Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy

Francesca Dominici, Jonathan M. Samet and Scott L. Zeger

Johns Hopkins University, Baltimore, USA

Summary. Reports over the last decade of association between levels of particles in outdoor air and daily mortality counts have raised concern that air pollution shortens life, even at concentrations within current regulatory limits. Criticisms of these reports have focused on the statistical techniques that are used to estimate the pollution–mortality relationship and the inconsistency in findings between cities. We have developed analytical methods that address these concerns and combine evidence from multiple locations to gain a unified analysis of the data. The paper presents log-linear regression analyses of daily time series data from the largest 20 US cities and introduces hierarchical regression models for combining estimates of the pollution–mortality relationship across cities. We illustrate this method by focusing on mortality effects of PM_{10} (particulate matter less than 10 μm in aerodynamic diameter) and by performing univariate and bivariate analyses with PM_{10} and ozone (O_3) level. In the first stage of the hierarchical model, we estimate the relative mortality rate associated with PM_{10} for each of the 20 cities by using semiparametric log-linear models. The second stage of the model describes between-city variation in the true relative rates as a function of selected city-specific covariates. We also fit two variations of a spatial model with the goal of exploring the spatial correlation of the pollutant-specific coefficients among cities. Finally, to explore the results of considering the two pollutants jointly, we fit and compare univariate and bivariate models. All posterior distributions from the second stage are estimated by using Markov chain Monte Carlo techniques. In univariate analyses using concurrent day pollution values to predict mortality, we find that an increase of 10 μg m^{-3} in PM_{10} on average in the USA is associated with a 0.48% increase in mortality (95% interval: 0.05, 0.92). With adjustment for the O_3 level the PM_{10}-coefficient is slightly higher. The results are largely insensitive to the specific choice of vague but proper prior distribution. The models and estimation methods are general and can be used for any number of locations and pollutant measurements and have potential applications to other environmental agents.

Keywords: Air pollution; Hierarchical models; Log-linear regression; Longitudinal data; Markov chain Monte Carlo methods; Mortality; Relative rate

1. Introduction

In spite of improvements in measured air quality indicators in many developed countries, the health effects of particulate air pollution remain a regulatory and public health concern. This continued interest is motivated largely by recent epidemiological studies that have examined both acute and longer-term effects of exposure to particulate air pollution in various cities in the USA and elsewhere in the world (Dockery and Pope, 1994; Schwartz, 1995; American

Address for correspondence: Francesca Dominici, Department of Biostatistics, School of Hygiene and Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205-3179, USA.
E-mail: fdominic@jhsph.edu

Thoracic Society, 1996a, b; Korrick *et al.*, 1998). Many of these studies have shown a positive association between measures of particulate air pollution — primarily total suspended particles or particulate matter less than 10 μm in aerodynamic diameter (PM_{10}) — and daily mortality and morbidity rates. Their findings suggest that daily rates of morbidity and mortality from respiratory and cardiovascular diseases increase with levels of particulate air pollution below the current national ambient air quality standard for particulate matter in the USA. Critics of these studies have questioned the validity of the data sets used and the statistical techniques applied to them; the critics have noted inconsistencies in findings between studies and even in independent reanalyses of data from the same city (Lipfert and Wyzga, 1993; Li and Roth, 1995). The biological plausibility of the associations between particulate air pollution and illness and mortality rates has also been questioned (Vedal, 1996).

These controversial associations have been found by using Poisson time series regression models fitted to the data by using generalized estimating equations (Liang and Zeger, 1986) or generalized additive models (Hastie and Tibshirani, 1990). Following Bradford Hill's criterion of temporality, they have measured the acute health effects, focusing on the shorter-term variations in pollution and mortality by regressing mortality on pollution over the preceding few days. Model approaches have been questioned (Smith *et al.*, 1997; Clyde, 1998), although analyses of data from Philadelphia (Samet *et al.*, 1997; Kelsall *et al.*, 1997) showed that the particle–mortality association is reasonably robust to the particular choice of analytical methods from among reasonable alternatives. Past studies have not used a set of communities; most have used data from single locations selected largely on the basis of the availability of data on pollution levels. Thus, the extent to which findings from single cities can be generalized is uncertain and consequently for the 20 largest US locations we analysed data for the population living within the limits of the counties making up the cities. These locations were selected to illustrate the methodology and our findings cannot be generalized to all of the USA with certainty. However, to represent the nation better, a future application of our methods will be made to the 90 largest cities. The statistical power of analyses within a single city may be limited by the amount of data for any location. Consequently, in a comparison with analyses of data from a single site, pooled analyses can be more informative about whether an association exists, controlling for possible confounders. In addition, a pooled analysis can produce estimates of the parameters at a specific site, which borrow strength from all other locations (DuMouchel and Harris, 1983; DuMouchel, 1990; Breslow and Clayton, 1993).

One additional limitation of epidemiological studies of the environment and disease risk is the measurement error that is inherent in many exposure variables. When the target is an estimation of the health effects of personal exposure to a pollutant, error is well recognized to be a potential source of bias (Lioy *et al.*, 1990; Mage and Buckley, 1995; Wallace, 1996; Ozkaynak *et al.*, 1996; Janssen *et al.*, 1997, 1998). The degree of bias depends on the correlation of the personal and ambient pollutant levels. Dominici *et al.* (1999) have investigated the consequences of exposure measurement errors by developing a statistical model that estimates the association between personal exposure and mortality concentrations, and evaluates the bias that is likely to occur in the air pollution–mortality relationships from using ambient concentration as a surrogate for personal exposure. Taking into account the heterogeneity across locations in the personal–ambient exposure relationship, we have quantified the degree to which the exposure measurement error biases the results towards the null hypothesis of no effect and estimated the loss of precision in the estimated health effects due to indirectly estimating personal exposures from ambient measurements. Our approach is

an example of regression calibration which is widely used for handling measurement error in non-linear models (Carroll *et al*., 1995). See also Zidek *et al*. (1996, 1998), Fung and Krewski (1999) and Zeger *et al*. (2000) for measurement error methods in Poisson regression.

The main objective of this paper is to develop a statistical approach that combines information about air pollution–mortality relationships across multiple cities. We illustrated this method with the following two-stage analysis of data from the largest 20 US cities.

   (a) Given a time series of daily mortality counts in each of three age groups, we used generalized additive models to estimate the relative change in the rate of mortality associated with changes in the air pollution variables (relative rate), controlling for age-specific longer-term trends, weather and other potential confounding factors, separately for each city.

   (b) We then combined the pollution–mortality relative rates across the 20 cities by using a Bayesian hierarchical model (Lindley and Smith, 1972; Morris and Normand, 1992) to obtain an overall estimate, and to explore whether some of the geographic variation can be explained by site-specific explanatory variables.

This paper considers two hierarchical regression models — with and without modelling possible spatial correlations — which we referred to as the 'base-line' and the 'spatial' models.

In both models, we assumed that the vector of the estimated regression coefficients obtained from the first-stage analysis, conditional on the vector of the true relative rates, has a multivariate normal distribution with mean equal to the 'true' coefficient and covariance matrix equal to the sample covariance matrix of the estimates. At the second stage of the base-line model, we assume that the city-specific coefficients are independent. In contrast, at the second stage of the spatial model, we allowed for a correlation between all pairs of pollutant and city-specific coefficients; these correlations were assumed to decay towards zero as the distance between the cities increases. Two distance measures were explored.

Section 2 describes the database of air pollution, mortality and meteorological data from 1987 to 1994 for the 20 US cities in this analysis. In Section 3, we fit the log-linear generalized additive models to produce relative rate estimates for each location. The semiparametric regression is conducted three times for each pollutant: using the concurrent day's (lag 0) pollution values, using the previous day's (lag 1) pollution levels and using pollution levels from 2 days before (lag 2).

Section 4 presents the base-line and the spatial hierarchical regression models for combining the estimated regression coefficients and discusses Markov chain Monte Carlo methods for model fitting. In particular, we used the Gibbs sampler (Geman and Geman, 1993; Gelfand and Smith, 1990) for estimating parameters of the base-line model and a Gibbs sampler with a Metropolis step (Hastings, 1970; Tierney, 1994) for estimating parameters of the spatial model. Section 5 summarizes the results, compares between the posterior inferences under the two models and assesses the sensitivity of the results to the choice of lag structure and prior distributions.

2.   Description of the databases

The analysis database included mortality, weather and air pollution data for the 20 largest metropolitan areas in the USA for the 7-year period 1987–1994 (Fig. 1 and Table 1). In several locations, we had a high percentage of days with missing values for PM_{10} because it is generally measured every 6 days. The cause-specific mortality data, aggregated at the level of counties, were obtained from the National Center for Health Statistics. We focused on daily death counts

Fig. 1. Map of the 20 cities with largest populations including the surrounding country: the cities are numbered from 1 to 20 following the order in Table 1

for each site, excluding non-residents who died in the study site and accidental deaths. Because mortality information was available for counties but not for smaller geographic units to protect confidentiality, all predictor variables were aggregated to the county level.

Hourly temperature and dewpoint data for each site were obtained from the EarthInfo compact disc database. After extensive preliminary analyses that considered various daily summaries of temperature and dewpoint as predictors, such as the daily average, maximum and 8-h maximum, we used the 24-h mean for each day. If a city has more than one weather-station, we took the average of the measurements from all available stations. The PM_{10} and ozone (O_3) data were also averaged over all monitors in a county. To protect against outliers, a 10% trimmed mean was used to average across monitors, after correction for yearly averages for each monitor. This yearly correction is appropriate since long-term trends in mortality are also adjusted in the log-linear regressions. See Kelsall *et al.* (1997) for further details. Aggregation strategies based on Bayesian and classical geostatistical models as suggested by Handcock and Stein (1993), Cressie (1994), Kaiser and Cressie (1993) and Cressie *et al.* (1999) and Bayesian models for spatial interpolation (Le *et al.*, 1997; Gaudard *et al.*, 1999) are desirable in many contexts because they provide estimates of the error associated with exposure at any measured or unmeasured locations. However, they were not applicable to our data sets because of the limited number of monitoring stations that are available in the 20 counties.

3. City-specific analyses

In this section, we summarize the model used to estimate the air pollution–mortality relative rate separately for each location, accounting for age-specific longer-term trends, weather and

Table 1. Summary by location of the county population Pop, percentage of days with missing values P_{missO_3} and $P_{missPM_{10}}$, percentage of people in poverty $P_{poverty}$, percentage of people older than 65 years $P_{>65}$, average of pollutant levels for O_3 and PM_{10}, $\bar{X}_{O_3}$ and $\bar{X}_{PM_{10}}$, and average daily deaths $\bar{Y}$

Location (state)	Label	Pop	P_{missO_3}	$P_{missPM_{10}}$	$P_{poverty}$ (%)	$P_{>65}$ (%)	$\bar{X}_{O_3}$ (parts per billion)	$\bar{X}_{PM}$ ($\mu g\ m^{-3}$)	$\bar{Y}$
Los Angeles	la	8863164	0	80.2	14.8	9.7	22.84	45.98	148
New York	ny	7510646	0	83.3	17.6	13.2	19.64	28.84	191
Chicago	chic	5105067	0	8.2	14.0	12.5	18.61	35.55	114
Dallas–Fortworth	dlft	3312553	0	78.6	11.7	8.0	25.25	23.84	49
Houston	hous	2818199	0	72.9	15.5	7.0	20.47	29.96	40
San Diego	sand	2498016	0	82.2	10.9	10.9	31.64	33.63	42
Santa Ana–Anaheim	staa	2410556	0	83.6	8.3	9.1	22.97	37.37	32
Phoenix	phoe	2122101	0.1	85.1	12.1	12.5	22.86	39.75	38
Detroit	det	2111687	36.3	53.9	19.8	12.5	22.62	40.90	47
Miami	miam	1937094	1.4	83.4	17.6	14.0	25.93	25.65	44
Philadelphia	phil	1585577	0.7	83.1	19.8	15.2	20.49	35.41	42
Minneapolis	minn	1518196	100	5.4	9.7	11.6	—	26.86	26
Seattle	seat	1507319	37.3	24.5	7.8	11.1	19.37	25.25	26
San Jose	sanj	1497577	0	67.7	7.3	8.6	17.87	30.35	20
Cleveland	clev	1412141	41.4	55.6	13.5	15.6	27.45	45.15	36
San Bernardino	sanb	1412140	0	81.6	12.3	8.7	35.88	36.96	20
Pittsburg	pitt	1336449	1.3	0.8	11.3	17.4	20.73	31.61	38
Oakland	oakl	1279182	0	82.6	10.3	10.6	17.24	26.31	22
San Antonio	sana	1185394	0.1	77.1	19.4	9.8	22.16	23.83	20
Riverside	river	1170413	0	81.3	14.8	11.3	33.41	51.99	20

day of the week. The core analysis for each city is a log-linear generalized additive model that accounts for smooth fluctuations in mortality that potentially confound estimates of the pollution effect and/or introduce autocorrelation in mortality series.

This is a study of the acute health effects of air pollution on mortality. Hence, we modelled daily expected deaths as a function of the pollution levels on the same or immediately preceding days, not of the average exposure for the preceding month, season or year as might be done in a study of chronic effects. We built models which include smooth functions of time as predictors as well as the pollution measures to avoid confounding by influenza epidemics which are seasonal and by other longer-term factors.

To specify our approach more completely, let y_{at}^c be the observed mortality for each age group a ($\leqslant 65$, 65–75, $\geqslant 75$ years) on day t at location c, and let $\mathbf{x}_{at}^c$ be a $p \times 1$ vector of air pollution variables. Let $\mu_{at}^c = E(y_{at}^c)$ be the expected number of deaths and $v_{at}^c = \mathrm{var}(y_{at}^c)$. We used a log-linear model $\log(\mu_{at}^c) = \mathbf{x}_{at}^{c'}\boldsymbol{\beta}^c$ for each city c, allowing the mortality counts to have variances v_{at}^c that may exceed their means (i.e. be overdispersed) with the overdispersion parameter ϕ^c also varying by location so that $v_{at}^c = \phi^c \mu_{at}^c$.

To protect the pollution relative rates $\boldsymbol{\beta}^c$ from confounding by longer-term trends due, for example, to changes in health status, changes in the sizes and characteristics of populations, seasonality and influenza epidemics, and to account for any additional temporal correlation in the count time series, we estimated the pollution effect using only shorter-term variations in mortality and air pollution. To do so, we partial out the smooth fluctuations in the mortality over time by including arbitrary smooth functions of calendar time $S^c(\mathrm{time}, \lambda)$ for each city. Here, λ is a smoothness parameter which we prespecified, on the basis of prior epidemiological knowledge of the timescale of the major possible counfounders, to have 7 degrees of freedom per year of data so that little information from timescales longer than approximately 2 months is included when estimating $\boldsymbol{\beta}^c$. This choice largely eliminates expected confounding from seasonal

influenza epidemics and from longer-term trends due to changing medical practice and health behaviours, while retaining as much unconfounded information as possible. We also controlled for age-specific longer-term and seasonal variations in mortality, adding a separate smooth function of time with 8 degrees of freedom for each age group.

To control for weather, we also fitted smooth functions of the same day temperature ($temp_0$), the average temperature for the three previous days ($temp_{1-3}$), each with 6 degrees of freedom, and the analogous functions for dewpoint (dew_0 and dew_{1-3}), each with 3 degrees of freedom. In the US cities, mortality decreases smoothly with increases in temperature until reaching a relative minimum and then increases quite sharply at higher temperature. 6 degrees of freedom were chosen to capture the highly non-linear bend near the relative minimum as well as possible. Since there are missing values of some predictor variables on some days, we restricted analyses to days with no missing values across the full set of predictors.

In summary, we fitted the following log-linear generalized additive model (Hastie and Tibshirani, 1990) to obtain the estimated pollution log-relative-rate $\hat{\beta}^c$ and the sample covariance matrix V^c at each location:

$$\log(\mu_{at}^c) = \mathbf{x}_{at}^{c'}\boldsymbol{\beta}^c + \gamma^c \, \mathrm{DOW} + S_1^c(\mathrm{time}, 7/\mathrm{year}) + S_2^c(temp_0, 6) + S_3^c(temp_{1-3}, 6)$$
$$+ S_4^c(dew_0, 3) + S_5^c(dew_{1-3}, 3) + \mathrm{intercept \ for \ age \ group} \ a$$
$$+ \mathrm{separate \ smooth \ functions \ of \ time} \ (8 \ \mathrm{degrees \ of \ freedom}) \ \mathrm{for \ age \ group} \ a, \quad (1)$$

where DOW are indicator variables for the day of the week. Samet *et al.* (1995, 1997) and Kelsall *et al.* (1997) give additional details about choices of functions used to control for longer-term trends and weather. Alternative modelling approaches that consider different lag structures of the pollutants and of the meteorological variables have been proposed (Davis *et al.*, 1996; Smith *et al.*, 1997, 1998). More general approaches that consider non-linear modelling of the pollutant variables have been discussed by Smith *et al.* (1997) and by Daniels *et al.* (2000).

Because the functions $S^c(x, \lambda)$ are smoothing splines with fixed λ, the semiparametric model described above has a finite dimensional representation. Hence, the analytical challenge was to make inferences about the joint distribution of the β^cs in the presence of finite dimensional nuisance parameters, which we shall refer to as $\boldsymbol{\eta}^c$.

We separately estimated three semiparametric regressions for each pollutant with the concurrent day (lag 0), prior day (lag 1) and 2 days prior (lag 2) pollution predicting mortality. The estimates of the coefficients and their 95% confidence intervals for PM_{10} alone and for PM_{10} adjusted by O_3 level are shown in Figs 2 and 3. Cities are presented in decreasing order by the size of their populations. The pictures show substantial between-location variability in the estimated relative rates, suggesting that combining evidence across cities would be a natural approach to explore possible sources of heterogeneity, and to obtain an overall summary of the degree of association between pollution and mortality. To add flexibility in modelling the lagged relationship of air pollution with mortality, we could have used distributed lag models instead of treating the lags separately. Although desirable, this is not easily implemented because many cities have PM_{10} data available only every sixth day.

To test whether the log-linear generalized additive model (1) has taken appropriate account of the time dependence of the outcome, we calculate, for each city, the autocorrelation function of the standardized residuals. Fig. 4 displays the 20 autocorrelation functions; they are centred near zero, ranging between -0.05 and 0.05, confirming that the filtering has removed the serial dependence.

We also examined the sensitivity of the pollution relative rates to the degrees of freedom used in the smooth functions of time, weather and seasonality by halving and doubling each
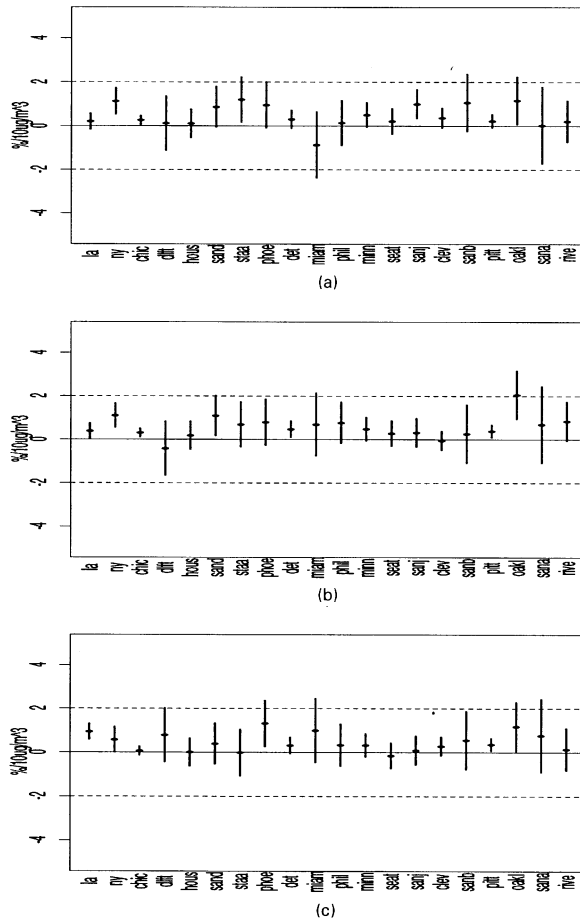
Fig. 2. Results of regression models for the 20 cities by selected lag ($\hat{\beta}^c$ and 95% confidence intervals of $\hat{\beta}^c \times 1000$ for PM_{10}; cities are presented in decreasing order by population living within their county limits; the vertical scale can be interpreted as the percentage increase in mortality per 10 μg m^{-3} increase in PM_{10}): the results are reported (a) using the concurrent day (lag 0) pollution values to predict mortality, (b) using the previous day's (lag 1) pollution levels and (c) using pollution levels from 2 days before (lag 2)

of them. The relative rates changed very little as these parameters are varied over this fourfold range (the data are not shown).

4. Pooling results across cities

In this section, we present hierarchical regression models designed to pool the city-specific pollution relative rates across cities to obtain summary values for the 20 largest US cities. Hierarchical regression models provide a flexible approach to the analysis of multilevel data. In this context, the hierarchical approach provides a unified framework for making estimates of the city-specific pollution effects, the overall pollution effect and of the within- and between-cities variation of the city-specific pollution effects.

The results of several applied analyses using hierarchical models have been published. Examples include models for the analysis of longitudinal data (Gilks *et al.*, 1993), spatial data
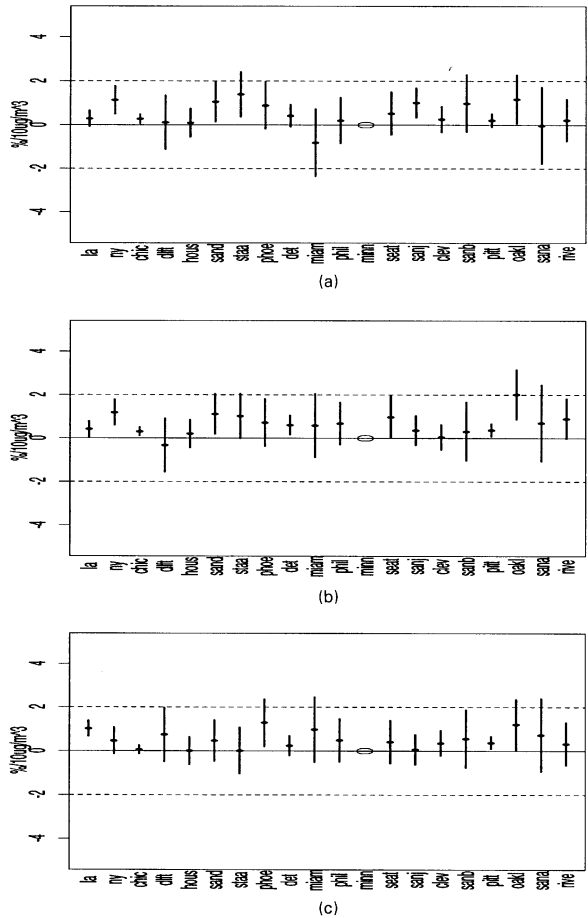
Fig. 3. Results of regression models for the 20 cities by selected lag ($\hat{\beta}^c$ and 95% confidence intervals of $\hat{\beta}^c \times 1000$ for PM_{10} adjusted by O_3 level; cities are presented in decreasing order by population living within their county limits; the empty symbol at Minneapolis represents the missingness of the ozone data in this city; the vertical scale can be interpreted as the percentage increase in mortality per 10 μg m^{-3} increase in PM_{10}): the results are reported (a) using the concurrent day (lag 0) pollution values to predict mortality, (b) using the previous day's (lag 1) pollution levels and (c) using pollution levels from 2 days before (lag 2)

(Breslow and Clayton, 1993) and health care utilization data (Normand *et al.*, 1997). Other modelling strategies for combining information in a Bayesian perspective are provided by Du Mouchel (1990), Skene and Wakefield (1990), Smith *et al.* (1995) and Silliman (1997). Recently, spatiotemporal statistical models with applications to environmental epidemiology have been proposed by Wikle *et al.* (1997) and Wakefield and Morris (1998).

In Section 4.1 we present an overview of our modelling strategy. In Sections 4.2 and 4.3, we consider two hierarchical regression models with and without modelling of the possible spatial autocorrelation among the β^cs which we refer to as the base-line and spatial models respectively.

4.1. Modelling approach

The modelling approach comprises two stages. At the first stage, we used the log-linear generalized additive model (1) described in Section 3:
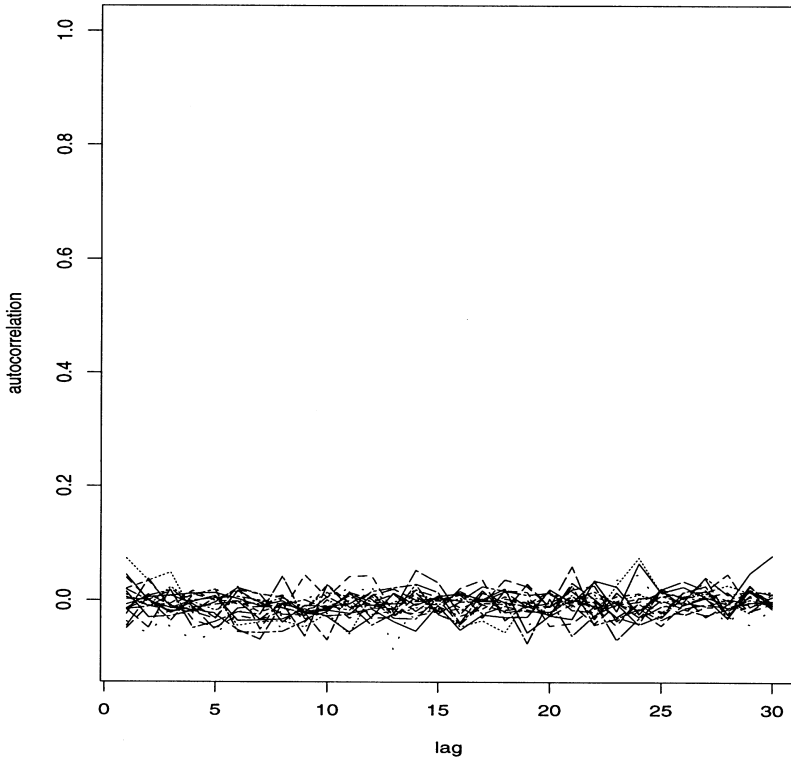
Fig. 4.   Plots of city-specific autocorrelation functions of standardized residuals r_t, where $r_t = (Y_t - \hat{Y}_t)/\sqrt{\hat{Y}_t}$ and $\hat{Y}_t$ are the fitted values from log-linear generalized additive model (1)

$$y_t^c | \beta^c, \eta^c \sim \text{Poisson}\{\mu_t(\beta^c, \eta^c)\}$$

where $y_t^c = (y_{\leqslant 65t}^c, y_{65\text{-}75t}^c, y_{\geqslant 75t}^c)$. The parameters of scientific interest are the mortality relative rates β^c, which for the moment are assumed not to vary across the three age groups within a city. The vector η^c of the coefficients for all the adjustment variables, including the splines in the semiparametric log-linear model, is a finite dimensional nuisance parameter.

The second stage of the model describes variation among the β^cs across cities. We regressed the true relative rates on city-specific covariates z^c to obtain an overall estimate, and to explore the extent to which the site-specific explanatory variables explain geographic variation in the relative risks. In epidemiological terms, the covariates in the second stage are possible effect modifiers. More specifically, we assumed

$$\beta^c | \alpha, \Sigma \sim N_p(z^c \alpha, \Sigma)$$

where p is the number of pollutant variables that enter simultaneously in model (1). Here the parameters of scientific interest are the vector of the regression coefficients, α, and the overall covariance matrix Σ. Unlike the overall air pollution effect α, we are not interested in estimating overall non-linear adjustments for trend and weather; therefore we assume that the nuisance parameters η^c are independent across cities. Our goal is to make inferences about the parameters of interest — the β^cs, α and Σ — in the presence of nuisance parameters η^c. To estimate an exact Bayesian solution to this pooling problem, we could analyse the joint

posterior distributions of the parameters of interest, as well as of the nuisance parameters, and then integrate over the η^c-dimension to obtain the marginal posterior distributions of the β^cs. Although possible, the computations become extremely laborious and are not practical for either this analysis or a planned model with 90 or more cities.

Given the large sample size at each city (T ranges from 550 to 2550 days), accurate approximations to the posterior distribution can be obtained by using the normal approximation of the likelihood (Le Cam and Yang, 1990). If the likelihood function of β^c and η^c is approximated by a multivariate normal distribution with mean equal to the maximum likelihood estimates $\hat{\beta}^c$ and $\hat{\eta}^c$ and covariance matrices V_β and V_η, then by definition the marginal likelihood of β^c has a multivariate normal distribution with mean $\hat{\beta}^c$ and covariance matrix V_β. We then replaced the first stage of the model with a normal distribution with mean and variance equal to the maximum likelihood estimates of the parameter. Recently it has been shown that the strategy based on the normal approximation of the likelihood gives an alternative two-stage model that well approximates the original model and leads to more efficient simulation from the posterior (Daniels and Kass, 1998).

To check whether inferences based on the normal approximation of the likelihood are proper, we compared our approach with the implementation of the full Markov chain Monte Carlo approach for a few cities with sample sizes ranging from 2000 in Pittsburgh to 545 in Riverside. Fig. 5 shows the histogram of samples for Riverside from $p(\beta^c|\text{data})$—obtained by implementing a Gibbs sampler that simulates from $p(\beta^c|\eta^c, \text{data})$ and $p(\eta^c|\beta^c, \text{data})$ and approximate

$$p(\beta^c|\text{data}) = \int p(\beta^c, \eta^c|\text{data})\, d\eta^c$$

—with samples from $N(\hat{\beta}^c, V^c)$ (full curve). The two distributions are very similar.

4.2. Base-line model

Let $\beta^c = (\beta^c_{\text{PM}_{10}}, \beta^c_{\text{O}_3})'$ be the log-relative-rate associated with PM_{10} and O_3 level at city c. We considered the hierarchical model

$$\left.\begin{aligned}
\hat{\beta}^c|\beta^c &\sim N_2(\beta^c, V^c), \\
\beta^c_{\text{PM}_{10}} &= \mathbf{z}^{c'}_{\text{PM}_{10}}\boldsymbol{\alpha}_{\text{PM}_{10}} + \epsilon^c_{\text{PM}_{10}}, \\
\beta^c_{\text{O}_3} &= \mathbf{z}^{c'}_{\text{O}_3}\boldsymbol{\alpha}_{\text{O}_3} + \epsilon^c_{\text{O}_3}, \\
\boldsymbol{\epsilon}^c|\Sigma &\sim N_2(0, \Sigma)
\end{aligned}\right\} \tag{2}$$

where $\mathbf{z}^c_{\text{PM}_{10}} = (1, P^c_{\text{poverty}}, P^c_{>65}, \bar{X}^c_{\text{PM}_{10}})'$, $\mathbf{z}^c_{\text{O}_3} = (1, P^c_{\text{poverty}}, P^c_{>65}, \bar{X}^c_{\text{O}_3})'$, $\boldsymbol{\alpha}_{\text{PM}_{10}}$ and $\boldsymbol{\alpha}_{\text{O}_3}$ are 4×1 vectors and finally $\boldsymbol{\epsilon}^c = (\epsilon^c_{\text{PM}_{10}}, \epsilon^c_{\text{O}_3})'$, $c = 1, \ldots, 20$. This model specification allowed a dependence between the relative rates associated with PM_{10} and O_3 level, but implied independence between the relative rates of cities c and c'.

Under this model, the true PM_{10} and O_3 log-relative-rates in city c were regressed on predictor variables including the percentage of people in poverty (P^c_{poverty}) and the percentage of people older than 65 years ($P^c_{>65}$), and on the average of the daily values of PM_{10} and O_3 level over the period 1987–1994 in location c ($\bar{X}^c_{\text{PM}_{10}}$ and $\bar{X}^c_{\text{O}_3}$). If we centred the predictors about their means, the intercepts $\alpha_{0,\text{PM}_{10}}$ and α_{0,O_3} can be interpreted as overall effects for a city with mean predictors. A simple pooled estimate of the pollution effect is obtained by setting all covariates to 0. To compare the consequences of considering two pollutants
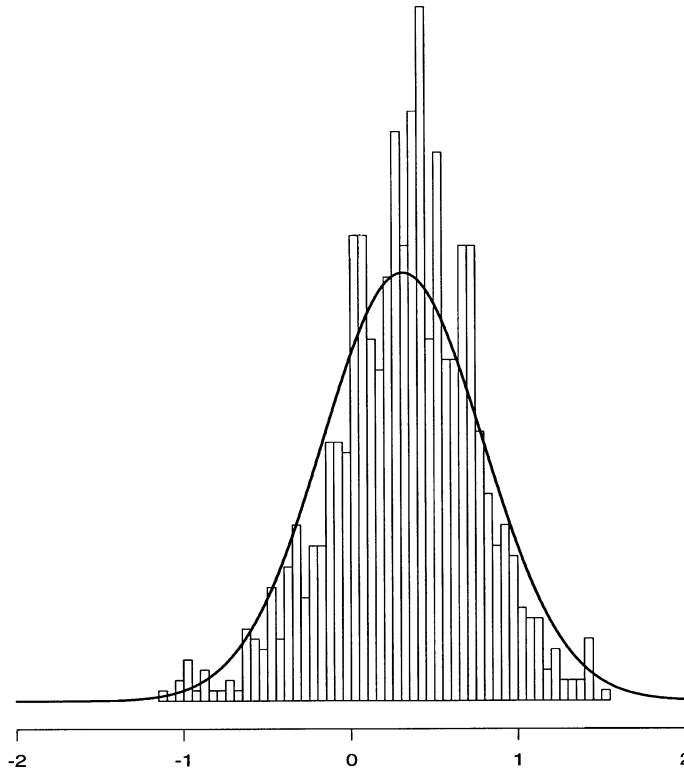
Fig. 5. Comparison between the normal approximation of the likelihood of β^c and the marginal posterior distribution of β^c: ———, normal density $N(\hat{\beta}^c, V^c)$ where $\hat{\beta}^c$ and V^c are the maximum likelihood estimates of a semiparametric Poisson regression model; histogram, marginal posterior distribution of β^c obtained by implementing a full Gibbs sampler for the parameter of interest β^c and for the coefficients of the natural cubic splines η^c

independently and jointly in the model, we fit a base-line–univariate model—i.e. Σ assumed diagonal—and a base-line–bivariate model—i.e. Σ assumed to have non-zero off-diagonal elements.

Inference on the parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{\mathrm{PM}_{10}}, \boldsymbol{\alpha}_{\mathrm{O}_3})'$ and Σ represents a synthesis of the information from the 20 cities; for example the parameters α_{0j}, $[\Sigma]_{jj}$, $j = \mathrm{PM}_{10}$, O_3, determine the overall level and the variability of the relative change in the rate of mortality associated with changes in the jth pollutant level on average over all the cities.

The Bayesian formulation was completed by specifying dispersed but proper base-line prior distributions and then supplementing the base-line analysis with additional sensitivity analysis. *A priori*, we assumed that the joint prior is the product of the marginals for $\boldsymbol{\alpha}$ and Σ. The following base-line prior specifications for the marginals are used:

$$\text{overall log-relative-rates } \boldsymbol{\alpha} \sim N_{p(k+1)}(\mathbf{m}, V_\alpha),$$

$$\text{overall covariance matrix } \Sigma \sim \mathrm{IW}_p(\mathrm{df}, D)$$

where $\mathrm{IW}_p(\mathrm{df}, D)$ denotes the inverse Wishart distribution with df degrees of freedom and scale matrix D, a $p \times p$ positive definite matrix, whose density is proportional to

$$\frac{D^{(\mathrm{df}+p-1)/2}}{|\Sigma|^{(\mathrm{df}+2p)/2}} \exp\left\{-\frac{1}{2}\operatorname{tr}(D\Sigma^{-1})\right\}.$$

Here p denotes the number of pollutant variables entering the model simultaneously and k the number of city-specific covariates. We select $\mathbf{m}$ equal to a vector of 0s, V_α equal to a diagonal matrix, with diagonal elements equal to 100, $\mathrm{df} = 3$ and D a diagonal matrix with diagonal elements equal to 3. In the univariate case we denote Σ by σ^2. These prior hyper-parameters lend prior 95% support to the overall effect, the city-specific effects and the correlation between the PM_{10} and the O_3 log-relative-rates equal to $(-15, 15)$, $(-4, 4)$ and $(-0.85, 0.85)$ respectively. This prior specification was selected because it did not impose too much shrinkage of the study-specific parameters towards their overall means, while specifying a reasonable range for the unknown parameters *a priori*. A sensitivity analysis is presented in Table 4 in Section 5.

Given these prior assumptions, we can draw inferences on the unknown parameters by using the posterior distribution

$$p(\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^{20}, \boldsymbol{\alpha}, \Sigma | \hat{\boldsymbol{\beta}}^1, \ldots, \hat{\boldsymbol{\beta}}^{20}, V^1, \ldots, V^{20}). \tag{3}$$

To do this, we implemented a Markov chain Monte Carlo algorithm with a block Gibbs sampler (Gelfand and Smith, 1990) in which the unknowns are partitioned into the groups $\boldsymbol{\beta}^c$, $\boldsymbol{\alpha}$ and Σ. Each group is sampled in turn, given all others. The full conditional distributions were available in closed form. Their derivation was routine (Bernardo and Smith, 1994) and is not detailed here. Because of the normality assumptions at the first and second stage of the hierarchical model, computations of the posterior distributions of all the unknowns under a univariate model can be performed via direct simulation following the factorization above:

$$p(\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^{20}, \boldsymbol{\alpha}, \sigma^2 | \mathrm{data}) = p(\sigma^2 | \mathrm{data})\, p(\boldsymbol{\alpha} | \sigma^2, \mathrm{data}) \prod_c p(\boldsymbol{\beta}^c | \boldsymbol{\alpha}, \sigma^2, \mathrm{data}).$$

The first step, simulating σ^2, can be performed numerically (using the inverse cumulative density function method, for example). The second and third steps can be done easily by sampling from normal distributions. This strategy can be conveniently implemented only for the univariate base-line model.

4.3. Spatial model

The assumption of independence of the city-specific coefficients that is made in the base-line model can be relaxed to a more general model in which the correlation between $\boldsymbol{\beta}^c$ and $\boldsymbol{\beta}^{c'}$ decays as either a smooth or step function to 0 as the distance between the two cities, c and c', increases. In this section, we consider a hierarchical model in which the inferences allow for the possible spatial correlation among the $\boldsymbol{\beta}^c$s. We only considered univariate models given the small number of cities; an extension to multivariate models is straightforward but requires a larger data set.

At the second stage of the spatial model, we assumed that there is a systematic variation in the air pollution–mortality relationship from pollutant to pollutant as specified in the base-line model (2). We expressed the degree of similarity of the relative rates in locations c and c' as a function of an (arbitrary) distance between c and c', by assuming $\rho(c, c') = \operatorname{corr}(\boldsymbol{\beta}^c, \boldsymbol{\beta}^{c'}) = \exp\{-\theta\, d(c, c')\}$. We considered two distance measures, the Euclidean distance between the cities c and c' in the longitude and latitude co-ordinates and a step function such

that $d(c, c') = 1$ if locations c and c' are within a common 'region' and $d(c, c') = \infty$ if not. To make the results of these two models comparable we rescaled the Euclidean distance such that it ranges between 0 and 4 with median equal to 0.64. The spatial model with $(1, \infty)$ distance can also be specified as a three-stage hierarchical model where the first stage is as the base-line model (2), the second stage describes the heterogeneity of the estimates across cities within regions and the third stage describes the heterogeneity of the estimates across regions. For this regional model, we have clustered the 20 cities in the following three regions: north-east, south-east and west coast. Thus, if we indicate by τ^2 the variability of the estimates across regions and by σ^2 the variability of the estimates within regions, then the correlation of the log-relative-rates for locations c and c' within a common region is $\tau^2/(\tau^2 + \sigma^2)$. Alternative definitions of distance can be incorporated easily into the model as appropriate.

The spatial model specification is completed with the elicitation of the prior distribution. For α and σ^2 we choose the same prior specified in Section 4.2. For the parameter θ under the spatial model with Euclidean distance, we choose a log-normal prior with mean 0.2 and standard deviation 0.5. Let $\tilde{d}$ be the median of the distribution of all distances; this specification leads to a prior distribution of the correlation $\exp(-\theta\tilde{d})$ having mean 0.45 (95% interval: 0.11, 0.74). For the parameter τ^2 under the spatial model with step distance, we chose an inverse gamma prior $IG(A, B)$ with parameters $A = 5$ and $B = 8.5$. This specification leads to a prior distribution for τ having mean 1.35 (95% prior interval: 0.9, 2.2) and a prior distribution for the correlation $\tau^2/(\tau^2 + \sigma^2)$ having mean 0.45 (95% prior interval: 0.13, 0.77).

In the spatial model, the full conditionals for β^c, α and σ^2 are all available in closed form. In contrast, to sample from the full conditional distribution of θ, we used a Metropolis–Hastings algorithm with a gamma proposal distribution having mean equal to the current value of θ and fixed variance. The spatial model with a step distance can be more efficiently sampled with a block Gibbs sampler because the full conditional distributions of all the unknown parameters are available in closed form.

5. Results

We ran the Gibbs sampler for 3000 iterations for both the base-line and the spatial models, ignoring the first 100. The autocorrelation, computed from a random sample of the $\alpha_{0,PM_{10}}$, is negligible at lag 5 so we sampled every fifth observations for posterior estimation. The acceptance probabilities for the Metropolis algorithm averaged between 0.3 and 0.5. Convergence diagnosis was performed by implementing Raftery and Lewis's (1992) methods in CODA (Best *et al.*, 1995) which reported the minimum number of iterations $N_{\min}$ needed to estimate the variable of interest with an accuracy of ± 0.005 and with probability of attaining this degree of accuracy equal to 0.95. $N_{\min} \simeq 2000$ are proposed.

Fig. 6 summarizes results of the pooled analyses under the univariate–base-line model. It displays the posterior distributions of city-specific regression coefficients β^c associated with changes in PM_{10}-measurements for the 20 cities at the current day, 1-day lag and 2-day lag. The marginal posterior distribution of the overall effect $(\alpha_{0,PM_{10}})$ is displayed at the far right-hand side. Cities are ordered by the decreasing size of their populations. At the current day, the highest relative rate for the PM_{10}-variable occurs in New York with a 1.05% increase in mortality (95% interval: 0.5, 1.6) per 10 $\mu g\ m^{-3}$ increase in PM_{10}. Overall, we found that a 10 $\mu g\ m^{-3}$ increase in PM_{10} is associated with an estimated 0.48% increase in mortality (95% interval: 0.05, 0.92).

Fig. 7 summarizes the results of the pooled analyses under the bivariate–base-line model. When PM_{10} and O_3 level are combined in the same model, we estimated that 10-unit
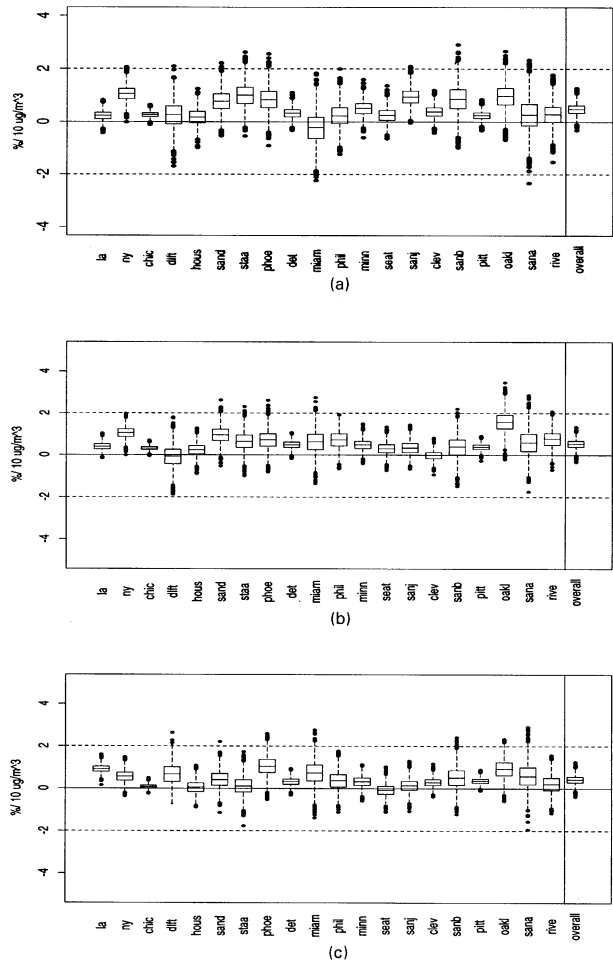
Fig. 6. Results of pooled analyses under the univariate–base-line model (PM$_{10}$ entered independently in the model) (box plots of samples from the posterior distributions of city-specific regression coefficients β^c associated with changes in PM$_{10}$-measurements; for comparison, samples from the marginal posterior distribution of the corresponding overall effects are displayed at the far right-hand side; the vertical scale can be interpreted as the percentage increase in mortality per 10 μg m^{-3} increase in PM$_{10}$): the results are reported (a) using the concurrent day (lag 0) pollution values to predict mortality, (b) using the previous day's (lag 1) pollution levels and (c) using pollution levels from 2 days before (lag 2)

increments in PM$_{10}$ adjusted by O$_3$ are associated with mortality increases of 0.52% (95% interval: 0.16, 0.85).

The marginal posterior distribution of the overall regression effect combined and synthesized the information from the 20 locations. Fig. 8 shows the marginal posterior distributions of the overall pollution relative rates at the current day, 1-day and 2-day lags obtained from the base-line–univariate, base-line–bivariate and spatial models. At the top right-hand side are summarized the posterior probabilities that the overall effects are larger than 0 for each lag specification. In the univariate and bivariate analyses, we found significant effects of PM$_{10}$.

Results of the adjusted analyses under the univariate–base-line model are shown in Table 2. Here we summarize the posterior means and the 95% posterior support intervals for the
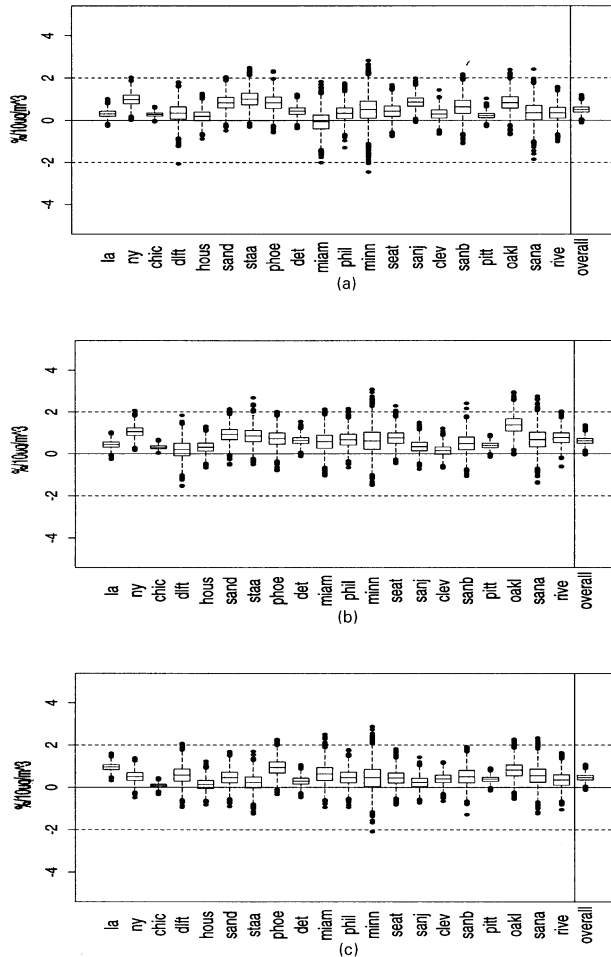
Fig. 7. Results of pooled analyses under the bivariate–base-line model (PM_{10} and O_3 level entered simultaneously in the model) (box plots of samples from the posterior distributions of city-specific regression coefficients β^c associated with changes in PM_{10} adjusted by O_3 measurements; for comparison, samples from the marginal posterior distribution of the corresponding overall effects are displayed at the far right-hand side; the vertical scale can be interpreted as the percentage increase in mortality per 10 μg m^{-3} increase in PM_{10} respectively): the results are reported (a) using the concurrent day (lag 0) pollution values to predict mortality, (b) using the previous day's (lag 1) pollution levels and (c) using pollution levels from 2 days before (lag 2)

relationship between the mean of the city-specific coefficients and the percentage in poverty, the percentage of people older than 65 years and the mean level of the pollutant. Here the intercept α_0 denotes the overall effect of PM_{10} with mean predictors. None of these variables are found to predict the PM_{10} relative rate.

An interaction of the pollution effects and age could be detected by the coefficient of the variable $P_{>65}$ in the second-stage regression model. A more direct approach was to estimate a separate pollution relative rate for each age stratum in the first-stage log-linear models and then to pool the trivariate vector $(\hat{\beta}_{<65}, \hat{\beta}_{65-75}, \hat{\beta}_{>75})$ across cities. When we did so, the estimates of the overall effect of PM_{10} for the three age groups have posterior means 0.63 (95% interval: 0.24, 1.05), 0.26 (95% interval: -0.14, 0.67) and 0.46 (95% interval: 0.04, 0.83).
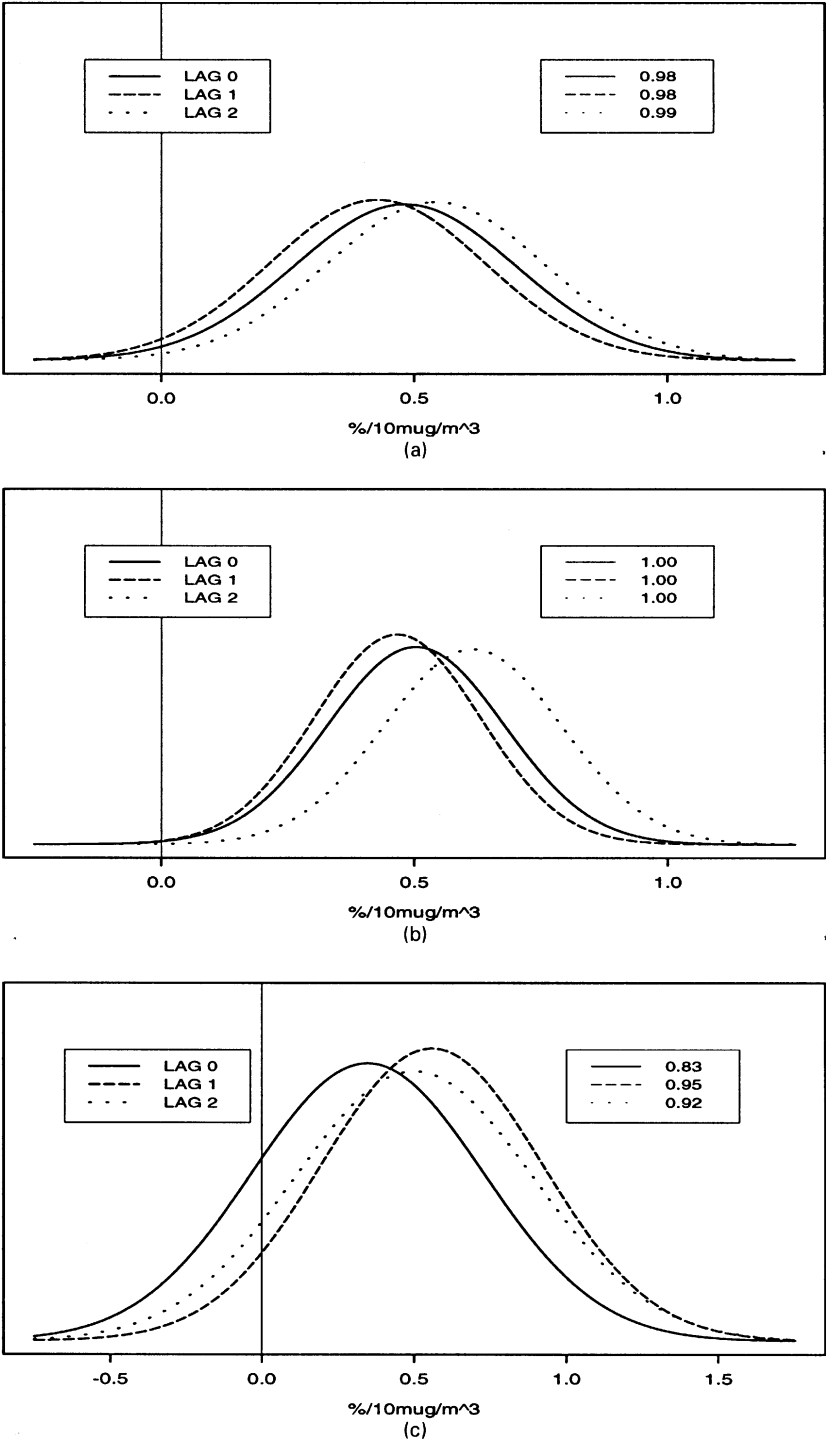
Fig. 8. Results of pooled analyses under (a) the univariate–base-line, (b) bivariate–base-line and (c) spatial models (marginal posterior distributions of the overall effects, $\alpha_{0,PM_{10}}$, for various lags; at the top right-hand side are specified the posterior probabilities that the overall effects are larger than 0)

Table 2. Results of the second-stage analyses under the base-line–univariate model (PM$_{10}$ entered independently in the model)†

City-specific covariate	Posterior means and support intervals for the following lags:		
	Lag 0	Lag 1	Lag 2
Overall PM$_{10}$	0.40 (−0.06, 0.85)	0.52 (0.06, 0.98)	0.43 (−0.03, 0.87)
$P_{poverty}$ (%)	−0.08 (−0.21, 0.04)	−0.01 (−0.14, 0.11)	−0.01 (−0.11, 0.12)
$P_{>65}$ (%)	−0.01 (−0.19, 0.17)	0.03 (−0.15, 0.20)	0.00 (−0.16, 0.17)
$\bar{X}^c_{PM_{10}}$ (μg m^{-3})	0.02 (−0.05, 0.08)	−0.01 (−0.07, 0.06)	0.01 (−0.05, 0.07)

†Posterior means and 95% posterior support intervals of the coefficients for the relationship between the true relative rate β^c, the percentage in poverty $P_{poverty}$, the percentage of people older than 65 years $P_{>65}$ and the mean level of the pollutant $\bar{X}_{PM_{10}}$. The results are reported using the concurrent day (lag 0) pollution values to predict mortality, using the previous day's (lag 1) pollution levels and using pollution levels from 2 days before (lag 2).

Table 3. Posterior means and 95% support intervals of the elements of Σ under the three models (univariate, bivariate and spatial)

Model	Posterior means and support intervals for the following effects†:		
	std of PM$_{10}$ effects	std of O$_3$ effects	corr of PM$_{10}$ and O$_3$ effects
Base-line–bivariate	0.36 (0.17, 0.75)	0.91 (0.33, 2.01)	−0.09 (−0.5, 0.22)
Base-line–univariate	0.76 (0.41, 1.37)	1.28 (0.69, 2.28)	
Spatial	0.71 (0.38, 1.27)	1.21 (0.61, 2.32)	

†std of PM$_{10}$ effects, standard deviation across locations of the $\beta^c_{PM_{10}}$; std of O$_3$ effects, standard deviation across locations of the $\beta^c_{O_3}$; corr of PM$_{10}$ and O$_3$ effects, correlation between the $\beta^c_{PM_{10}}$ and $\beta^c_{O_3}$.

These results suggest that there is no trend in the pollution relative rates with age as is suggested by the second-stage regression results in Table 2.

The variability of the regression coefficients, on average, over all the locations was captured by the matrix Σ. Marginal posterior means and 95% posterior support intervals are summarized in Table 3. A large diagonal element signified large variability over cities in the corresponding coefficient, whereas a large off-diagonal element signifies strong correlation between the PM$_{10}$- and O$_3$ coefficients. Table 3 shows the results. Under the base-line–univariate model, the standard deviation of the true coefficients across cities was estimated to be 0.76 (95% interval: 0.41, 1.37) which is about twice as large as the overall estimate of the pollution effect. Hence, in univariate analyses, the variability in the PM$_{10}$-coefficient is non-negligible. The posterior distribution of the off-diagonal elements of Σ indicates a negative mean correlation between the effects of the two pollutants, but with a large standard deviation.

From the posterior samples of θ in the spatial model, we could easily calculate the marginal posterior distributions of the correlation coefficient $\rho(c, c') = \exp\{-\theta\, d(c, c')\}$ for each distance $d(c, c')$. For the cities having median distance, the posterior mean correlation between β^c and $\beta^{c'}$ was 0.61 (95% interval: 0.3, 0.8). Consider the 25% and 75% quantiles of the distribution of all distances. Each of these quantiles has an associated correlation coefficient. The posterior means of these two correlation coefficients were 0.86 (95% interval: 0.68, 0.93) and 0.3 (95% interval: 0.05, 0.58), both larger than the corresponding prior means.

Under the regional model, with distance equal to a step function, the posterior mean of the within-region correlation of the city-specific relative rates $\tau^2/(\tau^2 + \sigma^2)$ was 0.68 (95% interval: 0.42, 0.86). Results for the PM$_{10}$ effects under the two spatial models were similar

Table 4. Posterior probabilities that the overall effects of PM_{10} are larger than 0 by lag and by three prior distributions under the three models (univariate, bivariate and spatial)

Model	Posterior probabilities for the following priors and lags†:								
	Prior 1			Prior 2			Prior 3		
	Lag 0	Lag 1	Lag 2	Lag 0	Lag 1	Lag 2	Lag 0	Lag 1	Lag 2
Base-line–univariate	0.98	0.98	0.99	0.98	0.96	0.98	0.95	0.96	0.93
Base-line–bivariate	1	1	0.97	1	0.99	0.99	0.98	1	0.93
Spatial	0.83	0.95	0.92	0.83	0.93	0.91	0.78	0.89	0.85

†The three prior specifications have the following 95% support intervals of the overall effects, the city-specific effects and of the spatial correlation for the relative rates of the two closest cities with median distance: prior 1, $(-15, 15)$, $(-4, 4)$, $(0.11, 0.74)$; prior 2, $(-4, 4)$, $(-4, 4)$, $(0.11, 0.74)$; prior 3, $(-4, 4)$, $(-7, 7)$, $(0, 0.9)$.

qualitatively. The posterior means and interquartile range for the regional effects β^{east}, β^{south} and β^{west} are 0.40 $(-0.22, 1.03)$, -0.06 $(-0.96, 0.93)$ and 0.69 $(0.07, 1.35)$, revealing that the adverse health effects of PM_{10} on mortality in the west of the USA is larger than in the east and south.

We have assessed the robustness of the results with respect to choices of the model (univariate, bivariate and spatial), of the lag structure (lag 0, lag 1 and lag 2) and of the prior distributions. Our sensitivity analysis compared 27 alternative scenarios (three for model choice, three for lag structures and three for prior distributions). For these scenarios we compare the posterior probability that the overall effect of PM_{10} is larger than 0. The consequences of these choices are shown in Table 4. Significant effects of PM_{10} on total daily mortality are observed in all three models (but weaker under a spatial model with current day pollution predicting mortality). When both pollutants are included in the model, adverse effects of PM_{10} became stronger. Spatial analyses attenuate the effects.

6. Discussion

We have developed a statistical model for obtaining a national estimate of the effect of urban air pollution on daily mortality using data for the 20 largest US cities. The raw data comprised publicly available listings of individual deaths by day and location, and hourly measurements of pollutants and weather variables. Substantial preprocessing of the nearly 1 Gbyte of information is necessary to create daily time series of mortality, pollutants and weather for each of the 20 cities.

Because the estimation of a national pollution relative rate is the primary objective of this study, a two-stage approach was developed that allowed the modelling effort to focus on combining information across cities. In the first stage, a log-linear regression is used to estimate a pollution relative rate for each city while controlling for the city-specific longer-term time trends and weather effects. Because we had no specific scientific interest in the time or weather effects, no effort is made to impose modelling assumptions to enable borrowing strength across cities when estimating the effects on mortality of these variables.

In the second stage, we regressed the true relative rates on city-specific covariates to obtain an overall estimate, and to estimate the variation among the coefficients across cities. We then generated posterior estimates of the overall pollution effect and of the city-specific effects by using Markov chain Monte Carlo methods. Four models for combining relative rates of mortality for PM_{10} across cities were used. In the first, relative rates from different cities are treated as independent of one another. In the second, relative rates from different cities are treated as independent of one another, but are adjusted by O_3 level. In the third and fourth

models the possibility of geographic correlation between the true coefficients is allowed. Results under the four models are similar: bivariate analyses give slightly higher effects and spatial analyses slightly attenuate the effects. Results under different models, lag specifications and priors are summarized in Fig. 8 and Table 4. Note that the variance of the posterior distribution of the overall relative rate in the spatial models is somewhat sensitive to the prior specification for the between-region variance or equivalently within-region correlation since, with our 20 cities, we have only three regions and hence limited information. A similar analysis of the 90 largest cities will provide more precise information about variation across regions.

These analyses demonstrated that there was a consistent association of particulate air pollution PM_{10} with daily mortality across the 20 largest US cities leading to an overall effect, which was positive with high probability. Our overall estimate was that an increase of 10 μg m^{-3} in particulate level is associated with a roughly 0.48% increase in daily mortality on that day or the next day.

Another multicity study of air pollution and mortality is the multicentre European study, 'Air pollution and health: a European approach' (Katsoyanni *et al.*, 1997; Toulomi *et al.*, 1997). The cities were selected from across Europe, although not systematically. Data on particulate air pollution and daily mortality are analysed from 12 cities from western and central Europe according to a standardized protocol. Model estimates from the individual cities are pooled as the weighted means of the regression coefficients and heterogeneity among cities is explored using a random-effects model. For particulate matter, the findings differed between the western and central Europe cities, with a fivefold greater effect in the western cities (Katsoyanni *et al.*, 1997). A similar approach is applied to the six selected cities with data available on O_3. A significant effect of O_3 is found, after controlling for levels of black smoke and an index of particulate matter (Toulomi *et al.*, 1997).

Although it is only a first step, the modelling described here establishes a basis for carrying out national surveillance for effects of air pollution and weather on public health. The analyses could be easily extended to studies of cause-specific mortality and other pollutants. Monitoring efforts using models like that described here would be appropriate given the important public health questions that they can address and the considerable expense to government agencies for collecting the information that forms the basis for this work.

An alternative modelling strategy would have been to use one large Markov chain Monte Carlo method to estimate simultaneously the parameters in the log-linear models within each city, the overall estimate of the pollutant and all the nuisance parameters, borrowing strength across cities to obtain more precise estimates of the nuisance functions for each city. This type of approach would be necessary if there were limited information about the nuisance parameters within each city as, for example, in the Neyman and Scott problem (Neyman and Scott, 1960). As this is not the case in our investigation, we focused the modelling and computing effort on combining city-specific relative rate estimates to obtain a national average relative rate.

If the likelihood function for the pollution relative rate and the nuisance parameters is well approximated by a Gaussian distribution, then our approach will give a close approximation to the posterior distribution from a Markov chain Monte Carlo sample that simulated both the parameters of interest and the nuisance parameters. We compared the marginal posterior of the β^c obtained by using a full Markov chain Monte Carlo procedure with our normal approximation for a few cities; they are indistinguishable.

The approach of taking a weighted average of the city-specific estimates to obtain an estimate of the overall effect, as for example suggested by DerSimonian and Laird (1986), is a

simplified version or approximation to the use of hierarchical models with a Gibbs sampler. Under the weighted average approach for a random-effect model, the weights of the city-specific estimates are modified to take into account the variability between locations, say σ^2, and an estimate of this variance is included. Rather than including a single estimate of σ^2, the Bayesian method permits incorporating the whole posterior distribution of σ^2. In this way, all the information about the variability between studies is considered. In addition, the Bayesian method provides estimates of the posterior distribution of the city-specific relative rates and of the national estimate, and it easily lends itself to generating ranking probabilities as, for example, P(overall log-relative-rate $\geqslant$ 0|data). In addition, the Gibbs sampler is necessary for approximating the posterior distributions under the spatial model.

These analyses alone cannot establish that increased levels of particulate air pollution as measured by PM_{10} cause an increase in mortality. They do, however, establish that there is a consistent association between shorter-term variations in PM_{10} and shorter-term variations in mortality, and that this association is very unlikely to be explained by the effects of longer-term confounders such as a change in medical practice, influenza epidemics or seasonality, which have been controlled for by using a city-specific adjustment for longer-term trends. Nor can these associations be explained by confounding effects of temperature or dewpoint temperature, which again have been controlled for by using city-specific adjustment methods.

Acknowledgements

References

American Thoracic Society (1996a) Health effects of outdoor air pollution, part 1. *Am. J. Resp. Crit. Care Med.*, **153**, 3–50.
——— (1996b) Health effects of outdoor air pollution, part 2. *Am. J. Resp. Crit. Care Med.*, **153**, 477–498.
Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory*. New York: Wiley.
Best, N. G., Cowles, M. K. and Vines, K. (1995) CODA: convergence diagnostics and output analysis software for Gibbs sampling output, version 0.30. *Technical Report*. Medical Research Council Biostatistics Unit, Cambridge.
Breslow, N. and Clayton, D. (1993) Approximation inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.
Carroll, R. J., Ruppert, D. and Stefanski, L. (1995) *Measurement Error in Nonlinear Models*. New York: Chapman and Hall.
Clyde, M. (1998) Bayesian model averaging and model search strategies. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 83–102. Oxford: Oxford University Press.
Cressie, N. (1994) *Statistics for Spatial Data*. New York: Wiley.
Cressie, N., Kaiser, M., Daniels, M., Aldworth, J., Lee, J., Lahiri, S. and Cox, L. (1999) Spatial analysis of particulate matter in an urban environment. In *Proc. GeoEnv 98* (eds J. Gomez-Hernandez, A. Soares and R. Froidevaux).
Daniels, M., Dominici, F. and Samet, J. (2000) Estimating PM_{10}-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *Am. J. Epidem.*, to be published.
Daniels, M. and Kass, R. (1998) A note on first-stage approximation in two-stage hierarchical models. *Sankhya* B, **60**, 19–30.
Davis, J., Sacks, J., Saltzman, N., Smith, R. and Styer, P. (1996) Airborne particulate matter and daily mortality in Birmingham, Alabama. *Technical Report*. National Institute of Statistical Science, Research Triangle Park.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Contr. Clin. Trials*, **7**, 177–188.

Dockery, D. and Pope, C. (1994) Acute respiratory effects of particulate air pollution. *A. Rev. Publ. Hlth*, **15**, 107–132.

Dominici, F., Zeger, S. and Samet, J. (2000) A measurement error correction model for time-series studies of air pollution and mortality. *Biostatistics*, **2**, in the press.

DuMouchel, W. H. (1990) Bayesian metaanalysis. In *Statistical Methodology in the Pharmaceutical Sciences* (ed. D. A. Berry), pp. 509–529. New York: Dekker.

DuMouchel, W. H. and Harris, J. E. (1983) Bayes methods for combining the results of cancer studies in humans and other species. *J. Am. Statist. Ass.*, **78**, 293–308.

Fung, Y. and Krewski, D. (1999) On measurement error adjustment methods in poisson regression. *Environmetrics*, **10**, 213–224.

Gaudard, M., Karson, M., Linder, E. and Sinha, D. (1999) Bayesian spatial prediction. *Environ. Ecol. Statist.*, **6**, 147–171.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.

Geman, S. and Geman, D. (1993) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *J. Appl. Statist.*, **20**, 25–62.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D. and Kirby, A. J. (1993) Modelling complexity: applications of Gibbs sampling in medicine. *J. R. Statist. Soc.* B, **55**, 39–52.

Handcock, M. and Stein, M. (1993) A bayesian analysis of kriging. *Technometrics*, **35**, 403–410.

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

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Janssen, N., Hoek, G., Brunekreef, B., Harssema, H., Mensink, I. and Zuidhof, A. (1998) Personal sampling of particles in adults: relation among personal, indoor, and outdoor air concentrations. *Am. J. Epidem.*, **147**, 537–544.

Janssen, N., Hoek, G., Harssema, H. and Brunekreef, B. (1997) Childhood exposure to PM10: relation between personal, classroom, and outdoor concentrations. *Occup. Environ. Med.*, **54**, 1–7.

Kaiser, M. and Cressie, N. (1993) The construction of multivariate distributions from markov random fields. *J. Multiv. Anal.*, **2**, 35–54.

Katsoyanni, K., Toulomi, G., Spix, C., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., Schouten, J., Ponka, A. and Anderson, H. R. (1997) Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *Br. Med. J.*, **314**, 1658–1663.

Kelsall, J., Samet, J. and Zeger, S. (1997) Air pollution, and mortality in Philadelphia, 1974–1988. *Am. J. Epidem.*, **146**, 750–762.

Korrick, S., Neas, L., Dockery, D. and Gold, D. (1998) Effects of ozone and other pollutants on the pulmonary function of adult hikers. *Environ. Hlth Perspect.*, **106**, 93–99.

Le, N. D., Sun, W. and Zidek, J. V. (1997) Bayesian multivariate spatial interpolation with data missing by design. *J. R. Statist. Soc.* B, **59**, 501–510.

Le Cam, L. and Yang, G. L. (1990) *Asymptotics in Statistics: Some Basic Concepts*. New York: Springer.

Li, Y. and Roth, H. (1995) Daily mortality analysis by using different regression models in Philadelphia county, 1973–1990. *Inhaln Toxicol.*, **7**, 45–58.

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

Lindley, D. V. and Smith, A. F. M. (1972) Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc.* B, **34**, 1–41.

Lioy, P., Waldman, J., Buckley, T., Butler, J. and Pietarinen, T. (1990) The personal, indoor, and outdoor concentrations of PM10 measured in an industrial community during the winter. *Atmos. Environ.*, **24**, 57–66.

Lipfert, F. and Wyzga, R. (1993) Air pollution and mortality: issues and uncertainty. *J. Air Waste Mangmnt Ass.*, **45**, 949–966.

Mage, T. and Buckley, T. J. (1995) The relationship between personal exposures and ambient concentrations of particulate matter. In *Air and Waste Management 88th A. Meet., Pittsburgh*, pp. 1–16.

Morris, C. N. and Normand, S.-L. (1992) Hierarchical models for combining information and for meta-analysis. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 321–344. Oxford: Oxford University Press.

Neyman, J. and Scott, E. L. (1960) Correction for bias introduced by a transformation of variables. *Ann. Math. Statist.*, **31**, 643–655.

Normand, S., Glickman, M. and Gatsonis, C. (1997) Statistical methods for profiling providers of medical care: issues and applications. *J. Am. Statist. Ass.*, **92**, 803–814.

Ozkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E. and Jenkins, P. (1996) Personal exposure to airborne particles and metals: results from the particle team study in Riverside, California. *J. Expos. Anal. Environ. Epidem.*, **6**, 57–78.

Raftery, A. and Lewis, S. (1992) How many iterations in the gibbs sampler? In *Bayeisan Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 763–774. Oxford: Oxford University Press.

Samet, J., Zeger, S. and Berhane, K. (1995) *The Association of Mortality and Particulate Air Pollution*. Cambridge: Health Effects Institute.

Samet, J., Zeger, S., Kelsall, J., Xu, J. and Kalkstein, L. (1997) Air pollution, weather and mortality in Philadelphia. 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*. Cambridge: Health Effects Institute.

Schwartz, J. (1995) Air pollution and daily mortality in Birmingham, Alabama. *Am. J. Epidem.*, **137**, 1136–1147.

Silliman, N. (1997) Hierarchical selection models with applications in meta-analysis. *J. Am. Statist. Ass.*, **92**, 926–936.

Skene, A. M. and Wakefield, J. C. (1990) Hierarchical models for multicentre binary response studies. *Statist. Med.*, **9**, 919–929.

Smith, R., Davis, J., Sacks, J., Speckman, P. and Styer, P. (1997) Assessing the human health risk of atmospheric particle. *Proc. Environ. Statist. Sect. Am. Statist. Ass.*

————(1998) Air pollution and mortality in Birmingham, Alabama: a reappraisal. *Technical Report*. National Institute of Statistical Science, Research Triangle Park.

Smith, T. C., Spiegelhalter, D. J. and Thomas, A. (1995) Bayesian approaches to random-effects meta-analysis: a comparative study. *Statist. Med.*, **14**, 2685–2699.

Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, **22**, 1701–1762.

Toulomi, G., Katsoyanni, K., Zmirou, D. and Schwartz, J. (1997) Short-term effects of ambient oxidant exposure on mortality: a combined analysis within the APHEA project. *Am. J. Epidem.*, **146**, 177–183.

Vedal, S. (1996) Ambient particles and health: lines that divide. *J. Air Waste Mangmnt Ass.*, **47**, 551–581.

Wakefield, J. and Morris, S. (1998) Spatial dependence and errors-in-variables in environmental epidemiology: investigating the relationship between magnesium and myocardial infarction. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 510–516. Oxford: Oxford University Press.

Wallace, L. (1996) Indoor particles: a review. *J. Air Waste Mangmnt Ass.*, **46**, 98–126.

Wikle, C., Berliner, M. and Cressie, N. (1997) Hierarchical Bayesian space-time models. *Technical Report*. Iowa State University, Ames.

Zidek, J., Wong, H. and Burnett, R. (1998) Including structural measurement errors in the nonlinear regression analysis of clustered data. *Can. J. Statist.*, **26**, 537–548.

Zidek, J., Wong, H., Le, N. and Burnett, R. (1996) Causality, measurement error and multicollinearity in epidemiology. *Environmetrics*, **7**, 441–451.

Discussion on the paper by Dominici, Samet and Zeger

David Clayton (*Medical Research Council Biostatistics Unit, Cambridge*)
I have particularly enjoyed reading this paper and I can think of no higher praise than to say that it made me think — particularly about the role of statistical models in this important and interesting problem. Since I have not worked in environmental epidemiology for many years, my remarks will largely concern the statistical methodology. However, I shall start by remarking on the choice of data to be analysed. My memories of my days in this field are that analyses of all-cause mortality contribute little to the unravelling of aetiology and the establishment of causality. For this, we need to get closer to the pathological processes and to look at cause-specific rates. The place of all-cause analyses is more in the assessment of public health *impact* after causality has been reasonably established. This raised an immediate question in my mind about what this analysis hoped to advance — scientific understanding or political policy? This question persists in my methodological questions, which concern not so much the choice of model but the choice of parameters which have been singled out for special attention.

The data involve observations at three levels:

(a) the person–day,
(b) the city–day and
(c) the city.

Although the interest is in causal relationships at the person level, most explanatory variables are aggregate summaries at higher levels. The difficulties and pitfalls of analyses using such 'compositional covariates' are well described in the extensive literature surrounding the 'ecological fallacy' (see, for example, Greenland and Robins (1994)) and I do not propose to venture onto this well-trodden ground. It should be noted, however, that even at the person level there are two distinct types of relationship between air pollution and mortality:

(a) the variation in the risk of death within subject over time in response to temporal variation in pollution levels and
(b) the variation between subjects' long-term risk in response to their aggregate exposure.

These may be very different relationships and we must beware of extrapolating from one to the other. The target of this analysis is the former relationship and this is arguably of rather less public health importance than the latter. Corresponding to the hierarchical nature of the data, the analysis also has three stages:

  (a) the *summary* of person level data into daily age-specific rates,
  (b) the *analysis* of age-specific daily rates at the city level, using generalized additive models with Poisson error structure, and
  (c) *meta-analysis* of the city level analyses using a Gaussian mixed model.

I have used the term 'meta-analysis' to describe the highest level analysis, whereas the authors have tended to describe it as 'pooling'. Although one aim of meta-analysis is the pooling of information on parameters that are imperfectly estimated at the lower level, it is by no means the only aim. Indeed, in epidemiology it may not even be the primary aim since a heterogeneity of effect can easily occur owing to methodological biases and confounding, and the *consistency* of findings is widely regarded as central to the establishment of causal relationships (Clayton, 1991).

The analysis presented by Dominici and co-workers does provide information on this point, although it is not greatly stressed. However, their interpretation of the analysis in this respect seems to differ from mine! Whereas they conclude that

'. . . there is a consistent association between PM_{10} and shorter-term variations in mortality'

I am not at all sure that the results of the analysis justify this statement. The estimate of mean and standard deviation of the city-specific slopes β^c are 0.48 and 0.76 respectively. On this basis, air pollution would be *protective* in 25% of US cities! In the presence of such heterogeneity of effect, the α-parameter representing the mean effect is of much less interest than those relating the variation of effect to characteristics of cities. Indeed, it is very doubtful whether this 'overall effect' parameter has any useful interpretation; in particular I am rather unsure in what sense it can be regarded as the 'national effect'.

A focus on scientific rather than regulatory aims makes the study of effect modification of more importance than the estimation of an overall summary of a heterogeneous effect. In this respect there is a role for pooling estimates of interaction terms which may be imprecisely estimated in individual within-city analyses.

These considerations also lead me to question the role of spatial modelling here. Whereas it is nice that we can add spatially autocorrelated terms in such analyses, what do we gain by so doing? The addition of random terms to a model explains nothing — it merely better characterizes that which remains to be explained! Thus, it might be of interest to know whether the heterogeneity follows broad geographical trends, since this might indicate confounding or effect modification by climatic variables, but the Bayesian method used here does not simply allow a direct *test* of this. The regional analysis is more informative, although one must question the fitting of a random effect for a factor on only three levels!

Finally I must question the use of Markov chain Monte Carlo sampling. Is this used because the authors wish to adopt an explicitly Bayesian posture or purely for computational convenience? Since the city level analysis is based on large amounts of data, the model may be treated as a pure Gaussian problem:

$$\hat{\beta}^c \sim N(\beta^c, V^c),$$
$$\beta^c \sim N(\alpha x, \Sigma_\theta).$$

The likelihood has a closed form (Harville, 1977) involving matrices which, in this application, are far from unmanageable. The parameters α and θ can be fitted by maximum likelihood and restricted maximum likelihood respectively and the β^cs can be estimated by empirical Bayes estimates. The only benefit that Markov chain Monte Carlo sampling brings is the ability to calculate interval estimates for β^c which reflect uncertainties in the estimates of θ. But, again, it is open to question whether these parameters hold any great interest in themselves. Since heterogeneity of effect is, in all probability, attributable to methodological artefacts and confounding it would not seem very important to explore the uncertainty of city-specific slopes very carefully; it is to be hoped that no-one will wish to construct a league table of these indices!

M. J. Campbell (*University of Sheffield*)
I would like to congratulate the authors on a veritable cornucopia of modern statistical methods, from log-linear generalized additive models to Bayesian hierarchical models and Gibbs sampling. Just now David Clayton mentioned that he was revisiting his past. I have also been revisiting my past since my first job was working on air pollution at the Medical Research Council's pneumoconiosis unit, albeit on occupational air pollution.

For those people who are not familiar with air pollution research, it may help to set the scene. PM_{10} is particulate matter less than 10 μm aerodynamic diameter (or, more strictly, particles which pass through a size-selective inlet with a 50% efficiency cut-off at 10 μm aerodynamic diameter). In the past particulate pollution was called black smoke, the major component being black coal smoke, and was measured by how much it darkened filter paper. Nowadays the black component is largely from diesel vehicles (Committee on the Medical Effects of Air Pollutants, 1995). This is accompanied by sulphates and nitrates generated from the oxidation of sulphur dioxide (SO_2) and nitrogen oxides (NO_x) together with industrially emitted particles and wind-blown soil and dust. Understanding the whole mixture is very important, not least to the road transport lobby because of the association with diesel exhausts. Some gases such as ozone (O_3) do not contribute to PM_{10} which is perhaps why the authors include it in the analysis. Disentangling the effect of gases such as SO_2 and NO_x from the effect of PM_{10} is very difficult and is the subject of much current research, often using cohort or laboratory studies. There is still much controversy over the relative contributions of different particles and gases to the effect of air pollution on health. As the authors state, a huge amount of data was processed, 7 years of daily data from 20 metropolitan areas, nearly 1 Gbyte in total. As anyone who has looked at time series knows, as soon as you start to lag covariates a huge number of data points ensue. Of necessity much of the detail is not available here. In the end the results are neatly summarized in Fig. 8, and the fact that a 10 $\mu g\,m^{-3}$ increment in PM_{10} is associated with daily mortality increases of 0.48% (95% interval 0.05–0.92) and adjusted for O_3 it becomes 0.52% (95% interval 0.16–0.85). Considering the number of assumptions underlying this estimate it is surprisingly close to the estimate of 0.42% given by Katsoyanni *et al.* (1997) for five western European cities for the best 1-day effect (I interpolated the results from coefficients for 50 $\mu g\,m^{-3}$), although the confidence interval (0.42–0.59) is much narrower in the European study. This possibly reflects a closer homogeneity of the European cities and the fact that because of the lack of heterogeneity a fixed effects model was fitted in the European study. I agree with David Clayton that it is silly to produce an overall average in the presence of heterogeneity, but I can understand that this is what the regulatory authorities require. One could add that the interval in the US study is disappointingly wide in terms of an interpretation for public health. For example the Quantification of Air Pollution Risk Committee in their report to the UK Department of Health used a World Health Organization figure of 0.74%, which, although it is well within the 95% interval supplied here, implies approximately double the effect suggested in this paper, a difference which would have major public health consequences (Committee on the Medical Effects of Air Pollutants Quantification of Air Pollution Risk Sub-Group, 1997).

There has been widespread criticism of the use of time series models in this area, particularly to investigate causality (Rushton, 1999; Gamble and Lewis, 1998). The authors here have confined themselves to a description of the models and methods of estimation of the parameters and, wisely in my view, steered clear of controversy in terms of an interpretation. In particular, it is difficult to extrapolate the relative risks to measure the numbers of lives that might be saved if pollution were reduced. There is evidence that the deaths are simply 'brought forward', possibly by only a few days, a point that the Committee on the Medical Effects of Air Pollutants Quantification of Air Pollution Risk Sub-Group (1997) was keen to make.

One way of reassuring the reader about causality would have been to include a lag -1 in Fig. 8. If they could have shown that the effect definitely does not precede the cause then causality is rather more plausible (Campbell and Tobias, 2000).

The authors assume a log-linear model between mortality and pollution, but they allow all other effects on mortality to be approximated by a generalized additive model. Although this helps the interpretation and enables us to pool different cities' results, is the model truly linear? All the other confounders, temperature in particular, are known to have a non-linear effect. I would imagine that linearity only holds true within a fairly narrow range of pollution levels, and I would like to have seen more evidence to support the assumption.

Turning to the time series aspect, one point that I noted, which concurred with my own experience (Campbell, 1997) is that, when confounding factors are correctly accounted for in this type of

environmental time series, the serial correlation of the residuals disappears. In general, since most of the deaths are not caused by infection, they are only related by common environmental factors, and conditionally on these factors they are independent. Thus the usual problem of time series regression with correlated errors is removed! However, although I was pleased to see Fig. 4, as presented with so much blank space it is largely uninformative. Also it would have been nice to have been reassured by a Ljung–Box-type test that there were no hidden signals in the residuals.

There is clearly much further research to be carried out in this area. To misquote Churchill, it is not the end, nor even the beginning of the end, but it is perhaps the end of the beginning of research in this area!

It gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Nicholas T. Longford (*De Montfort University, Leicester*)
The authors should be congratulated on employing the impressive apparatus of Markov chain Monte Carlo sampling with integrity and rigour. I would like to mention a non-Bayesian alternative to hierarchical modelling which may be easier to trace back to basic principles. Simply, the unbiased large variance estimator from the city's data, $\hat{\theta}_c$, is combined with the biased small variance national estimator $\hat{\theta}$ from the entire database:

$$(1 - b_c)\hat{\theta}_c + b_c\hat{\theta},$$

with b_c chosen to minimize the expected mean-squared error. The Bayesian prior can be incorporated similarly. See Longford (2000) for details. Since there is much background information, I find that imposing the uninformative prior is contrary to both the Bayes spirit and the spirit implied by the title of the paper.

The study conducted is observational, without a random allocation of subjects to cities, cities to pollution levels and the like. Therefore the estimated quantities are adjusted differences, not *effects* that would allow a causal interpretation. We are exposed to air pollution throughout our lives and, given resources, we take active measures to reduce its effect — by a choice of housing, climate control (air-conditioning) or migration. So the population implied by the study is highly selective, with an excess of the immobile poor and, depending on administrative boundaries, a shortfall of the affluent commuters. In the USA, a common pattern of migration among the mature affluent is to abandon the city or suburbs at the end of the career of employment. The extent of pollution before death is bound to become less relevant as more deaths occur in hospital intensive care units and other controlled environments with discountable acute effects of air pollution. The cities studied, and Los Angeles in the extreme, cover large areas with diverse patterns of air pollution which are inadequately captured by a single daily quantity.

Separating the chronic and acute effects of air pollution requires an intricate analysis, with plenty of leeway allowed, e.g. by means of sensitivity analysis (Rosenbaum, 1995), for an imperfect understanding of their effects on human health. The mechanistic interpretation offered by the authors' analysis is akin to inquiring about patients dying of lung cancer or cirrhosis of the liver how much they smoked or drank on the few preceding days. The exposure in the recent past is of distinctly secondary importance to the long-term exposure to the identified hazard; the authors assume that the two kinds of exposure are orthogonal.

Ben Armstrong (*London School of Hygiene and Tropical Medicine*)
I found this paper and presentation rich in many respects. I hope that the authors will forgive me for making a comment, not on the hierarchical part of the model, but on the city-specific modelling of seasonal variation. For this Dominici and co-workers, along with most recent investigators of daily mortality–air pollution data, have chosen to depend entirely on flexible general purpose smooth functions of time. I wonder whether this is optimal.

In my mind 'season' refers to patterns that repeat year after year, following the rotation of the earth about the sun. The reason that season, thus defined, is suspected as a confounder here is that there is ample evidence that it is associated with both mortality and with air pollution. Much of the seasonal variation in mortality can usually be explained by models for measured temperature and humidity such as those used by Dominici, but not all of it can be. General purpose smooth functions in time, especially

those as amply parameterized as in Dominici's model, would be expected to mop up much remaining seasonal variation, but again not necessarily all of it. Subtle seasonal effects would be less well picked up by such functions than by models that are specifically designed to identify annually repeated patterns, such as Fourier terms.

There are, of course, long-term variations in mortality that would not be repeated year after year, and thus need to be modelled in their own right. My point is that lumping these in with season may not be the most effective way of controlling for confounding by both. This line of argument would suggest including *both* a specifically seasonal component of the model (e.g. Fourier terms) *as well as* general smooth functions of time. The numbers of parameters given to the two could be informed by the usual empirical model choice criteria. In my limited experience with mortality series, this generally leads to a mixture.

I would be interested in the authors' thought on this issue, which may be one that they have considered on their path to their current choice. I would especially like to know whether they have tried adding Fourier terms or other explicitly seasonal terms to their model, and if so whether this improved the fit or changed the pollution parameters.

Sylvia Richardson (*Institut National de la Santé et de la Recherche Médicale, Villejuif*)
I would like to congratulate the authors on this impressive and timely paper which has clearly required considerable work in model development, sensitivity analysis and data handling.

My brief comments concern the specification of the second-stage models used to synthesize the results over the 20 cities. As the authors comment, there is substantial heterogeneity in the relative rate estimates. Faced with this, and with the small size of the estimates, we can suspect that there is a potential interplay between the specification of the between-city distribution and the posterior estimates of the overall effect and its variability. Indeed, there is some evidence of this in the different amount of shrinkage obtained by the spatial analysis in comparison with the non-spatial analysis.

Faced with heterogeneity, the usual approach is to try to explain it by including relevant regression variables and/or to use a flexible statistical model to make the inference robust to a misspecification of the between-city distribution. The authors have addressed this delicate issue by

(a) including a few city-specific covariates (but it turns out that they are not useful),
(b) making a Gaussian assumption on the between-city distribution of the β^cs and
(c) assuming independent (exchangeable) or spatially dependent (partially exchangeable) β^cs.

It would be interesting to see the authors investigate the appropriateness of the Gaussian assumption, for example, by including in their sensitivity analysis a heavier-tail distribution like the t or alternatively a mixture of Gaussian distributions, an option which is sometimes used for random-effect modelling when there is substantial heterogeneity (Watier *et al.*, 1999). It is not easy to see from Figs 2 or 3 whether there is evidence for outliers or subgroups in the β^cs, though the results from the regional spatial model indicate that there may be. An exploration of subgroups will be particularly interesting when the analysis is extended to the 90 cities. It would also be helpful if the assumption made on the inverse Wishart distribution for Σ, with 3 degrees of freedom and scale matrix D chosen to be diagonal with elements equal to 3, could be made more transparent in terms of the prior range for the variability of the β^cs, and other settings included in the sensitivity analysis. Maybe the flexible approach used by Spiegelhalter and Marshall (1999) in their investigation of a hierarchical model for institutional comparisons is relevant here.

Lesley Rushton (*Medical Research Council Institute for Environment and Health, Leicester*)
I welcome the development of statistical methods for combining findings from ecological studies such as time series. These methods have potentially many applications in the field of the environment and health. My comments will focus on the interpretation and use of the results from the combined and individual studies, rather than on the actual statistical methodology.

The authors use the phrase 'controversial associations' to describe the results from the many studies of air pollution and mortality. This controversy arises because the association has been interpreted by some as implying causation, and concerns raised include biological plausibility and mechanisms, inconsistencies across studies and the lack of data on individual exposures and other characteristics. The authors state that their primary objective was to estimate a national relative rate of pollution. However, perhaps a more useful aspect of their methodology is the potential to explore the factors influencing the heterogeneity of results from individual studies. This can facilitate the development of risk reduction strategies.

Time series studies are useful to inform 'population-based' risk reduction approaches, in which a shift in the distribution of risk is sought, so that the overall mean risk is reduced. This is a traditional public health approach but may have little obvious benefit to an individual. In the air pollution situation, it is unclear whether this mass environmental control will have the desired effect on public health.

A different approach, e.g. the use of cohort studies, in which data on participants' health and life style characteristics are available and for which an attempt is made to estimate individual exposure to pollutants, is needed to quantify individual risk, to identify high risk susceptible individuals and to inform individual-based risk reduction strategies.

The limitations that are inherent in the individual air pollution time series studies will to some extent be carried over into any pooled analysis. In particular, data on many city-specific variables, such as data on particle sources and pollutant mixtures, which would aid an exploration of heterogeneity are not readily available. The Health Effects Institute and European Union have included many of these gaps and limitations as high priority areas for future research. Health effects from exposure to particles are likely to remain a key public health issue and this paper has made a useful contribution to the debate.

Gillian M. Raab (*Napier University, Edinburgh*)
The analyses carried out in this paper are made at the aggregate level for large cities. Yet the inferences drawn from them are implicitly assumed to apply at the level of the individual. There is now a substantial literature that investigates when such ecological inference may be misleading. There are two factors that, taken together, can distort the individual relationship when it is viewed at the aggregate level (see for example Greenland and Robins (1994)). These are

(a) spatial heterogeneity of the exposure source and
(b) non-linearity of the exposure–outcome relationship.

It is likely that these two factors may be operating within cities in the results presented here and may be part of the explanation of the variability in the apparent effects of the pollutants between cities. For example, the true exposure relationship at the individual level involved might be a threshold above which mortality was increased and the mean level of pollutants might differ across a city. Then changes in short-term mortality would show up only in those parts of the city where the mean levels were close to the threshold. This might introduce the sort of between-city heterogeneity on the effects of pollutants that is seen here. I can see two ways in which some clarification about whether this might be happening could be obtained in the context of the present study. You mention in Section 2 that pollution data were obtained by averaging across monitors. To what extent did these values vary systematically within cities? Also, have any analyses been carried out using units smaller than complete counties? You have introduced a hierarchical model that takes you up to a higher level of aggregation than the simple analyses you present in Section 3. For all its complexity this does not seem to add any new insights to the interpretation of the data. Going down to a lower level in the hierarchy might prove more interesting.

Michael Quinn (*Office for National Statistics, London*)
I have one point and one query for clarification. I would like to pick up David Clayton's first comment related to the aetiology and to ask why are the authors looking at all-cause mortality?

It is difficult to believe that the PM_{10}-level or any other pollutant will affect deaths from cancer unless it is exactly the bring-forward effect that was mentioned earlier—or for that matter suicide, unless people become depressed when they cannot see the view, or even accidents, especially those outside the home, unless the pollutants are so dense that people cannot see where they are going and there are extra road traffic accidents. There would be more variability to explain if the analyses were restricted to the sort of deaths that it is felt might be more directly related to the pollution.

There are probably two more interesting questions than those being addressed by the paper. First, does the overall pollution affect, or lead to higher, death-rates? To determine that, as Michael Campbell mentioned, we need to look at time series, to look over a long period. If the overall level of pollution is going down, are the deaths related to that pollution also going down? This cannot be determined from the sort of analysis in the paper.

The second important question to address is whether short-term pollution incidents indeed have this bring-forward effect. This is seen with severe weather conditions: there is always a peak, followed later by a corresponding shortfall. It should be possible to look at the same sort of patterns relative to pollution incidents.

My query is in Section 3 where the authors say

'Since there are missing values of some predictor variables on some days, we restricted analyses to days with no missing values across the full set of predictors'.

Later in the same section they say 'many cities have PM_{10} data available only every sixth day'. The authors do not say how many cities. Does this mean that for many cities only a sixth of the mortality data were actually used? Also, when fitting the lag models, does it mean that the deaths at lag 0 are different from the deaths used in the model for lag 1, which are again different from the deaths used in the model for lag 2?

J. A. Nelder (*Imperial College of Science, Technology and Medicine, London*)
On the ninth page, the authors set out what is essentially a Poisson–normal two-stage model. Then, on the next page, by an approximation they turn this into a normal–normal two-stage model. With respect to the second model, it was pointed out by Mr Clayton that negative αs suggest that pollution is actually beneficial in those places. However, this is not true if we go back to the original model where all these effects are on a log-scale. The fact that there are quite different implications for negative αs from the two models makes me wonder exactly what is the connection between the normal approximation on the tenth page and the original Poisson–normal model on the ninth page.

Martin Bland (*St George's Hospital Medical School, London*)
I would like to thank Francesca Dominici for a stimulating presentation. The department in which I work is collaborating with her and her colleagues in European studies of pollution and health (to which Michael Campbell alluded). I thought that there was no doubt now that there is a relationship between daily changes in air pollution and daily mortality and morbidity, particularly among the elderly. (David Clayton doubts it, so that is not true—there is doubt about it.) However, if we accept for the moment that these relationships exist, I would like to ask whether this is the important effect. Is this what we actually want to estimate?

I would like to know the effect of chronic exposure to air pollution, not just over days or months but over many years. What, for example, has been the effect on me (I like to make these things personal) of living in central London for 34 of the past 35 years, 25 of those years living just off the A23 Brighton Road and inhaling vast amounts of air pollution from traffic going past?

I appreciate that this is difficult to determine, which is why we look at the very short-term effects of daily variation in air pollutants or short episodes of very high levels of pollution. However, I think that the effects at which the authors are looking are merely surrogates for the effects of long-term exposure to air pollution which *must* be the important thing. These estimates do not measure the effect of air pollution but are looking at what can only be one small part of the effect of air pollution. Therefore, going to great lengths to produce improved interval estimates and so on for them is beside the point. They merely show that air pollution has some relationship to health; they are not the whole or even the main story.

I would be interested in any ideas which Francesca Dominici and her colleagues may have about how we can start to estimate the real effects of air pollution, of which what we are looking at here is only a small part.

Alison Macfarlane (*National Perinatal Epidemiology Unit, Oxford*)
I worked in air pollution research in the mid-1970s. The period when the peaks in mortality were so huge that associations between daily deaths and air pollution could be detected simply by using a pencil and graph paper had recently ended. The disappearance of these peaks led some people to conclude that no further research was needed (Holland *et al.*, 1979). As a result, most research in the 1980s that used data from England took place in the USA (Mazumdar *et al.*, 1982; Ostro, 1984; Thurston *et al.*, 1989; Schwartz and Marcus, 1990).

As the paper has shown, the much more powerful techniques used in recent work show that associations between air pollution and daily mortality can still be detected. Nevertheless, I agree with previous speakers that many important questions remain unanswered, and I would like to add to the list.

The paper analyses all-cause mortality, but there is a long tradition of using cause-specific mortality (Russell, 1924, 1926; Martin and Bradley, 1960; Lawther *et al.*, 1970; Macfarlane, 1977). In particular, deaths attributed to cardiovascular and respiratory causes have been extensively analysed. In addition, North American research found associations between homicides and high temperature (Rogot *et al.*,

1976). Underlying causes of death alone can give an incomplete picture, however. Using all the causes mentioned on death certificates can yield fuller information, although they still leave vast gaps in our knowledge about people's backgrounds and circumstances. They may also reflect differences in certification practice.

The question of whether exposure to air pollution simply hastens death by a few days is an old one and dates back at least to the report of the 1952 smog (Ministry of Health, 1954). A question that has been less fully considered is that of the medium term. For example, if the illest sections of the population have just been exposed to influenza and have survived it, does this then alter their susceptibility to subsequent short-term variations in temperature or air pollution (Macfarlane, 1984)?

In the context of a country like the UK that does not usually experience extremes of temperature in winter or summer and is therefore unprepared for them, a further question arises. Are our ill or elderly people more susceptible when extremes do occur compared with countries with more widespread access to air-conditioning and central heating?

Finally, compared with air pollution, influenza epidemics and extremes of temperature seem to have more marked and visible associations with short-term changes in mortality, yet they are usually relegated to the status of confounders. Could they be more important in public health terms? Studies of daily mortality can shed some light on this question, but other research is needed as well.

Ian L. Dryden (*University of Nottingham*)
This is an interesting and thought-provoking paper. It would have been good to see the box plots for the regression coefficient posteriors given for each city under the spatial model (similar to Figs 6 and 7). Presumably these posteriors have substantially higher variances than under the univariate and bivariate base-line models, judging from the pooled analysis of Fig. 8. So, the effect of air pollution seems to be much less significant under the spatial model.

As the spatial correlation appears to make quite a difference it is important to check whether or not the spatial model is reasonable. What model checks have been carried out to check whether the exponential correlation with either choice of distance is satisfactory? The model might be more reasonable for southern Californian cities but less so for say New York–Philadelphia or Chicago–Detroit I would have thought. A non-stationary model within and between regions would seem more natural and of course there is a large amount of data for estimating correlations between specific pairs of cities. In this particular application spatial prediction between sites does not seem particularly important and so a model may only need to be developed at the specific sites rather than in continuous space.

A non-separable spatial–temporal covariance model would also be an appropriate avenue for further exploration and may be worth examining with the enlarged data set of 90 cities.

R. I. Smith (*Institute of Terrestrial Ecology, Penicuik*)
This paper focuses on one aspect of the important question of detecting effects of air pollution on human health. I welcome it and would like to comment on two issues.

First, in common with several other studies, the authors use data from a single air pollution monitor, or an average of two or three, to reflect some surrogate for personal exposure in a city. In the famous 1952 London smog, health problems were triggered across a large area as pollution from coal combustion, emitted from domestic and industrial chimneys, was mixed throughout the city's atmosphere. To some extent air pollution studies have continued to assume this model. This paper, however, focuses on fine particulates where a substantial fraction of the pollution comes from road traffic. The emission is much closer to the ground and the atmospheric mixing can be quite different. Concentrations are more spatially variable with high concentrations on a main road rapidly and substantially reduced down side-streets. Even the temporal pattern of concentration can change over short distances. As the continuous monitoring systems, which provide the appropriate data for this type of health effects study, are relatively expensive, the characterization of the spatial and temporal variability is often put to one side. Single monitors, often deliberately located to detect high concentrations, may or may not represent large areas depending on their precise location and the particular urban landscape. In comparing across cities, this unquantified within-city variability can be sufficiently large to mask any differences in health outcomes.

Secondly, on a related aspect of the work, the authors use a spatial correlation based on the distance between cities. Broad regional differences across the USA may be important, as indeed the authors show. However, it is not obvious that the distance between cities is the relevant measure of similarity. The physical structure of the environment is very important and the pollution climate in cities with

similar structure may be closely related however distant the cities. Some attempt to characterize the type of city by its structure may be more useful. There is an increasing temptation to use a spatial correlation structure because it can easily be set up, but that does not mean that we should necessarily do it without supporting evidence.

The following contributions were received in writing after the meeting.

Kiros Berhane and Duncan C. Thomas (*University of Southern California, Los Angeles*)
We congratulate the authors for writing such a seminal and academically stimulating paper. The ideas that they present with outstanding clarity will, no doubt, shape the way that we do research on the health effects of daily air pollution for many years to come. Our comments will focus on two aspects of the paper, namely the use of generalized additive models to dealing with autocorrelation in the data and ways of dealing with the lag structure of the pollution effect. We fully agree with the authors on their approaches to dealing with the lag structure and their choices of the amount of smoothing done to account for the long-term time trends and autocorrelation in their data. Our comments are mainly intended to make the methods proposed more globally applicable to related, yet different, applications. For example, we are currently analysing data on acute effects and/or short-term chronic effects of pollution on daily school absences and chronic effects of pollution on longitudinal yearly measurements of pulmonary functions from a study of 12 southern California, USA, communities. It would be nice also to apply the methods proposed in this paper to such data.

(a) The authors have demonstrated (Fig. 4) that the amount of smoothing that they do adequately takes care of the autocorrelation in their data. However, there is no clear guideline on how to choose the amount of smoothing when applying the methods in a different context. We hope that the authors will comment on two issues related to this. Is there a way to produce an automated procedure to select the amount of smoothing that needs to be done to choose the smoothing parameter value that takes care of the autocorrelation? Such an approach would probably need to take the dependence in the data into account in the style of Altman (1990), Hurvich and Zeger (1990) and Diggle and Hutchinson (1989). Is it possible to examine analytically the correspondence between the methods proposed and methods that directly account for the autocorrelation through an appropriate filter? We believe that the two questions are interrelated and the answer to the latter could probably help in tackling the former.

(b) The use of methods such as the polynomial distributed lag model may not be possible in the authors' application since PM_{10} data are only available every sixth day. But, would it be possible to apply the methods proposed in the distributed lag framework, if such data had been available?

We thank the authors again for their excellent contribution and the Editors for giving us the chance to comment on the paper.

John W. Cotton (*University of California, Santa Barbara*) **and Abdul R. Othman** (*Universiti Sains Malaysia, Pulan Pinang*)
This impressive paper provides important information on possible deleterious effects of air pollution. We wonder whether the authors' use of linear fits of pollution variables to log-deaths and smoothed function additive fitting of contributions from nuisance variables rather than generalized additive modelling throughout is process driven. Would employing a full additive model require less developed Bayesian methods or using some other hierarchical technique (Hastie and Tibshirani (1990), pages 156 and 266–271)? Also, do two counter-intuitive results, a non-significant decline in deaths in Miami with increases in PM_{10} (Figs 2(a) and 3(a)) and a slightly negative posterior mean for PM_{10} effects in the south, suggest the need for further inquiry?

Implications for social science
The current methods may be useful with multiperiod longitudinal behavioural studies or their special case, crossover designs in which time-varying covariates are controlled experimentally.

(a) In criminology one might treat the number of murders in a city per day as a Poisson variate to be predicted from the maximum daily temperature and possibly copy Dominici and co-workers with a semiparametric log-linear analysis and an intercity hierarchical regression model.

(b) In much psychophysical research, few people (subjects) are employed, the response measure is ordinal and there are few between-subjects covariates. An experiment on the temperature

perceived when a metal bar is touched by a human finger (Refinetti, 1989) yielded interesting data from seven subjects, with perhaps only three available for further analysis (Cotton and Othman, 1991). There were 90 temperature judgments, three for each of 30 temperature values ranging from 33 to 42 °C. At the beginning of each session an extra 33 °C stimulus was presented as a reference point; the subject was told to call it a '20'. No other constraint was placed on responses, which ranged from 14 to 22 for one subject and from 0 to 115 for another.

Refinetti (1989) fitted linear and power functions to 52 subjects' combined data for only 30 judgments each and power functions to all individuals' data. Gender, age and body temperature were available as between-subjects covariates. Refinetti's analyses for 90-observation data need to be supplemented by something comparable with the methods of Dominici and co-workers, with temperature and prior temperature (carry-over) as predictors and period as a nuisance factor, possibly employing proportional odds measures and a version of hierarchical analysis illustrated by Ribaudo *et al.* (1999), assuming that the sample of subjects for level 2 is not too small. Other relevant analysis methods include Kalman filtering and covariance structure modelling (Smith, 1996).

Brent A. Coull and Maura Mezzetti (*Harvard School of Public Health, Boston*) **and Louise M. Ryan** (*Harvard School of Public Health and Dana-Farber Cancer Institute, Boston*)
We congratulate the authors on this excellent paper and appreciate the opportunity to provide some brief comments. Historically, the default approach in environmental risk assessment has been to choose a single study as the basis on which to develop regulatory standards. We agree with the authors that a careful synthesis of information over several different studies may provide a more reliable and appropriate basis for risk assessment.

We found the authors' use of an approximate two-stage Bayesian analysis to be particularly interesting. The approach is heuristically appealing and is in fact similar to the *ad hoc* two-stage algorithm that was often used to fit linear growth curve models before the advent of programs such as SAS PROC MIXED (see, for example, Laird (1990)). We appreciated the reference to the theoretical justification given by Daniels and Kass (1998) and were impressed by the empirical evidence of how well the approach works in the current setting.

Another advantage of the two-stage approach is its potential for application in settings where the original study-specific data are unavailable. For example, we might have access to only summary measures (e.g. dose–response slopes and standard errors) based on published results. We have encountered exactly this situation in a current project assessing the effect of *in utero* methylmercury exposure on multiple neurological outcomes. To be precise, our data comprise a series of estimated slopes $\hat{\beta}_{ij}$ for study $i = 1, \ldots, I$ and outcome within study $j = 1, \ldots, J_i$, along with corresponding standard errors $\hat{\sigma}_{ij}$. Our approach involves fitting a hierarchical model to the $\hat{\beta}_{ij}$ in BUGS (Spiegelhalter *et al.*, 1995), with levels corresponding to study and outcome within study. The authors have provided a useful framework for evaluating our approach. It would be useful to develop guidelines, for example, on the number of studies and the sample sizes within study required to yield reliable results.

There are many broader challenges in addition to these interesting technical ones. For example, how should we allow for varying quality of studies, especially when using summarized, rather than original, data. One approach would be simply to inflate the study-specific standard errors $\hat{\sigma}_{ij}$, although it is not clear how to do this appropriately.

Kate Cowles (*University of Iowa, Iowa City*)
I am delighted to see this application of Bayesian hierarchical analysis to increase our substantive understanding of an important and complex public health issue.

The appropriate use of asymptotic normal approximations at the first stage (where the city-specific sample sizes are extremely large) makes the computing required to fit the hierarchical model straightforward and feasible. In fact, with the exception of the spatial model using Euclidean distance, the WinBUGS software (Spiegelhalter *et al.*, 1999) could fit the models proposed, thus making the strategy convenient and accessible to epidemiological researchers.

The method used to assess convergence of the Markov chain Monte Carlo sampler is not clear. As reported by the Raftery and Lewis convergence diagnostic in the CODA software, the quantity '$N_{\min}$' is the number of *independent* samples required to obtain a specified probability of estimating a specified quantile with a specified degree of accuracy. The number of (dependent) sampler iterations required is called 'total (N)'.

Because the conclusions regarding the relationship between same-day PM_{10} and mortality are much weaker under the spatial model than under the other models (Table 4), a formal Bayesian model comparison would be helpful. Bayes factors are feasible for this because all priors are proper. The method of Chib (1995) could be used to compute the required marginal likelihoods for the models for which all full conditionals are standard forms. For the spatial model, in which there is only a single scalar parameter for which a Metropolis–Hastings step is required, the kernel density estimation method of Chib and Greenberg (1998) is likely to provide an accurate estimate of the marginal likelihood. The auxiliary simulation should not be excessive for a model with only 20 cities.

Christopher Cox (*University of Rochester Medical Center*)
I would like to congratulate Dr Dominici, Dr Samet and Dr Zeger for their outstanding work in developing this Bayesian approach to combining results from separate analyses of data from individual cities. A real strength of the approach is the development of a two-stage model. The first stage of the model seems exactly right, and the authors were wise to preserve this aspect in developing their extension. And, although I am not really a Bayesian, we all seem to be increasingly appreciative of the conceptual merits of the Bayesian framework. Thus a hierarchical modelling strategy for the second stage of the model seems both reasonable and elegant.

I would like to add a brief comment growing out of my own involvement in an observational study of environmental exposures, which may have some relevance in the present context. I have become convinced that what might be called confirmatory observational studies of toxic exposures, where results often have political and economic as well as scientific consequences and opinions can be sharply divided, should aspire to the standard set by clinical trials in terms of preplanning. This applies particularly to the statistical analysis, and I believe that a prespecified analysis plan should be an essential component of the study design. In this spirit I suggest that it might have further strengthened the final conclusions to develop the approach using the data from the 20 cities without specifically addressing the association between particle levels and mortality. This could have been reserved for the full analysis involving 90 cities. Indeed I found it somewhat difficult to separate the evaluation of the approach from the interpretation of the results. Generally, in evaluating results I would like to hear additional testimony from the original data regarding the estimated effects. For example, although the second stage of the model provides the final synthesis, much can be learned by examining the first-stage analysis of the individual city data. Figs 2 and 3 display individual confidence intervals; it would also be useful to provide an idea of the consistency of the data with the Poisson regression model. Possible examples include plots of observed and predicted deaths over time, as well as something like partial residual or regression plots to show the consistency of the data with the estimated effects. This information does not seem critical for the explication of the model, but it might help in evaluating the results.

Lawrence H. Cox (*US Environmental Protection Agency, Research Triangle Park*)
Environmental science and management are fed by individual studies of pollution effects, often focused on single locations. Data are encountered data, typically from multiple sources and on different time-scales and spatial scales. Statistical issues including publication bias and multiple comparisons are often present but unaddressed. Policy makers must pool individual studies to infer and understand pollution–health effects relationships and trends and act on them. Inference and pooling cannot rely on traditional design-based methods or meta-analytic techniques, and model-based, hierarchical, Bayesian methods are needed. The authors have contributed to environmental epidemiology and environmetrics by providing a general hierarchical methodology for pooling estimates of pollution health effects, demonstrated for the important case of particulate mortality effects.

Contributions of this work are to enable scientific investigation on a regional or national scale, and to borrow strength between multiple studies, some of which may be too small (in observations) or too coarse (in frequency) to provide reliable estimates at all locations. Potential advantages are to enable a critical examination of the problem at single locations, e.g. to assess bias or to explore additional covariates. The latter is particularly important for particulate matter health effects, as questions remain regarding the relative mortality effects of size, chemical composition, shape and number of particles. Effects of multiple pollutants are also important, including fine (2.5 μm or less) and coarse (2.5–10 μm) particles.

Some observations

The examination of bias introduced by measurement error is important and was raised in National Research Council (1998) on setting research priorities for particulate matter studies. The authors selected a two-stage, not fully Bayesian, model. Pooled meteorological adjustments might capture otherwise ignored regional effects and improve estimates of local relative mortality effects. Some regional effects are captured by the spatial analysis, based on distance. Direction may also be influential; geometric anisotropy (Cressie (1994), page 64) could be incorporated into the spatial model. A fully Bayesian approach enables assessing uncertainty due to model selection (Clyde, 1998). Despite evidence of a negligible difference between the two approaches for these data, a general methodology for national surveillance of effects of air pollution and weather on public health would probably benefit from a fully Bayesian approach. The authors did not set out to compare US national air quality standards for particulate matter with epidemiological evidence of particulate matter health effects but observe that estimated relative effects obtain above and below the regulatory standard. A more meaningful regulatory comparison would be based on the maximum daily concentration.

The views expressed are solely those of the author and should not be interpreted as representing the policies or practices of the US Environmental Protection Agency.

Sudhir Gupta and Balakrishna S. Hosmane (*Northern Illinois University, DeKalb*)
This is a nice application of two-stage methods which uses a mixture of classical and Bayesian methodologies to an important environmental issue concerning the effects of pollution on mortality. Although a full log-linear model which incorporates the intracity and intercity variability for estimating the air pollution mortality relative rates would be a natural choice, it is nice to see that the two-stage modelling approach gives indistinguishable results.

As we know, patients develop resistance to treatments over time: something similar could occur with pollution. Probably, the spline function used for the time variable in the log-linear model would account for the plateau effect due to acclimatization to pollution.

To adjust the air pollution mortality relative rate for the age effect, it might be worth including it either at the first stage with interaction or only at the second stage without interaction to avoid overadjusting for it. Also, are we losing information by discretizing the age variable? Is there something to gain by using different discretizations of the age variable at the two stages? The final results may remain unchanged, but some discussion on this aspect might be helpful.

The random-error variance within each city may also be influenced by the way that the data have been collected. The cities could be classified into two or three groups on the basis of the percentage of days with missing values on particulate matter or the widths of the confidence intervals displayed in Figs 2 or 3. At the second stage of the modelling, the heterogeneity between these groups may be accounted for in the mean model itself or in the specification of the covariance structure. This might lead to the same conclusions as evidenced by the indistinguishable results obtained by the authors using four different models.

It may be worthwhile to know how the model fitting methods of Katsoyanni *et al.* (1997) and Toulomi *et al.* (1997) contrast with the authors' methodology and the approach of others.

Lyle C. Gurrin (*King Edward Memorial Hospital, Perth*)
The authors are to be commended for the clear presentation of their efforts in modelling the relationship between air pollution and daily mortality counts in US cities. They have taken some trouble to justify the assumptions and simplifications (in particular the two-stage modelling procedure) that have allowed the development of statistical models based on a large, complex data set. Subject-matter specialists will no doubt be pleased that a rigorous statistical analysis can produce a quantitative relationship between levels of particles in outdoor air and the risk of death that has a straightforward interpretation.

Analyses were restricted to days with no missing values across the full set of predictors, which despite the size of the complete data set results in a substantial reduction in the data available for analysis in some cities; 10 of the 20 cities have PM_{10}-values that are missing for more than 80% of the days covered during the study period. Moreover, for those cities where the PM_{10}-values were available only every sixth day, mutually exclusive subsets of the complete mortality data for that city will result when the pollution variable is lagged by 1, 2 or 3 days and days with no matching lagged pollution values are subsequently excluded from the analysis. Alternatively, we might express this phenomenon by regarding

the days on which the pollution measurements are available as fixed and advancing the 'mortality' counts by 1, 2 or 3 days.

Is it the case that complete daily data on PM_{10}-levels are available for some time periods even in those cities that usually record pollution levels only every sixth day? Do the authors think that the distinct data sets generated from the lagging process contribute to differences between the results of the analyses at each value of the lag?

A. N. Pettitt (*Queensland University of Technology, Brisbane*)
I congratulate the authors on an excellent paper which is a *tour de force* in terms of the analysis of a large data set. The authors combine modern techniques and models to find answers for a very significant scientific problem.

First, I have some remarks about the analyses in particular and, secondly, some general remarks about the methodology which has been adopted.

For the city–age group time series analyses, there is a similarity with the work of Campbell (1994), on daily sudden infant death syndrome deaths and temperature, and our own work, on weekly cases of pneumonia and the use of an antibiotic in a hospital (Hay and Pettitt, 2000). In these papers, an explicit low order time series structure is introduced into the linear predictor, the authors' equation (1), where a smoothing spline is used. However, given the incompleteness of the authors' data, their approach using a spline representation for the time series is probably very efficient. But, I would have liked to have read the authors' general views of the various pros and cons of the time series modelling approaches.

One wonders, incidentally, what, if any, bias might be introduced by the selection of days with a full set of responses and predictors. For example, PM_{10} is only measured for fewer than 30% of days for several cities.

In the second level of the hierarchy, the authors adopt a normal distribution. As a safeguard against outliers and the assumption of normality, one wonders what differences in posterior distributions might have been obtained if a *t*-distribution with a small number of degrees of freedom had been adopted.

As a comment on the general methodology, that proposed by the authors appears to offer practical solutions to problems with several layers of complexity or heterogeneity. The first-level summary analysis allows checking of assumptions and the second-level analysis allows the pooling of information without large data sets being involved. We are currently using a similar strategy with a large spatial data set involving a binary response and many covariates at each of many sites. Building a large Markov chain Monte Carlo code might be possible but obtaining error-free code, convergent runs and analysing the results would be a far more complex activity than following the strategy suggested by the authors here.

Marc Saez (*Universitat de Girona*)
When I read in the synopsis that the authors have developed and applied a hierarchical regression model I expected to find a mixed model, i.e. a model that fits both fixed effects and random effects. What the authors show in their paper, however, is a meta-analysis, a Bayesian meta-analysis if you like, but a genuine one. As Hedges and Olkin (1985) defined it, meta-analysis refers to the pooling of results (as in Section 4 of the paper) of separate studies (see Section 3), all of which are concerned with the same research hypothesis. I do not want to say that a meta-analysis would not be valid, especially when, because of the large sample size at each city, the computations became extremely cumbersome.

Meta-analyses, although practical and simple, present some limitations that, I think, have not been explicitly indicated in the paper. In the context of the present paper, the most important limitation of meta-analyses is that the final combined evidence is extremely dependent on the results to be pooled. In the second stage the authors pooled the results (city-specific pollution relative rates) obtained in the log-linear generalized additive model of the first stage. The problem is that I am not very sure that all these generalized additive models would be correctly specified and, therefore, it is not clear that estimates would be the true (log-) relative rates that the authors try to infer.

Among others, I mention the following probable misspecifications.

(a) It seems that the authors do not consider the possibility of interaction between weather variables or between weather variables and air pollutants, particularly ozone.
(b) The authors say nothing about the lags, if any, considered for the ozone measurements, when they adjust for this pollutant.
(c) I am not very comfortable with the, non-statistical, choice of the degree of smoothness 'pre-specified, on the basis of prior epidemiological knowledge'.

(d) It is probable that residual autocorrelation would not remain, because the authors controlled for both observed and unobserved confounders. However, they do not show a correct diagnostic to demonstrate the absence of such autocorrelation. Because they allowed for overdispersion, standardized residuals should be computed as $r_t = (Y_t - \hat{Y}_t)/\sqrt{(\hat{\phi}\hat{Y}_t)}$ rather than $r_t = (Y_t - \hat{Y}_t)/\sqrt{\hat{Y}_t}$ as shown in Fig. 4.

Summing up, the final results and, consequently, the evidence that could be drawn depend on the correct specification of all the models in the first stage. Unfortunately, I am not sure that this would be the case for all the models.

Lianne Sheppard and Thomas Lumley (*National Research Center for Statistics and the Environment, and University of Washington, Seattle*)
We commend the authors for their important contribution to epidemiologic methods for short-term air pollution health effects. Their strategy is appropriately conservative. They divide the information in the data into components that provide reliable information and those that provide potentially unreliable information and simply discard the latter. These choices reduce the precision of the air pollution health effect estimates but, more importantly, remove potential biases. More extensive modelling might extract more information and give tighter confidence intervals but at the risk of bias.

By performing the same analysis on each city and combining the results in a prespecified way, the authors avoid the publication and model selection biases that probably pervade the applied statistics literature. However, the limited flexibility in models between cities may mean that the best models for relating PM_{10} and mortality are missed.

It is important to emphasize that this analysis examines only within-city air pollution health effects. The only point where between-city pollution differences enter the model is in the inclusion of mean pollution in the second stage. We believe that this only assesses linearity in the within-city pollutant effect. The investigators have chosen not to use the substantial between-city information about health effects. Although the risk of introducing bias cannot be ignored, we feel that it would be valuable to explore alternatives.

We are concerned about the potential effect of missing exposure data. Unlike cities with daily monitoring, cities with 6-day monitoring frequencies use completely different data sets for different lags. Likewise comparisons between pollutants use different time periods in different cities when some are measured seasonally (e.g. ozone in Seattle). Both missingness patterns will at least introduce heterogeneity into the relative risk estimates. Then the second-stage analysis will combine genuine variation between cities with design variation due to different monitoring practices. Further, when we average unadjusted measurements across monitors and there is both heterogeneity in the monitor-specific pollutant distributions and missing data, the monitor averages will have an additional component of variability beyond random measurement error. This could alter the health effect estimates (e.g. Clyde and DeSimone-Sasinovska (1999)). Finally, the structure of the missing information should affect the choice of smoothing parameters. When a pollutant is measured throughout the year, but only on one day in six, all 7 degrees of freedom for that year should still be used. However, when a pollutant is measured daily for a single season, a *pro rata* reduction in the degrees of freedom is needed to achieve the same effective smoothing bandwidth.

The **authors** replied later, in writing, as follows.

We are grateful for the many and diverse contributions to the discussion. Our reply is offered in five sections, covering the principal themes of the discussants: causality and public policy implications of the results, heterogeneity, mortality displacement and measurement error in the exposure variables, adjustment for trend and seasonality factors, and finally some further comments on the modelling assumptions and estimation methods.

Causality and public policy implications of the health effects of air pollution
Mr Clayton pin-points the two key time frames of the relationships between air pollution and mortality, short term and long term, and the possibly different public health and hence policy implications of effects on these distinct time frames. Although our analysis addresses only the short-term associations, positive findings from daily time series analyses have been one of the principal reasons for proceeding with new air quality standards for particulate matter, although the degree of life shortening underlying the effect is still highly uncertain.

Mr Clayton and Professor Raab also note that, although scientific interest is on the causal relationship of air pollution with mortality at the individual level, most explanatory variables are aggregate summaries at a higher level. Any interpretation of the time series analyses needs to be placed in the broader context of our understanding of the toxicity of particulate matter, including other observational data and also experimental evidence. Although we can debate whether the models are properly interpreted as showing or not showing 'causal effects', regulators must combine results from studies of association like ours with expert judgment to make policy decisions. Our paper shows a consistent association of particulate air pollution with daily mortality across the 20 largest US cities, that the overall effect is positive with high probability even when we allow for substantial heterogeneity across cities and that this association is very unlikely to be explained by longer-term confounders, or seasonality or by confounding effects of temperature or humidity. Our results thus provide a useful summary for policy makers.

We concur with Dr Rushton's view that an interpretation of findings remains daunting in this area, even as we develop increasingly sophisticated statistical methodology. A better understanding of heterogeneity, particularly if based in links to specific particle characteristics, could inform mitigation and risk reduction. The US National Research Council and others have placed understanding of toxicity determining characteristics as a high priority research topic (National Research Council, 1998).

Heterogeneity

One key issue for public health implication is whether there is one or multiple dose–response relationships between the cities. We were enlightened by the degree of attention in the discussion to the issue of heterogeneity. In our analysis, we imposed the prior assumption that there is heterogeneity. In fact, our prior specification for the variance of the random effects $\sigma^2 \sim \mathrm{IG}(3, 6)$ is quite informative because it gives little prior weight to the possibility of near homogeneity in the pollution effects. We made this assumption to be sufficiently vague about city-specific estimates, to explore covariates that might explain the variation across cities, and to be conservative in our confidence regions for the overall effects, the widths of which are sensitive to the prior specification for the random-effects variance.

Our original report did not adequately emphasize the degree to which the posterior estimates for the city-specific log-relative-risks and for the variance of the city-specific estimates are sensitive to the prior. With only 20 imperfectly estimated values of β^c, the data provide only limited evidence about σ^2.

To illustrate this important point, we have refitted the basic model with a prior distribution for σ^2 that has the same median as we used originally but which puts much more weight near zero (half-normal for σ^2 with standard deviation 1). The overall estimate is unchanged, but its 95% posterior region changes from (0.05–0.92) to (0.29–0.64) which is more consistent with the results of the 'Air pollution and health: a European approach' study where a fixed effect model was used (Katsoyanni *et al.*, 1997). Under the new prior, no cities have negative posterior estimates as opposed to one of 20 (Atlanta) which had a negative posterior estimate originally.

How much heterogeneity is there in the true relative risks between cities? The profile likelihood for σ^2 is maximized near 0. An interval for σ in which the profile likelihood is within 1/32 of its maximum is roughly (0, 0.4). Hence the data suggest less heterogeneity than was assumed *a priori* in the analyses reported here. We thank Mr Clayton for raising the questions that led us to clarify this important point. Some heterogeneity of effect across US cities is plausible, given variation in pollution sources, and other city-specific characteristics, and we agree that a further exploration of heterogeneity is needed.

Mortality displacement and measurement error in the exposure variable

Professor Campbell, Professor Bland and Ms Macfarlane proposed that the interpretation of the findings of the daily time series studies for public health purposes has been constrained by the difficulty in estimating the actual degree of life shortening. The unfortunate term 'harvesting' has been given to the possibility that the associations in the daily data only reflect brief life shortening of already frail individuals. We and others have developed methods to assess the extent to which the harvesting hypothesis explains the evidence to date (Zeger *et al.*, 1999; Schwartz, 2000). The results, along with those from several key prospective cohort studies (Dockery *et al.*, 1993), suggest that the associations do not reflect harvesting alone. We would like to be able to answer Professor Bland's personal question also, although we have no particular data for the A23 Brighton Road. The daily time series data have been extensively analysed as they are readily available and hence abundant; not many studies that are now under way are likely to be informative about long-term effects. Although we concur with Professor Bland's emphasis on the need for long-term studies, the short-term data have motivated much of the current public health concern with regard to particulate air pollution. Our approach does offer a

summary measure and provides a picture of variation across the country. This represents an advance over the haphazard assortment of cities now reported in the literature, as investigators have taken their choice of locations for uncertain reasons. To address Professor Bland's important question more directly, the unit of the analysis must be the individual rather than the city, and key person level confounders, such as smoking, will need to be assessed.

Professor Campbell reminds us that particles are only one component of the complex contaminant mixture found in urban air. Much work is focused on disentangling the effects of separate pollutants, even though they may be acting in concert to produce adverse effects. The form of our air pollution regulations provides one, perhaps inappropriate, justification as individual pollutants are regulated, rather than some overall measure of mixture toxicity. Dr Longford alludes to the complexities of investigating air pollution in the face of control measures that range from societal to personal levels. We cannot agree with his argument that the population under investigation is highly selective. Our sample size is large and the cities included are diverse in their characteristics.

Professor Smith, Dr Sheppard, Dr Lumley and Dr Quinn comment on the potential limitation introduced by using data from only a small number of monitors. We are also concerned with the potential problem of measurement errors and have explored this issue, as reported elsewhere (Zeger *et al.*, 2000; Dominici *et al.*, 2000). For fine particles there tends to be a surprising homogeneity of concentrations, reflecting the ubiquitous nature of one of the key sources, motor vehicles, as well as the regionalization of air pollution in the USA. Road traffic largely generates particles in the coarse mode. This is certainly an important topic for further research.

A concern raised by Dr Gurrin, Professor Pettitt, Dr Gupta, Dr Sheppard and Dr Lumley is the potential effect of missing exposure data on the lagging process. In most cities PM_{10}-measurements are available only every 6 days. Hence we have considered models with a single PM_{10}-predictor at a range of lags by shifting the daily mortality record with respect to the partially observed PM_{10}-record. Two possible effects of this strategy are

(a) to introduce additional heterogeneity across cities due to different lag structures and
(b) to underestimate (or perhaps to overestimate if harvesting is more important than current research indicates) the total effect of PM_{10} relative to when all lags are taken into account.

It is also possible to regress mortality on PM_{10} lagged by 0, 6, 12, . . . days using a parametric model for the lag and to estimate the cumulative effect over all lags, an approach which we are now pursuing.

Adjustment for confounding factors
Dr Armstrong and Dr Saez asked whether smooth functions of time can properly take into account the confounding effect of long-term trend and seasonality. Initially we also included harmonic terms to adjust for seasonality and used a smaller number of degrees of freedom in the smooth function of time (Kelsall *et al.*, 1997). Because the seasonal behaviour of the mortality time series does not have a constant phase and amplitude, the inclusion of harmonic functions of time in the model might not provide enough flexibility. In our current model formulation, the overall effect of PM_{10} is relatively robust to different specifications of the number of degrees of freedom in the smooth function of time, temperature and dewpoint. Fig. 9 shows the posterior mean and 95% posterior intervals for eight other models in which the degrees of freedom for the smooth function of time, temperature and dewpoint are halved, kept constant and doubled. There is little or no change in the findings over this substantial variation of nuisance parameter values.

Dr Berhane and Professor Thomas asked whether it may be useful to consider an automated procedure to select the smoothing parameter value that controls the autocorrelation. This is an idea which is worth pursuing. However, if we adjust for the long-term trend and seasonality differently across cities, then the city-specific relative rates will represent the association of PM_{10} and mortality at different sets of timescales. In addition, an automated criterion might select degrees of freedom that are discordant with our prior scientific understanding of the timescales (e.g. seasons) on which confounding is most likely to occur.

Issues on modelling assumptions and estimation methods
We agree with Mr Clayton, Professor Cowles and Professor Pettitt that the likelihood is reasonably manageable, and in some cases the implementation of Markov chain Monte Carlo sampling would not be needed. However, the Gibbs sampler is necessary for evaluating the posterior distributions under the spatial model. In our approach, the base-line model is the starting-point for additional variations and extensions, and we thus can benefit from using a common computational approach such as Markov
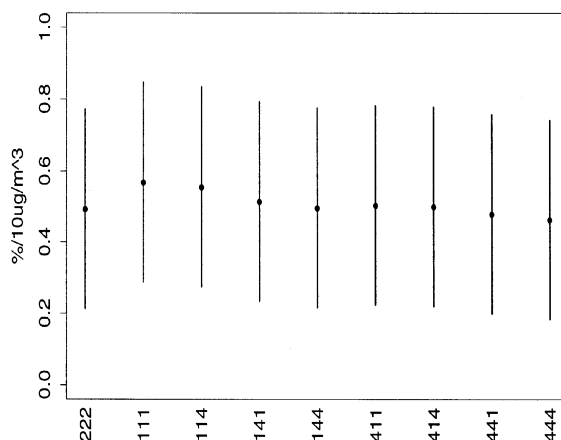
Fig. 9. Point estimates and 95% confidence intervals of the overall effect of PM_{10} at lag 0 for nine models, in which the degrees of freedom for the smooth functions of time, temperature and dewpoint are halved (denoted by 1), kept constant (denoted by 2) or doubled (denoted by 4) (the confidence intervals are calculated on the basis of a random-effects model in which the variance of the effects across cities, σ^2, is set to its posterior mean ($\hat{\sigma}^2 = 0.76$))

chain Monte Carlo sampling. Professor Campbell and Professor Raab raised the point that a potential limitation of our findings is their dependence on the assumption that the relationship between air pollution and log-mortality is linear. We have further investigated this issue by developing a threshold and a spline dose–response model to estimate threshold PM_{10}-levels and PM_{10}-mortality dose–response curves and applied these models to the same database. For total mortality, our findings confirm that the linear model is well supported by the data (Daniels *et al.*, 2000). Dr Richardson and Professor Pettitt mentioned the appropriateness of the Gaussian distribution to describe the heterogeneity of the city-specific estimates across cities. Recall that we also put a prior on the variance of this distribution, allowing for wider tails than would result from a Gaussian distribution with fixed variance. More flexible second-stage distributions, such as a mixture of Gaussian distributions, might result in differential shrinkage of outlying city-specific relative rates towards the overall mean. We are planning to relax this assumption in our analysis of data from the largest 90 US cities. Mr Clayton and Professor Smith, Professor Dryden and Dr L. Cox question the use of a spatial correlation based on the distance between cities. We agree that our assumption on the spatial correlation of the estimates is simple and somewhat arbitrary. We consider it to be a sensible first deviation from the naïve assumption of spatial independence. Planned analyses of the 90-cities data will further investigate the amount and possible sources of heterogeneity across locations by including city-specific characteristics as effect modifiers, and by developing regional models.

Dr Coull, Dr Mezzetti and Professor Ryan question how a two-stage approach allows for the varying quality of studies. We generally agree that measures of the quality of a study can be important variables in meta-analysis. However, in our case the same data sources and statistical methodology were used for all 20 cities. Hence the statistical variances probably capture main differences between cities in the information that is available. But there may be a substantial between-city variability in the quality of the exposure measurements. The two-stage model that we are developing for the 90 cities allows for varying quality of exposure data by including, as an effect modifier, a measure of the quality of the exposure data for that location.

Again, we thank the discussants for their contributions to this important public health issue.

References in the discussion

Altman, N. S. (1990) Kernel smoothing of data with correlated errors. *J. Am. Statist. Ass.*, **85**, 749–759.
Campbell, M. J. (1994) Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *J. R. Statist. Soc.* A, **157**, 191–208.

————(1997) Time series regression. In *Encyclopaedia of Biostatistics* (eds P. Armitage and T. Colton), pp. 4536–4538. Chichester: Wiley.

Campbell, M. J. and Tobías, A. (2000) Causality and temporality in the study of short term effects of air pollution on health. *Int. J. Epidem.*, **29**, 271–273.

Chib, S. (1995) Marginal likelihood from the Gibbs output. *J. Am. Statist. Ass.*, **90**, 1313–1321.

Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Clayton, D. (1991) The EURODEM collaborative re-analysis of case-control studies of Alzeimer's disease: some methodological considerations. *Int. J. Epidem.*, **20**, S62.

Clyde, M. (1998) Bayesian model averaging and model search strategies. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 83–102. Oxford: Oxford University Press.

Clyde, M. and DeSimone-Sasinovska, H. (1999) Does particulate matter particularly matter?: accounting for model uncertainty in Poisson regression models. *Discussion Paper*. Duke University, Durham.

Committee on the Medical Effects of Air Pollutants (1995) *Non-biological Particles and Health*. London: Her Majesty's Stationery Office.

Committee on the Medical Effects of Air Pollutants Quantification of Air Pollution Risk Sub-Group (1997) *Quantification of the Effects of Air Pollution on Health in the United Kingdom*. London: Stationery Office.

Cotton, J. W. and Othman, A. R. (1991) Modeling perception of temperature change using a generalized additive model. *Psychonomic Society Meet, San Francisco, Nov. 23rd.*

Cressie, N. (1994) *Statistics for Spatial Data*. New York: Wiley.

Daniels, M., Dominici, F., Samet, J. and Zeger, S. L. (2000) Estimating PM_{10}-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. *Am. J. Epidem.*, to be published.

Daniels, M. and Kass, R. (1998) A note on first-stage approximation in two-stage hierarchical models. *Sankhya* B, **60**, 19–30.

Diggle, P. J. and Hutchinson, M. F. (1989) On spline smoothing with autocorrelated errors. *Aust. J. Statist.*, **31**, 166–182.

Dockery, D., Pope, C. I. and Xu, X. (1993) An association between air pollution and mortality in six U.S. cities. *New Engl. J. Med.*, **329**, 1753–1759.

Dominici, F., Zeger, S. and Samet, J. (2000) A measurement error correction model for time-series studies of air pollution and mortality. *Biostatistics*, **2**, in the press.

Gamble, J. F. and Lewis, R. J. (1998) PM_{25} and mortality in long-term prospective cohort studies: cause-effect or statistical associations? *Environ. Hlth Perspect.*, **106**, 535–549.

Greenland, S. and Robins, J. (1994) Ecological studies—biases, misconceptions and counterexamples. *Am. J. Epidem.*, **139**, 747–760.

Harville, D. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Ass.*, **72**, 320–340.

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

Hay, J. L. and Pettitt, A. N. (2000) Bayesian analysis of a time series of counts with covariates: an application to the control of an infectious disease. *Biostatistics*, to be published.

Hedges, L. V. and Olkin, I. O. (1985) *Statistical Methods for Meta Analysis*. Orlando: Academic Press.

Holland, W. W., Bennett, A. E., Cameron, I. R., Florey, C. du V., Leeder, S. R., Schilling, R. S. F., Swan, A. V. and Waller, R. E. (1979) Health effects of particulate pollution: reappraising the evidence. *Am. J. Epidem.*, **110**, 527–659.

Hurvich, M. H. and Zeger, S. L. (1990) A frequency domain selection criterion for regression with autocorrelated errors. *J. Am. Statist. Ass.*, **85**, 705–714.

Katsoyanni, K., Toulomi, G., Spix, C., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., Schouten, J., Ponka, A. and Anderson, H. R. (1997) Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *Br. Med. J.*, **314**, 1658–1663.

Kelsall, J., Samet, J. and Zeger, S. (1997) Air pollution, and mortality in Philadelphia, 1974-1988. *Am. J. Epidem.*, **146**, 750–762.

Laird, N. M. (1990) Analysis of linear and non-linear growth models with random parameters. In *Advances in Statistical Methods for Genetic Improvement of Livestock* (eds D. Gianola and K. Hammond), pp. 329–343. Berlin: Springer.

Lawther, P. J., Waller, R. E. and Henderson, M. (1970) Air pollution and exacerbations of chronic bronchitis. *Thorax*, **25**, 525–539.

Longford, N. T. (2000) Multivariate synthetic estimation. To be published.

Macfarlane, A. J. (1977) Daily mortality and climate in English conurbations: 1, Air pollution, low temperature, and influenza in Greater London. *Br. J. Prev. Socl Med.*, **31**, 54–61.

————(1984) A time to die? *Int. J. Epidem.*, **1**, 38–44.

Martin, A. E. and Bradley, W. H. (1960) Mortality, fog and air pollution—an investigation during the winter of 1958-59. *Mthly Bull. Min. Hlth Publ Hlth Lab. Serv.*, **19**, 56–72.

Mazumdar, S., Schimmel, H. and Higgins, I. T. T. (1982) Relation of daily mortality to air pollution: an analysis of 14 London winters, 1958/59-1971/72. *Arch. Environ. Hlth*, **37**, 213–220.

Ministry of Health (1954) Mortality and morbidity during the London fog of December 1952. *Reports on Public Health and Medical Subjects 95*. London: Her Majesty's Stationery Office.

National Research Council (1998) *Research Priorities for Airborne Particulate Matter: I, Immediate Priorities and a Long-range Research Portfolio*. Washington DC: National Academy Press.

Ostro, B. (1984) A search for a threshold in the relationship of air pollution to mortality: a reanalysis of data on London winters. *Environ. Hlth Perspect.*, **58**, 397–399.

Refinetti, R. (1989) Magnitude estimation of warmth: intra- and intersubject variability. *Percept. Psychophys.*, **46**, 81–84.

Ribaudo, H. J., Bacchi, M., Bernhard, J. and Thompson, S. G. (1999) A multilevel analysis of longitudinal ordinal data: evaluation of the level of physical performance of women receiving adjuvant therapy for breast cancer. *J. R. Statist. Soc.* A, **162**, 349–360.

Rogot, E., Fabsitz, R. and Feinleib, M. (1976) Daily variation in USA mortality. *Am. J. Epidem.*, **103**, 198–211.

Rosenbaum, P. R. (1995) *Observational Studies*. New York: Springer.

Rushton, L. (1999) Overview and critique of the air pollution and health: a european approach (APHEA) project. *Report 99/54*. CONCAWE, Brussels.

Russell, W. T. (1924) The influence of fog on mortality from respiratory diseases. *Lancet*, ii, 335–339.

———(1926) The relative influence of fog and low temperature on the mortality from respiratory disease. *Lancet*, ii, 1128–1130.

Schwartz, J. (2000) Harvesting and long term exposure in the relation between air pollution and mortality. *Am. J. Epidem.*, **151**, 440–448.

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

Smith, M. D. (1996) The design and analysis of crossover trials with many periods. *PhD Thesis*. University of Kent at Canterbury, Canterbury.

Spiegelhalter, D. J. and Marshall, C. E. (1999) Inference-robust institutional comparisons: a case study of school examination results. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford: Oxford University Press.

Spiegelhalter, D., Thomas, A. and Best, N. (1999) *WinBUGS User Manual*. Cambridge: Medical Research Council Biostatistics Unit.

Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. (1995) *BUGS: Bayesian Inference using Gibbs Sampling, Version 0.50*. Cambridge: Medical Research Council Biostatistics Unit.

Thurston, G. D., Ito, K., Lippmann, M. and Hayes, C. (1989) Reexamination of London, England, mortality in relation to exposure to acidic aerosols during 1963-1972 winters. *Environ. Hlth Perspect.*, **79**, 73–82.

Toulomi, G., Katsoyanni, K., Zmirou, D. and Schwartz, J. (1997) Short-term effects of ambient oxidant exposure on mortality: a combined analysis within the APHEA project. *Am. J. Epidem.*, **146**, 177–183.

Zeger, S., Dominici, F. and Samet, J. (1999) Harvesting-resistant estimates of pollution effects on mortality. *Epidemiology*, **89**, 171–175.

Zeger, S., Thomas, D., Dominici, F., Cohen, A., Schwartz, J., Dockery, D. and Samet, J. (2000) Measurement error in time-series studies of air pollution: concepts and consequences. *Environ. Hlth Perspect.*, **108**, in the press.