

# Bayesian hierarchical distributed lag models for summer ozone exposure and cardio-respiratory mortality

Yi Huang<sup>1</sup>, Francesca Dominici<sup>1,\*†</sup> and Michelle L. Bell<sup>2</sup>

<sup>1</sup>*Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205-3179, U.S.A.*

<sup>2</sup>*Forestry and Environmental Studies, Yale University, 205 Prospect Street, New Haven, CT 06511, USA*

## SUMMARY

In this article we develop Bayesian hierarchical distributed lag models for estimating associations between daily variations in summer ozone levels and daily variations in cardiovascular and respiratory (CVDRESP) mortality counts for 19 large U.S. cities included in the National Morbidity, Mortality and Air Pollution Study (NMMAPS) for the summers of 1987–1994.

In the first stage, we define a semi-parametric distributed lag Poisson regression model to estimate city-specific relative rates of CVDRESP mortality associated with short-term exposure to summer ozone. In the second stage, we specify a class of distributions for the true city-specific relative rates to estimate an overall effect by taking into account the variability within and across cities. We perform the calculations with respect to several random effects distributions (normal, *t*-student, and mixture of normal), thus relaxing the common assumption of a two-stage normal–normal hierarchical model. We assess the sensitivity of the results to: (i) lag structure for ozone exposure; (ii) degree of adjustment for long-term trends; (iii) inclusion of other pollutants in the model; (iv) heat waves; (v) random effects distributions; and (vi) prior hyperparameters.

On average across cities, we found that a 10ppb increase in summer ozone level over the previous week is associated with a 1.25 per cent increase in CVDRESP mortality (95 per cent posterior regions: 0.47, 2.03). The relative rate estimates are also positive and statistically significant at lags 0, 1 and 2. We found that associations between summer ozone and CVDRESP mortality are sensitive to the confounding adjustment for PM<sub>10</sub>, but are robust to: (i) the adjustment for long-term trends, other gaseous pollutants (NO<sub>2</sub>, SO<sub>2</sub> and CO); (ii) the distributional assumptions at the second stage of the hierarchical model; and (iii) the prior distributions on all unknown parameters.

Bayesian hierarchical distributed lag models and their application to the NMMAPS data allow us to estimate of an acute health effect associated with exposure to ambient air pollution in the last few days on average across several locations. The application of these methods and the systematic assessment of the sensitivity of findings to model assumptions provide important epidemiological evidence for future air quality regulations. Copyright © 2005 John Wiley & Sons, Ltd.

**KEY WORDS:** Bayesian hierarchical model; distributed lag model; ozone; cardiovascular and respiratory mortality

\*Correspondence to: F. Dominici, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205-3179, U.S.A.

†E-mail: fdominic@jhsph.edu.

Contract/grant sponsor: EPAN; contract/grant number: 3D-6867-NAEX.

Contract/grant sponsor: NIEHS RO1; contract/grant number: ES012054-01.

Contract/grant sponsor: NIEHS Center in Urban Environmental Health; contract/grant number: P30 ES 03819.

## 1. INTRODUCTION

The health effects associated with short-term exposure to ambient ozone are well documented in air pollution studies (Galizia and Kinney, 1999; Devlin *et al.*, 1991; Folinbee *et al.*, 1988; Kinney *et al.*, 1996; Stern *et al.*, 1994; Levy *et al.*, 2001; Kelsall *et al.*, 1997; Moolgavkar, 2000; Thurston and Ito, 2001; Goldberg *et al.*, 2001). Exposure to ozone levels less than 100ppb can irritate the respiratory tract, particularly for sensitive populations, concentrations above 300ppb can cause nose and throat irritation, and levels above 1000ppb can cause breathing problems and headaches. In addition, long-term exposure to low ozone concentrations may damage sensitive areas of skin, causing premature aging of lung tissue (Turco, 1997).

Although several time series studies in single cities have shown significant positive associations between mortality/morbidity and summer ozone levels (Kinney and Ozkaynak, 1991; Kinney *et al.*, 1996; Burnett *et al.*, 1997a, 1997b 1994; Moolgavkar, 2000; Goldberg *et al.*, 2001; Kelsall *et al.*, 1997), few multi-site time series studies have been used to estimate an average short-term effect of ozone on mortality and morbidity by combining information across several locations (Stern *et al.*, 1994; Burnett *et al.*, 1997a, 1997b; Thurston and Ito, 2001). Critics of single-site studies focus on: (i) inconsistencies of findings among studies; (ii) arbitrariness in the selection of the cities to be studied; and (iii) model choices that might lead to relative rates estimates biased upward. Multi-site time series studies address these criticisms because: (i) site-specific data are assembled under a common framework; (ii) the same statistical model is fitted to all the cities; and (iii) the average effect is estimated by taking into account the within-site and across-sites variability by use of hierarchical models (Dominici, 2002).

In this article we extend previous modelling approaches developed in NMMAPS for  $PM_{10}$  (Dominici *et al.*, 2000) to take into account the following characteristics of the summer ozone data (from 1 June to 30 September). Unlike  $PM_{10}$ , summer ozone levels are measured daily, thus allowing estimation of health effects associated with exposure to ozone in the past few days instead of health effects associated with a one day impulse in air pollution concentrations. In addition, the estimated city-specific relative rates for ozone show some outliers, indicating that a normal random effects distribution might be inappropriate.

We develop Bayesian hierarchical distributed lag models for estimating the percentage increase in cardiovascular and respiratory (CVDRESP) mortality associated with short-term changes in summer ozone levels on average across 19 cities in the United States. The distributed lag models have been used in economics after Almon (Almon, 1995) popularized them in 1965. Then, Pope, Schwartz, and others (Pope and Schwartz, 1996; Zanobetti *et al.*, 2000; Schwartz, 2000) used these methods in environmental health studies. Building upon these methods, at the first stage, we estimate city-specific relative rates of CVDRESP mortality associated with summer ozone levels using a distributed lag Poisson regression model with natural cubic splines to take into account non-linear confounding by long-term trend, and weather. At the second stage, we combine the information across the 19 cities by assuming that the true city-specific effects are normally distributed with unknown mean and variance (Gelman *et al.*, 1995). In addition to the usual assumption of normality, at the second stage of the hierarchical model we also assume that the 'true' city-specific effects are distributed as a *t*-student and as a mixture of two normal distributions. Third, we assign prior distributions to all the unknown parameters. Our model specification is then applied to 19 large cities of the National Morbidity, Mortality and Air Pollution Study (NMMAPS) (Samet *et al.*, 2000a, 2000b; Dominici *et al.*, 2003) for which ozone data were available.

We choose CVRESP mortality as the outcome to capture a population subgroup at higher susceptibility than the general population (Burnett *et al.*, 1997a, 1997b, 1994; Kinney and Ozkaynak,

1992; Kinney *et al.*, 1996; Galizia and Kinney, 1999). We choose ambient summer ozone levels (from 1 June to 30 September) as the exposure, since summer ozone is highest among all seasons and measured more frequently.

Controlling for confounding bias and exploring effect modification in time series studies of air pollution and mortality is challenging (Samet *et al.*, 2000a, 2000b; Dominici *et al.*, 2004), and requires a systematic assessment of the sensitivity of findings to model specification. For example, temperature is both a potential confounder and effect modifier of the ozone-mortality association. Ozone and mortality tend to be higher at elevated temperatures, and the health effects of ozone could be exacerbated at extreme temperatures. In addition, ozone levels can be highly correlated with other pollutants, making it difficult to disentangle its effect from those of other environmental agents. Other long-term varying factors such as influenza epidemics and weather patterns are also determinants of mortality and ozone levels, and their incorrect assessment could lead to spurious associations. We systematically explore the sensitivity of the results to the adjustment for confounding bias due to long-term trends, other pollutants, and heat waves.

## 2. DATA

We used data from NMMAPS including mortality counts, weather, and air pollutant concentrations for the 19 large metropolitan areas in the U.S. from 1987 to 1994. Data are available on the internet at <http://ihapss.biostat.jhsph.edu/data/data.htm>. See also Samet *et al.* (2000a, 2000b); Dominici *et al.* (2003) for additional details.

Cause-specific mortality data, excluding non-residents and accidental deaths, were obtained from the National Center for Health Statistics. Because mortality information contained residence information at county level only to protect confidentiality, all predictor variables were aggregated at county level as well. Mortality data were classified by age ( $< 65$ , 65 to 74, and  $\geq 75$  years), and by cause according to the International Classification of Disease Ninth Revision. The CVDRESP mortality used in this analysis include the following ICD-9 codes: cardiovascular (codes 390–448), respiratory, including chronic obstructive pulmonary disease and related disorders (codes 490–496), influenza (code 487) and pneumonia (codes 480–486, 507). The study outcome is the daily counts of CVDRESP mortality from 1 June to 30 September for 1987–1994 (eight years of summer months).

Hourly temperature and dew point data for each site were obtained from the Earth-Info CD database (<http://www.earthinfo.com>). After extensive preliminary analyses that considered various daily summaries of temperature and dew point, we used 24-h averages. If a county has more than one weather station, we used the average of the measurements from all available stations.

Daily time series of ambient air pollution levels for each county were obtained from the Aerometric Information Retrieval Service (AIRS) database maintained by the U.S. Environmental Protection Agency. The daily ozone levels used in analysis for each county are obtained by the following: (i) for each day and each monitor in a county, calculating the mean ozone concentration with moving window of 365 days centering around that day, called annual mean; (ii) calculating daily residuals from the annual mean for each monitor; and (iii) to protect against outliers, calculating the 10 per cent trimmed mean of the daily residuals across all monitors in this county. Other daily pollutant data are obtained in a similar fashion. Summer daily ozone levels from 1 June to 30 September for 1987–1994 (eight years of summer months) are the exposure measures of interest.

### 3. METHODOLOGY

In this section we describe the two-stage hierarchical model used for estimating an overall relative rate of CVDRESP mortality associated with short-term summer ozone exposure. Let  $Y_t^c$  be the daily count of CVDRESP deaths, and  $x_{t-\ell}^c$  be the ambient ozone levels at lag  $\ell$  in city  $c$ . We assume that  $Y_t^c$  has a over-dispersed Poisson distribution with mean  $\mu_t^c$  and over-dispersion parameter  $\phi^c$  and that:

$$\begin{aligned} \log \mu_t^c = & \text{age-specific intercept} + \sum_{\ell=0}^6 \beta_{\ell}^c x_{t-\ell}^c \\ & + \text{long-term trend} + \text{long-term trend specific to each age group} + \text{day of week} \\ & + \text{temperature and dew point} + \text{copollutants, or heat wave} \end{aligned} \quad (1)$$

We first provide the interpretation of the parameters of interest, that is the summer ozone health effects. At the end of the section, we describe all the potential confounders and how we adjust for confounding in our model specification. Let  $\hat{\beta}^c = [\hat{\beta}_0^c, \dots, \hat{\beta}_6^c]$  be the vector of the estimated city-specific estimates of the lag coefficients, and let  $V^c$  be the corresponding statistical covariance matrix. Let  $\hat{\theta}^c = \sum_{\ell=0}^6 \hat{\beta}_{\ell}^c$  be the estimated city-specific cumulative effect and let  $\hat{v}^c$  be the corresponding statistical variance calculated as  $\mathbf{1}' \hat{V}^c \mathbf{1}$ , where  $\mathbf{1}$  is a vector of ones. We assume:

$$\begin{aligned} \hat{\theta}^c & | \theta^c, v^c \sim N(\theta^c, v^c) \\ \theta^c & | \mu, \tau^2 \sim g(\mu, \tau^2) \end{aligned} \quad (2)$$

The parameter  $\theta^c = \sum_{\ell=0}^6 \beta_{\ell}^c$  denotes the true city-specific cumulative log-relative rate of CVDRESP mortality associated with a 10ppb increase in summer ozone level over the previous week. The detailed explanation of this interpretation is listed in Appendix 1. As an alternative, we could specify a multivariate normal distribution for the vector of parameters  $\beta^c = (\beta_0^c, \dots, \beta_6^c)$  at the second stage, and then calculate  $\theta^c = \sum_{\ell=0}^6 \beta_{\ell}^c$ , rather than modelling  $\theta^c$  directly. However, the high collinearity of the ozone levels at different lags creates numerical problems in the lag-specific effect estimations.

The parameters  $\mu$  and  $\tau^2$  denote the true average cumulative log-relative rate across the 19 cities and the between-city variance of the  $\theta^c$ s. Our goal is to approximate the marginal posterior distributions of  $\mu$  and  $\tau^2$ .

We perform the calculations under the following assumptions for the random effects distribution  $g$ :

1.  $g$  is normal with mean  $\mu$  and variance  $\tau^2$
2.  $g$  is a  $t$ -student with 3 degrees of freedom and with mean  $\mu$  and variance  $\tau^2$
3.  $g$  is a mixture of two normals with unknown weights, that is:  
 $g(\mu, \tau^2) = (1 - p) N(\lambda_1, \tau^2) + p N(\lambda_2, \tau^2)$ . Then  $\mu = (1 - p)\lambda_1 + p\lambda_2$ .

We fit model (2) by use of Monte Carlo Markov chain methods (MCMC) (Gelman *et al.*, 1995) implemented by use of Winbugs (Spiegelhalter *et al.*, 2000). As prior distributions we use the ‘default non-informative priors’: that is  $\mu \sim N(0, 10000)$  and  $1/\tau^2 \sim \text{Gamma}(0.01, 0.01)$ . In the next section, we report the sensitivity of the posterior inference on  $\mu$  to the specification of the prior distribution for  $\tau^2$  and to the choice of the prior hyperparameters. Winbugs software is detailed in Appendix 2.

Day-to-day associations between air pollution levels and mortality counts can be confounded by time-varying factors that vary in a similar manner as pollution and mortality. Below are listed the potential confounding factors for the associations between summer ozone and CVDRESP mortality, and how they are included into the regression model to adjust relative rates estimates for confounding bias.

**Age group:** We have adjusted for age by adding age-specific intercepts which allow for age-specific baseline mortality rate. The variable age is a categorical variable with 3 levels: < 65, 65 to 74, and  $\geq 75$  years.

**Yearly trend:** The number of deaths for cardiovascular and respiratory mortality and ozone levels have both declined in the past few years because of improvements in medical practice and air quality regulations. We have employed an 8 degree of freedom piecewise-cubic spline over 8 years (with 4 summer months in each year and knots at the separation of each summer) to remove the yearly trend. This form of smoothing spline adjustment will allow for increases or decreases over time within each summer, which is a more flexible way of time adjustment than using yearly dummies. In addition, to further remove trends in the summer mortality time series specific to each age group, we add interaction terms between smooth functions (natural cubic spline) of time with 2.4 degrees of freedom, and the second and third of the three age groups.

**Day of the week:** Day of the week is a surrogate for traffic patterns and daily activity. In addition there is evidence that the daily number of deaths and pollution levels are different for weekdays versus weekends. We adjust for the potential confounding effect of day of the week by including the model a categorical variable that allows for different baseline mortality counts on the different days of the week.

**Temperature and dew point temperature:** Since the chemistry of ozone formation is temperature and sunlight driven (Seinfeld and Pandis, 1998), ozone levels are correlated with temperature. In addition, epidemiological studies (Curriero *et al.*, 2002) have reported associations between summer temperatures and mortality. To account for potential confounding by weather, we included in the model natural cubic splines of daily temperature, dew point temperature, and adjusted average temperatures and dew point temperatures over the three previous days (both adjusted for current-day temperature and dew point) with six degrees of freedom.

**Heat waves:** Correlations between ozone and temperature and between mortality and temperature can be even higher for a subset of days with extreme temperatures (heat waves). We further explored the sensitivity of the short-term effects of summer ozone on CVDRESP mortality to the potential confounding of heat-waves. We modelled a heat-wave as a natural cubic spline of the interaction between current temperature and average temperature in the three previous days with 2 degrees of freedom.

**Other pollutants:** Because previous studies have found evidence of a short-term association between PM<sub>10</sub> and mortality (Samet *et al.*, 2000a, 2000b; Dominici *et al.*, 2003), and because both PM<sub>10</sub> and O<sub>3</sub> levels are correlated with ultra-fine particles, PM<sub>10</sub> might be a confounder of the ozone-mortality association. Similar arguments apply to other pollutants, including carbon monoxide (CO), sulfur dioxide (SO<sub>2</sub>), and nitrogen dioxide (NO<sub>2</sub>). Sensitivity analyses of the short-term effects of summer ozone on CVDRESP mortality to inclusion of copollutants were performed. To control for their potential confounding effect, we fit several multi-pollutant models within each city having as predictors the current day co-pollutant concentrations. Notice that, since PM<sub>10</sub> measurements are available only every 1 to 6 days, a distributed lag model is not appropriate to examine the potential confounding effect of PM<sub>10</sub>. Therefore we explore the sensitivity of the ozone health effects to the inclusion of PM<sub>10</sub> by use of a single-lag model (we use ambient ozone exposure at lag 2). Note that only the days with available PM<sub>10</sub> measures are included in this analysis.

In summary, this article introduces the modelling approach with a focus on summer data and explores the sensitivity of the results to model choices (adjustment for confounding, random effects distributions, and prior hyperparameters). The design of the sensitivity of results to random effects

distributions and prior hyperparameters is listed at the beginning of this section. Other sensitivity studies include: (i) lag structure for ozone exposure: distributed lag model versus single lag models where we included only one air pollution predictor at lags 0, 1 or 2; (ii) degree of adjustment for long-term trends: monitor the changes of model inferences to various degrees of freedoms in the natural cubic splines of time, ranging from 4 to 16 with step = 2; (iii) inclusion of other pollutants in the model; and (iv) heat waves.

#### 4. RESULTS

Figure 1 shows the box-plot of summer ozone levels in the 19 large U.S. cities. Daily levels of ambient ozone vary from 0 to 100ppb. Figure 2 shows the box-plot of the logarithm of the daily counts of CVDRESP mortality in the 19 large cities. The CVDRESP mortality counts in Los Angeles, New York and Chicago are much larger than others. Figure 3 shows city-specific estimates and 95 per cent posterior regions of the percentage increases CVDRESP mortality associated with a 10ppb increase in summer ozone levels at lags 0, 1, 2, and under a distributed lag model (cumulative effect in the last week). At the far right are shown the posterior means and 95 per cent posterior intervals of the corresponding overall relative rates. City-specific and overall relative rates estimates under a distributed lag model and their 95 per cent posterior regions are also reported in Table 1. Figure 4 (a and b) shows the posterior distributions of the average relative rate ( $\mu$ ) and of the between-city standard deviation ( $\tau$ ).

On average across the 19 cities, we found strong evidence of an association between summer ozone levels and CVDRESP mortality under a distributed lag model, and at lags 0, 1 and 2. For example, we

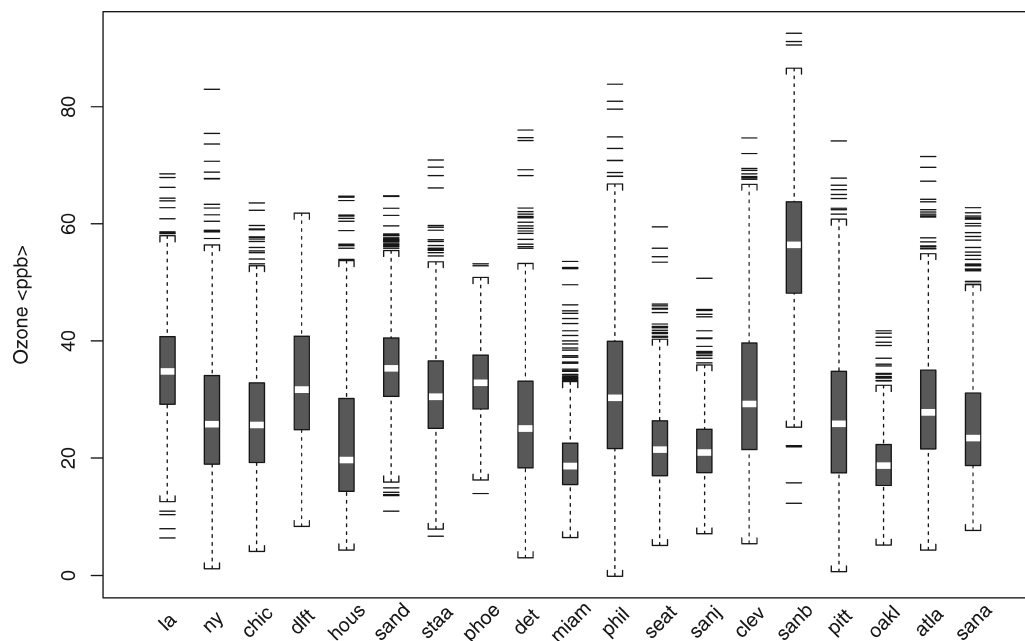


Figure 1. Box-plots of summer ozone levels for 19 large U.S. cities in NMMAPS for the period 1987–1994

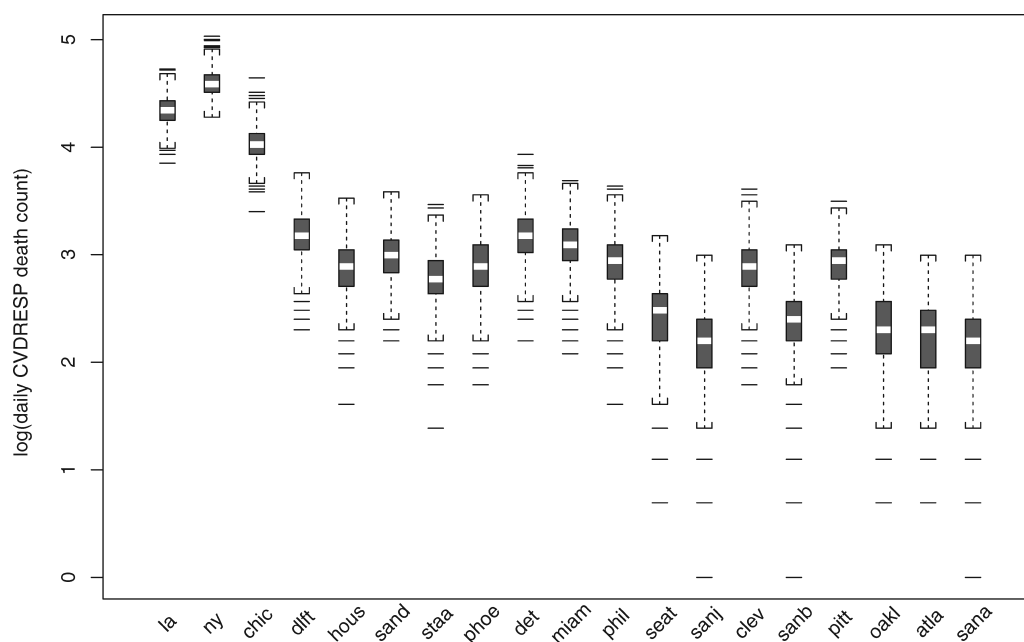


Figure 2. Box-plots of log(daily CVDRESP death count) for 19 large U.S. cities in NMMAPS in summer months of 1987–1994

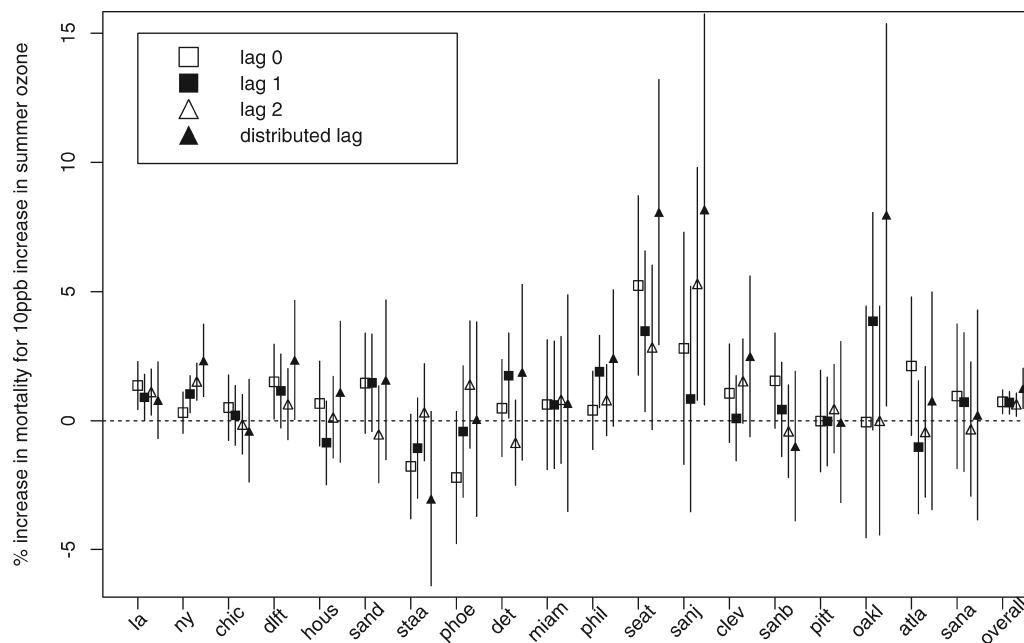


Figure 3. City-specific relative rates estimates (percent increases in CVDRESP mortality associated with 10ppb increase in ozone) and their 95 per cent confidence intervals, with overall ozone effect and its 95 per cent posterior regions at the end. These effects are estimated by use of a distributed lag model for one week and at lags 0, 1 and 2 days



Table 1. City-specific estimates of the percentage increase in CVDRESP mortality for a 10ppb increase in summer ozone levels over the previous week obtained under a distributed lag model (95 per cent confidence intervals). Last row denotes the posterior mean and the 95 per cent posterior regions of the overall effect

Los Angeles	0.79	(−0.69, 2.28)
New York	2.33	(0.93, 3.73)
Chicago	−0.39	(−2.38, 1.60)
Dallas/Fort Worth	2.35	(0.05, 4.66)
Houston	1.11	(−1.62, 3.84)
San Diego	1.58	(−1.50, 4.67)
Santa Ana/Anaheim	−3.03	(−6.40, 0.35)
Phoenix	0.05	(−3.71, 3.81)
Detroit	1.88	(−1.52, 5.27)
Miami	0.68	(−3.51, 4.86)
Philadelphia	2.42	(−0.22, 5.07)
Seattle	8.08	(2.94, 13.21)
San Jose	8.18	(0.61, 15.75)
Cleveland	2.49	(−0.62, 5.60)
San Bernardino	−0.98	(−3.87, 1.91)
Pittsburgh	−0.05	(−3.17, 3.07)
Oakland	7.97	(0.58, 15.36)
Atlanta	0.77	(−3.45, 4.99)
San Antonio	0.22	(−3.84, 4.28)
Overall	1.25	(0.47, 2.03)

found that, on average across the 19 large cities, a 10ppb increase in summer ozone levels over the previous week is associated with a 1.25 per cent increase in CVDRESP mortality (95 per cent posterior regions: 0.47 per cent, 2.02 per cent). The average effects for ozone exposure at lags 0, 1, and 2 are 0.73 (95 per cent posterior regions: 0.27, 1.19), 0.70 (95 per cent posterior regions 0.26, 1.12) and 0.64 (95 per cent posterior regions 0.17, 1.07), respectively.

Notice that in several cities the estimated effects show monotonicity, with the smallest effect at lag 0 and the largest effect corresponding to the cumulative relative rate under the distributed lag model. This monotonicity suggests a consistency of findings across lag specifications with associations between ozone and mortality that persist for lags of at least 6 days. Posterior means of the heterogeneity parameter  $\tau$  are similar across lag specifications with wider posterior intervals for the distributed lag model.

Table 2 summarizes the posterior means and 95 per cent posterior regions of the overall relative rates  $\mu$  with respect to: (i) distributional assumptions for the random effects distribution (normal,  $t$ -student, and mixture of two normal); (ii) prior distributions on the heterogeneity variance  $\tau^2$  (Inverse Gamma, Jeffrey prior, uniform on the shrinkage factor); and (iii) choice of the prior hyperparameters. Posterior inference on  $\mu$  are very robust to these model choices.

Figure 5 shows the robustness of the city-specific estimates to the inclusion of other pollutants in the model. At each panel, two sets of city-specific estimates for ozone at lag 2 are plotted against each other. On the  $x$ -axes, the estimates do not include any adjustment for other pollutants, whereas the estimates on the  $y$ -axes are adjusted for one additional pollutant. The cross mark corresponds to the overall estimates and their 95 per cent posterior regions. Notice that, except when  $\text{PM}_{10}$  is included in the model, the two sets of estimates and their averages are very close to the diagonal line, indicating very little sensitivity of the ozone effect to the inclusion of gaseous pollutants into the model.



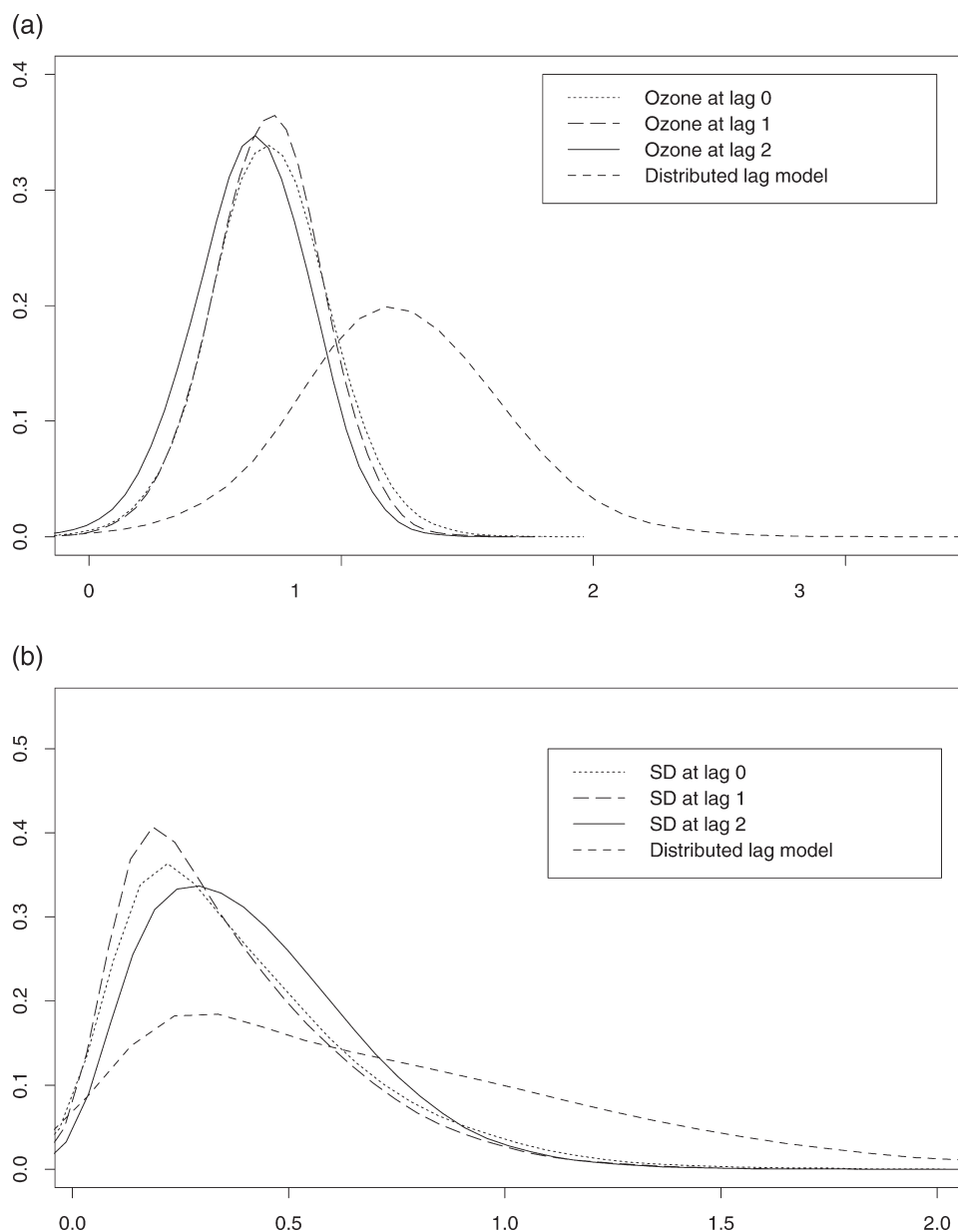


Figure 4. (a) Posterior distribution of average relative rate ( $\mu$ ) and (b) of the heterogeneity parameter ( $\tau$ ) obtained by combining city-specific relative rates estimates under a distributed lag model and single-lag models

City-specific ozone estimates vary with the adjustment for  $PM_{10}$ ; however, on average the ozone effect estimate remains positive and significant. Since  $PM_{10}$  measurements are available only every 1 to 6 days and only the days with available  $PM_{10}$  measures are included in the analysis, the sensitivity of the ozone health effects to the inclusion of  $PM_{10}$  in the model might indicate a true confounding effect or

Table 2. Posterior mean and 95 per cent posterior region of the overall effect  $\mu$  under alternative specifications for the random effects distributions

Random effect distribution	National average estimate
Normal & $1/\tau^2 \sim G(0.01, 0.01)$	1.25 (0.47, 2.03)
Normal & $1/\tau^2 \sim G(0.001, 0.001)$	1.20 (0.47, 1.93)
Normal & $1/\tau^2 \sim G(0.0001, 0.0001)$	1.20 (0.51, 1.89)
<i>t</i> -student with 3 df & $1/\tau^2 \sim G(0.001, 0.001)$	1.26 (0.50, 2.02)
Mixture of two normals & $1/\tau^2 \sim G(0.001, 0.001)$	1.33 (0.51, 2.15)
Normal & $1/\tau^2 \propto \text{constant}$	1.29 (0.35, 2.23)
Normal & $1/\tau^2 / (1/\tau^2 + 1/\nu^c) \sim U[0, 1]$	1.27 (0.37, 2.17)

might be attributed to higher statistical uncertainty due to the lack of daily PM<sub>10</sub> measurements. The average relative rates under the several bi-pollutant models are summarized in Table 3.

To assess the confounding effect of heat waves, we fit the distributed lag model within each city by including a smooth function of interaction between current day temperature and average temperature in the last three days. The average relative rate is robust to the inclusion of this interaction term and moved from 1.25, 95 per cent posterior regions (0.47, 2.03) to 1.11, 95 per cent posterior regions (0.38, 1.86), as listed in Table 4.

Figure 6 shows the robustness of ozone effects at lag 2 to the degrees of freedom in the smooth functions of time. More specifically, the midline of each box denotes the relative rate estimates as we vary the number of degrees of freedom in the smooth function of time from 4 to 16 with increasing step = 2, and the polygon captures the corresponding confidence intervals. The polygon at the far right shows the posterior means of the overall relative rates and their 95 per cent posterior regions. The picture shows that the average relative rate estimates are somewhat sensitive to the selection of the number of degrees of freedom in the smooth function of time.

In summary, we found strong evidence of associations between short-term summer exposure to ozone and CVDRESP mortality on average across 19 large cities in the U.S. We found that this association persists at different lags and it is robust to the adjustment for long-term trend, temperature, heat waves, and other pollutants.

## 5. DISCUSSION

In this article, we extend modelling approaches developed as part of the NMMAPS by introducing Bayesian hierarchical distributed lag models with non-normal random effects distributions for estimating associations between short-term variations in ozone and CVDRESP mortality. On average across the 19 large cities in the U.S., we found that a 10ppb increase in summer ozone levels over the previous week is associated with a 1.25 per cent increase in CVDRESP mortality (95 per cent posterior region: 0.47, 2.03). Results were robust to: (i) distributional assumptions on the random effects distribution; (ii) inclusion of other gaseous pollutants (CO, SO<sub>2</sub> and NO<sub>2</sub>) in the model; (iii) potential confounding effects of long-term trend, temperatures, and heat waves; (iv) lag specification; and (v) prior distributions.

Our methods have recently been applied to a more extensive NMMAPS analysis on ozone and mortality, which includes 95 large cities in the U.S. for the period 1987–2000 (Bell *et al.*, 2004). While this article focuses on summer mortality, Bell *et al.* (2004) focuses on yearly mortality. These more extensive ozone analyses found a strong association between short-term variations in ozone and mortality. In fact, Bell *et al.* (2004) report a national average equal to 0.64 per cent (95 per cent

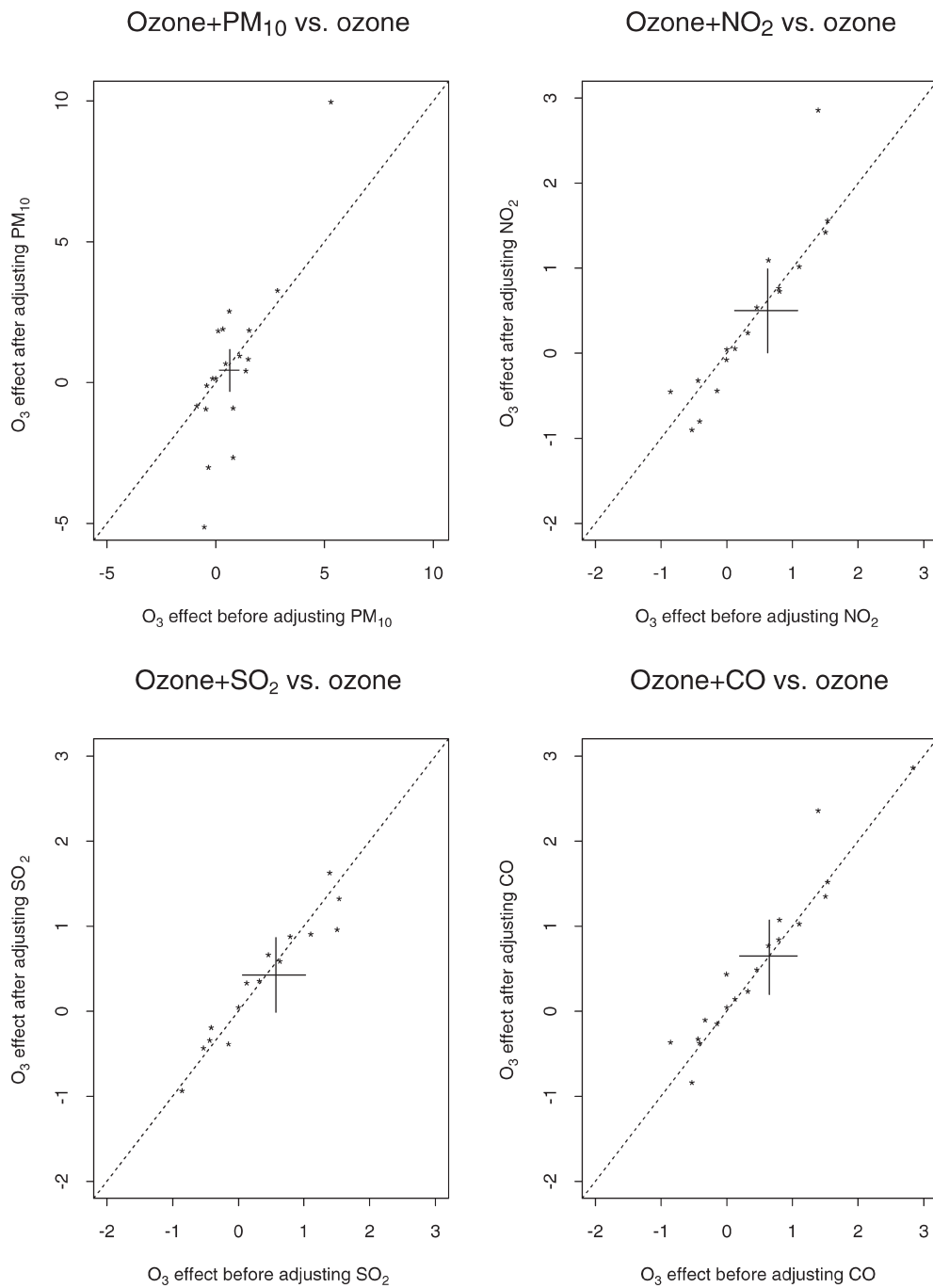


Figure 5. City-specific ozone relative rates estimates at lag 2 adjusted by another pollutant (PM<sub>10</sub>, NO<sub>2</sub>, SO<sub>2</sub> or CO), plotted against ozone relative rates estimates at lag 2. The cross is plotted corresponding to the two overall relative rates (adjusted versus unadjusted) and their 95 per cent posterior regions

Table 3. Percentage increase in CVDRESP mortality associated with a 10ppb increase in ozone, adjusting for co-pollutants (95 per cent posterior regions)

O <sub>3</sub> effect	O <sub>3</sub>	O <sub>3</sub> + PM <sub>10</sub>	O <sub>3</sub> + NO <sub>2</sub>	O <sub>3</sub> + SO <sub>2</sub>	O <sub>3</sub> + CO
O <sub>3</sub> lag 0	0.73 (0.27, 1.19)	0.74 (−0.33, 1.72)	0.60 (0.11, 1.10)	0.51 (0.01, 1.03)	0.69 (0.20, 1.18)
O <sub>3</sub> lag 1	0.70 (0.26, 1.12)	0.69 (−0.01, 1.38)	0.63 (0.15, 1.12)	0.49 (0.06, 0.95)	0.65 (0.21, 1.08)
O <sub>3</sub> lag 2	0.64 (0.17, 1.07)	0.46 (−0.32, 1.17)	0.50 (0.01, 0.99)	0.42 (−0.01, 0.86)	0.65 (0.20, 1.07)

*Note:* All pollutants and ozone are at the same lag structure. The availability of PM<sub>10</sub> measurements is only every 1 to 6 days; Seattle and San Antonio do not have any NO<sub>2</sub> measurements; Miami, Seattle, San Jose, Oakland and San Antonio do not have any SO<sub>2</sub> measurements

Table 4. Percentage increase in CVDRESP mortality associated with a 10ppb increase in ozone, adjusting for heat wave (95 per cent posterior regions)

O <sub>3</sub> effect	O <sub>3</sub>	O <sub>3</sub> + heat wave
Distributed lag (DL)	1.25 (0.47, 2.03)	1.11 (0.38, 1.86)

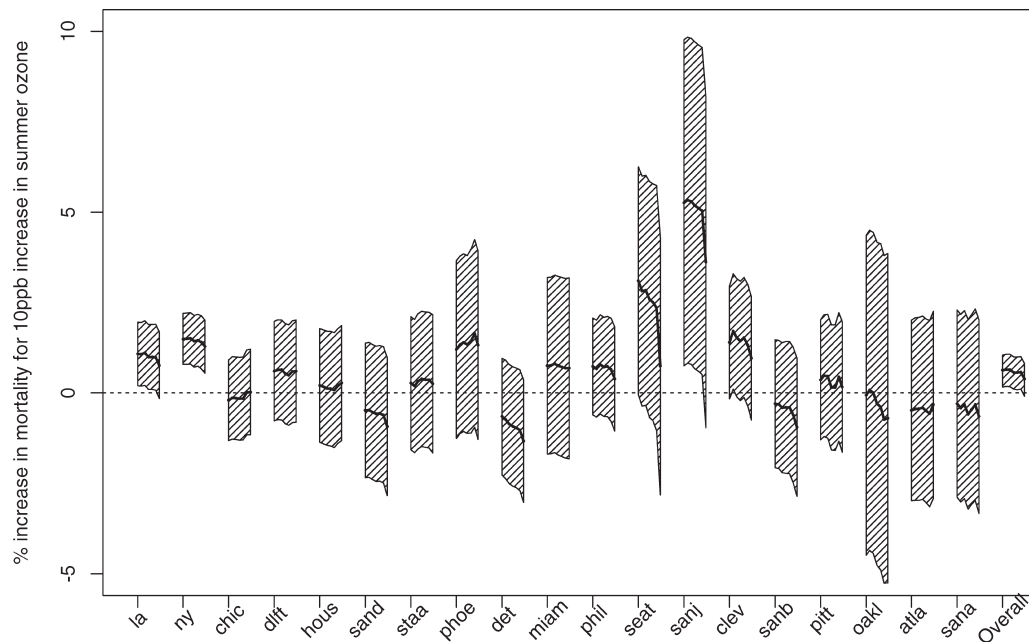


Figure 6. Sensitivity analysis of the city-specific effects at lag 2 with respect to the degrees of freedom in the smooth functions of time. The number of degrees of freedom vary from 4 to 16 with increasing step = 2. Each polygon captures the change in the city-specific relative rates estimates with their corresponding 95 per cent confidence intervals. The last polygon captures the change of overall ozone effect estimates with its 95 per cent posterior regions

posterior region: 0.31, 0.98) in CVDRESP mortality for a 10ppb increase in the preceding week's ozone levels, using a constrained distributed lag model.

Our results provide additional evidence in support of an association between ozone and mortality and are consistent with the literature. Epidemiological studies have found associations between low

ozone levels and: (i) decline in pulmonary functions (Galizia and Kinney, 1999; Folinbee *et al.*, 1988; Kinney *et al.*, 1996; Devlin *et al.*, 1991); (ii) increase in hospital admissions for asthma, especially among children (Stern *et al.*, 1994; Galizia and Kinney, 1999); and (iii) congestive heart failure among the elderly (Burnett *et al.*, 1997b). Positive associations between short-term 1-h maximum ozone levels and mortality, adjusted for the potential confounding effects of other pollutants and weather, have been found in Los Angeles and New York (Kinney and Ozkaynak, 1991; Kinney *et al.*, 1996). In addition, a multi-site time-series study of 16 Canadian cities showed a significant positive association between respiratory admissions and summer daily 1-h maximum ozone levels at lag 1 (Burnett *et al.*, 1997a).

The strengths of our approach are: (i) the use of routinely collected data on air pollution, mortality and weather; (ii) the implementation of Bayesian hierarchical models for combining information across locations and for estimating the average relative rate accounting for within-city statistical uncertainty and heterogeneity of the effects across locations; and (iii) the systematic assessment of the sensitivity of findings to adjustment for confounding bias and model choice.

This analysis has the following limitations. First, the 24-h daily average of hourly ozone measurements is used as the exposure variable. Because of the volatile nature of ozone, with the maximum generally appearing in the early afternoon and lowest levels at night, the 24-h average might not be an adequate summary measure of the ozone exposure. Therefore additional work is needed to explore the sensitivity of our results with respect to other summary measures, such as the daily 1-h and 8-h maxima. Bell *et al.* (2004) investigated multiple metrics of ozone concentration, and found all to be significantly associated with mortality. Second, we use ambient levels of ozone as a surrogate for exposure without taking into account potential exposure measurement error. Third, our model-based methods for adjusting for confounders, although extensive, cannot completely rule out confounding bias in the estimated associations. In this article we used 1 degree of freedom per summer to adjust for trend. Therefore the lack of a more severe adjustment for long-term trend might have biased some of our relative rate estimates. However, when we repeated the analyses by using 2 degrees of freedom per summer, we still found a positive association between ozone and mortality with a national average estimate equal to 0.93 per cent (95 per cent posterior region:  $(-0.06, 1.99)$ ). Unfortunately, no data-driven method exists to estimate an optimal number of degrees of freedom for removing confounding due to long-term trends. This issue has been discussed extensively by Dominici *et al.* (2004). Fourth, the association between ozone and mortality might also be confounded by other pollutants which are not currently measured in NMMAPS. For example, ozone contributes to the secondary formation of organic aerosols, which are a large source of ultra-fine particles (Andreae and Crutzen, 1997; Seinfeld and Flagan, 1997).

These results provide evidence to policy-makers on the harmful effects of ozone on human health. Such epidemiological evidence is central to the regulatory process, including the setting of the National Ambient Air Quality Standards (NAAQS) for ozone. Regulations are designed to protect human health with an adequate margin of safety for susceptible populations, such as those with cardiovascular or respiratory disease. This research extends earlier time-series analysis and indicates that summer ozone levels adversely impact cardiovascular and respiratory mortality.

#### ACKNOWLEDGEMENTS

Funding for Michelle Bell and Francesca Dominici was provided by a grant from the Environmental Protection Agency (EPA 3D-6867-NAEX). Funding for Francesca Dominici was also provided by NIEHS RO1 grant (ES012054-01) and by the NIEHS Center in Urban Environmental Health (P30 ES 03819).

## REFERENCES

- Almon S. 1965. The distributed lag between capital appropriations and expenditures. *Econometrica* **33**(1): 178–196.
- Andreae M, Crutzen P. 1997. Atmospheric aerosols: biogeochemical sources and role in atmospheric chemistry. *Science* **276**: 1052–1058.
- Bell M, McDermott A, Zeger S, Samet J, Dominici F. 2004. Ozone and mortality in 95 U.S. urban communities from 1987 to 2000. *Journal of American Medical Association* **292**(19): 2372–2378.
- Burnett R, Dales R, Raizenne M, Krewski D, Summers P, Roberts G, Raad-Young M, Dann T, Brook J. 1994. Effects of low ambient levels of ozone and sulfates on frequency of respiratory admissions to Ontario hospitals. *Environmental Research* **65**: 172–194.
- Burnett R, Brook J, Yung W, Dales R, Krewski, D. 1997a. Association between ozone and hospitalization for respiratory diseases in 16 Canadian cities. *Environmental Research* **72**: 24–31.
- Burnett R, Dales R, Brook J, Raizenne M, Kreswski, D. 1997b. Association between ambient carbon monoxide levels and hospitalizations for congestive heart failure in the elderly in 10 Canadian cities. *Epidemiology* **8**: 162–167.
- Curriero F, Heiner K, Samet J, Zeger S, Strug L, Patz J. 2002. Temperature and mortality in 11 cities of the Eastern United States. *American Journal of Epidemiology* **155**: 80–87.
- Devlin R, McDonnell W, Mann R, Becker S, House D, Schreinemachers D, Koren H. 1991. Exposure of humans to ambient levels of ozone for 6.6 hours causes cellular and biochemical changes in the lung. *American Journal of Respiratory Cell and Molecular Biology* **4**: 72–81.
- Dominici F. 2002. Air pollution and health: what can we learn from an hierarchical approach? Invited commentary. *American Journal of Epidemiology* **1**: 11–15.
- Dominici F, Samet J, Zeger S. 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* **163**: 263–302.
- Dominici F, McDermott A, Zeger S, Samet J. 2003. National maps of the effects of PM on mortality: exploring geographical variation. *Environmental Health Perspect* **111**: 39–43.
- Dominici F, McDermott A, Hastie T. 2004. Improved semi-parametric time series models of air pollution and mortality. *Journal of the American Statistical Association*, to be published.
- Folinbee L, McDonnell W, Horstman D. 1988. Pulmonary function and symptom responses after 6.6-hour exposure to 0.12 ppm ozone with moderate exercise. *Journal of Air Pollution Control Association* **38**: 28–35.
- Galizia, A, Kinney P. 1999. Long-term residence in areas of high ozone: associations with respiratory health in a nationwide sample of nonsmoking young adults. *Environmental Health Perspect* **107**: 675–679.
- Gelman A, Carlin J, Stern H, Rubin D. 1995. *Bayesian Data Analysis*. Chapman & Hall.
- Goldberg MS, Burnett RT, Brook J, Bailar JC, Valois MF, Vincent R. 2001. Associations between daily cause specific mortality and concentrations of ground-level ozone in Montreal, Quebec. *American Journal of Epidemiology* **154**: 817–826.
- Kelsall J, Samet J, Zeger S, Xu J. 1997. Air pollution and mortality in Philadelphia, 1974–1988. *American Journal of Epidemiology* **146**: 750–762.
- Kinney P, Ozkaynak H. 1991. Associations of daily mortality and air pollution in Los Angeles county. *Environmental Research* **54**: 99–120.
- Kinney P, Ozkaynak H. 1992. Associations between ozone and daily mortality in Los Angeles and New York city. *American Review Respiratory Disease* **145**: A95.
- Kinney P, Thurston G, Raizenne M. 1996. The effects of ambient ozone on lung functions in children: a reanalysis of six summer camp studies. *Environmental Health Perspect* **104**: 170–174.
- Levy J, Carrothers T, Tuomisto J, Hammitt J, Evans J. 2001. Assessing the public health benefits of reduced ozone concentrations. *Environmental Health Perspect* **109**: 1215–1226.
- Moolgavkar S. 2000. Air pollution and daily mortality in three U.S. counties. *Environmental Health Perspect* **108**: 777–784.
- Pope CA, Schwartz J. 1996. Time series for the analysis of pulmonary health data. *American Journal of Respiratory and Critical Care Medicine* **154**: 229–233.
- Samet J, Dominici F, Zeger S, Schwartz J, Dockery, D. 2000a. *The National Morbidity, Mortality, and Air Pollution Study—Part I: Methods and Methodologic Issues*. Health Effects Institute.
- Samet J, Zeger S, Dominici F, Curriero F, Coursac I, Dockery DW, Schwartz J, 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.
- Schwartz J. 2000. The distributed lag between air pollution and daily deaths. *Epidemiology* **11**: 320–326.
- Seinfeld J, Flagan R. 1997. 1997 progress report: atmospheric transformation of volatile organic compounds: gas-phase photo-oxidation and gas-to-particle conversion. URL: <http://es.epa.gov/ncercqa/progress/centers/airborne/seinfeld97.html>.
- Seinfeld J, Pandis S. 1998. *Atmospheric Chemistry and Physics, from Air Pollution to Climate Change*. John Wiley & Sons, Inc.
- Spiegelhalter D, Thomas A, Best N. 2000. *Winbugs Version 1.3*. Software.
- Stern B, Raizenne M, Burnett R, Jones L, Kearney J, Franklin C. 1994. Air pollution and childhood respiratory health exposure to sulfate and ozone in 10 Canadian rural communities. *Environmental Research* **66**: 125–142.

- Thurston G, Ito K. 2001. Epidemiological studies of acute ozone exposures and mortality. *Journal of Exposure Analysis and Environmental Epidemiology* **11**: 286–294.
- Turco R. 1997. *Earth Under Siege*. Oxford University Press.
- Zanobetti A, Wand MP, Schwartz J, Ryan LM. 2000. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics* **1**(3): 279–292.
- Zeghnoun A, Czernichow P, Beaudou P, Hautemanire A, Froment L, Le Tertre A, Qunel P. 2001. Short-term effects of air pollution on mortality in the cities of Rouen and Le Havre, France, 1990–1995. *Archives of Environmental Health* **56**(4): 327–335.

## APPENDIX 1

To illustrate why the log-relative rate  $\theta^c$  can be interpreted as the percentage increase in CVDRESP associated with a 10ppb increase in summer ozone levels over the previous week, we consider the following:

$$\begin{aligned}\log \mu_t^c &= \sum_{l=0}^6 \beta_l^c x_{t-l}^c + \text{all other covariates} \\ \log \mu_t^{c*} &= \sum_{l=0}^6 \beta_l^c (x_{t-l}^c + 1) + \text{all other covariates} \\ \log \mu_t^{c*} - \log \mu_t^c &= \log \frac{\mu_t^{c*}}{\mu_t^c} = \sum_{l=0}^6 \beta_l^c = \theta^c \\ \frac{\mu_t^{c*}}{\mu_t^c} &= e^{\theta^c} \\ &\approx 1 + \theta^c, \text{ 1st-order approximation using Taylor expansion} \\ \frac{\mu_t^{c*} - \mu_t^c}{\mu_t^c} &\approx \theta^c \\ 100 \times \frac{\mu_t^{c*} - \mu_t^c}{\mu_t^c} &\approx (1000 \times \theta^c)/10\end{aligned}$$

In summary,  $1000 \times \theta^c$ , which is consistently reported in the article, indicates the percentage increase in CVDRESP mortality associated with 10ppb increasing in summer ozone levels every day for the previous week. This interpretation used in this article is consistent with other time series studies published in the literature (Moolgavkar, 2000; Zeghnoun *et al.*, 1990; Goldberg *et al.*, 2001).

## APPENDIX 2

```
model {
  for (i in 1:N) {
    ozone[i] <- -y[i,1]
    sigma[i] <- -y[i,2]
    prec[i] <- -1/(sigma[i]*sigma[i])
  }
}
```



```

        ozone[i]~dnorm(beta[i],prec[i])
        beta[i]~dnorm(mu,tau)
    }

    mu~dnorm(0,0.0001)
    tau~dgamma(0.01,0.01)
    sig.0<-1/sqrt(tau)
}
##mixture of two normal distributions
model { for (i in 1:N) {
    ozone[i]<-y[i,1]
    sigma[i]<-y[i,2]
    prec[i]<-1/(sigma[i]*sigma[i])
    ozone[i]~dnorm(beta[i],prec[i])
    beta[i]~dnorm(mu[i], tau)
    mu[i]<-lambda[T[i]]
    T[i]~dcat(P[])
}
P[1:2]~ddirch(alpha[])
theta~dnorm(4, 1.0E-6)I(0.0,)
lambda[2]<-lambda[1]+theta
lambda[1]~dnorm(-2, 1.0E-6)
tau~dgamma(0.001, 0.001)
sig.0<-1/sqrt(tau)
tot<-P[1]*lambda[1]+P[2]*lambda[2]
}

```