and time trends, which do not appear to be directly related to either particles or meteorology. Schwartz concluded that these trends appear to follow a two-year cycle (based on four years' data), and modeled the cyclical trend by including 24 sine and 24 cosine terms in the regression equation, together with a linear trend term and an indicator variable for year. His principal fitting method was Poisson regression, using the generalized estimating equations approach developed by Liang and Zeger (1986). Numerous variants on this technique, including corrections for overdispersion and autocorrelation, robust estimation, least squares regression and nonlinear regression for the temperature and PM₁₀ components of the model, were used but apparently without greatly affecting the conclusions. The results consistently showed a statistically significant PM₁₀ effect. In Schwartz's papers this has often been characterized in terms of the relative risk associated with an increase in PM₁₀ levels of 100 $\mu\text{g}\cdot\text{m}^{-3}$. A typical result for this is 1.11, with 95% confidence interval of (1.02,1.20). Samet *et al.* (1995) obtained an exactly equivalent result based on refitting the same model as Schwartz to the same data, but without further consideration of model selection.

3 LINEAR REGRESSION ANALYSIS

3.1. Background

All analyses of the relationship between atmospheric pollution and human health outcomes rely on regression analysis in some form, but there are many variations in the precise methodology adopted. Since one of the purposes of this paper is to highlight how different approaches to the analysis may lead to very different conclusions, we begin with a brief overview of the main "issues" which arise in this kind of analysis, and how they have been resolved by earlier authors, before describing our own approach.

One of the first papers to discuss methodological issues in depth was the analysis by Schwartz and Marcus (1990) of London data from the 1960s. They highlighted such issues as

- correcting for long-term trends and seasonal variation,
- correcting for meteorological confounding,
- correcting for confounding by other atmospheric pollutants (in the case of particulate matter, the comparative effect of SO₂ is of particular interest),
- whether there is evidence of a nonlinear or "threshold" effect in the pollution-mortality relationship,
- the effect of autocorrelation in the residuals.

As interest in the subject developed, so did the methods become better fixed. Samet *et al.* (1995) considered a general structure of the form

$$\log \mu_t = \sum \beta_j x_{jt}, \tag{3.1}$$

where μ_t is the mean level of deaths on day t and the covariates $\{x_{jt}\}$ include terms representing trend and seasonality, meteorology and air pollution. Some of the models included nonlinear terms in the meteorological and air pollution variables. The obvious probabilistic model associated with (3.1) is a Poisson model in which the death count on day t is Poisson with mean μ_t , but Samet *et al.* (1995) rejected that because the data exhibit both overdispersion and autocorrelation. Instead, they proposed an “iteratively weighted and filtered least squares” (IWFLS) approach. In spite of their caution on this point, however, subsequent authors have often used Poisson maximum likelihood without finding overdispersion or autocorrelation to be of major concern.

Subsequently Samet *et al.* (1997) presented a more detailed analysis of one particular data set from Philadelphia, which included the use of smoothing splines to represent the long-term trend and seasonal components. In some analyses they also applied smoothing splines to nonlinear effects of meteorology and air pollution. In this analysis, they did not find any evidence of autocorrelation, which may have been because the smoothing splines approach is more sensitive to the true shape of the long-term trend than some of the approaches adopted in earlier papers. They considered synoptic approaches to the meteorological modeling, in which the weather variables were used to define a finite number of states representing typical weather scenarios, these states being then used as covariates in the analysis. However, their conclusion was that the synoptic approach is inferior to the more usual regression approach based on temperature and dewpoint as continuous covariates. Finally, they considered the joint effects of particulate matter and four other atmospheric pollutants (SO₂, NO₂, O₃ and CO), without finding strong evidence that particulate matter was the sole causative factor, as some other authors had claimed.

In a different and more exploratory spirit, Styer *et al.* (1995) also considered models of the form (3.1) as well as alternative linear regression models in which some of the nonlinear terms were represented nonparametrically as sample paths from stochastic processes. A particular feature of their approach was to examine the consistency of regression models fitted on either a seasonal or a monthly basis. For example, in analysis of data from Cook County, Illinois (in effect, the city of Chicago), they found a significant PM₁₀ effect when modeled as a single linear term, but when interactions between PM₁₀ and season were considered, they found strong interaction, with the PM₁₀ effect strong only in the fall. Such an odd conclusion is hard to reconcile with a strict causal interpretation. They also found no PM₁₀ effect at all in an analysis of data from Salt Lake County, Utah. On the other hand, Smith *et al.* (1999), re-analyzing the Chicago data using smoothing splines for the long-term trend, did not find the same strength of evidence for a seasonal interaction, which serves to highlight again the sensitivity of the results to different methods of analysis. There is also a question left open by the Styer *et al.* analysis, of whether fitting entirely separate regressions on a seasonal or even monthly basis is simply overfitting, with inevitable difficulties for interpretation of the results.

For the present analyses, we consider both Poisson maximum likelihood analyses based on (3.1), and a simpler approach based on ordinary linear regressions of the form

$$y_t = \sum \beta_j x_{jt} + \epsilon_t, \quad (3.2)$$

where y_t is some transformation (e.g. log or square root) of daily death count on day t , and $\{\epsilon_t\}$ are independent normal random errors with mean 0 and common variance σ^2 . Experience with both Birmingham and other data sets has shown that models based on (3.2) lead to virtually equivalent results to (3.1) after due allowance is made for the transformation. The square root transformation is natural in view of the well-known result that for Poisson random variables, this is variance stabilizing — if z_t is Poisson with mean ν_t then $y_t = \sqrt{z_t}$ has mean $\mu_t \approx \sqrt{\nu_t}$ and variance approximately $\frac{1}{4}$ regardless of the true value of ν_t .

We now turn to detailed specification of individual components of the model, using the Birmingham data to illustrate how the model is constructed in practice.

3.2. Trend and seasonality

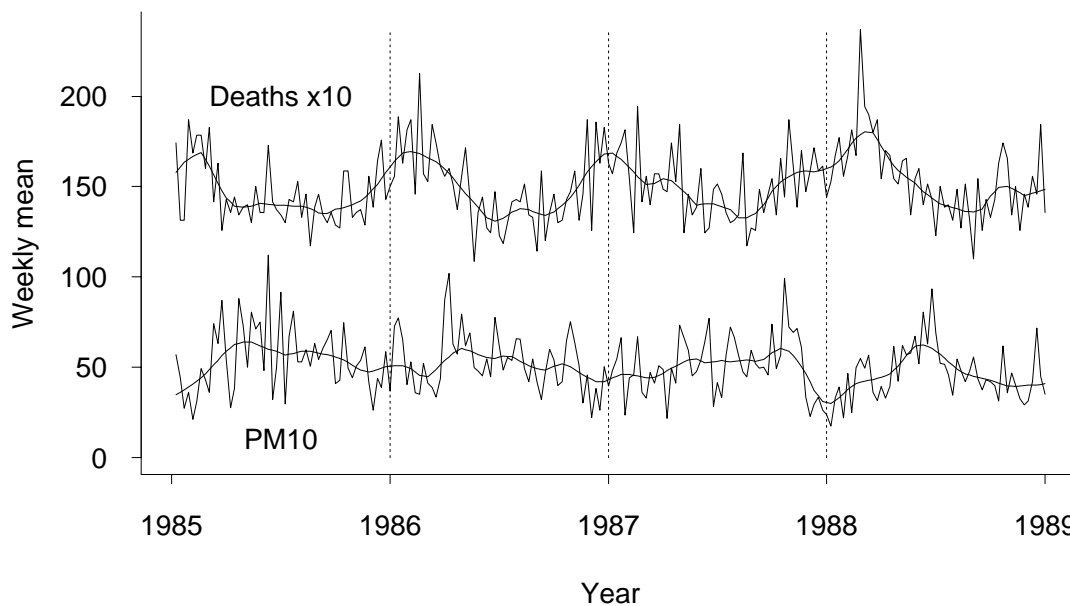


Figure 3.1. Weekly means of the daily nonaccidental death counts among the 65+ population in Birmingham (top plot) and weekly PM_{10} means (bottom plot), together with a scatterplot smoother for each.

Fig. 3.1 shows weekly deaths and weekly PM_{10} averages plotted throughout the four-year period, with a smoothed curve to represent the long-term trend (fitted by `lowess` in `S-PLUS`). The plots show clearly that there are seasonal effects, but also that they are irregular from year to year. For example, each year the deaths peak in the winter, but in 1985–6 and 1987–8, the peak occurred in February, whereas in 1986–7 the peak occurred at the beginning of January. These facts, which are almost certainly due to epidemics

rather than the effects of either meteorology or air pollution, highlight the need for careful modeling of the trend. Another feature of Fig. 3.1 is that PM_{10} appears to peak in the summer, so a naïve regression of mortality on PM_{10} might well lead to a negative effect (in fact, the correlation between the two series of weekly totals in Fig. 3.1 is -0.1) but of course this is also highly misleading given the number of other factors likely to explain seasonal variability.

To model the seasonal effect, Schwartz (1993) cited the well-known tendency of epidemics to follow a two-year cycle, and modeled the trend as such, using 24 sine and 24 cosine basis functions. The B-spline approach of the present paper is based on representations of the form

$$f(t) = \alpha_0 + \sum_{k=1}^K \alpha_k B \left\{ \frac{K}{T}(t - \tau_k) \right\}, \quad 0 \leq t \leq T, \quad (3.3)$$

where $f(t)$ is the trend at time t , $\tau_k = \frac{T}{K} (k - \frac{1}{2})$ is the k 'th “knot”, and $B(\cdot)$ is the B-spline basis function, see e.g. Green and Silverman (1994).

The key decision in (3.3) is the number of knots K , which controls the smoothness of the curve — the smaller K is, the smoother the curve. General considerations are that K should be large enough to pick up the kind of irregular seasonal behavior which we saw in Fig. 3.1, but not so large that the seasonal effect could be confounded with the PM_{10} effect. Since air pollution episodes frequently last for several days and possibly as long as two or three weeks, the latter consideration suggests that the frequency of knots $\{\tau_k\}$ should be no greater than about one per month.

3.3. Meteorological variables

On the selection of meteorological variables, there is again a diversity of views in the literature. Most authors agree on the need to include temperature and some measure of humidity and many also include pressure. For example, Styer *et al.* (1995) used temperature, specific humidity and station pressure, together with their one- and two-day lagged values, while Samet *et al.* (1995, 1997) based their analyses on temperature and dewpoint, the latter being another way to measure the effects of humidity. For Birmingham, Schwartz (1993) considered temperature and dewpoint as candidate meteorological variables, though the main model he fitted did not use dewpoint. Our own investigations suggest that it is highly desirable to include at least one of specific humidity and dewpoint in the model, along with temperature.

The main variables examined in the current study are daily maximum and minimum temperature, and one measure of humidity (either dewpoint or specific humidity). One might expect that the mortality vs. temperature relationship would be nonlinear (decreasing at low temperatures, increasing at high). This could be modeled using a quadratic or higher-order polynomial, or alternatively through a piecewise linear model of the form

$$f(x) = \beta_1 x + \beta_2 x I(x > x_0), \quad (3.4)$$

where x is temperature, I is the indicator function and x_0 is some threshold. Preliminary analysis suggested that an effective way to deal with nonlinearity was through a model of the form (3.4) for $x = \text{tmax}$, with $x_0 = 30$ °C. This motivated the definition of the variable tg30 in Table 2.1.

The question also arises whether the effect of specific humidity should be nonlinear and for this we find that a quadratic fit is good. Further details are given below.

Finally, all the meteorological variables were lagged up to 4 days to allow for effects that may persist over several days. The lag is indicated with a suffix; for example, tmax_2 means tmax lagged two days, and mnshsq_0 means the lag 0 (in other words, today's) value of mnshsq .

3.4. Selecting the meteorological and trend model

The first step of the analysis is to find the best fit to the data using just the trend and meteorological variables: later, we add terms involving PM_{10} individually to ascertain the significance of those terms.

For trend and seasonality, the approach taken here is first to fix K , the number of spline basis function for the trend, and then to select meteorological variables. The method of selection is backward selection using tests at significance level 0.1 to determine which variables to omit from the model. The results are repeated for several values of K , and using three transformations of daily death count (square root, logarithmic and no transformation) to determine y_t in (3.2). A scale correction was applied to ensure that the mean squared error results are directly comparable (Atkinson 1985, p. 86). Models are then compared by AIC. Results are in Table 3.1. Also included in this table are the corresponding results based on Schwartz's (1993) sine-cosine representation of the trend.

The results of Table 3.1 show convincingly that the square root transformation performs best among the three transformations considered, and also that $K = 12$ is best among the values of K considered. Although AIC is just one of several model selection criteria we could have considered, others lead to essentially the same conclusion. The results for $K = 0$ are included following a suggestion of the referee, who asked whether the whole of the seasonal variation could be modeled as a function of meteorology. The answer is no, though surprisingly, $K = 0$ performs second best (after $K = 12$) of the various values in Table 3.1. As a further check on this conclusion, significance tests for the alternative hypothesis $K = 12$ have been performed against the null hypothesis $K = 0$, using common sets of meteorological covariates to make the models nested. In every case tested, the result is significant with a p -value of the order of .01–.03. Thus we conclude that $K = 12$ is the best.

K	Square root trans.		Log trans.		No trans.	
	p	AIC	p	AIC	p	AIC
48	55	4037.87	55	4106.26	54	4066.87
40	47	4028.72	47	4096.27	46	4058.24
32	39	4022.41	39	4088.49	39	4054.08
24	30	4017.93	36	4083.99	28	4049.47
20	34	4018.03	34	4084.02	27	4048.98
16	23	4013.49	23	4079.76	23	4044.99
12	19	4007.64	19	4074.14	19	4038.97
8	15	4014.29	16	4078.89	19	4048.39
0	9	4013.46	9	4076.05	9	4048.11
SC	55	4053.96	55	4120.69	55	4084.60

Table 3.1: Number of parameters p and AIC values for trend+meteorology models in which trend is modeled by a B-spline representation with K knots and meteorological variables are selected by backward variable selection. Three transformations (square root, logarithmic and none) are considered. Also tabulated are the results for Schwartz's sine-cosine representation (SC).

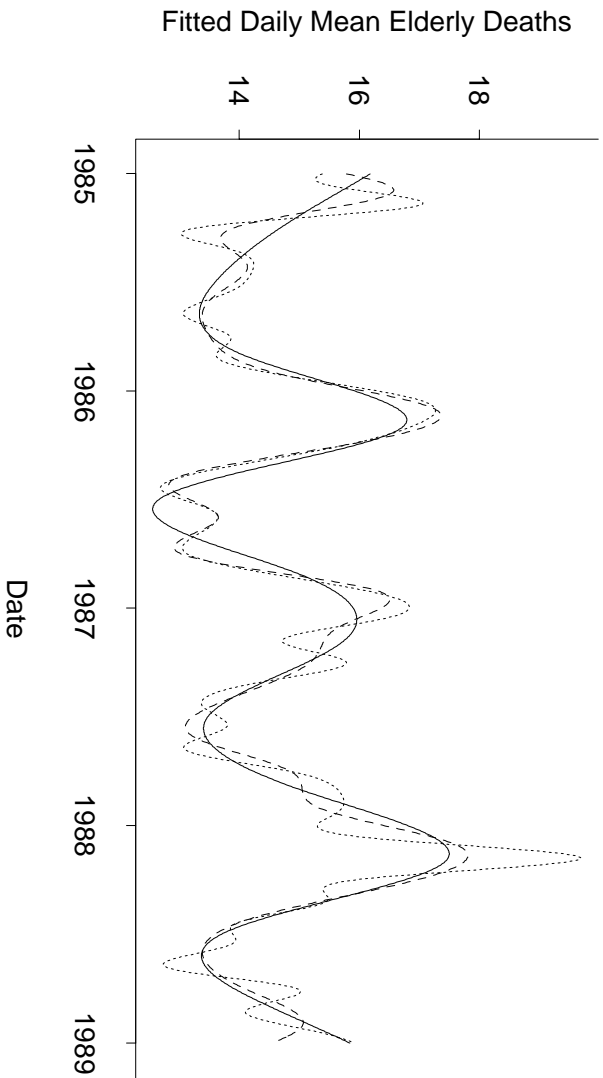


Figure 3.2. Modeling the trend in daily mean deaths (without meteorological component) using different values of K : $K = 12$ (solid curve), $K = 24$ (dashed curve), $K = 48$ (dotted curve). The curve for $K = 12$ shows a clear seasonal pattern with a well-defined peak in each year, whereas the others show local fluctuations which seem likely to be spurious.

In previous analyses of Birmingham, Smith *et al.* (1998, 1999) have used larger values of K so the present analysis represents a departure from that. As a further assessment, Fig. 3.2 shows the fitted smoothed trends by the B-spline method for three values of $K = 48, 24$ and 12 , in this case fitted without meteorological components (otherwise these curves are difficult to compare visually). The curves for $K = 24$ and $K = 48$ show numerous local fluctuations which seem likely to be spurious. This therefore provides some visual confirmation of the AIC-based results.

So far, although we have selected meteorological variables by backward selection, the class of meteorological variables considered for variable selection has remained fixed. However, other issues which arise in these analyses include whether humidity is best represented by specific humidity or dewpoint, or whether humidity variables should be included at all. To test this, Table 3.2 shows AIC values for the best model in four classes, (a) using temperature variables only, (b) using temperature plus linear terms in specific humidity, (c) using temperature variables plus dewpoint temperature, (d) using temperature variables plus linear and quadratic terms in specific humidity. All the models were obtained by backward variable selection of meteorological variables, with a fixed $K = 12$ knots for the B-spline representation of trend. As judged by AIC, model (d) is clearly the best of these. This highlights the need to include humidity as well as temperature.

Model	AIC	df
(a) $t_{\max_1}, t_{\min_3}, tg30_1$	4022.46	15
(b) $t_{\min_0}, t_{\min_3}, tg30_1, mnsh_0, mnsh_2, mnsh_3$	4016.67	18
(c) $t_{\min_3}, tg30_1, dptp_2, dptp_3$	4016.59	16
(d) $t_{\min_3}, mnsh_0, mnsh_1, mnsh_2, mnsh_3, mnshsq_0, mnshsq_1$	4007.64	19

Table 3.2: Best meteorological models in four classes

3.5. Adding particulate matter to the model

As remarked already, previous authors have used a variety of exposure measures for PM_{10} . For example Styer *et al.* (1995) used three day averages *including* the current day, while Schwartz (1993) used three day averages *excluding* the current day. In the following discussion, we attempt to be systematic in searching for a suitable variable or variables to represent PM_{10} . We define variables pm_0, pm_1, \dots, pm_4 to represent the one-day PM_{10} reading lagged up to 4 days, and we used pm_{mean} with appropriate lags to represent three-day means. Thus pm_{mean_0} denotes the average of pm_0, pm_1 and pm_2 , while pm_{mean_1} denotes the average of pm_1, pm_2 and pm_3 . The five-day average was also considered but is not reported since in no model was it found to be statistically significant. The analyses in this subsection are based on the same meteorological variables and B-spline basis functions as previously, but the analysis is restricted to the period August 3 1985 – December 31 1988 for which there is an almost continuous series of monitoring data.

Variable	Estimate	Standard error	t statistic
(i) pmmean ₀	.00136	.00082	1.65
(ii) pmmean ₁	.00192	.00082	2.34
(iii) pm ₀ ,...,pm ₄	.00108	.00130	0.83
(iv) pm ₀ , pm ₁ , pm ₃	.00117	.00106	1.10
(v) Poisson version of (ii)	.00098	.00040	2.44
(vi) Poisson version if (iv)	.00057	.00052	1.10

Table 3.3: Parameter estimates and standard errors due to PM₁₀ in the model based on temperature plus linear and quadratic terms in specific humidity.

Table 3.3 shows the coefficient, standard error and t statistic for several different ways of introducing PM₁₀ into the model of Table 3.2(d). These are based on one-day or three-day averages computed using the Birmingham monitor alone, as discussed earlier. It should be pointed out that a number of other combinations of the PM₁₀ variable were considered — single-day values, two-day averages and three-day averages at various lags — but no single variable produced as strong an association as pmmean₁ here, for which the t statistic is 2.34, a significant result if the model selection aspect of the analysis is ignored. Row (iii) shows the effect of putting in each of the five single-day variables pm₀,...,pm₄ and using the sum of the five coefficients along with the standard error of that sum as a measure of the overall PM effect. Measured this way, the overall effect is weaker than in row (ii), and not statistically significant. However within this analysis, pm₂ and pm₄ were not significant, the remaining statistically significant coefficients being pm₀ with a negative coefficient, and pm₁ and pm₃ with positive coefficients. Therefore, row (iv) repeats the analysis of row (iii) but based just on pm₀, pm₁ and pm₃. Again it does not show a statistically significant result. Finally, rows (v) and (vi) repeat the analysis of rows (ii) and (iv) but based on the Poisson regression model (3.1). The coefficients in this case are not directly comparable with those based on model (3.2) because the transformation of y is different (y_t in (3.2) is taken as square root of elderly nonaccidental deaths) but the two sets of results may be compared by translating them into relative risks.

To calculate relative risks associated with the parameters in rows (i)–(iv), the simplest way to illustrate is by example. The mean deaths per day are 15.055. Suppose the model of row (ii) holds and consider the effect of a 10 $\mu\text{g}/\text{m}^3$ rise in PM₁₀. According to the coefficient, this should lead to an increase of $10 \times .00192$ in mean square root of deaths. Thus the estimated relative risk is $(\sqrt{15.055 + .0192})^2/15.055 = 1.010$ and an associated approximate 95% confidence interval is (1.001, 1.018) after applying the same calculation to both endpoints of a 95% confidence interval for the coefficient. In contrast, the estimated relative risk based on model (v) is $e^{.0098} = 1.010$ with 95% confidence interval (1.002, 1.018), virtually the same. Similarly, the estimated risk for model (iv) is 1.006 with 95% confidence interval (.995, 1.017) and these are identical to the estimates based on model (vi).

Table 3.4 shows the results comparable to Table 3.3 in which the meteorological modeling is based on Table 3.2(a), i.e. temperature alone. This table was calculated because one of the conclusions from Schwartz’s (1993) analysis was that it does not make any difference whether humidity is included or not in the model (though Schwartz used dewpoint rather than specific humidity in making the assertion). Whether the differences between Tables 3.3 and 3.4 are of any practical significance is a matter for debate, but it can be seen that the coefficients in Table 3.4 are uniformly higher than in Table 3.3. In other words, omitting humidity may have had the effect of inflating Schwartz’s estimates.

Variable	Estimate	Standard error	<i>t</i> statistic
(i) pmmean ₀	.00158	.00085	1.87
(ii) pmmean ₁	.00217	.00088	2.46
(iii) pm ₀ ,...,pm ₄	.00163	.00132	1.23
(iv) pm ₀ , pm ₁ , pm ₃	.00139	.00108	1.29
(v) Poisson version of (ii)	.00111	.00043	2.61
(vi) Poisson version if (iv)	.00070	.00053	1.32

Table 3.4: Parameter estimates and standard errors due to PM₁₀ in the model based on temperature alone.

If these analyses are repeated using average PM₁₀ values from all the monitors rather than just the values from the Birmingham monitor, the PM coefficients are uniformly smaller. For example, the Poisson coefficients from rows (v) and (vi) of Table 3.3 become .00088 and .00049 (instead of .00098 and .00057), while those from Table 3.4 become .00100 and .00058 (instead of .00111 and .00070). Thus, the decision to use just the Birmingham monitor has had the effect of increasing the estimated effects compared with earlier analyses based on this data set.

Variable	Coefficient	S.E.	<i>t</i> statistic
tmin ₃	-.00270	.00330	-0.82
mnsh ₀	-.03309	.01315	-2.52
mnsh ₁	.00748	.01422	0.53
mnsh ₂	-.00807	.00682	-1.18
mnsh ₃	.00792	.00698	1.13
mnshsq ₀	.00184	.00070	2.61
mnshsq ₁	-.00068	.00072	-0.94
pm ₀	-.00070	.00039	-1.77
pm ₁	.00074	.00039	1.89
pm ₃	.00053	.00034	1.55

Table 3.5: Individual coefficients and standard error, Poisson model with individual coefficients for single-day PM₁₀ effects

Tables 3.5 and 3.6 give detailed regression results under the Poisson model for rows (vi) and (v) of Table 3.3. In Table 3.5, the coefficients of pm_0 , pm_1 and pm_3 are all near statistical significance, the coefficient of pm_0 being negative and the other two positive. When the three coefficients are added, the combined result is positive but not statistically significant, as is reflected in the RR results in rows (iv) and (vi) of Tables 3.3 and 3.4. It would appear that the reason why the analysis with $pmmean_1$ is the only one to give a statistically significant result is because this choice of exposure measure excludes the effect of day 0 which is negative.

An incidental comment about Tables 3.5 and 3.6 is that several of the meteorological variables do not appear significant here. However, the original meteorological variable selection was done without any PM terms present; we have not chosen to drop parameters which became insignificant at this stage of the analysis.

Variable	Coefficient	S.E.	<i>t</i> statistic
$tmin_3$	-.00351	.00281	-1.25
$mnsh_0$	-.03815	.01116	-3.42
$mnsh_1$.01947	.01195	1.63
$mnsh_2$	-.01047	.00578	-1.81
$mnsh_3$.01089	.00595	1.83
$mnshsq_0$.00249	.00060	4.15
$mnshsq_1$	-.00148	.00061	-2.41
$pmmean_1$.00098	.00040	2.44

Table 3.6: Individual coefficients and standard error, Poisson model based on three-day averages of PM_{10}

The differences among the estimates from the different models do not appear to result from whether we use normal or Poisson regression, which lead to very similar estimates for the relative risk, but on which variables are included in the regression. The most substantial source of discrepancy arises from whether we choose to use $pmmean_1$ as the single most significant measure of the pollution effect, or estimate separate one-day values as in Table 3.5. There is no clear-cut judgement about which is the right model, but it is important to understand that the conclusion is sensitive to this judgement. The latter model may be thought of as a crude form of “distributed lag” model, in which the time-dependent shape of the response curve is assumed *a priori* unknown, and ultimately we feel that models of this form are likely to be the way forward with this kind of analysis.

Our finding about pm_0 should not be interpreted as meaning that the current day’s PM_{10} has a protective effect. It seems much more plausible that it is an artifact caused by linear correlations among the different variables. Nevertheless it seems to be a persistent phenomenon. For example Roth and Li (1996) noticed the same thing, computing the pm_0 coefficient separately for each year of data from 1988 to 1993, and in only one of the six

years was it positive. However this has only a one-year overlap (1988) with the present data set.

One difficulty with including multiple PM_{10} effects in the analysis is that they magnify the problems of missing data. Out of 1,247 days covered by the data, 147 had missing single-day values, but a total of 300 are missing at least one of pm_0 , pm_1 or pm_3 and were therefore omitted from the analyses just reported. There is no evidence that this has any effect on the parameter estimates but it does mean that the degrees of freedom for error are smaller, and it makes other kinds of calculations more difficult, e.g. serial correlations among the residuals. For this reason and for better comparison with previously published results, much of the subsequent discussion is still based on the model using $pmmean_1$ as the exposure variable. The variable $pmmean_1$ is considered missing only if all three days in a row are missing, which occurred only four times.

3.6. Diagnostics

For the normal linear regression model (3.2), the residual for day t is defined in the usual way, as

$$e_t = y_t - \sum_j \hat{\beta}_j x_{jt} \quad (3.5)$$

where the $\hat{\beta}_j$ are estimated coefficients. For the Poisson model (3.1), if $\hat{\mu}_t$ is the estimated Poisson mean for day t and if y_t is the observed value, then we can define the transformation

$$u_t = \sum_{y=0}^{y_t-1} \hat{p}_t(y) + \frac{1}{2} \hat{p}_t(y_t) \quad (3.6)$$

where $\hat{p}_t(y) = \hat{\mu}_t^y e^{-\hat{\mu}_t} / y!$, the fitted Poisson probability distribution for day t . The formula (3.6) is intended as an approximation to the probability integral transformation, so that u_t are approximately uniformly distributed on $[0,1]$. By further transformation, we also obtain approximately normal residuals, analogous to (3.5). For the following discussion, we concentrate on the normal and Poisson models based on $pmmean_1$, as in rows (ii) and (v) of Table 3.3.

One issue that arises is *overdispersion*. If y_t has a Poisson distribution, then the variance of $\sqrt{y_t}$ is approximately $\frac{1}{4}$. This should also apply approximately to the residuals in (3.4) if the linear regression is a reasonable representation of reality. In the regression of Table 3.3(ii), which we believe to be representative of all the regressions we have considered, the standard unbiased estimator of σ^2 is $s^2 = 0.269$ with 1,222 degrees of freedom. This represents an approximately 7.5% overdispersion, and is not statistically significantly different from 0.25 based on the χ^2 distribution of s^2 . Based on this, we conclude that overdispersion is not a serious problem.

A second issue which sometimes arises in these kinds of analyses is the possibility that residuals may be serially correlated. In the present case, the first few serial correlation

coefficients of the residuals are $-.016, -.017, -.017, \dots$. These should be compared with the standard reference point for correlation coefficients, $2/\sqrt{N} = .057$. We conclude that there is no evidence of serial correlation. For the normal-transformed residuals from the Poisson model, the serial correlations are $-.014, -.015, -.015, \dots$ which leads to the same conclusion.

Roth and Li (1996) did find significant serial correlation in their residuals and accommodated this with autoregressive models of order up to 4. However their modeling of seasonal trend was less detailed than ours. In general, our experience with this and similar data sets has been that if the model for trend is inadequate to pick up all the fluctuations in the underlying death rate, then this leads to significant serial correlations. One advantage of the B-splines approach is that it seems to model trend sufficiently accurately to make autoregressive models unnecessary.

A third diagnostic issue is whether the assumed distribution (normal or Poisson) adequately fits the data. There is no contradiction in testing both distributions because the Poisson distribution for large enough mean would be hard to distinguish from a normal distribution, a direct consequence of the Central Limit Theorem. However, so far we have not presented any evidence that either distribution fits the data.

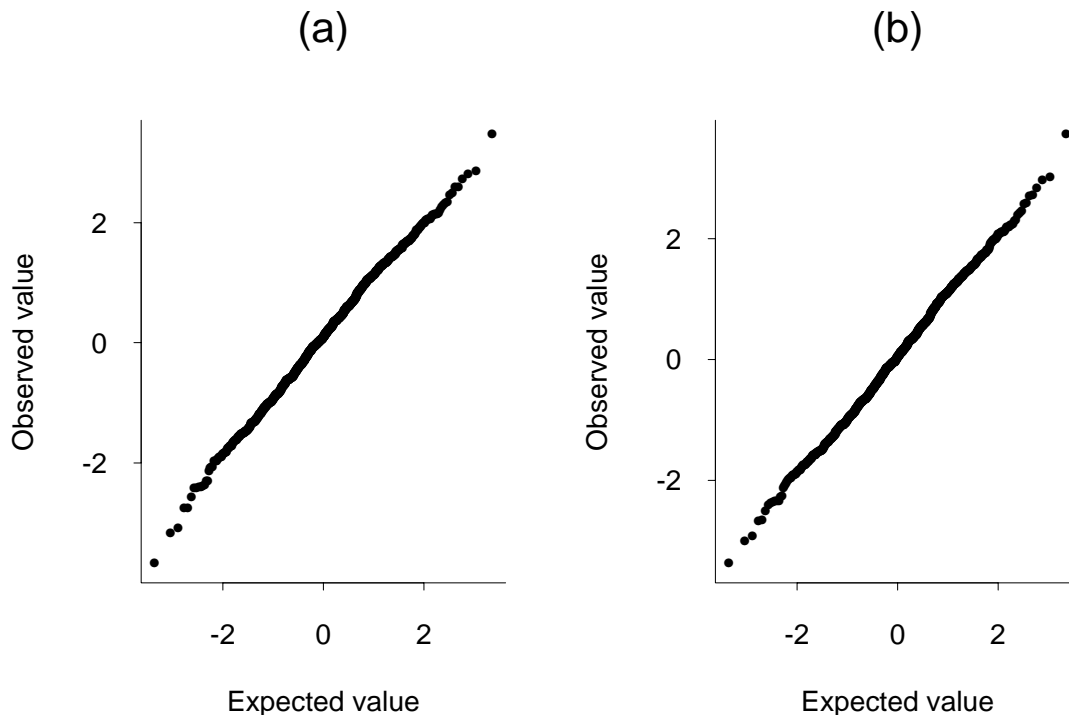


Figure 3.3. Probability plot for standardized residuals under (a) normal model, (b) Poisson model.

A standard method of deciding whether residuals follow a normal distribution is as a probability plot of the order statistics against their expected values. This is done in Fig.

3.3(a) for the OLS regression model and Fig. 3.3(b) for the Poisson regression model. In Fig. 3.3(a), the residuals are first standardized to have variance 1, while Fig. 3.3(b) is based on the normal-transformed version of the residuals from the Poisson model. In both plots the data are close to the solid line which is evidence that the model is indeed a good fit.

The final “diagnostic” issue we consider is whether the estimated PM_{10} coefficient varies from year to year or from season to season. Roth and Li (1996) found significant variations from year to year in the case of Birmingham. The possibility of seasonal variation was not investigated by them, but other studies of similar data sets have shown that it can be important, e.g. Styer *et al* (1995).

We confine our discussion here to the model based on $pmmean_1$, since in this case there is only one parameter to vary. Suppose the model of Table 3.6 is fitted, but with separate $pmmean_1$ coefficients for each year or for each season. Table 3.7 shows the results. There is some slight suggestion of a variation in the PM coefficient both from year to year and from season to season, but in neither case is the variability statistically significant when measured by a deviance test (deviance statistics 0.8 and 1.9 respectively for yearly and seasonal interactions, each with 3 d.f.). The one possibly interesting conclusion is that the PM_{10} effect does appear to be substantially weaker in the summer than in the other seasons, and it would be interesting to see whether this conclusion is sustained over a longer time span in Birmingham or in other cities with a similar climate.

	Coefficient	S.E.	t value
(a) Variation by year			
1985	0.00037	0.00108	0.35
1986	0.00074	0.00057	1.28
1987	0.00126	0.00062	2.03
1988	0.00115	0.00074	1.54
(b) Variation by season			
Spring	0.00110	0.00057	1.92
Summer	0.00050	0.00058	0.86
Fall	0.00111	0.00053	2.11
Winter	0.00106	0.00058	1.84

Table 3.7: Variation of the $pmmean_1$ coefficient by year or by season, for the model of Table 3.10

In summary, we have examined the fit of the linear regression models for four possible departures from the underlying assumptions: overdispersion, serial correlation, lack of fit of the distribution and seasonal or annual variation in the coefficients. In no case do we find any reason to reject the original assumptions.

4 THRESHOLD EFFECTS

A critical question for the whole particulates-mortality debate is whether there exists a threshold below which there is no effect. This question can only be addressed through some form of nonlinear modeling. Schwartz (1993, Fig. 6) estimated a smooth nonlinear curve for the PM₁₀-mortality relationship but did not calculate any confidence band. Here, we re-visit this question from three alternative points of view.

The first new analysis is based on a simple piecewise linear regression function for the dependence on particulates. An early example of such analysis was that of Ostro (1984), who used it to examine the possibility of a threshold in data from London.

We use the same idea, but for a sequence of thresholds. Consider the functions,

$$\begin{aligned} f_1(x) &= \begin{cases} x - u & \text{if } x < u \\ 0 & \text{if } x \geq u \end{cases}, \\ f_2(x) &= \begin{cases} 0 & \text{if } x < u \\ x - u & \text{if } x \geq u \end{cases}, \end{aligned} \tag{4.1}$$

and consider a function of response variable y (typically, square root of daily deaths) against PM₁₀ level x of the form $\beta_1 f_1(x) + \beta_2 f_2(x)$ plus other terms depending on other covariates. Thus, the PM₁₀ relationship is piecewise linear with a discontinuity in the slope (though not the function itself) at a threshold u . For this analysis, the PM₁₀ was represented by pmmean₁ and the other covariates were as in Table 3.6. Thus β_1 and β_2 represent slopes of the regression below and above the threshold.

Fig. 4.1 shows the resulting estimates of the PM₁₀ effect, with pointwise 95% confidence bands, for four possible choices of the threshold u . At $u = 40$, the results show no evidence of a change of slope either side of the threshold. At $u = 60$ and $u = 80$, the estimated slope is clearly higher above the threshold than below. The effect has disappeared again by $u = 100$, but for this high threshold, the standard error of the slope above the threshold is very large. Of the four threshold plots, only at $u = 100$ do the results below the threshold provide nearly significant evidence of an increasing slope in that region.

Fig. 4.2 shows an alternative treatment based on B-splines for the estimation of a nonlinear effect. The results are expressed as relative risks, relative to the long-term mean of PM₁₀, which is 50.5. In other words, the relative risk for the level 50.5 is defined to be 1, and everything else is defined from that. To achieve this, we represent the PM₁₀ effect in the form

$$f(x) = \sum_{k=1}^{K'} \beta_k \{B_k(x) - B_k(\bar{x})\}, \tag{4.2}$$

in which x denotes the measured variable (here pmmean₁), \bar{x} is the sample mean of all x values (here 50.5), K' is the number of knots in the B-spline representation for $f(x)$, and

$B_1, \dots, B_{K'}$ are centered and normalized B-spline basis functions, analogous to (3.3). This is then translated to relative risk via

$$RR = \frac{(\sqrt{15.055} + f(x))^2}{15.055}, \quad (4.3)$$

analogous to the calculation following Table 3.3. Pointwise confidence bands are calculated by calculating a 95% confidence interval on each $f(x)$ from (4.2), and translating to relative risks via (4.3). This leads to the plots in Fig. 4.2. Two plots are calculated, one using five basis functions with knots at 20, 50, 80, 110, 140, the other using seven basis functions with knots are 20, 40, 60, 80, 100, 120, 140.

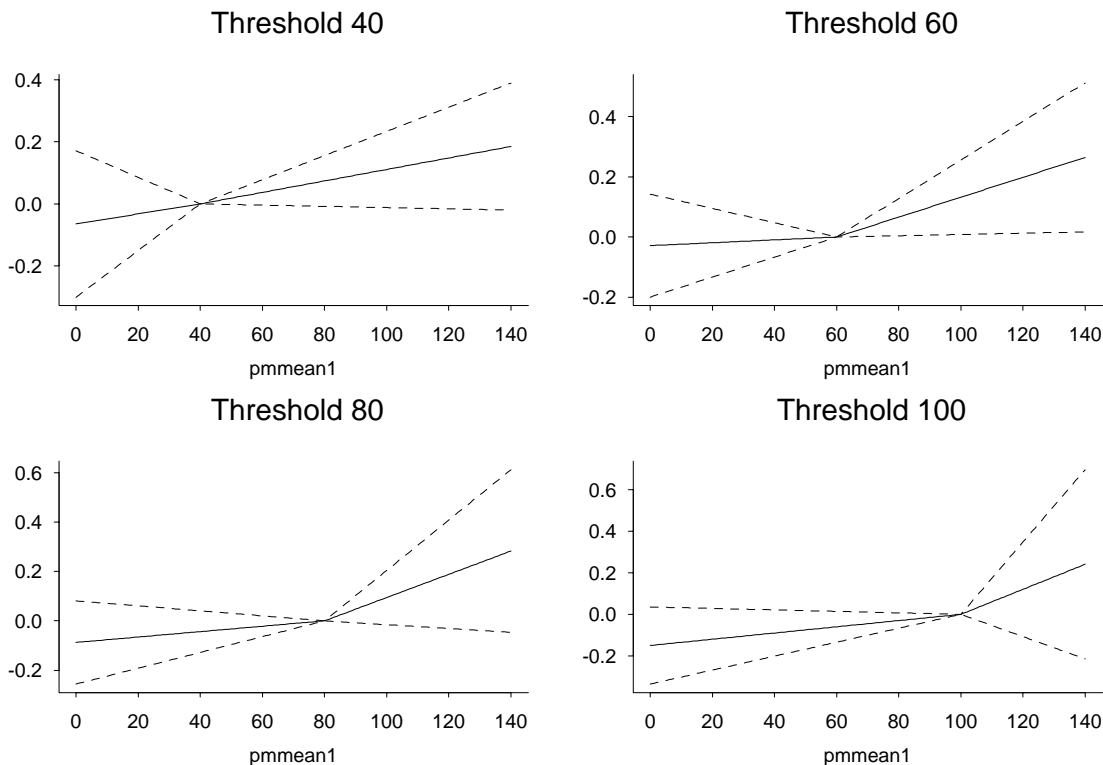


Figure 4.1. Piecewise linear effects for mortality on three-day-averaged PM₁₀, relative to four different thresholds, with pointwise 95% confidence bands.

The confidence bands calculated by this method are inevitably very wide, but they do give a rough indication of the statistical significance of the fluctuations in the function. In particular, the fluctuations in the observed function below the mean \bar{x} do not seem at all significant, whereas the increase in RR observed above the level of 60 $\mu\text{g}/\text{m}^3$ seems strong enough not to be entirely accidental.

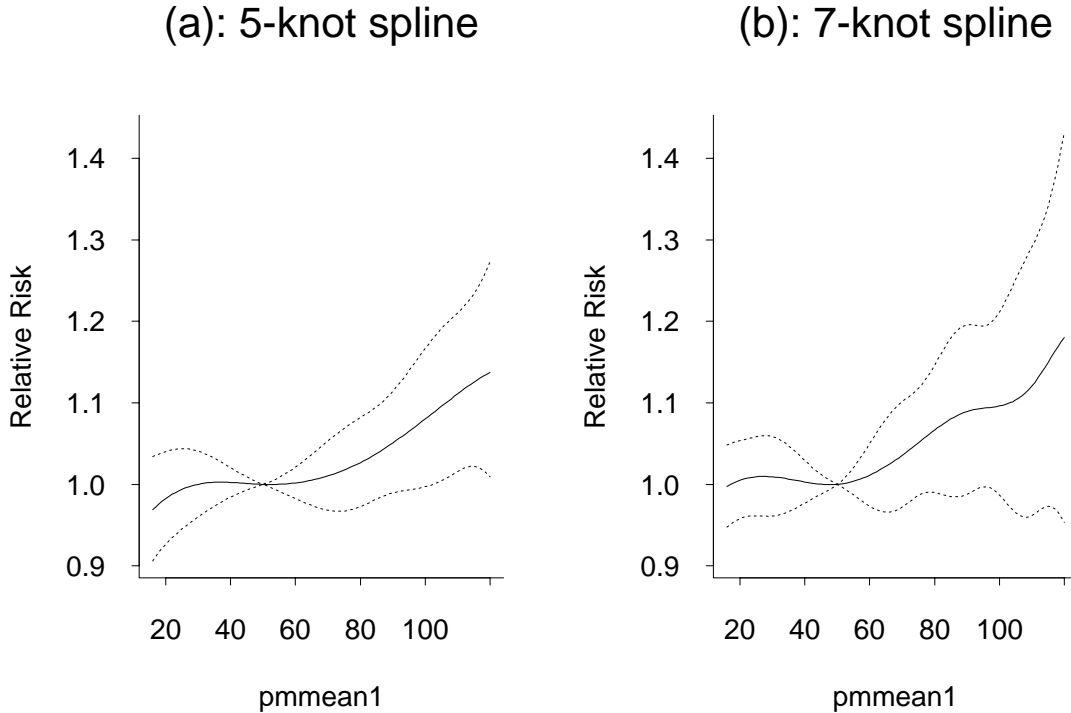


Figure 4.2. Smooth nonlinear curve for the relative risk of mortality on three-day-averaged PM_{10} , computed using either a 5-knot or a 7-knot cubic spline, with pointwise 95% confidence bands.

A third possible approach is a semi-Bayesian approach to “detection” of a threshold. Suppose we re-fit the model based on piecewise linear functions (4.1), but with f_1 having coefficient 0 — in other words, we are now assuming that there is no PM effect below the threshold u . Suppose, in (3.2), the vector of regression parameters conditional on a given value of u is denoted $\beta^{(u)}$, and the residual error variance is $\sigma^{2(u)}$. If we assume a joint prior density for $u, \beta^{(u)}, \sigma^{2(u)}$ of the form

$$\pi(u, \beta^{(u)}, \sigma^{2(u)}) \propto \frac{1}{\sigma^{2(u)}}, \quad 0 \leq u \leq u_{\max}, \quad \sigma^{2(u)} > 0, \quad (4.4)$$

with u_{\max} some prescribed upper limit on the values of u , then it may easily be verified that the marginal posterior density for u is of the form

$$\pi(u \mid \text{data}) \propto G(u)^{n-p}, \quad (4.5)$$

where n is the number of data points, p the number of regressors in the model (3.2), and $G^2(u)$ is the usual residual sum of squares after performing the regression for fixed u . This approach begs the question of what is really an appropriate prior distribution for u (that is why we call it only semi-Bayesian), but the posterior distribution based on (4.5) may provide a useful indication of the information available in the data to support different values of u .

Fig. 4.3 shows the posterior density for u calculated under this approach. This is based on assessing (4.5) at discrete values $u = 20, 22, 24, \dots, 120$ and normalizing so that the integral of the posterior density is 1. The result shows a peak in the posterior density near $u = 65$, but it is also noticeable that the posterior density does not tend to 0 as u approaches its lower endpoint (taken here to be $u = 20$ because of the very small number of data points for which $pmmean_1$ is below 20). In other words, the analysis provides support for a threshold in the region $55\text{--}75 \mu\text{g}/\text{m}^3$, but it also shows that the possibility of a threshold near 0 cannot be dismissed.

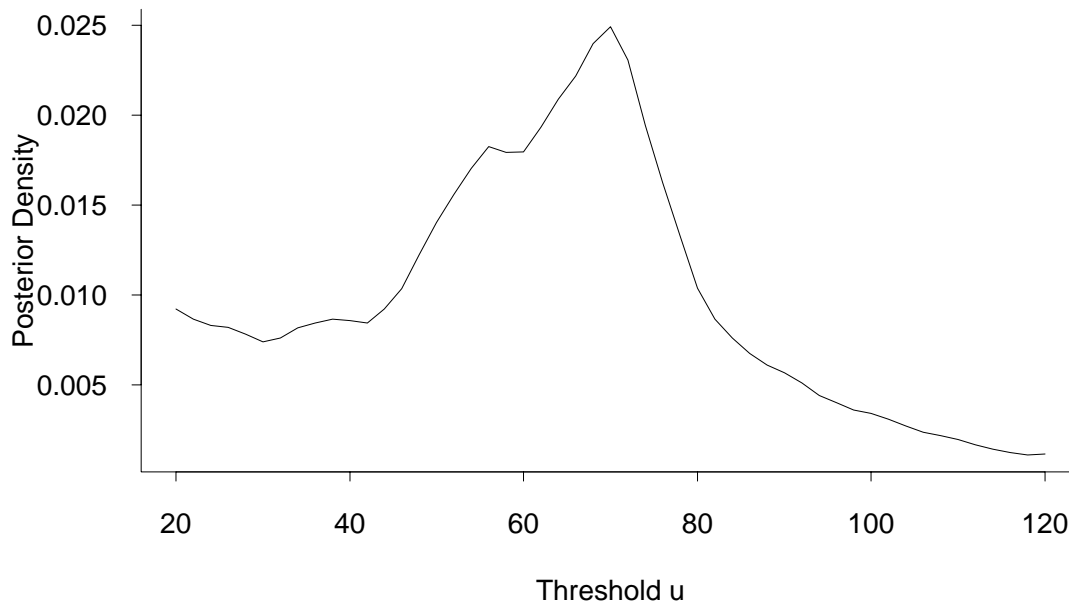


Figure 4.3. Bayesian calculation of the posterior probability for the location of a threshold.

In conclusion, the results of this section support the existence of a nonlinear effect or a threshold above the current annual mean NAAQS of $50 \mu\text{g}/\text{m}^3$. None of the results provide statistically significant “proof” either that the effect is nonlinear or that there is a threshold effect, if the question is formulated from the point of view of testing the significance of a nonlinear *vs.* a linear effect. However, we question whether that is the right formulation. Much of the current controversy concerns whether there is significant evidence of an adverse health effect *within the levels permitted by the current NAAQS*, and on the evidence presented here, for Birmingham, our conclusion is that there is not, even without taking account of the selection effects associated with the choice of $pmmean_1$ as the variable of interest.

5 GAM MODELING

As an alternative to the whole approach to regression taken in sections 3 and 4, the data were also analyzed using the generalized additive model (GAM) approach of Hastie and Tibshirani (1990). There were a number of reasons for pursuing this as an alternative approach, among them, that GAMs form a flexible class of models which automatically incorporate nonlinear effects, and that the methodology of GAMs is different from the spline basis function approach taken in the earlier sections of the paper and therefore provides an independent check on the results obtained through that approach.

The GAM is defined by the equation

$$y_t = \sum_j f_j(x_{jt}) + \epsilon_j \quad (5.1)$$

where x_{jt} is the value of the j 'th covariate on the t 'th day. As in ordinary least squares, $E\{\epsilon_j\} = 0$ and $\text{Var}\{\epsilon_j\} = \sigma^2$, but the f_j terms are arbitrary functions with an f_j modeled for each covariate. We follow the procedure used in S-PLUS in which the nonparametric f_j functions are modeled using the loess procedure (Cleveland and Devlin, 1988). A critical parameter is the *span*, the fraction of the data set used for fitting each function value.

In contrast with earlier analyses, where days with missing values of covariates were omitted from the analysis, we interpolated the PM₁₀ series (by linear interpolation between nearest available days) in order to have a complete set of data. Meteorological variables were tmax, mntp (mean daily temperature, average of tmax and tmin) and mnsh. Lagged values were included as in earlier analyses and the default span provided by S-PLUS (0.5) was used for all variables except the time trend, after visual inspection of the plots suggested that it provided satisfactory results. For the time trend, we also used a loess fit with span 0.05, a somewhat subjective choice based primarily again on visual inspection of plots. This leads to somewhat more irregular trends than in the earlier B-splines approach, but doing this provides a further check on the robustness of the results. The trend effect is a concomitant variable, and ideally, estimates of particulate effects should not be heavily dependent on precise modelling of the trend as long as there is little bias in the trend model. Experiments with span 0.10 produced almost no change in the results below and no change in the conclusions we draw. The response variable was the same as before, i.e. square root of elderly nonaccidental deaths.

The GAM stepwise algorithm identified the following variables as significant for the basic model: $\text{lo}(\text{tmax}_1)$, tmax_3 , $\text{lo}(\text{tmax}_4)$, mntp_2 , mntp_3 , $\text{lo}(\text{mnsh})$, $\text{lo}(\text{mnsh}_1)$, mnsh_2 , mnsh_3 , and a trend term, $\text{lo}(\text{time})$. Here $\text{lo}(\dots)$ indicates that a loess nonlinear function was fitted, while the remaining variables were all treated linearly. The dispersion parameter for the Gaussian family was estimated to be 0.271, i.e. an 8% increase compared with the value 0.25 that would be expected if the Poisson model were exact. This is consistent with estimates of overdispersion found in section 3. The fitted trend effect is displayed in the top left hand box of Fig. 5.1. Since this plot shows the trend effect after adjusting for all other meteorology variables in the model, it does not exactly match the curves in Fig. 3.2.

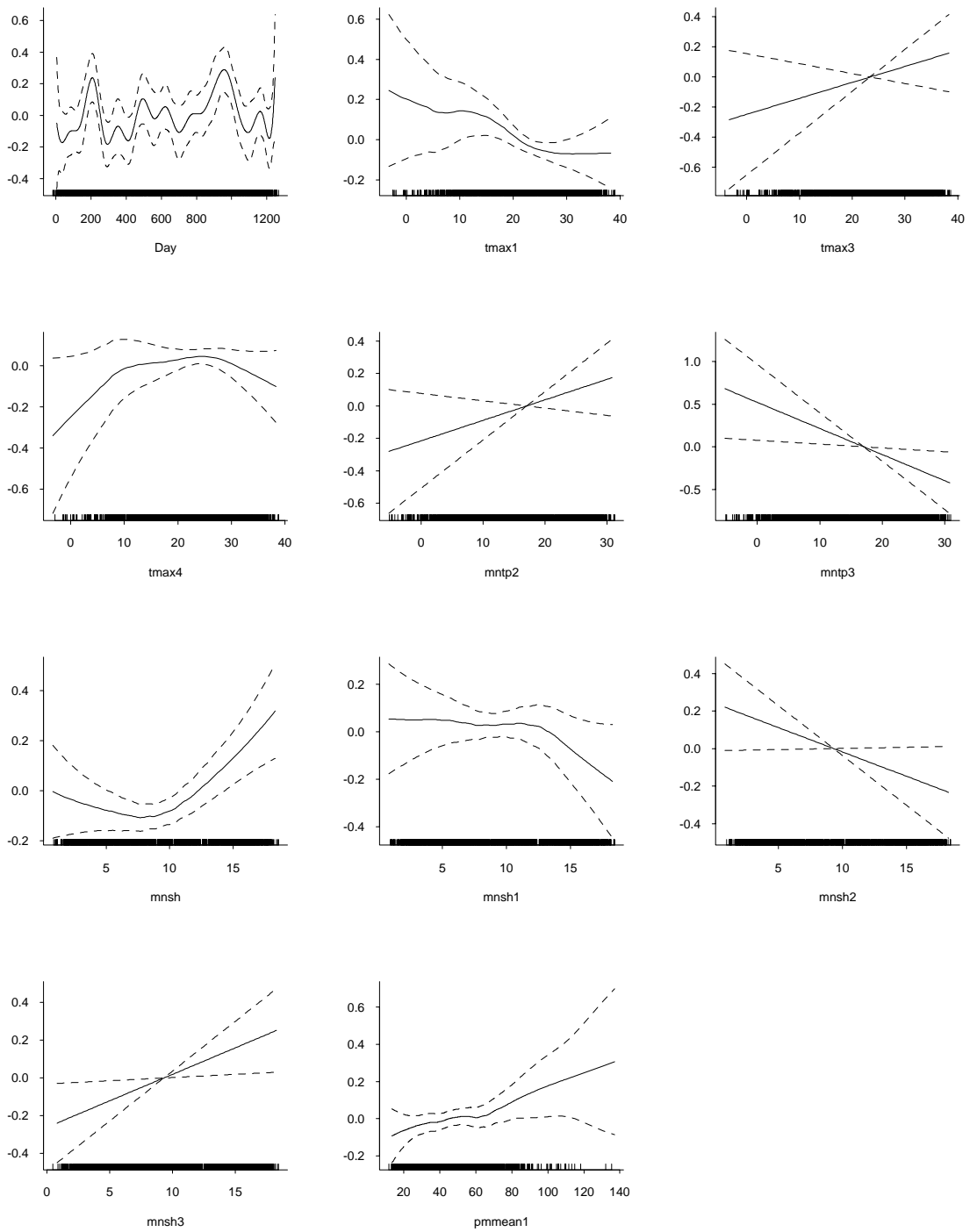


Figure 5.1. Estimated effects and 95% confidence bands for 11 variables under the generalized additive model of section 5.

When $\text{pm}_0, \dots, \text{pm}_4$ are added to the basic model, the analysis of deviance table shows a drop in deviance of 2.05 ($F=1.51$) with five degrees of freedom and an approximate p -value of 0.18. Since this is not significant, it suggests that subset selection is a potential problem in identifying the proper combination of particulate variables to place in the model. Similarly, when pmmean_0 was entered in the basic meteorology model, the drop in deviance was 0.89 ($F=3.29$) with a p -value of 0.07. This agrees with the result from section 3. Only for pmmean_1 (deviance 1.65, $F=6.08$, p -value .014) do we get a significant result.

A further analysis was performed using pm_0 , pm_1 and pm_3 as particulate matter variables, to give results comparable with those in Table 3.5. An initial fit again showed a negative coefficient for pm_0 and positive coefficients for pm_1 and pm_3 , with t values -1.09 , 2.48 and 1.10 respectively. (Here the approximate t values were computed using standard errors derived from the S-PLUS GAM plot function.) In further analysis, first pm_3 and then pm_0 were dropped from the model, leaving pm_1 as the only significant variable, for which the coefficient was .00164, standard error .00069, t value 2.37. The deviance was 1.39 with approximate $F = 5.11$ and p -value .024. (Because the F and t statistics are based on different approximations, the t value is not exactly the square root of F .) Thus the results differ somewhat from Table 3.5 in that with this version of the analysis, the negative coefficient for pm_0 is not statistically significant. The conclusion from the analysis based on one-day values is that only pm_1 is significant, and then only just so. The fact that the negative coefficient for pm_0 is no longer significant supports the interpretation that the relation is spurious and due to multicollinearity. With the richer GAM model, the effect of pm_0 is partially explained by the meteorology variables.

Once again the uncertainties associated with variable selection cast doubt on whether any of the PM_{10} -based variables are really significant. The final fit for a model based on pmmean_1 is presented in Fig. 5.1, where the 11 fitted $f_j(x_j)$ functions are plotted along with their approximate pointwise 95% confidence interval bounds. In this model, we fit the PM_{10} effect with a loess term for pmmean_1 in order to examine the question of whether or not a threshold effect exists. The fitted response for mnsh clearly shows the curvilinear relationship modeled by a quadratic term in Table 3.6. The other variables showing evidence of nonlinearity are tmax_1 , tmax_4 and mnsh_1 . The last panel in Fig. 5.1 shows the nonparametric fit for pmmean_1 with the basic meteorological model. The confidence band seems to confirm the conclusion that a linear fit is adequate. On the other hand, the confidence band is also apparently consistent with a hypothesized threshold of around 60–70 since below this threshold the confidence bands are consistent with the underlying relationship being constant.

6. CONCLUSIONS

The purpose of this paper has been neither to verify nor to disprove the notion that there is a strong association between deaths and particulate matter, but rather, to show that there are many possible interpretations of the data and no single conclusion is definitive. We have given particular attention to the paper by Schwartz (1993) since this is the one which has been most widely cited in the literature, though it should be pointed out that by now there have been a number of re-analyses of data from Birmingham, including those by Samet *et al.* (1995), Roth and Li (1996) and Clyde (2000), which have led to a wide range of conclusions. The main ones from our own study are:

1. The models are highly sensitive to the definition of an “exposure measure” for lagged PM_{10} values. Schwartz used $pmmean_1$ as an exposure measure and our analysis has confirmed that the PM_{10} effect is significant using this variable, but under other measures, it is not significant. In particular, when a combination of lags 0–4 was taken, the result was not statistically significant. It may well be that the apparently negative coefficient for lag 0 is spurious in some way, but the fact that its inclusion changes an apparently significant result into one which is not significant shows that its influence cannot be neglected.

2. On the question of whether the meteorology effects should include both temperature and humidity or only temperature, our finding is that humidity should be included, and that in that case, the resulting PM_{10} coefficients are somewhat smaller than if humidity is not included.

3. Even if we allow for $pmmean_1$ to be taken as the exposure measure of interest, there is little evidence that this has an effect at low levels; a threshold analysis suggests that the main effect is above $80 \mu g/m^3$.

4. Further analysis of nonlinear effects through generalized additive models reinforces the conclusion that the principal effect of $pmmean_1$ is at the upper end of the range.

The broader implication for particulate matter and health is that in a typical data set, there are many issues that need to be taken into account before a conclusion of a causal effect can be drawn. Crude analyses that do not take into account possible alternative interpretations of the data are of limited value in the context of a public debate over pollution regulations.

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. *American Journal of Respiratory and Critical Care Medicine* **159**, 373–382.

- Atkinson, A.C. (1985), *Plots, Transformations and Regression*. Oxford University Prss.
- Cleveland, W.S. and Devlin, S.J. (1988), Locally-weighted regression: an approach to regression analysis by local fitting. *J. Am. Statist. Assoc.* **83**, 597-610.
- Clyde, M. (2000), Model uncertainty and health effect studies for particulate matter. To appear, *Environmetrics*.
- 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.
- 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.
- Ostro, B. (1984), A search for a threshold in the relationship of air pollution to mortality: a re-analysis of data on London winters. *Environ. Health Perspect.* **58**, 397–399.
- Pope, C.A., Bates, D.V. and Raizenne, M.F. (1995a), Health effects of particulate air pollution: Time for reassessment? *Environmental Health Perspectives* **103**, 472–480.
- Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E. and Heath, C.W. (1995b), 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.
- Pope, C.A., Schwartz, J. and Ransom, 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. (1994), Air pollution and hospital admissions for the elderly in Birmingham, Alabama. *American Journal of Epidemiology* **139**, 589–598.
- Schwartz, J. and Dockery, D.W. (1992a), Increased mortality in Philadelphia associated with daily air pollution concentrations. *Am. Rev. Respir. Dis.* **145**, 600-604.

Schwartz, J. and Dockery, D.W. (1992b), Particulate air pollution and daily mortality in Steubenville, Ohio. *Am. J. Epidemiology* **135**, 12-19.

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. (1998), Airborne particles and mortality. Chapter 6 of *Case Studies in Environmental Statistics*, by L.H. Cox, D. Nychka and W.W. Piegorsch (eds.), Springer Lecture Notes in Statistics **132**, 91-120.

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.

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.