10-28-2003

Hierarchical Bivariate Time Series Models: A Combined Analysis of the Effects of Particulate Matter on Morbidity and Mortality

Francesca Dominici
The Johns Hopkins Bloomberg School of Public Health, fdominic@jhsph.edu

Antonella Zanobetti
Harvard School of Public Health, azanobet@hsph.harvard.edu

Scott L. Zeger
The Johns Hopkins Bloomberg School of Public Health, szeger@jhsph.edu

Joel Schwartz
Harvard School of Public Health, jschwrtz@hsph.harvard.edu

Jonathan M. Samet
The Johns Hopkins Bloomberg School of Public Health, cgerczak@jhsph.edu

Hierarchical Bivariate Time Series Models: A Combined Analysis of the Effects of Particulate Matter on Morbidity and Mortality

Dominici F., Zanobetti A., Zeger S.L., Schwartz J., and Samet J.M.

Abstract

In this paper we develop a hierarchical bivariate time series model to characterize the relationship between particulate matter less than 10 microns in aerodynamic diameter (PM_{10}) and both mortality and hospital admissions for cardiovascular diseases. The model is applied to time series data on mortality and morbidity for 10 metropolitan areas in the United States from 1986 to 1993. We postulate that these time series should be related through a shared relationship with PM_{10}.

At the first stage of the hierarchy, we fit two seemingly unrelated Poisson regression models to produce city-specific estimates of the log relative rates of mortality and morbidity associated with exposure to PM_{10} within each location. The sample covariance matrix of the estimated log relative rates is obtained using a novel generalized estimating equation approach that takes into account the correlation between the mortality and morbidity time series. At the second stage, we combine information across locations to estimate overall log relative rates of mortality and morbidity and variation of the rates across cities.

Using the combined information across the 10 locations we find that a 10 $\mu g/m^3$ increase in average PM_{10} at the current day and previous day is associated with a 0.26% increase in mortality (95% posterior interval $-0.37, 0.65$), and a 0.71% increase in hospital admissions (95% posterior interval $0.35, 0.99$). The log relative rates of mortality and morbidity have a similar degree of heterogeneity across cities: the posterior means of the between-city standard deviations of the mortality and morbidity air pollution effects are 0.42 (95% interval $0.05, 1.18$), and 0.31 (95% interval $0.10, 0.89$), respectively. The city-specific log relative rates of mortality and morbidity are estimated to have very low correlation, but the uncertainty in the correlation is very substantial (posterior mean $= 0.20$, 95% interval $-0.89, 0.98$).

With the parameter estimates from the model, we can predict the hospitalization log relative rate for a new city for which hospitalization data are unavailable, using that city's estimated mortality relative rate. We illustrate this prediction using New York as an example.

Key Words: Generalized Estimating Equations, Generalized additive models, Hierarchical models, Particulate matter, log relative rate, Air pollution.

Francesca Dominici, Scott L. Zeger, Department of Biostatistics, Jonathan M. Samet, Department of Epidemiology, all at the Johns Hopkins Bloomberg School of Public Health. Antonella Zanobetti, Department of Environmental Health, Joel Schwartz, Department of Environmental Health, all at Harvard School of Public Health. Correspondence may be addressed to Dr. Francesca

Dominici, Department of Biostatistics, Bloomberg School of Public Health, 615 N. Wolfe Street, The Johns Hopkins University, Baltimore, MD 21205-3179, USA. ph: 410-614-5107, fax: 410-955-0958, e-mail: fdominic@jhsph.edu.

Acknowledgments: Research described in this article was partially supported by a contract and grant from Health Effects Institute (HEI), an organization jointly funded by the Environmental Protection Agency (EPA R824835) and automotive manufacturers. The contents of this article do not necessarily reflect the views and policies of HEI, nor do they necessarily reflect the views and policies of EPA, or motor vehicles or engine manufacturers. Funding for Francesca Dominici was provided by a grant from the Health Effects Institute (Walter A. Rosenblith New Investigator Award) and by NIEHS RO1 grant (ES012054-01). Funding was also provided by the NIEHS Center in Urban Environmental Health (P30 ES 03819)

1 Introduction

The potential for air pollution at high concentrations to cause excess deaths and morbidity was firmly established in the mid-twentieth century by a series of well-documented air pollution "disasters" in the US and Europe. By the early 1990's, time series studies with data from single locations (Dockery et al., 1993; Schwartz, 1994; American Thoracic Society, 1996a; Pope, 2000), showed that air pollution, even at much lower concentrations than existed during the earlier disasters, was associated with increased rates of mortality and morbidity in cities in the United States, Europe, and other developed countries (Pope and Dockery, 1999). One key limitation of these studies was the use of data from a single, or at most a few locations of uncertain representativeness of broader geographic regions. The *National Morbidity, Mortality, and Air Pollution Study* (NMMAPS) addressed this limitation by assembling and analyzing a national data base that includes information on mortality, morbidity, weather and air pollution for numerous metropolitan areas in the US.

The NMMAPS mortality analyses estimated associations between all-cause and cause-specific mortality and particulate matter less than 10 microns in aerodynamic diameter (PM_{10}) for 90 cities in the U.S. (Samet et al., 2000b; Dominici et al., 2002). The NMMAPS morbidity analyses estimated associations between hospitalization in the elderly and PM_{10} for 14 cities in the U.S. (Samet et al., 2000b; Schwartz, 2000; Zanobetti et al., 2000a). Methodological approaches and substantive results of the separate mortality and morbidity analyses have been reported (Samet et al., 2000a; Daniels et al., 2000; Zeger et al., 1999; Schwartz and Zanobetti, 2000; Schwartz, 2000; Zanobetti et al., 2000b). The analyses showed that PM_{10} concentrations were positively associated with mortality and morbidity outcomes on average across locations (Samet et al., 2000b; Dominici et al., 2003a).

Poisson time series regression models (Liang and Zeger, 1986; Zeger and Liang, 1992; Fahrmeir and Tutz, 2001; McNeney and Petkau, 1994; Albert, 1999) or generalized additive models (Hastie and Tibshirani, 1990) have been widely used to analyze univariate time series data of air pollution and health in selected locations (Dockery and Pope, 1994; Schwartz, 1995; American Thoracic Society, 1996a,b; Korrick et al., 1998). Critics of single-site studies questioned the choice of particular cities and asked if models had been selected that gave estimates of effect that were biased upwards (Lipfert and Wyzga, 1993; Li and Roth, 1995). These criticisms have been addressed by using multi-site studies (Katsouyanni et al., 1997; Samet et al., 2000a; Hwang and Chan, 2001) in which site-specific data on air pollution and health are assembled under a common framework.

Hierarchical models (DuMouchel and Harris, 1983; DuMouchel, 1990; Breslow and Clayton, 1993; Carlin and Louis, 1996) are a suitable approach for analyzing univariate time series data from multiple locations (Dominici et al., 2000, 2002; Hwang and Chan, 2001). In comparison to analyses of data from a single site, pooled analyses can be more informative about whether an association exists, controlling for possible counfounders. In addition, pooled analyses can produce estimates of the parameters at a specific site, which borrow strength from all other locations.

Analyses with hierarchical univariate time series models to estimate associations between air pollution and health have focused mainly on the estimation of the overall log relative rate of mortality (or morbidity) associated with PM_{10}, and their heterogeneity across locations. The analyses have not explored whether cities where PM_{10} had greater or lesser effects on morbidity also tended to have a similar pattern of PM_{10} effects on mortality. A correlation between levels of effects for morbidity and mortality in a particular city would be anticipated if characteristics of inhaled particles of the populations influenced risks for morbidity and mortality in a similar fashion.

Our modeling approach extends previous work in two directions. First, within each city we extend Poisson regression approaches for univariate time series to bivariate time series and we estimate the log relative rates of mortality and morbidity by taking into account the correlation between the mortality and morbidity time series. Second, we extend this bivariate time series model in a hierarchical fashion, by combining relative rates of mortality and morbidity across locations in order to characterize their relationship.

More specifically, at the first stage, we fit two *seemingly unrelated* Poisson regression models to estimate the relative rates of mortality and hospitalization associated with exposure to PM_{10} $(\hat{\beta}_M^c,\ \hat{\beta}_H^c)$ within each location c. We define these two Poisson regression models as *seemingly unrelated* for the following two reasons. First, we estimate $\hat{\beta}_M^c$ and $\hat{\beta}_H^c$ under the working assumption that the daily mortality and hospitalization time series are independent. Secondly, to take into account the joint correlation function for the bivariate mortality and morbidity time series, we estimate the sample covariance between $\hat{\beta}_M^c$ and $\hat{\beta}_H^c$ by using a novel generalized estimating equations approach (Zeger et al., 1988).

At the second stage, we assume that the vector of true log relative rates of mortality and morbidity $(\beta_M^c,\ \beta_H^c)$ has a bivariate normal distribution with unknown means $(\alpha_M,\ \alpha_H)$ and unknown covariance matrix (Σ) which we then estimate using a Markov Chain Monte Carlo (MCMC) algorithm. Although the sample covariance between $\hat{\beta}_M^c$ and $\hat{\beta}_H^c$, and the correlation between the mortality and morbidity time series within each location are of interest, here we focus on the parameters at the second stage of the hierarchical model.

The hierarchical bivariate time series model discussed here can be used to facilitate prediction of the log relative rates of mortality and morbidity for cities other than the 10 included in the joint analysis. For example, consider New York for which we have mortality data, but we do not have morbidity data. We can approximate the posterior predictive distribution of the log relative rate of hospital admissions for New York (β_H^{NY}) conditional on the mortality data for New York, and the mortality and morbidity data for the other 10 cities. In addition, we estimate the reductions in the posterior variances of β_M^c and β_H^c obtained by using the time series data at location c relative to ignoring this information. We report these reductions in posterior variances for all 10 locations and New York.

A description of the database of air pollution, mortality, morbidity, and meteorological data for

the 10 U.S. cities in this analysis is included in Section 2. These 10 cities were selected from the larger group of cities in the NMMAPS data, because they have daily PM_{10} data as well as both mortality and hospitalization data. In section 3, we describe the two-stage model for combining the log relative rates of mortality and hospital admissions across locations. The generalized estimating equation approach for estimating the sample covariance matrix of the estimates is explained in the Appendix. Results and discussion follow in Sections 4 and 5.

2 Data

The database used for this analysis includes mortality, hospital admissions for cardiovascular disease, 24-hour average temperature, barometric pressure, relative humidity and 24-hour average PM_{10} concentrations for 10 metropolitan areas in the United States (See Table 1). The general observation period is 1986-1993, but varies across locations. The air pollution data were obtained from the the Aerometric Information System (AIRS) data base maintained by the US Environmental Protection Agency. The daily time series of PM_{10} used for these analyses are the same as those used for the morbidity analysis of the NMMAPS. Daily total mortality data, aggregated at the level of the county, were obtained from the National Center for Health Statistics (NCHS). Daily counts of hospital admissions were extracted from the files of the Health Care Financing Administration (HCFA). The hourly temperature, barometric pressure and relative humidity for each site were obtained from the Earth Info CD-ROM database (`www.sni.net/earthinfo`). A detailed description of the data base is given elsewhere (Samet et al., 2000b,c). We have focused on cardiovascular events because prior research has suggested these are the most strongly associated with variations in air quality (Dockery et al., 1993; Samet et al., 2000c). These 10 metropolitan areas were chosen from the 14 locations of the morbidity analyses with daily time series of mortality available for the same time period of the daily time series of morbidity. Table 1 summarizes for each city: the start and end dates of the PM_{10} monitoring; number of days with PM_{10} measurements; the 24-h average PM_{10} concentrations; mean numbers of hospital admissions; and mean numbers of deaths from cardiovascular diseases.

3 Methods

This section describes a two-stage hierarchical model for combining log relative rates of mortality and hospital admissions across locations, taking into account the joint correlation function for the bivariate mortality and morbidity time series when estimating the covariance of the two log relative rates. The goals of our analysis are to estimate: 1) overall log relative rates of mortality and hospital admissions from exposure to PM_{10}; 2) heterogeneity of the log relative rates across cities; 3) correlation between the two log relative rates across cities; and 4) log relative rate of hospital admissions for a city other than the 10 sampled using that city's mortality time series data. The

two stages of the hierarchical model are described below.

Within each city, two seemingly unrelated log-linear regressions are fitted to the mortality (M) and hospital admissions (H) data bases. We assume

$$\begin{aligned} E(y_t^M) &= \mu_t^M,\ var(y_t^M) = w_t^M = \phi^M \mu_t^M \\ \log \mu_t^M &= X_t^{M'} \boldsymbol{\theta}_M \\ E(y_t^H) &= \mu_t^H,\ var(y_t^H) = w_t^H = \phi^H \mu_t^H \\ \log \mu_t^H &= X_t^{H'} \boldsymbol{\theta}_H \end{aligned} \tag{1}$$

where: y_t^M and y_t^H are the mortality and hospital admissions daily time series; X_t^M and X_t^H are the design matrices including average lag 0 and lag 1 PM_{10} daily time series and potential confounding factors for the mortality and hospitalization data such as long-term trends, seasonality and weather (Samet et al., 1995, 1997; Kelsall et al., 1997; Dominici et al., 2000; Samet et al., 2000a). Note that X_t^M might be the same as X_t^H in some situations.

In this application, we specify model (1) as over-dispersed Poisson with a linear term for the average PM_{10} on day 0 and 1, and smooth functions (natural cubic splines) of calendar time, temperature and barometric pressure to adjust for time-varying confounding factors such as trend, seasonality and weather. The model specification, including confounding factors and the rationale for their inclusion in the model, is listed in Table 2.

Thus, the full vector of regression coefficients is denoted by $\boldsymbol{\theta}_M^c$ (or $\boldsymbol{\theta}_H^c$) in model (1), can be decomposed as $[\beta_M^c, \boldsymbol{\eta}_M^c]$ (or $[\beta_H^c, \boldsymbol{\eta}_H^c]$) where β_M^c (β_H^c) is the log relative rate of mortality (morbidity) for increases in PM_{10} and $\boldsymbol{\eta}_M^c$ (or $\boldsymbol{\eta}_H^c$) is the vector of nuisance parameters corresponding to the confounding factors listed in Table 2. Finally, the parameters ϕ^M and ϕ^H are overdispersion parameters.

Modeling strategies to reduce confounding bias in the air pollution effect estimates are among the most discussed statistical issues in time series analyses of air pollution and health. In particular the choice of the degrees of freedom (*df*) in the smooth functions of time and temperature is critical because it determines the residual temporal variability in the daily deaths and pollution levels used to estimate the pollution coefficient. As a baseline choice, we use 4 degrees of freedom per year to adjust for trend and seasonality, and 3 *df* to adjust for temperature and barometric pressure. These choices are made on the basis of our previously published results and on recent re-analyses and sensitivity analyses (Schwartz et al., 2003; Dominici et al., 2003a). In the results section we explore the sensitivity of the overall log relative rates to the *df* in the natural cubic splines of time, temperature and barometric pressure.

We estimate the log relative rate parameters $(\hat{\beta}_M^c, \hat{\beta}_H^c)$ for city c under the working assumption that the daily mortality and hospitalization series are independent. Hence, two separate log-linear regressions are estimated by maximum likelihood. This approach is sensible because our focus is on the association between the log relative rates, rather than the association between the daily counts. However, the correlation among the two series of counts will introduce correlation in the estimated

log relative rates $\hat{\beta}_M^c$ and $\hat{\beta}_H^c$ for a given city. We estimate the sample covariance matrix V^c, the within-city correlation $v_{MH}^c/\sqrt{v_M^c v_H^c}$ along with the overdispersion parameters using generalized estimating equations (GEE). To do so, we apply formula (7) detailed in the appendix for lag $L = 14$. The choice of $L = 14$ is based on the assumption that the mortality and morbidity time series are likely to be uncorrelated at lag 14. The estimated V^c were not sensitive to lag choices larger than 14.

Because the time series are relatively long (number of days with PM_{10} data available ≥ 1450), the estimates of the mortality and morbidity log relative rates are approximately bivariate normal:

$$\begin{bmatrix} \hat{\beta}_M^c \\ \hat{\beta}_H^c \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \beta_M^c \\ \beta_H^c \end{bmatrix}, V^c \right) \text{ where } V^c = \begin{bmatrix} v_M^c & v_{MH}^c \\ v_{MH}^c & v_H^c \end{bmatrix} \tag{2}$$

Dominici et al. (2001) have shown that this approximation to the likelihood has little impact on estimates of overall log relative rates on heterogeneity in rates across cities.

The second stage of the model describes variation among the true log relative rates β_M^c and β_H^c across cities. We assume:

$$\begin{bmatrix} \beta_M^c \\ \beta_H^c \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \alpha_M \\ \alpha_H \end{bmatrix}, \Sigma \right) \text{ where } \Sigma = \begin{bmatrix} \sigma_M^2 & \sigma_{MH} \\ \sigma_{MH} & \sigma_H^2 \end{bmatrix}. \tag{3}$$

Here α_M and α_H denote the overall log relative rates of mortality and hospital admissions from exposure to PM_{10}; σ_M^2 and σ_H^2 are the variances in β_M^c and β_H^c, and $\rho_{MH} = \sigma_{MH}/\sigma_M\sigma_H$ denotes the correlation across cities between β_M^c and β_H^c. Larger values of ρ indicate that cities with higher log relative rates of mortality are also more likely to have higher log relative rates of hospital admissions.

The specification of this Bayesian hierarchical model is completed by assigning prior distributions for the parameters. For the mean parameters (α_M, α_H), we assume vague normal priors having mean 0 with and large variance. Under the two-stage multivariate normal model (3), a natural choice for the prior distribution on the covariance matrix is the conjugate prior inverse Wishart distribution. Although the Inverse Wishart distribution is mathematically convenient for the implementation of the simulation-based techniques (Gilks et al., 1996), this distribution is not flexible enough to elicit non-informative priors on the variances and on the correlation coefficient (Daniels and Kass, 1999; Daniels, 1999). Instead of the conjugate prior for the entire covariance matrix, we assume that the two variance components σ_M^2 and σ_H^2 have a priori a half-normal distribution on $(0, \infty)$ with mean zero and a large variance (here chosen to be 10), and that the correlation coefficient ρ has a priori a uniform distribution in $[-1, 1]$.

In Section 4 we explore the consequences of this assumption using a sensitivity analysis of the posterior results to the prior specification.

4 Results

We apply the methods described in the Appendix to estimate the sample covariance matrix V^c of the log relative rate estimates, $\hat{\beta}_M^c$ and $\hat{\beta}_H^c$, within each city. Posterior distributions of all parameters of interest are approximated by simulation-based techniques (Gilks et al., 1996).

Figure 1 shows the 10% highest likelihood density regions (solid lines) and 10% highest posterior regions (shaded regions) of the mortality (x-axis) and hospitalization log relative rates (y-axis) for each of the 10 cities. The Bayesian estimates were obtained under our "baseline prior" for Σ specified in section 3.

Maximum likelihood and Bayesian estimates of the log relative rates are connected by arrows. The shapes of the likelihood density regions indicate that within-city estimated statistical correlations ($v_{HM}^c/\sqrt{v_{HH}^c v_{MM}^c}$) are small. The sample correlations range from -0.05 in Pittsburgh to 0.34 in Colorado Springs. Low values of the sample correlations indicate that the mortality and hospital admissions time series are only weakly correlated. The city-specific estimates of the overdispersion parameters $(\hat{\phi}_M, \hat{\phi}_H)$ and of the within-city statistical correlations are listed on Table 3.

Maximum likelihood and Bayesian estimates of the log relative rates of mortality and hospital admissions for the 10 locations with their 95% confidence intervals and 95% posterior regions are also listed in Table 4. Results are also reported under a "separate analysis" which assumes that mortality (morbidity) data do not provide any information on the log relative rates of hospital admissions (mortality), i.e. $\rho = 0$. Note that the Bayesian estimates of the city-specific log relative rates under a joint analysis are very similar to the estimates under a separate analysis, suggesting that ρ is very small and/or poorly estimated.

Because of the small number of cities, inferences about the degree of heterogeneity in pollution effects among cities are likely to be sensitive to the prior assumptions about Σ. Our strategy for investigating the impact of the prior distribution on our results is based on inspecting the posterior distributions of the parameters of interest under the following prior distributions for Σ: 1) Uniform prior on Σ, i.e. uniform prior on all the entries of Σ with the condition that Σ is positive definite; 2) Jeffreys prior on Σ, i.e. $p(\Sigma) \propto |\Sigma|^{-3/2}$; and 3) Uniform prior on the shrinkage matrix B_0, where $B_0 = V_0^{1/2}(V_0 + \Sigma)^{-1}V_0^{1/2}$, $V_0 = \frac{1}{10}\sum_{c=1}^{10} V^c$, i.e. again uniform prior on all the entries of the matrix B. Additional details on these prior distributions, including the definitions of the prior densities and software implementations are in (Everson and Morris, 2000).

The posterior distributions of the overall log relative rates of mortality and hospital admissions (α_M, α_H) (first row), and of the between-city standard deviations (σ_M, σ_H) (second row) under our baseline prior defined in Section 3, and under the alternative "non informative" prior distributions are shown in Figure 2. We found that the estimated overall relative rate of hospital admissions associated with PM_{10} expressed as percentage increase in mortality per 10 $\mu g/m^3$ PM_{10} increase is 0.71 percent (95 percent posterior interval, 0.35 to 0.99). The estimated overall log relative rate of mortality was 0.26 percent per 10 $\mu g/m^3$ PM_{10} (95 percent posterior interval, -0.37 to 0.65).

Posterior distributions of the among-city standard deviations indicate the degree of heterogeneity of the mortality and morbidity log relative rates across cities. The distribution of the standard deviation for mortality is similar to that for morbidity, with the posterior means of σ_M and of σ_H equal to 0.42 (95 percent posterior interval, 0.05 to 1.18), and 0.31 (95 percent posterior interval, 0.10 to 0.89), respectively.

The posterior distributions of the between-city correlation coefficient $\rho = \sigma_{MH}/\sigma_M\sigma_H$ are shown on Figure 3. We found that log relative rates of mortality and morbidity are weakly correlated and this correlation has large statistical uncertainty. The posterior mean of the correlation ρ between the true β_M^c and β_H^c is 0.20 with 95% interval $(-0.89, 0.98)$. Assuming a Uniform prior on Σ leads to larger posterior means and variances of σ_M^2 and σ_H^2, and to larger posterior variances of α_M and α_H, but has little effect on ρ. The Jeffreys prior on Σ, and the Uniform prior on B_0 give nearly identical posterior inferences as we obtained with our baseline prior on Σ (half normal on the variances and uniform on the correlation as specified in Section 3).

Generally, mortality time series data can be more readily assembled from publicly available data bases than morbidity series. Therefore, it may be desirable to predict the log relative rate of hospital admission for a city (other than the 10 sampled) which has mortality but not hospital admission data available. We consider New York as an example. Using the model, we can also estimate reductions in the posterior variances of the log relative rates of mortality and hospital admissions in New York (β_M^{NY} and β_H^{NY}) and compare the values with and without use of the mortality time series for New York.

Figure 4 (left) shows the marginal posterior distribution of β_M^{NY} using the New York mortality data (solid line) and the posterior predictive distribution of β_M^{NY} ignoring the NY mortality data. The marginal posterior distribution of β_M^{NY} (i.e. using the mortality data for NY) is obtained by sampling from an univariate normal distribution with mean $\left(1/\sigma_M^{2(j)} + 1/v_M^{NY}\right)\left(\alpha_M^{(j)}/\sigma_M^{2(j)} + \hat{\beta}_M^{NY}/v_M^{NY}\right)$ where $\boldsymbol{\alpha}^{(j)}$ and $\Sigma^{(j)}$ are the samples from the marginal posterior distribution of $\boldsymbol{\alpha}$ and Σ, and $\hat{\beta}_M^{NY}$ and v_M^{NY} are the maximum likelihood estimate of the log relative rate of mortality and the sample variance for NY. The predictive distribution of $(\beta_M^{NY}, \beta_H^{NY})$ (ignoring the NY mortality data) is obtained by sampling from the bivariate normal distribution $N(\boldsymbol{\alpha}^{(j)}, \Sigma^{(j)})$. The predictive distribution of β_H^{NY} (including the NY mortality data) is obtained by sampling from a normal distribution with mean $\alpha_H^{(j)} + \sigma_{MH}^{2(j)}/\sigma_M^{2(j)}(\beta_M^{NY^{(j)}} - \alpha_M^{(j)})$ where $\beta_M^{NY^{(j)}}$ is a sample from the marginal posterior distribution of β_M^{NY}. As expected, use of data from New York improves the estimate of β_M^{NY} with a reduction in posterior variance of 65% (see also Table 5).

Figure 4 (right) shows the posterior predictive distribution of the hospital admission log relative rate β_H^{NY} using the NY mortality data (solid line) and the posterior predictive distribution of β_H^{NY} ignoring the NY mortality data (dotted line). In this case, the reduction in the posterior variance of β_H^{NY} obtained by taking into account the mortality data in New York is much smaller, and equal to 15% (see also Table 6). The modest gain in precision of the Bayesian estimate of β_H^{NY} obtained by using the NY mortality data versus ignoring such information is due to the high imprecision in

the estimation of the correlation coefficient ρ.

Table 5 shows reductions in posterior variances of β_M^c and β_H^c when including versus ignoring time series data for city c. Percentage reductions in posterior variances of the log relative rates of morbidity and mortality are slightly larger under the combined analysis than under a separate analysis. This occurs because in the combined analysis, we use data for both the mortality and morbidity to approximate the marginal posterior distribution of β_M^c (β_H^c), while under the separate analysis only the mortality data (or morbidity data) are used.

Reductions in posterior variances of the log relative rates of morbidity and mortality are also larger in the presence of greater heterogeneity across cities. This pattern is found because, when the variance across cities is large then the Bayesian estimate of a city-specific log relative rate draws more heavily on the data from that city, and therefore a larger reduction in the posterior variances of the two log relative rates is obtained. For example, in New York under a uniform prior on B_0 which leads to larger estimates of σ_M^2 and σ_H^2, the reductions in the posterior variances of β_M^c and β_H^c, are 93% and 32%, as compared to 65% and 15% under the baseline prior for which the estimates of the heterogeneity are smaller. Finally, percentage reductions in posterior variances of the log relative rates of morbidity and mortality are more substantial in cities with smaller statistical variances, $\text{var}(\hat{\beta}_H^c \mid \beta_H^c)$ say, of the relative rates. For example in Detroit, where $\text{var}(\hat{\beta}_H^c \mid \beta_H^c) = 0.22$, we estimate a 86% reduction in the posterior variance of β_H^c whereas in Canton where $\text{var}(\hat{\beta}_H^c \mid \beta_H^c) = 1.04$, we estimate a 26% reduction in the posterior variance of β_H^c, because of the high statistical uncertainty in $\hat{\beta}_H^c$.

Finally, to explore the sensitivity of our results to the adjustment for confounding factors, we estimated overall log relative rates of mortality and morbidity (α_M, α_H) corresponding to five alternative scenarios for adjustment for confounding factors. The five scenarios were defined by multiplying the number of degrees of freedom of the smooth functions of time, temperature, and barometric pressure (natural cubic splines as defined in Table 2), by a calibration parameter δ which assumes values $1/3, 1/2, 1, 2$ and 3 respectively. Note that $\delta = 1$ is our baseline model and $\delta = 1/3$ and $\delta = 3$ represents less and more dramatic adjustment for trend, seasonality, and weather. Results for the mortality and morbidity analyses are shown in Figure 5. We found that the overall log-relative rate of mortality is sensitive to the degree of adjustment for confounding factors and loose significance when $\delta = 2$ and $\delta = 3$. In contrast , the log relative rate of morbidity is robust to the specification of δ. The rationale behind the choice of these scenarios, and more extensive results of the sensitivity of the overall log relative rate of mortality for the largest 90 cities, are discussed and summarized in (Dominici et al., 2003b).

5 Discussion

While understanding of pathogenesis remains limited, abundant evidence indicates that current levels of airborne particulate matter are associated with mortality counts and various indexes

of morbidity (Pope and Dockery, 1999; Environmental Protection Agency, 2001). Time series analyses have been carried out to characterize the effect of particulate matter on a variety of health outcomes including mortality, hospitalization, emergency room visits, and clinic or physician visits. In general, there is evidence linking particulate matter to increased risk for each of these outcome measures. There is some overlap among the cities included in each of these different sets of analyses; by design, some of the same cities are included in the NMMAPS and APHEA (Katsouyanni et al., 1997, 2001) studies. However, patterns of correlation of effects among the cities for different health outcomes have not yet been examined.

There are numerous hypotheses with regard to the nature of the processes underlying these associations and with regard to characteristics of particles and their potential to initiate local and systemic injury. In general, the same pathogenic mechanisms have been considered as responsible for effects on either mortality or morbidity. Additionally, the same populations have been considered as susceptible to the effects of particles, namely infants, the elderly and persons with chronic cardiac and respiratory diseases. For these susceptible individuals, air pollution has been postulated as worsening clinical status, and thereby increasing risk for hospitalization and ultimately death. These biomedical considerations imply that levels of effect of particulate air pollution on morbidity and mortality might be correlated. Unfortunately, the present analyses provide insufficient evidence to test for these hypothesized correlations and methods should be applied to longer time series for a larger number of cities.

Motivated by these general pathogenic considerations, we have developed a hierarchical bivariate time series model to jointly assess the relationships between mortality and morbidity in 10 U.S. cities. These cities were selected on the basis of data availability for PM_{10}, hospitalization, and mortality and they were not intended to be representative either of the NMMAPS data nor of the United States. Nonetheless, the data came from cities of varying characteristics scattered across the United States and we were unable to gain insights concerning the correlation between log relative rates of morbidity and mortality among cities.

Our modeling approach extended Poisson regression analyses of univariate time series data on air pollution and health to multivariate health outcomes. Within each city, we fitted two seemingly unrelated Poisson regression models to estimate log relative rates of mortality and morbidity $(\hat{\beta}^c_H, \hat{\beta}^c_M)$. In addition, we have developed a novel generalized estimating equations approach to estimate the sample covariance matrix of the relative rates (V^c) by using the bivariate time series on hospital admission and mortality $[y^c_{tH}, y^c_{tM}]$, directly. We then extended this bivariate Poisson time series model in an hierarchical fashion to combine the vector of the city-specific estimates of the relative rates of mortality and morbidity across cities. Although it is important to take into account the correlation between the mortality and morbidity time series within each city, we focused our analysis on making inferences on the parameters at the second stage of the hierarchical models and on approximating the marginal posterior distributions of the overall log relative rates of mortality and morbidity, their between-city variances, and their correlation across cities.

The combined analysis approach has several useful features: 1) by estimating the covariances between the log relative rates (v^c_{HM}), it takes into account correlation between the mortality and morbidity time series; 2) it provides more efficient estimates of the relative rates than would separate analyses, because it uses data for both mortality and morbidity to approximate the marginal posterior distribution of β^c_H (β^c_M), while under the separate analysis only the morbidity (mortality) data are used; 3) it can be used for prediction of a hospitalization log relative rate for an additional city using mortality data which might be needed for policy purposes; and finally 4) for a city of interest $c^{'}$, it quantifies the reduction in the variance of $\beta^{c^{'}}_H$ and $\beta^{c^{'}}_M$ which can be obtained by collecting time series data for $c^{'}$ with respect to predicting $\beta^{c^{'}}_H$ and $\beta^{c^{'}}_M$ based on the data from the other cities.

The application of our two-stage bivariate normal-normal model to daily time series data on mortality and morbidity in 10 cities are consistent with results of previous studies of morbidity and mortality separately. Overall log-relative rates of mortality and morbidity obtained by combining information across the 10 cities were similar to those reported in the recent NMMAPS re-analysis for 90 and 14 cities (Dominici et al., 2003a), but with larger posterior intervals due to the smaller number of cities analyzed here. As expected, the overall log-relative rate of morbidity was larger and less heterogeneous than the overall log-relative rate of mortality. Unfortunately because of the large statistical uncertainty within each city the correlation coefficient ρ was estimated very poorly, thus providing very weak information on the overall association between log-relative rates of mortality and morbidity. With the methods developed we should gain further insights on this issue by applying our modeling strategy to longer time series on pollution, mortality and morbidity and to a larger number of cities. Because of the strong biological basis for postulating a correlation between morbidity and mortality effects, such additional exploration is needed, given our initial findings.

Recent contributions on semi-parametric regressions could also be used to extend our modeling approach. For example, we could have used a Bayes approach via MCMC sampling for inference in generalized additive models with city-specific random effects as suggested by Fahrmeir and Lang (2001). This approach would have avoided our reliance on the assumption of normality at the first stage of the hierarchical model. At the other end, to estimate properly the sample covariance between the estimated log relative rates for mortality and morbidity, further method development is needed to extend Generalized Additive Models with random effects to multivariate outcomes.

6 Appendix: estimating the sample covariance matrix

In this section we describe the estimation procedure for the sample covariance matrix V^c. In what follows, we drop the index c for notational convenience.

Let $\boldsymbol{\xi} = [\boldsymbol{\theta}_M, \boldsymbol{\theta}_H]$ be the full vector of the coefficients, where $\boldsymbol{\theta}_M$ and $\boldsymbol{\theta}_H$ are q dimensional vectors, $\boldsymbol{\theta}_M = [\beta_M, \boldsymbol{\eta}_M]$ and $\boldsymbol{\theta}_H = [\beta_H, \boldsymbol{\eta}_H]$. Our goal is to estimate the $2q \times 2q$ covariance matrix

$Var(\hat{\boldsymbol{\xi}})$, where $\hat{\boldsymbol{\xi}}$ is the full vector of the maximum likelihood estimates (MLE) estimates of the coefficients. Note that an estimate of the sample covariance matrix V can be obtained by taking the $[1,1]$-th, $[1,q]$-th and $[q,q]$-th elements of $\widehat{Var}(\hat{\boldsymbol{\xi}})$.

Assume that $\boldsymbol{\theta} = \boldsymbol{\theta}_M$ (or $\boldsymbol{\theta}_H$), and let $U(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left(\frac{\partial \mu_t}{\partial \boldsymbol{\theta}}\right)' w_t^{-1}(y_t - \mu_t)$ be the estimating function for $\boldsymbol{\theta}$ (Zeger et al., 1988). Under an overdispersed Poisson model we have $(\partial \mu_t/\partial \boldsymbol{\theta})' = w_t \phi^{-1} X_t'$, leading to:

$$U(\boldsymbol{\theta}) = \phi^{-1} \sum_{t=1}^{T} X_t'(y_t - \mu_t)$$

where by definition $U(\hat{\boldsymbol{\theta}}) = 0$.
The first order Taylor series expansion of $U(\boldsymbol{\theta})$ about the MLE $\hat{\boldsymbol{\theta}}$ is:

$$U(\boldsymbol{\theta}) \simeq U(\hat{\boldsymbol{\theta}}) + \frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

from which we obtain:

$$\sqrt{T}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \;\; \simeq \;\; \left(\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}}/T\right)^{-1} U(\boldsymbol{\theta})/\sqrt{T}$$

For $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, the true value $U(\boldsymbol{\theta}_0)/\sqrt{T}$ converges to a Gaussian random variable with mean 0 and variance, denoted by B, which converges to a constant. Hence $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges to a normal random variable with mean zero and variance:

$$\begin{aligned} Var(\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)) \;\; &\rightarrow \;\; \left(E\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}}/T\right)^{-1} Var(U(\boldsymbol{\theta})/\sqrt{T}) \left(E\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}}/T\right)^{-1} = \\ &\rightarrow \;\; T\left(E\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}}\right)^{-1} Var(U(\boldsymbol{\theta})) \left(E\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}}\right)^{-1} = \\ &\rightarrow \;\; TA^{-1}BA^{-1} \qquad \text{for} \qquad T \rightarrow \infty \end{aligned} \tag{4}$$

From the asymptotic result (4) it follows that:

$$Var(\hat{\boldsymbol{\theta}}) \;\; \rightarrow \;\; A^{-1}BA^{-1} \tag{5}$$

where:

- $A = E\frac{\partial U(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\hat{\boldsymbol{\theta}}} = -\phi^{-1} \sum_{t=1}^{T} X_t' \mu_t X_t$, i.e. the Fisher information under the independence assumption.
- $B = Var(U(\boldsymbol{\theta})) = \phi^{-2} \sum_{t=1}^{T} X_t' Var(y_t - \mu_t) X_t$.

Let $A_{MM}, A_{HH}, B_{MM}, B_{HH}, B_{MH}$ be the matrices defined above as functions of the mortality (M) and hospital admission time series (H). From $\boldsymbol{\xi} = [\boldsymbol{\theta}_M, \boldsymbol{\theta}_H]$ and (5), the covariance matrix $Var(\hat{\boldsymbol{\xi}})$ is equal to:

$$\begin{aligned} Var(\hat{\boldsymbol{\xi}}) \;\; &= \;\; \begin{bmatrix} A_{MM}^{-1} & 0 \\ 0 & A_{HH}^{-1} \end{bmatrix} \begin{bmatrix} B_{MM} & B_{MH} \\ B_{HM} & B_{HH} \end{bmatrix} \begin{bmatrix} A_{MM}^{-1} & 0 \\ 0 & A_{HH}^{-1} \end{bmatrix} = \\ &= \;\; \begin{bmatrix} A_{MM}^{-1}B_{MM}A_{MM}^{-1} & A_{MM}^{-1}B_{MH}A_{HH}^{-1} \\ A_{HH}^{-1}B_{HM}A_{MM}^{-1} & A_{HH}^{-1}B_{HH}^{-1}A_{HH}^{-1} \end{bmatrix} \end{aligned} \tag{6}$$

Finally, we need to estimate the A and B matrices by using the output of the Poisson regressions (1). For A_{MM} and B_{MH} we use the estimators:

- $\widehat{A}_{MM} = \left(\hat{\phi}^M\right)^{-1} \sum_{t=1}^{T} X_t^{M'} \hat{\mu}_t X_t^M$
- $\widehat{B}_{MH} = \left(\hat{\phi}^M \hat{\phi}^H\right)^{-1} E\left[\sum_s X_s^{M'}(y_s^M - \hat{\mu}_s^M) \sum_l X_s^{H'}(y_l^H - \hat{\mu}_l^H)\right]$

where $\hat{\mu}_t^M$ ($\hat{\mu}_t^H$) are the fitted values from the Poisson models applied to the mortality (M) and hospital admissions (H) time series, and $\hat{\phi}^M$ ($\hat{\phi}^H$) are the estimates of the overdispersion parameters. We estimate $E\left[\sum_s X_s^{M'}(y_s^M - \hat{\mu}_s^M) \sum_l X_s^{H'}(y_l^H - \hat{\mu}_l^H)\right]$ by using:

$$\sum_s \sum_{l=(s-L)}^{s+L} X_s^{M'} X_l^H h\left((y_s^M - \hat{\mu}_s^M)(y_l^H - \hat{\mu}_l^H)\right), \tag{7}$$

where we assume that $Cov(y_t^M, y_t^H) = 0$ for $|s-l| > L$ and $h\left((y_s^M - \hat{\mu}_s^M)(y_l^H - \hat{\mu}_l^H)\right)$ is a smooth function applied to the cross products.

References

Albert, P. S. (1999). Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine* **18**, 1707–1732.

American Thoracic Society, Bascom, R. (1996a). Health effects of outdoor air pollution, Part 1. *American Journal of Respiratory and Critical Care Medicine* **153**, 3–50.

American Thoracic Society, Bascom, R. (1996b). Health effects of outdoor air pollution, Part 2. *American Journal of Respiratory and Critical Care Medicine* **153**, 477–498.

Breslow, N. and Clayton, D. (1993). Approximation inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, New York.

Daniels, M. (1999). A prior for the variance in hierarchical models. *Canadian Journal of Statistics* **27**, 569–580.

Daniels, M., Dominici, F., Samet, J. M., and Zeger, S. L. (2000). Estimating PM10-mortality dose-response curves and threshold levels: An analysis of daily time series for the 20 largest US cities. *American Journal of Epidemiology* **152**, 397–412.

Daniels, M. and Kass, R. (1999). Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association* **94**, 1254–1263.

Dockery, D. and Pope, C. A. (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health* **15**, 107—132.

Dockery, D., Pope, C. A., Xu, X., Spengler, J., Ware, J., Fay, M., Ferris, B., and Speizer, F. (1993). An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine* **329**, 1753–1759.

Dominici, F., Daniels, M., Zeger, S. L., and Samet, J. M. (2002). Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association* **97**, 100–111.

Dominici, F., McDermott, A., Daniels, M., Zeger, S. L., and Samet, J. M. (2003a). *A Special Report to the Health Effects Institute on the Revised Analyses of the NMMAPS II Data*. The Health Effects Institute, Cambridge, MA.

Dominici, F., McDermott, A., and Hastie, T. (2003b). *Improved Semiparametric Regression Models of Air Pollution and Mortality*. Technical report, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins.

Dominici, F., Samet, J. M., and Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the twenty largest US cities: A hierarchical modeling strategy (with discussion). *Royal Statistical Society, Series A, with discussion* **163**, 263–302.

DuMouchel, W. (1990). *Bayesian Metaanalysis*. Marcel Dekker.

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

Environmental Protection Agency (2001). Air Quality Criteria for Particulate Matter: Second External Review Draft March 2001. *US Environmental Protection Agency, Office of Research and Development* .

Everson, P. and Morris, C. (2000). Inference for multivariate normal hierarchical models. *Journal of the Royal Statistical Society, series B* **62**, 399–412.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive models based on markov random field. *Journal of the Royal Statistical Society Series C* **50**, 201–221.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer-Verlag.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall, New York.

Hwang, J. and Chan, C. (2001). Air pollution effects on daily clinic visits for lower respiratory illness. *American Journal of Epidemiology (in press)* .

Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., LeTertre, A., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., and Anderson, H. R. (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 european cities within the APHEA2 project. *Epidemiology* in press.

Katsouyanni, K., Touloumi, 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. *British Medical Journal* **314**, 1658–1663.

Kelsall, J., Samet, J. M., and Zeger, S. L. (1997). Air pollution, and mortality in Philadelphia, 1974-1988. *American Journal of Epidemiology* **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. *Environmental Health Perspectives* **106**, 93–99.

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

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

Lipfert, F. and Wyzga, R. (1993). Air pollution and mortality: Issues and uncertainty. *Journal Air Waste Manage. Assoc* **45**, 949—966.

McNeney, B. and Petkau, J. (1994). Overdispersed Poisson regression models for studies of air pollution and human health. *The Canadian Journal of Statistics* **22**, 421–440.

Pope, A. C. and Dockery, D. W. (1999). Epidemiology of particle effects. In: *Air Pollution and Health*, 673–705. San Diego: Academic Press.

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

Samet, J. M., Dominici, F., Curriero, F., Coursac, I., and Zeger, S. L. (2000a). Fine particulate air pollution and mortality in 20 U.S. cities: 1987-1994. *New England Journal of Medicine (with discussion)* **343**(24), 1742–1757.

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

Samet, J. M., Zeger, S. L., Dominici, F., Curriero, F., Coursac, I., Dockery, D., Schwartz, J., and Zanobetti, A. (2000b). *The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality from Air Pollution in the United States*. Health Effects Institute, Cambridge, MA.

Samet, J. M., Zeger, S. L., Dominici, F., Dockery, D., and Schwartz, J. (2000c). *The National Morbidity, Mortality, and Air Pollution Study Part I: Methods and Methodological Issues*. Health Effects Institute, Cambridge, MA.

Samet, J. M., Zeger, S. L., 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*. Health Effects Institute, Cambridge, MA.

Schwartz, J. (1994). Air pollution and daily mortality: A review and meta analysis. *Environmental Research* **64**, 36–52.

Schwartz, J. (1995). Air pollution and daily mortality in Birmingham, alabama. *American Journal of Epidemiology* **137**, 1136–1147.

Schwartz, J. (2000). Assessing confounding, effect modification, and thresholds in the associations between ambient particles and daily deaths. *Environmental Health Perspective* **108**, 563–568.

Schwartz, J. and Zanobetti, A. (2000). Is there a threshold in the relation between particulate air pollution and daily deaths: a meta-smoothing approach. *Epidemiology* **6**, 666– 672.

Schwartz, J., Zanobetti, A., and Bateson, T. (2003). *A Special Report to the Health Effects Institute on the Revised Analyses of the NMMAPS II Data: Morbidity and Mortality among Elderly Residents of Cities with Daily PM Measurements.* The Health Effects Institute, Cambridge, MA.

Zanobetti, A., Schwartz, J., and Dockery, D. (2000a). Airborne particles are a risk factor for hospital admissions for heart and lung disease. *Environmental Health Perspective* **108**, 1071–1077.

Zanobetti, A., Wand, M., Schwartz, J., and Ryan, L. (2000b). Generalized additive distributed lag models. *Biostatistics* **1**, 279–292.

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

Zeger, S. L., Liang, K., and Albert, P. (1988). Models for longitudinal data: A generalised estimating equation approach. *Biometrics* **44**, 1049–1060.

Zeger, S. L. and Liang, K.-Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine* **11**, 1825–1839.

Table 1: *Start and end dates of the PM_{10} monitoring, number of days with PM_{10} samples, 24-h average PM_{10}, mean hospital admissions, and mean mortality for cardiovascular diseases by city (number).*

Cities	Start Date	End Date	# days PM_{10} available	PM_{10} average	CVD (H)	CVD (M)
Birmingham (1)	3/1/87	12/31/93	2485	34.8	24	6
Canton (2)	1/1/88	12/31/93	1750	28.4	10	3
Colorado Springs (3)	7/1/87	12/31/93	2310	27.5	3	2
Minneapolis/Saint Paul (4)	2/1/87	12/31/93	2488	28.1	22	10
Seattle (5)	1/1/86	12/31/93	2913	32.2	20	9
Spokane (6)	1/1/86	12/31/93	2778	42.9	6	3
Chicago (7)	1/1/88	12/31/93	2058	36.3	114	48
Detroit (8)	4/1/86	12/31/93	2517	36.7	53	23
New Haven (9)	1/1/88	12/31/91	1450	28.6	19	8
Pittsburgh (10)	1/1/86	12/31/93	2918	36.0	51	16
New York	1/1/87	12/31/94	489	28.8	NA	108

Table 2: *Model specification for estimating city-specific log relative rates associated with current day and previous day particulate air pollution levels, including potential confounding factors and the rationale for their inclusion in the model. We used an overdispersed Poisson regression model and specified the smooth functions of time and temperature variables as natural cubic splines.*

Predictors	Primary reasons for inclusion
Average PM_{10} at lag 0 and at lag 1 (linear term)	To estimate log-log relative rates of mortality associated with short-term increase in air pollution levels
Indicator variables for the day of the week (linear terms)	Allow different baseline log mortality rate within each day of the week
Smooth functions of time ($4df \times$ year)	To adjust for long term trend and seasonality
Smooth functions of temperature at lag 0 and lag 1 ($3df$)	To control for the known effects of weather on mortality
Smooth functions of barometric pressure and relative humidity ($3df$)	To control for the known effects of humidity on mortality

Table 3: *Estimates of the city-specific overdispersion parameters and city-specific statistical correlations between* $\hat{\beta}_M^c$ *and* $\hat{\beta}_H^c$.

City	$\hat{\phi}^M$	$\hat{\phi}^H$	$v_{MH}^c / \sqrt{v_M^c v_H^c}$
Birmingham (1)	0.84	0.93	0.11
Canton (2)	0.82	0.96	0.07
Chicago (3)	0.91	1.21	0.34
Colorado Spring (4)	0.89	0.89	-0.02
Detroit (5)	0.83	1.01	0.14
Minneapolis (6)	0.82	0.94	0.18
New Haven (7)	0.84	0.89	0.17
Pittsburgh (8)	0.86	1.02	-0.02
Seattle (9)	0.85	0.97	0.10
Spokane (10)	0.88	0.87	-0.05

Table 4: *Maximum Likelihood and Bayesian estimates (posterior means) of the log relative rates of mortality and hospital admissions for cardiovascular diseases in the 10 locations. Between parentheses () are the* 95% *confidence intervals and* 95% *posterior regions, respectively. Results are reported for the combined analysis and for the separate analysis which assumes that* $\rho = 0$.

	Log Relative Rates of Mortality			**Log Relative Rates of Morbidity**		
Cities	MLE	Bayes (combined)	Bayes (separate)	MLE	Bayes (combined)	Bayes (separate)
1	-0.13 (-1.38,1.12)	0.17 (-0.57,0.90)	0.20 (-0.52,0.92)	0.28 (-0.28,0.84)	0.55 (0.08,1.02)	0.48 (-0.01,0.96)
2	0.86 (-2.00,3.72)	0.31 (-0.72,1.34)	0.32 (-0.58,1.23)	0.59 (-1.46,2.64)	0.70 (0.00,1.39)	0.67 (-0.13,1.47)
3	0.44 (-0.13,1.02)	0.35 (-0.09,0.79)	0.36 (-0.07,0.78)	0.99 (0.50,1.48)	0.84 (0.49,1.19)	0.85 (0.45,1.25)
4	0.11 (-2.71,2.92)	0.23 (-0.80,1.26)	0.27 (-0.58,1.13)	0.47 (-1.51,2.45)	0.70 (0.01,1.39)	0.67 (-0.13,1.46)
5	0.33 (-0.20,0.87)	0.33 (-0.10,0.75)	0.33 (-0.09,0.74)	0.63 (0.15,1.11)	0.69 (0.33,1.04)	0.66 (0.28,1.04)
6	1.07 (0.03,2.10)	0.56 (-0.19,1.32)	0.56 (-0.18,1.30)	0.32 (-0.60,1.24)	0.63 (0.08,1.17)	0.54 (-0.07,1.15)
7	0.07 (-1.65,1.78)	0.23 (-0.62,1.08)	0.23 (-0.57,1.02)	1.36 (0.26,2.47)	0.87 (0.32,1.41)	0.89 (0.18,1.60)
8	0.36 (-0.30,1.03)	0.36 (-0.12,0.83)	0.33 (-0.16,0.82)	0.91 (0.48,1.35)	0.82 (0.50,1.13)	0.84 (0.47,1.21)
9	0.30 (-0.44,1.04)	0.30 (-0.22,0.82)	0.28 (-0.23,0.80)	0.71 (0.10,1.33)	0.73 (0.33,1.12)	0.70 (0.23,1.17)
10	-0.29 (-1.32,0.73)	0.07 (-0.66,0.80)	0.09 (-0.63,0.81)	0.14 (-0.64,0.93)	0.54 (-0.07,1.15)	0.47 (-0.13,1.06)
New York	0.70 (-0.18,1.58)	0.52 (-0.10,1.14)	0.46(-0.19,1.12)	-	0.61 (-0.33,1.55)	-
overall		0.26 (-0.37,0.65)	0.28 (-0.12,0.63)		0.71 (0.35,0.99)	0.69 (0.33,1.06)

Table 5: *Percent reductions in the posterior variances of the log relative rates of mortality and morbidity under a combined analysis for mortality and morbidity:* $\mathtt{d}$ *denotes the MLEs* $\hat{\boldsymbol{\beta}}^c = \left[\hat{\beta}^c_M, \hat{\beta}^c_H\right]$ *and the sample covariance matrices* V^c *for the 10 cities;* $\mathtt{d}^{-c}$ *is the same as* $\mathtt{d}$*, but without the estimates for city* c*;* $\mathtt{d}_M$ *(*$\mathtt{d}_H$*) denotes the MLEs* $\hat{\beta}^c_M$ *(*$\hat{\beta}^c_H$*) and the sample variances* v^c_M *(*v^c_H*) for the 10 cities;* $\mathtt{d}^{-c}_M$ *and* $\mathtt{d}^{-c}_H$ *are the same as* $\mathtt{d}_M$ *and* $\mathtt{d}_H$*, but without the estimates for city* c*;* $\mathtt{d}^{NY}_M$ *denotes the MLE* $\hat{\beta}^{NY}_M$ *and the sample variance* v^{NY}_M *for New York.*

	Combined Analysis	
	Mortality	Morbidity
Cities	$1 - \mathrm{v}(\beta^c_M \mid \mathtt{d})/\mathrm{v}(\beta^c_M \mid \mathtt{d}^{-c})$	$1 - \mathrm{v}(\beta^c_H \mid \mathtt{d})/\mathrm{v}(\beta^c_H \mid \mathtt{d}^{-c})$
1	0.59	0.68
2	0.23	0.26
3	0.86	0.83
4	0.34	0.26
5	0.87	0.81
6	0.62	0.60
7	0.51	0.56
8	0.84	0.84
9	0.80	0.76
10	0.66	0.38
	$1 - \mathrm{v}(\beta^{NY}_M \mid \mathtt{d}^{NY}_M, \mathtt{d})/\mathrm{v}(\beta^{NY}_M \mid \mathtt{d})$	$1 - \mathrm{v}(\beta^{NY}_H \mid \mathtt{d}^{NY}_M, \mathtt{d})/\mathrm{v}(\beta^{NY}_H \mid \mathtt{d})$
NY (Baseline)	0.65	0.15
NY (Uniform on Σ)	0.92	0.45
NY (Jeffrey)	0.94	0.49
NY (Uniform on B_0)	0.93	0.32

Table 6: *Percent reductions in the posterior variances of the log relative rates of mortality and morbidity under a separate analysis for mortality and morbidity:* $\mathtt{d}$ *denotes the MLEs* $\hat{\boldsymbol{\beta}}^c = \left[\hat{\beta}^c_M, \hat{\beta}^c_H\right]$ *and the sample covariance matrices* V^c *for the 10 cities;* $\mathtt{d}^{-c}$ *is the same as* $\mathtt{d}$*, but without the estimates for city c;* $\mathtt{d}_M$ *(*$\mathtt{d}_H$*) denotes the MLEs* $\hat{\beta}^c_M$ *(*$\hat{\beta}^c_H$*) and the sample variances* v^c_M *(*v^c_H*) for the 10 cities;* $\mathtt{d}^{-c}_M$ *and* $\mathtt{d}^{-c}_H$ *are the same as* $\mathtt{d}_M$ *and* $\mathtt{d}_H$*, but without the estimates for city c;* $\mathtt{d}^{NY}_M$ *denotes the MLE* $\hat{\beta}^{NY}_M$ *and the sample variance* v^{NY}_M *for New York.*

	Separate Analysis	
	Mortality	Morbidity
Cities	$1 - \mathrm{v}(\beta^c_M \mid \mathtt{d}_M)/\mathrm{v}(\beta^c_M \mid \mathtt{d}^{-c}_M)$	$1 - \mathrm{v}(\beta^c_H \mid \mathtt{d}_H)/\mathrm{v}(\beta^c_H \mid \mathtt{d}^{-c}_H)$
1	0.52	0.72
2	0.16	0.22
3	0.81	0.81
4	0.24	0.31
5	0.82	0.85
6	0.47	0.58
7	0.45	0.48
8	0.74	0.84
9	0.74	0.77
10	0.49	0.56
	$1 - \mathrm{v}(\beta^{NY}_M \mid \mathtt{d}^{NY}_M, \mathtt{d}_M)/\mathrm{v}(\beta^{NY}_M \mid \mathtt{d}_M)$	$1 - \mathrm{v}(\beta^{NY}_H \mid \mathtt{d}_H)/\mathrm{v}(\beta^{NY}_H \mid \mathtt{d}_H)$
NY (Baseline)	0.67	0

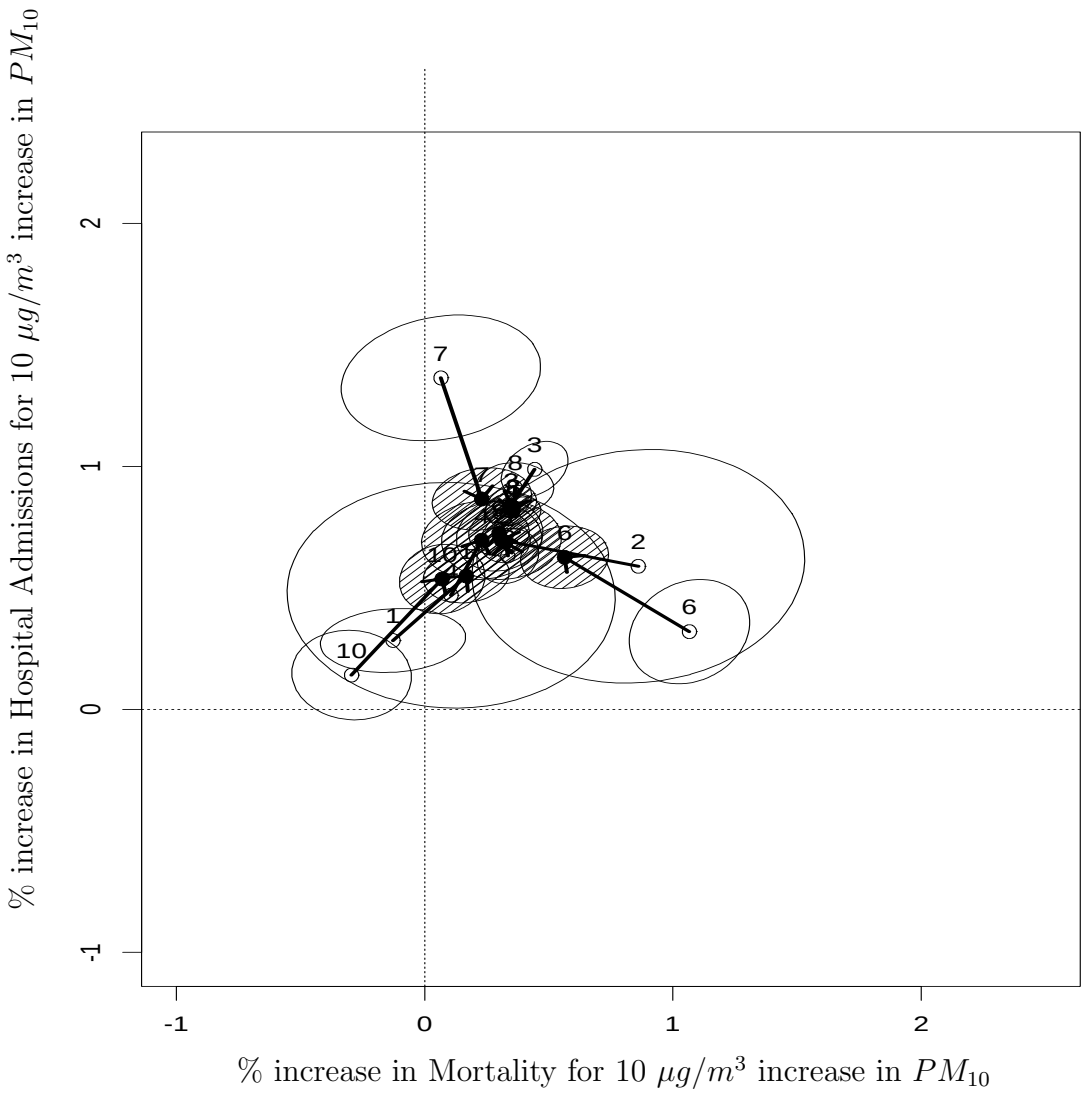

Figure 1: 10% *Highest Likelihood Density regions (solid lines) and* 10% *Highest Posterior Density regions (shaded regions) of the log relative rates of Total mortality and Hospital admissions for cardiovascular diseases. Maximum Likelihood estimates and Bayesian estimates are connected with arrows. For cities, refer to numbers given in Table 1. The Bayesian estimates were obtained under our "baseline prior" for the covariance matrix* Σ.

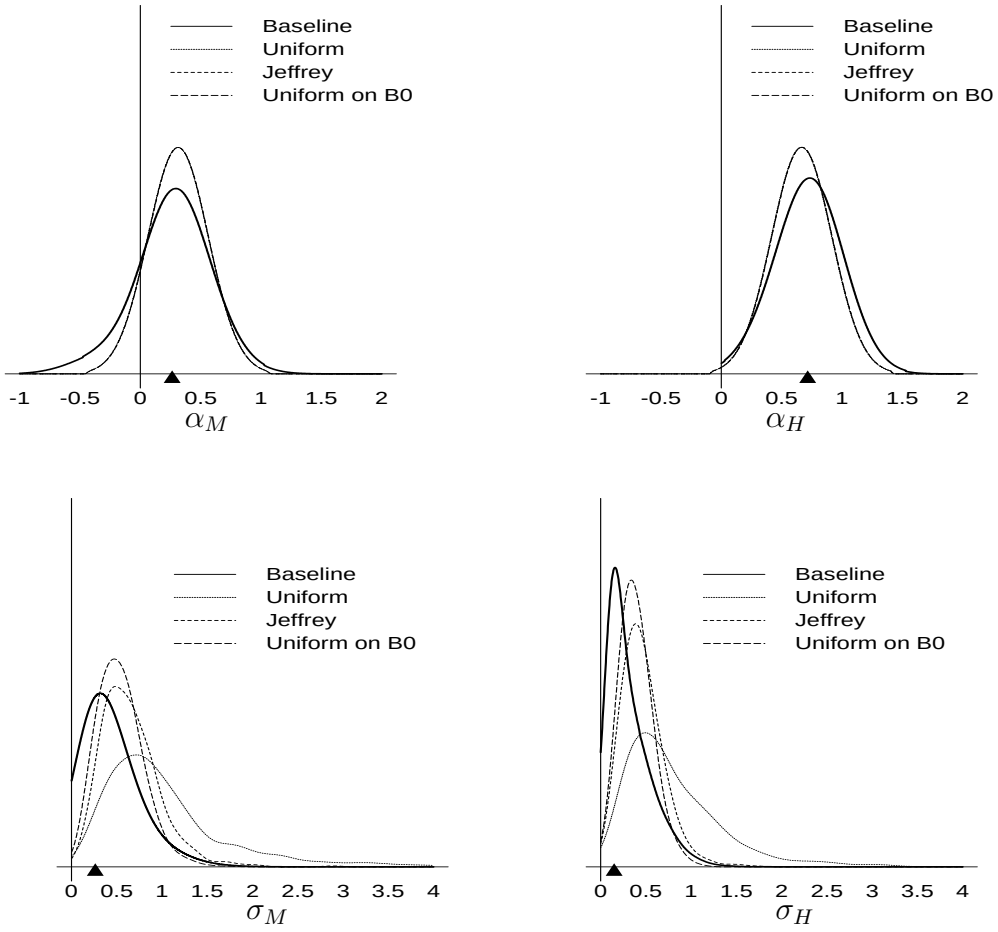

Figure 2: *Top left: marginal posterior distributions of the overall log relative rates of mortality (α_M) and overall log relative rate of hospital admissions (α_H). Top right: marginal posterior distributions of standard deviations (σ_M) and (σ_H). The filled triangles are placed at the posterior means.*

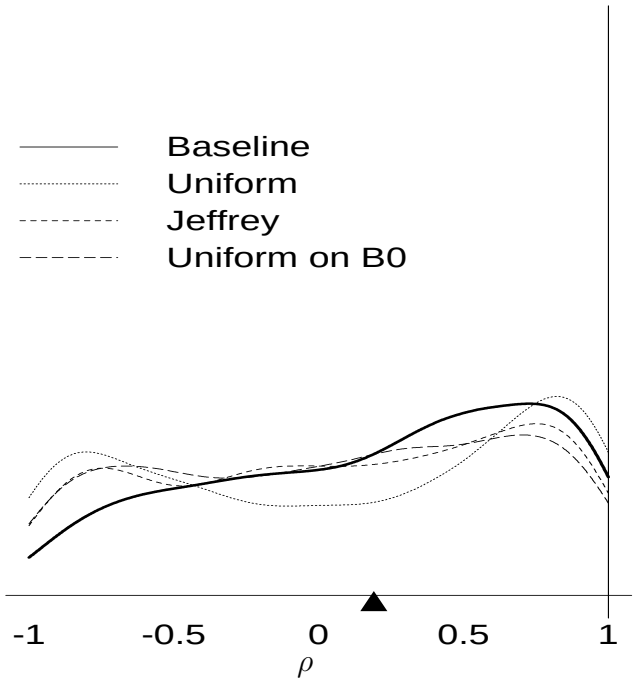

Figure 3: *Marginal posterior distribution of the correlation coefficient ($\rho = \sigma_{MH}/\sigma_M\sigma_H$). The filled triangle is placed at the posterior mean.*

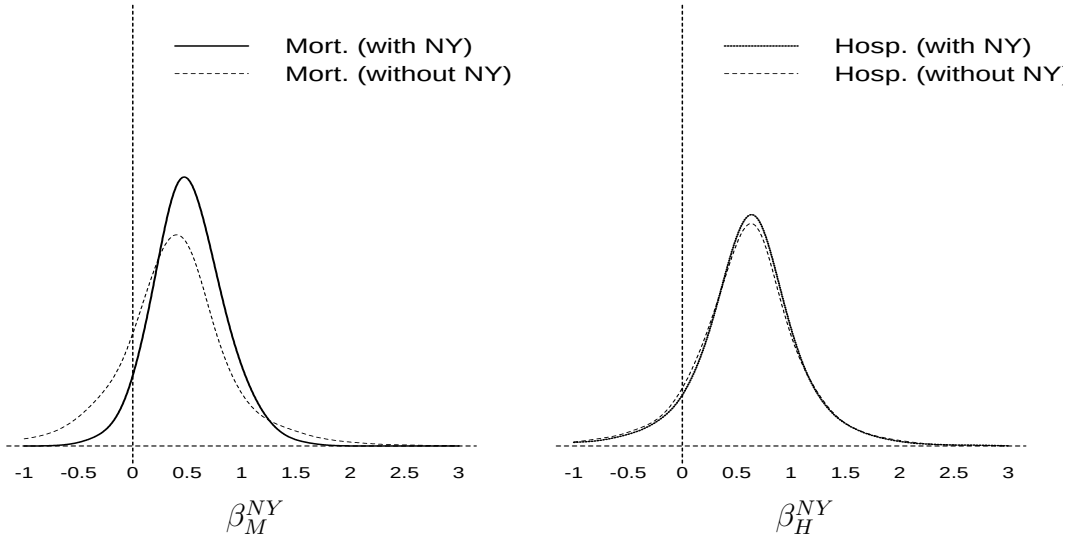

Figure 4: *Left: marginal posterior distribution of β_M^{NY} considering the NY mortality data (solid line) and posterior predictive distribution of β_M^{NY} ignoring the NY mortality data. Right: predictive distribution of β_H^{NY} considering the NY mortality data (solid line) and posterior predictive distribution of β_H^{NY} ignoring the NY mortality data (dotted line).*

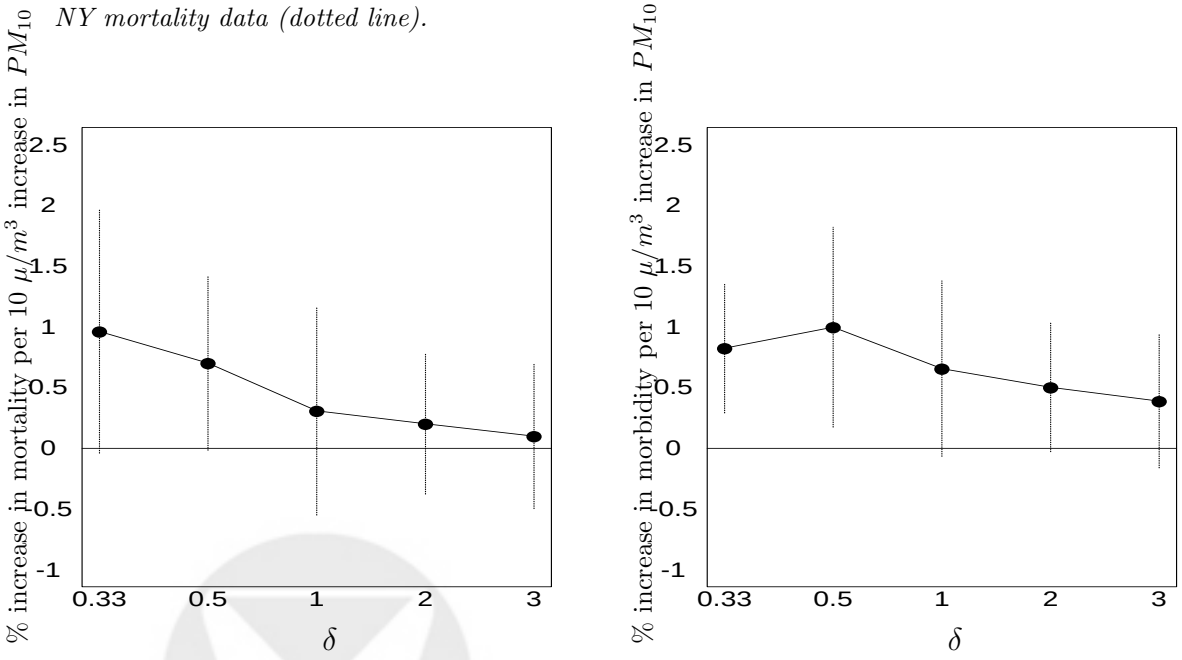

Figure 5: *Left: Overall log relative rates of mortality and hospital admissions (α_H, α_H) plotted in correspondence of five alternative scenarios for adjustment for confounding factors. On the x-axis are plotted the values of calibration parameter (δ) which multiply all df in the smooth functions of time, temperature, and barometric pressure. Our "baseline" model corresponds to $\delta = 1$.*