



## Invited Commentary: Timescale-dependent Mortality Effects of Air Pollution

Richard L. Smith

From the Department of Statistics, University of North Carolina, Chapel Hill, NC.

Received for publication January 10, 2002; accepted for publication March 13, 2002.

Abbreviations: NMMAPS, National Morbidity, Mortality, and Air Pollution Study;  $PM_{10}$ , particulate matter with an aerodynamic diameter  $\leq 10 \mu m$ ; TSP, total suspended particulates.

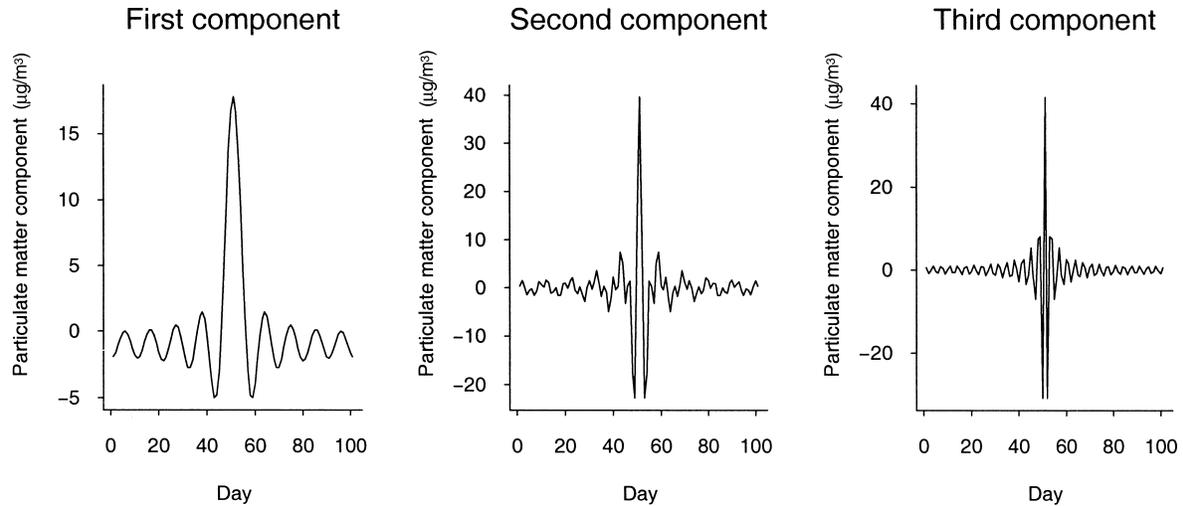
The concept of “harvesting,” by whatever name, dates back to the early days of research on the health effects of air pollution. For example, Martin (1), summarizing the results of several London smog studies from the 1950s, noted the short timescale of response to high-air-pollution events. Referring specifically to the December 1952 event, Martin noted that “no cases of sudden death were found which could not have been explained by previous respiratory or cardiovascular lesions” (1, p. 969). Later in the same article, the author stated, “[S]ome of [those who died] might in the natural course have died within a few days. This hypothesis is supported by the compensatory fall of the mortality curve to subnormal levels often noticed after a pollution incident” (1, p. 974). Evidently, Martin was talking about what we now call harvesting, but in a context where both the air pollution and the consequent mortality were much higher than in modern US and European cities. The question is whether similar effects persist under present-day conditions. The question is important, as I perceive it, not because proof of a harvesting effect would in any way lessen the need to take action against air pollution, but because such an effect, if present, would help us better understand the true meaning of studies such as the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), and could help us determine which pollution control or other kinds of public health strategies are most likely to be effective in reducing deaths due to air pollution.

Numerous investigators have looked for a harvesting (or mortality displacement) effect, using a variety of methodologies, but without finding any convincing evidence in favor of this hypothesis. Zeger et al. (2) and Smith et al. (3) independently proposed a compartment model in which people pass from a “healthy” state to a “frail” state in which they eventually die, but this model, even though it is almost certainly too simple to represent any true harvesting phenomenon, is extremely hard to fit. Smith et al. (3)

proposed a Markov chain Monte Carlo estimation scheme that appeared to show evidence of harvesting in Chicago, Illinois, but when the same model was fitted to simulated data series without harvesting, in 19 out of 20 cases the method appeared to provide evidence in favor of harvesting. Thus, the Markov chain Monte Carlo method does not seem to be effective in discriminating between a harvesting model and a standard regression model, and other authors have rejected this kind of approach in favor of one that examines timescale effects. Zeger et al. (2) proposed an approach based on frequency decompositions of both the mortality and air pollution series, while Schwartz (4) used nonparametric smoothing techniques to decompose the series into long, intermediate, and short timescale series and then looked for pollution-mortality relations in the intermediate timescale series. Both Zeger et al. and Schwartz concluded that air pollution effects exist on a timescale of weeks rather than days and argued that this was inconsistent with a simple harvesting effect.

In the current issue of the *Journal*, Dominici et al. (5) develop a new approach based on a spectral decomposition of only the pollution series, where different components correspond to effects at different timescales. This is simpler both conceptually and for practical implementation than the procedure of Zeger et al. (2). Dominici et al. have helpfully provided software with which to perform the decomposition, and the rest of the analysis uses Poisson regression techniques that are by now totally standard in this field. They have applied this method to four of the series in the NMMAPS database—the only four for which daily readings of particulate matter with an aerodynamic diameter of  $10 \mu m$  or less ( $PM_{10}$ ) are available—as well as some older data from Philadelphia, Pennsylvania. They conclude that the strongest effects are those associated with timescales on the order of 1–2 months. This reinforces the conclusions of the earlier

Correspondence to Dr. Richard Smith, Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260 (e-mail: rls@email.unc.edu).



**FIGURE 1.** Timescale decomposition for a 101-day series of particulate matter measurements in which the level of particulate matter with an aerodynamic diameter  $\leq 10 \mu\text{m}$  is  $100 \mu\text{g}/\text{m}^3$  on day 51 and 0 on every other day.

papers and again seems to contradict the notion of a short-term harvesting effect.

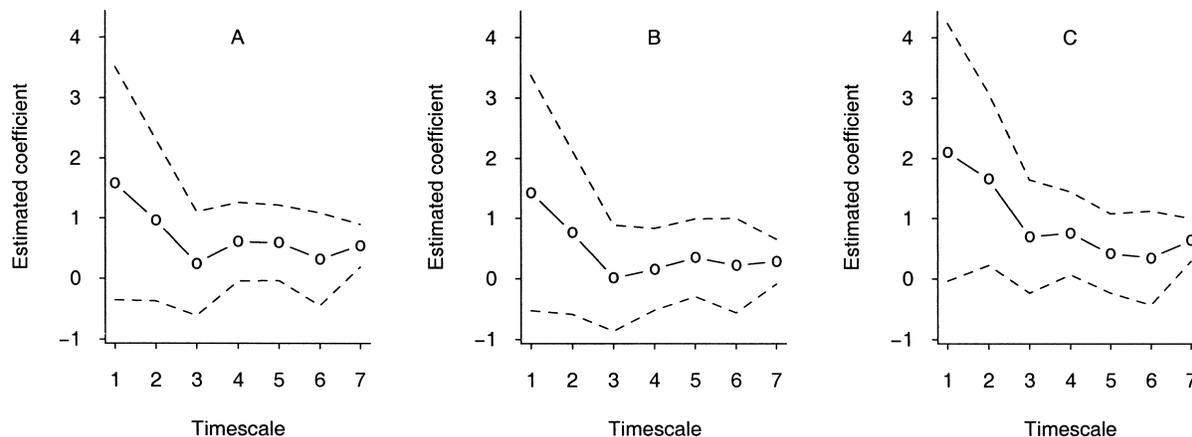
But does it? Although I applaud both the originality of the concept and the elegance of its implementation, I am less convinced about its practical interpretation in the context of public policy decisions on air pollution.

One issue is the variability of the regression coefficients, both within each city and in the combined data. Looking at the individual-city results in Dominici et al.'s table 1, I do not see any consistent pattern in the coefficients. Only when the results are combined, in the "pooled" column of table 1, does an interpretable pattern emerge. However, the results shown here are for  $\tau = 0$  ( $\tau$  is the assumed between-city standard deviation of the true regression coefficient, allowing that it may be different in different cities), and when the analyses are repeated using different values of  $\tau$  derived from previous hierarchical analyses of multicity data (Dominici et al.'s table 4), the widths of the confidence intervals cast doubt on the statistical significance of any of the differences between mortality effects at different timescales. On the basis of current knowledge, the balance of evidence seems to favor a smaller value of  $\tau$  rather than a larger value, but this is clearly one respect in which the results are sensitive to model assumptions.

I am also concerned about how to interpret the frequency-based decomposition in the context of response to specific air pollution events. To illustrate the difficulty, I considered the hypothetical air pollution pattern of a time series of length 101 days in which the  $\text{PM}_{10}$  level is  $100 \mu\text{g}/\text{m}^3$  on day 51 and 0 on every other day. This therefore represents the response to a single high-air-pollution event. The frequency decomposition from the authors' website was applied with breaks at 1, 10, 30, and 50 cycles, 50 being the maximum number of cycles in this instance. This decomposition represents the observed signal as a linear combination of cosine

waves, representing three timescales, which are shown in figure 1. These curves are symmetric about the date of the air pollution event, so the analysis combines forward-time and backward-time effects with no discrimination between the two. Dominici et al. say that this "does not complicate our inferences" (5, p. 1064), noting, in particular, that the backward-time component of the effect is not physically realistic and therefore presumably can be ignored in the interpretation. I am far from convinced: Does it really make sense to build a regression model that forces the inclusion of physically unrealistic components? This is a different issue from running backward-time analyses as a sensitivity check on forward-time results, as Chock et al. (6) have done, for instance. In the present context, it seems to me that adoption of an unrealistic model could both bias the regression coefficients and increase their standard errors to an unacceptable extent.

To gain further insight, I used the Philadelphia data set to conduct both some sensitivity analyses of the models themselves and a simulation study. Part A of figure 2 shows the results of a sensitivity analysis of the dependence of total mortality on total suspended particulates (TSP) in the Philadelphia data set, including a day-of-the-week effect and nonlinear regression components for the current day's temperature, the current day's dew point, and day, modeled using natural splines (the "ns" function in S-Plus) with 6, 6, and 90 degrees of freedom, respectively, and fitting the model as a maximum likelihood Poisson regression with the S-Plus "glm" (generalized linear model) function. Note that this method of fitting the model avoids some of the problems of establishing suitable convergence criteria that have arisen in connection with the alternative "gam" (generalized additive model) function (7); the results given here are robust to the use of different convergence criteria and have also been verified using a Newton-Raphson iteration programmed in



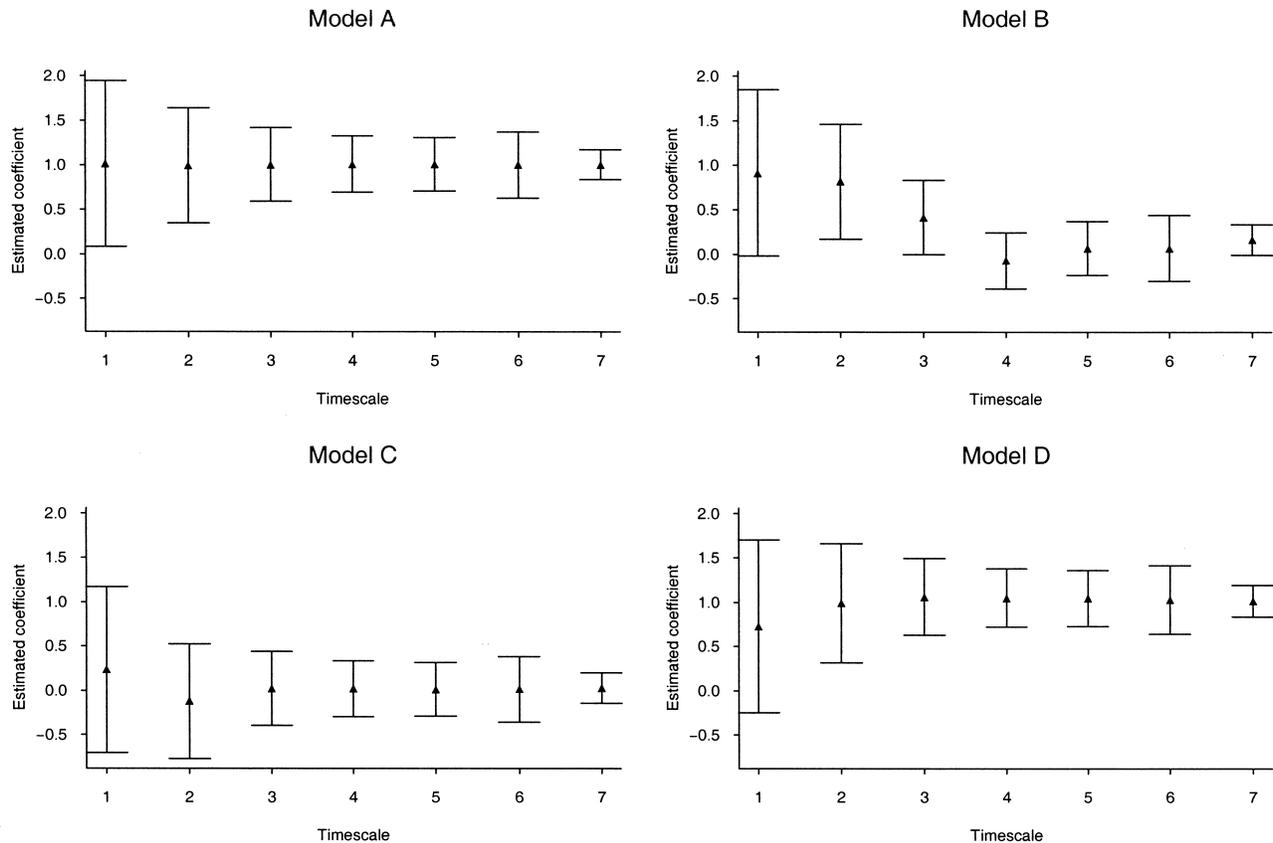
**FIGURE 2.** Alternative estimates (percentage increase in mortality associated with a  $10\text{-}\mu\text{g}/\text{m}^3$  rise in total suspended particulate levels) and 95% confidence bands (area between dashed lines) for the effect of total suspended particulates on total mortality in Philadelphia, Pennsylvania, 1973–1988. Timescales 1–6 correspond to the timescale effects 60–<90, 30–<60, 14–<30, 7–<14, 3.5–<7, and 1–<3.5 days, respectively. Timescale 7 is from a separate analysis based on a single linear regression coefficient for current-day total suspended particulates. All models included meteorologic and trend effects modeled by natural splines. Part A, model using just current-day values of temperature and dew point; part B, model using current-day values plus averages of the previous 3 days; part C, model including frequency decomposition of temperature and dew point data analogous to that used for total suspended particulates.

FORTTRAN. The “breaks” in the authors’ “decompose” function have been set at 1, 61, 91, 183, 391, 783, 1,565, and 10,958; except for the first and last breaks, these are defined as  $5,479/r$ , where 5,479 is the length of the series and  $r = 90, 60, 30, 14, 7, 3.5$  defines the timescales of interest (the first component is supposed to represent the contribution from  $\geq 90$  days, the second the contribution from 60–<90 days, etc.). The coefficients and associated approximate 95 percent confidence intervals are plotted for timescales of 60–<90 days (timescale 1), 30–<60 days (timescale 2), and so on, down to 1–<3.5 days (timescale 6); also shown as “timescale 7” is the regression coefficient from a separate analysis using just the current-day TSP. Then, in parts B and C of figure 2, the same analysis has been repeated using different forms of the meteorologic terms: Part B includes the current day’s values for temperature and dew point and also 3-day averages of lagged values (similar to the authors’ analyses of the NMMAPS cities), while part C uses frequency decompositions of temperature and dew point, employing the same breaks as those used for TSP and again treating each term nonlinearly using natural splines.

The purpose of these comparisons was to examine the sensitivity of the results to different treatments of the meteorologic components. In particular, if meteorology is a confounder for particulate matter, then a frequency decomposition of meteorologic data (figure 2, part C) might radically change the estimated signal for particulate matter. In fact, the  $\text{PM}_{10}$  coefficients *increase* when frequency-decomposed meteorologic data are included, which suggests that this analysis actually helps us separate the time-lagged  $\text{PM}_{10}$  effect from the time-lagged meteorology effect. Apart from that, it is noticeable that all three response curves in figure 2 are of the same general shape. The actual analysis

depicted in figure 5 of Dominici et al.’s paper is analogous to figure 2, part A, but it excludes the  $\geq 65$ -year age category. As a matter of fact, any of the curves shown in figure 2 provide clearer evidence of a statistically significant effect than Dominici et al.’s figure 5.

I also used the Philadelphia data set as the starting point for a simulation study. First, I reanalyzed the data to obtain estimates of the meteorologic (current-day) and long-term trend effects, omitting TSP. I then simulated Poisson count data from four different models including these meteorologic and trend components, plus a TSP effect. In each case, the assumed regression coefficient was 1 (percent rise in mortality corresponding to a  $10\text{-}\mu\text{g}/\text{m}^3$  rise in TSP). The four models were: model A, Poisson regression using current day’s TSP as the exposure variable; model B, Poisson regression using the average of the last 7 days’ TSP; model C, Poisson regression using the average of the last 30 days’ TSP; and model D, a compartment model similar to those described by Zeger et al. (2) and Smith et al. (3), with a mean time to death in the frail state of 30 days. The last model is the only one of these four that could truly be called a harvesting model. For each model, I simulated and analyzed 1,000 replications of the series using the frequency-decomposition software provided by Dominici et al. I calculated the mean of each estimated regression coefficient, averaged over the 1,000 replications for each of the four models, and the mean of the estimated standard errors for each regression coefficient and each model. These means and standard errors were used to calculate “typical confidence intervals” in figure 3. In any *single* simulation, the estimated regression coefficient could lie (with 95 percent probability) anywhere within the indicated confidence



**FIGURE 3.** Simulated results of 1,000 replications for each of four models of the air pollution-mortality relation. Plotted are the mean parameter estimates (percentage increase in mortality associated with a  $10\text{-}\mu\text{g}/\text{m}^3$  rise in total suspended particulate levels) from all 1,000 simulations, together with 95% confidence intervals (bars) derived from the mean standard errors for a single simulation. Timescales 1–6 correspond to the timescale effects 60–<90, 30–<60, 14–<30, 7–<14, 3.5–<7, and 1–<3.5 days, respectively. Timescale 7 is from a separate analysis based on a single linear regression coefficient for current-day total suspended particulates.

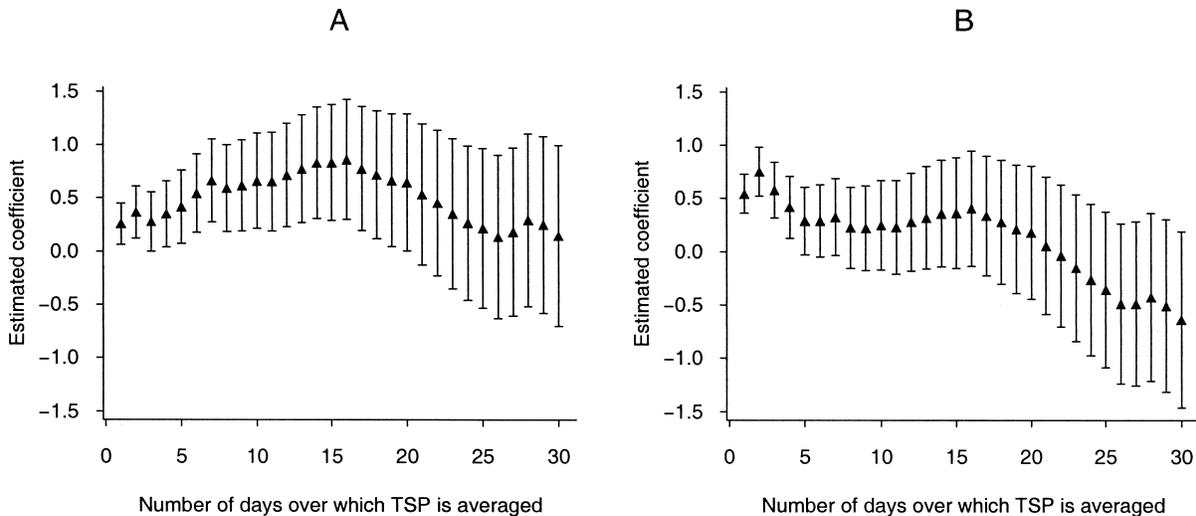
interval, while the width of the confidence intervals should be of the same order as those in figure 3.

For model A, figure 3 shows that the actual frequency-decomposition coefficients are very nearly constant across the different frequency ranges, even though the true effect in this instance is truly “short-term.” For models B and C, figure 3 shows some evidence that the  $\text{PM}_{10}$  effect is stronger at longer time lags, but not in a way that intuitively corresponds to the true timescale effect. Finally, for model D, the only one of the four derived from a genuine harvesting model, figure 3 shows coefficients that are fairly constant except for the longest time lag (60–90 days). On the basis of these simulations, it seems to me that one would be hard pressed in practice to determine the true timescale dependence from the coefficients estimated by the frequency decomposition method.

Where does this leave us? Perhaps with the need to perform simpler analyses. One possibility is simply to repeat the standard regressions but with different covariates to represent exposure to particulate matter. In figure 4, I have attempted this using “exposure measures” based on  $k$ -day

averages of TSP, for  $k$  values from 1 to 30. However, I have also repeated the analysis for two treatments of the meteorologic components, including only current day’s temperature and dew point in part A of figure 4 and adding to these the averages of the past 3 days in part B. As in earlier analyses, the model included a day-of-the-week effect, and the temperature and dew point effects were modeled nonlinearly using natural splines, as was the long-term trend. Both sets of estimates suggest a peak in the time-dependent TSP effect at approximately 15 days, but beyond that there are substantial differences between the two plots.

All of my comments should be tempered by the observation that they are based on a single data set. One would need to see similar behavior in many different data sets before the results could be accepted as generic, but with that caveat it seems justified to draw the following two conclusions. First, it is very hard to associate the frequency-decomposition regression coefficients with any specific shape of time-dependent response to a high-air-pollution event, or to distinguish true “harvesting” from other forms of time-lagged response to an air pollution event. Second, the results



**FIGURE 4.** Results from an alternative analysis of Dominici et al.'s (5) data from Philadelphia, Pennsylvania (1973–1988), in which the usual long-term trend and meteorologic effects are included, together with levels of total suspended particulates averaged over  $k$  days. The figure shows estimates (percentage increase in mortality associated with a  $10\text{-}\mu\text{g}/\text{m}^3$  rise in total suspended particulate levels) and 95% confidence intervals (bars) for all  $k$ 's between 1 and 30. Part A, model using just current-day values of temperature and dew point; part B, model using current-day values together with the averages of values from the previous 3 days. TSP, total suspended particulates.

are sensitive to the way in which meteorologic confounders are brought into the analysis.

As an observer of the work of this group of Johns Hopkins researchers over a number of years, I have been enormously impressed by the way they have brought some of the most modern innovations in statistical methodology—such as hierarchical models, Bayesian Monte Carlo estimation techniques, and spatial statistics—to the analysis of data on air pollution and health. However, one should not be misled by the elegance of their analysis into overinterpreting the results. Dominici et al. (5) have introduced another very striking innovation to the analysis of epidemiologic data sets, but the practical problems of drawing statistically valid conclusions that have meaningful public health implications remain as difficult as ever.

#### ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation grants DMS-9971980 and DMS-0084375.

The author thanks Dr. Francesca Dominici for supplying the Philadelphia data and for carrying out independent data analyses and simulations not reported in the paper.

#### REFERENCES

1. Martin AE. Mortality and morbidity statistics and air pollution. *Proc R Soc Med* 1964;57:969–75.
2. Zeger SL, Dominici F, Samet J. Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology* 1999;10:171–5.
3. Smith RL, Davis JM, Speckman P. Human health effects of environmental pollution in the atmosphere. In: Barnett V, Stein A, Turkman F, eds. *Statistics in the environment 4: statistical aspects of health and the environment*. Chichester, United Kingdom: John Wiley and Sons Ltd, 1999:91–115.
4. Schwartz J. Harvesting and long-term exposure effects in the relation between air pollution and mortality. *Am J Epidemiol* 2000;151:440–8.
5. Dominici F, McDermott A, Zeger SL, et al. Airborne particulate matter and mortality: timescale effects in four US cities. *Am J Epidemiol* 2003;157:1055–65.
6. Chock DP, Winkler SL, Chen C. A study of the association between daily mortality and ambient air pollution concentrations in Pittsburgh, Pennsylvania. *J Air Waste Manage Assoc* 2000;50:1481–500.
7. Dominici F, McDermott A, Zeger SL, et al. On the use of generalized additive models in time-series studies of air pollution and health. *Am J Epidemiol* 2002;156:193–203.