Biologically Plausible Particulate Air Pollution Mortality Concentration–Response Functions

Steven Roberts

School of Finance and Applied Statistics, Australian National University, Canberra, Australia

In this article I introduce an alternative method for estimating particulate air pollution mortality concentration–response functions. This method constrains the particulate air pollution mortality concentration–response function to be biologically plausible—that is, a nondecreasing function of the particulate air pollution concentration. Using time-series data from Cook County, Illinois, the proposed method yields more meaningful particulate air pollution mortality concentration–response function estimates with an increase in statistical accuracy. *Key words:* air pollution, daily mortality, dose–response relationship, models, particulate matter. *Environ Health Perspect* 112:309–313 (2004). doi:10.1289/ehp.6428 available via *http://dx.doi.org/* [Online 24 November 2003]

Numerous community time-series studies on the effect of particulate air pollution on mortality have been conducted (Dominici et al. 2000, 2002a; Ito et al. 1995; Samet et al. 2000; Schwartz and Zanobetti 2000; Smith et al. 1999). These studies typically fit a generalized additive model (GAM) (Hastie and Tibshirani 1990) or generalized linear model (GLM) (McCullagh and Nelder 1989) to time series of daily mortality, particulate air pollution, meteorological covariates, and other air pollutants. The fitted models are then used to quantify the effect on mortality of particulate air pollution.

Typically, these community particulate air pollution mortality time-series studies assume that the effect of particulate air pollution on the logarithm of daily mortality is a linear function of the particulate air pollution concentration. Some recent studies (Daniels et al. 2000; Dominici et al. 2002a; Schwartz and Zanobetti 2000; Schwartz et al. 2001; Smith et al. 2000a, 2000b) investigate the shape of the particulate air pollution concentration–response function by modeling the city-specific concentration–response functions using piecewise linear or smooth functions. Schwartz (2000) looks at the possibility of thresholds in the air pollution mortality concentration–response function by limiting his analysis to days with low air pollution concentrations. A potential problem with these studies is that the resulting concentration–response functions may not be biologically plausible. I use the term "biologically plausible" to mean a concentration–response function that is a nondecreasing function of the particulate air pollution concentration. A concentration–response function that is not nondecreasing would indicate that a higher level of particulate air pollution is less harmful to health than a lower level of particulate air pollution.

In this article I introduce a new method for fitting particulate air pollution mortality concentration–response functions. This method constrains the concentration–response function to be biologically plausible. Constraining the concentration–response function to be nondecreasing rules out the possibility of hormesis. However, some recent findings suggest that even low-level exposure to particulate air pollution may pose health risks [Health Effects Institute (HEI) Perspectives 2002], suggesting that hormesis is unlikely.

A simulation study shows that constraining the concentration–response function to be biologically plausible increases the statistical estimation precision. I use data from Cook County, Illinois, to illustrate the methodology of constrained versus unconstrained concentration–response function estimation.

Data and Methods

Data. The data used to illustrate the methods of this article are concurrent daily time series of mortality, weather, and particulate air pollution from Cook County for 1987–1994. These are the same Cook County data used by Samet et al. (2000).

The mortality time series are nonaccidental daily deaths of individuals ≥ 65 years old. The mortality time-series data are available from the National Center for Health Statistics (Hyattsville, MD). The weather time series are the daily 24-hr mean temperature and mean dew point temperature. The temperature and dew point temperature are 24-hr averages of hourly observations taken at O'Hare International Airport (Chicago, IL). The weather time-series data are available from the EarthInfo (2003) database.

The air pollution time series is based on the ambient 24-hr concentration of particulate matter < 10 μm in diameter (PM_{10}). The PM_{10} data are available from the U.S. Environmental Protection Agency (2004) Aerometric Information Retrieval System (AIRS; now Air Quality System Database). The PM_{10} time series used here is the average of the PM_{10} concentration of the current day and PM_{10} concentration of the previous day.

If either the current or previous day's PM_{10} concentration was missing, the average was set equal to the nonmissing value.

The largest 1% of PM_{10} concentrations was removed from the study to avoid the potential for these points to have an undue influence on the estimated concentration–response functions. This leaves 2,838 days of data for Cook County, with a maximum PM_{10} concentration of 89 μg/m^3.

Methods. In many community time-series studies on the effect of PM_{10} on mortality, an additive Poisson log-linear model is fit to the time series of observed mortality. Under this model, the daily mortality counts are modeled as independent Poisson random variables with a time varying mean $\log(\mu_t)$, where

$$\log(\mu_t) = \mu + \text{confounders}_t + \beta PM_t. \quad [1]$$

Here, the subscript t refers to the day of the study; μ_t is the mean number of deaths on day t; μ is the intercept term; confounders$_t$ represents other time-varying variables related to daily mortality. Typical confounders include temperature, humidity, longer-term mortality trends, and seasonality; PM_t is the time series of PM_{10} concentrations; and β is the effect of PM_{10} on mortality. It gives the increase in $\log(\mu_t)$ per unit increase in PM_{10}.

I allow the effect of PM_{10} on the logarithm of daily mortality to be a piecewise linear function with one or two change points. At the location of the change points, the slope of the piecewise linear concentration–response function is allowed to change. The difference between this study and other studies is that the piecewise linear functions are constrained to ensure the resulting concentration–response function is biologically plausible.

Piecewise linear functions are commonly used to model the PM_{10} mortality concentration–response function. These are useful for investigating the possibility of thresholds and

other nonlinearities in the concentration–response function. This could yield important information about the effect of particulate air pollution on mortality and be helpful to regulators. For example, a piecewise linear function with one change point allows for the following forms of the concentration–response function:

- Zero slope before the change point and positive slope after the change point. This would indicate that there is a threshold below which PM_{10} does not affect mortality.
- Positive slope before the change point and zero slope after the change point. This would indicate that there is a saturation point above which there is no incremental effect of PM_{10} on mortality.
- Positive slope before the change point and smaller positive slope after the change point. This would indicate that there is a point above which the incremental effect of PM_{10} on mortality is reduced.
- Positive slope before the change point and larger positive slope after the change point. This would indicate that there is a point above which the incremental effect of PM_{10} on mortality is increased.

A piecewise linear function with two change points allows even more flexibility. For example, it can approximate the logistic curve that has been suggested as a plausible shape for the particulate air pollution concentration–response function (Schwartz and Zanobetti 2000).

To allow for more general forms of the PM_{10} concentration–response function, model 1 may be extended:

$$\log(\mu_t) = \mu + \text{confounders}_t + g(PM_t). \quad [2]$$

Where $g(PM_t)$ is some function of PM_{10}, two forms of $g(PM_t)$ will be used in this article:
- Piecewise linear function with one change point:

$$g(PM_t) = \begin{cases} \beta_1 PM_t, & \text{if } PM_t < \theta \\ \beta_1\theta + \beta_2(PM_t - \theta), & \text{if } PM_t \geq \theta \end{cases}$$

β_1 and β_2 are constrained to be non-negative. These constraints ensure that the piecewise linear function is nondecreasing. θ is the change point and gives the PM_{10} concentration at which the slope of the piecewise linear concentration–response function is allowed to change. This form of $g(PM_t)$ is referred to as C1. This model fit without constraints on β_1 and β_2 is referred to as UC1.
- Piecewise linear function with two change points:

$$g(PM_t) = \begin{cases} \beta_1 PM_t, & \text{if } PM_t < \theta_1 \\ \beta_1\theta_1 + \beta_2(PM_t - \theta_1), & \text{if } \theta_1 \leq PM_t \leq \theta_2 \\ \beta_1\theta_1 + \beta_2(\theta_2 - \theta_1) + \\ \beta_3(PM_t - \theta_2), & \text{if } PM_t > \theta_2 \end{cases}$$

β_1, β_2, and β_3 are constrained to be non-negative. These constraints ensure that the piecewise linear function is nondecreasing. θ_1 and θ_2 are the change points and give the PM_{10} concentrations at which the slope of the piecewise linear concentration–response function is allowed to change. This form of $g(PM_t)$ is referred to as C2. This model fit without constraints on β_1, β_2, and β_3 is referred to as UC2.

Other forms of the concentration–response function such as smooth functions represented by cubic or natural cubic splines could also be considered. Additional constraints would be imposed on the parameters to ensure that the estimated concentration–response function is nondecreasing.

Once the form of $g(PM_t)$ is chosen, the model parameters are fit using a constrained maximum likelihood. I obtained the constrained maximum likelihood estimates iteratively in two repeated steps. The first step fixes the parameters corresponding to the PM_{10} terms at their current values and estimates the parameters corresponding to the confounders. This step can be performed using the GLM function available in S-Plus (Insightful, Seattle, WA, USA). The GLM function allows the parameters corresponding to the confounders to be estimated while keeping the PM_{10} parameters fixed at their current values. The second step fixes the parameters corresponding to the confounders at their current values and estimates the parameters corresponding to the PM_{10} terms. This step can be performed using constrained optimization software. These two steps are iterated to obtain the final parameter estimates. GLM estimation was chosen to avoid the recent convergence criteria problems associated with the S-Plus GAM function (Dominici et al. 2002b).

Simulation Study

I use simulations to explore the bias and mean squared error properties of the constrained, or biologically plausible, concentration–response functions compared with unconstrained concentration–response functions. Mortality time series are generated using two forms of the PM_{10} concentration–response function:
- Null concentration–response. For the null concentration–response, PM_{10} has no effect on mortality. In this situation, the log mean of the simulated Poisson mortality time series is given by $\log(\mu_t) = \log(83)$. The average daily mortality in Cook County was 83.
- Dual change-point concentration–response. This is the concentration–response found from fitting model C2 to the actual Cook County data (see "Application"). In this situation, the log mean of the simulated Poisson mortality time series is given by $\log(\mu_t) = \log(83) + f(PM_t)$, where $f(PM_t)$ is the estimated C2 function for Cook County.
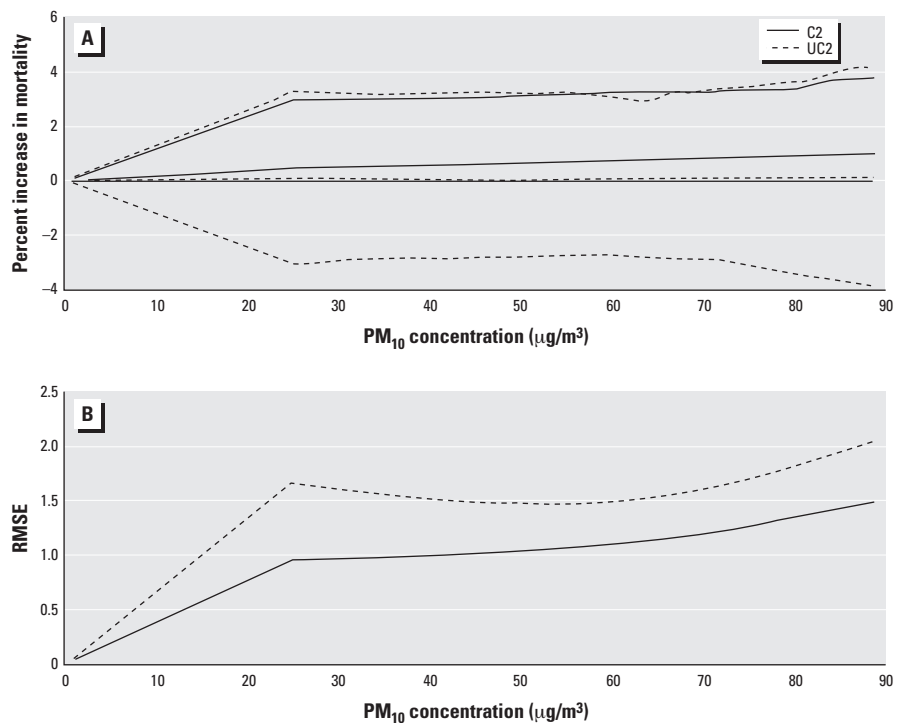
Figure 1. Results of simulations using the null concentration response to generate mortality. (*A*) Summary of the 400 concentration–response estimates from model C2 and model UC2 estimation. The solid lines correspond to the mean value and 97.5 and 2.5 percentiles of the 400 concentration–response estimates using model C2; the dotted lines are similarly defined for estimates using model UC2. (*B*) RMSE at each PM_{10} concentration.

For simplicity, confounders were not included in the simulations. Using the observed Cook County PM_{10} time series for both the null and dual change-point concentration response, 400 mortality time series were generated. Models C2 and UC2 were estimated from the simulated mortality time series. This was done by fitting the Poisson log linear models $\log(\mu_t) = \mu + g(PM_t, C2)$ and $\log(\mu_t) = \mu + g(PM_t, UC2)$ to each simulated mortality time series, where $g(PM_t, C2)$ and $g(PM_t, UC2)$ correspond to models C2 and UC2, respectively.

Figure 1 gives the results of the simulations where models C2 and UC2 were used to estimate the null concentration response. The curve corresponding to the mean estimate indicates that the bias of model C2 is modest. The curves corresponding to the 97.5 and 2.5 percentiles indicate that model C2 has a lower estimation variance than model UC2. The root mean squared error (RMSE) was used to compare the statistical accuracy of the two methods. Figure 1B gives the RMSE of the 400 concentration–response function estimates at each PM_{10} concentration. The RMSE for model C2 is always lower than the RMSE for model UC2.

Figure 2 gives the results of the simulations where models C2 and UC2 were used to estimate the dual change-point concentration response. The curve corresponding to the mean estimate for model C2 showed little

bias (the mean estimate for model UC2 is an unbiased estimate of the dual change-point concentration response). The curves corresponding to the 97.5 and 2.5 percentiles indicate that model C2 has a lower estimation variance than model UC2. The RMSE for model C2 was again lower than for model UC2.

The simulations showed that the bias of the constrained concentration–response functions was not large and that the bias is a small part of the RMSE. The bias of the constrained concentration–response functions is more than compensated for by a decrease in estimation variance. The smaller RMSE for the constrained concentration–response functions indicates that they offer an improvement in statistical accuracy over unconstrained concentration–response functions. The constrained concentration–response functions have the added advantage of giving biologically plausible concentration–response function estimates.

Application

In this section the Cook County data illustrate the methodology of constrained versus unconstrained concentration–response function estimation. Using data from only one city is adequate for this purpose.

A Poisson log-linear model similar to those used in Dominici et al. (2002b) was fit to the data:

$$
\begin{aligned}
\log(\mu_t) = \mu &+ S_{t1}(\text{time}, 8/\text{year}) \\
&+ S_{t2}(\text{temp}_0, 6) \\
&+ S_{t3}(\text{temp}_{1-3}, 6) \\
&+ S_{t4}(\text{dew}_0, 3) \\
&+ S_{t5}(\text{dew}_{1-3}, 3) \\
&+ \gamma DOW_t + g(PM_t). \quad [3]
\end{aligned}
$$

Here,
• The subscript t refers to the day of the study
• μ_t is the mean number of deaths on day t
• μ is the intercept term
• $S_{t1}(\text{time}, 8/\text{year})$ is a smooth function of time with 8 degrees of freedom per year
• $S_{t2}(\text{temp}_0, 6)$ and $S_{t3}(\text{temp}_{1-3}, 6)$ are smooth functions with a total of 6 degrees of freedom. temp_0 is the current day's mean 24-hr temperature, and temp_{1-3} is the average of the previous 3 days' 24-hr mean temperatures. $S_{t4}(\text{dew}_0, 3)$ and $S_{t5}(\text{dew}_{1-3}, 3)$ are similar functions for the 24-hr mean dew point temperature
• DOW_t is a set of indicator variables for the day of the week
• γ is a vector of coefficients that contains the mortality adjustments for the day of the week
• $g(PM_t)$ is the PM_{10} concentration–response function.

The smooth functions, $S_{ti}()$, were represented using natural cubic splines. The use of GLMs and natural cubic splines was one of the methods used to adjust for confounders in the Revised Analyses of the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), Part II (Health Effects Institute 2003). Another common method to adjust for confounders is the use of GAMs, with smoothing splines (Green and Silverman 1994) used to represent the smooth functions $S_{ti}()$. In light of the revised NMMAPS study, the GAM models are now fit with stricter convergence criteria. GLM estimation with natural cubic splines is essentially a fully parametric version of GAM estimation with smoothing splines. Currently, there is no firm evidence that favors the use of GLM or GAM estimation (Health Effects Institute 2003).

Single change-point models C1 and UC1 were fit to the Cook County data using a sequence of change-point values (θ) ranging from 5 to 90 $\mu g/m^3$, using increments of 5 $\mu g/m^3$. As the maximum PM_{10} concentration is 89 $\mu g/m^3$, fitting model C1 and UC1 with a change point of 90 $\mu g/m^3$ is identical to fitting a constrained linear or unconstrained linear concentration–response function, respectively. For model C1 the optimal change-point value for Cook County occurs at 60 $\mu g/m^3$. For model UC1 the optimal change-point value for Cook County occurs at 65 $\mu g/m^3$.

A check of the parameter estimates obtained from the constrained iterative estimation procedure is possible using only the S-Plus GLM function. This can be done by creating
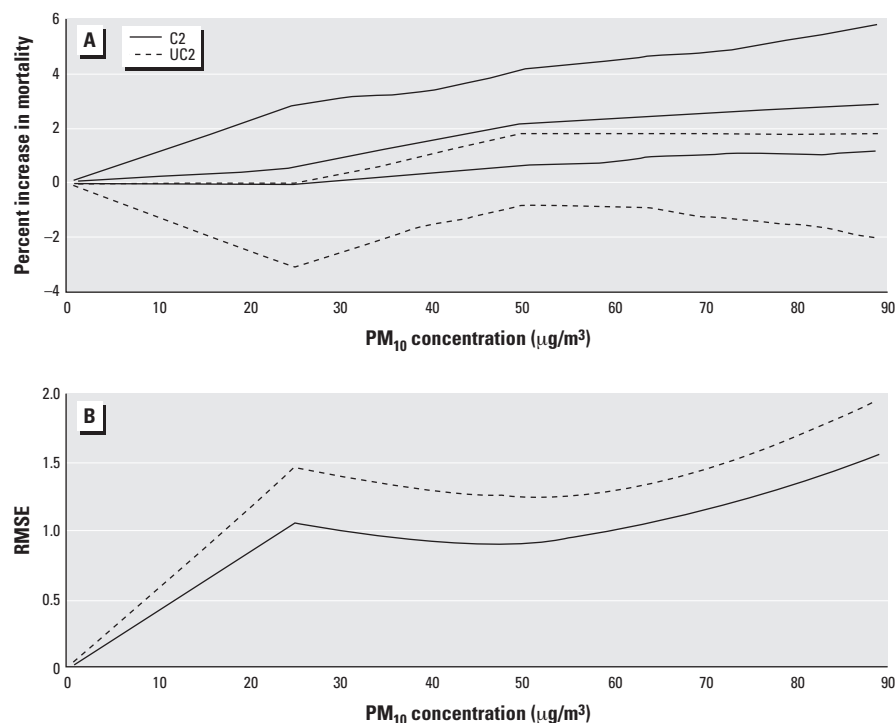

Figure 2. Results of simulations using the dual change-point concentration response to generate mortality. (*A*) Summary of the 400 concentration–response estimates from the model C2 and model UC2 estimation. The solid lines correspond to the mean value and 97.5 and 2.5 percentiles of the 400 concentration–response estimates using model C2; the dotted lines are similarly defined for estimates using model UC2. (*B*) RMSE at each PM_{10} concentration.

two new PM_{10} time series PML and PMU. PML is equal to PM_{10} if PM_{10} is $< \theta$ and θ otherwise, and PMU is equal to 0 if PM_{10} is $< \theta$ and ($PM_{10}-\theta$) otherwise. The parameters from the GLM fit corresponding to PML and PMU give the slope of the concentration–response function before and after the change point, respectively. For example, to check the estimate of the constrained iterative estimation procedure for $\theta = 60$, the GLM model can be fit with PML being the only PM_{10} term included in the model. Not including PMU in the model corresponds to the estimated concentration–response function having zero slope after the change point. If the iterative estimation procedure is giving appropriate estimates, the PM_{10} parameter estimates from the two methods should be the same. This check was used for each value of the change point θ. In each case the PM_{10} parameter estimates from the two methods agreed.

Figure 3 contains a plot of the estimated C1 function for Cook County and a 95% pointwise confidence interval. The estimated UC1 function is included for comparison. Model C1 says that the effect of PM_{10} on mortality increases until the change point of 60 $\mu g/m^3$, but that there is no incremental effect beyond 60 $\mu g/m^3$. If an actual concentration–response function were shaped like the estimated UC1 function, the effect of PM_{10} on mortality would decrease for concentrations above 65 $\mu g/m^3$. This is not biologically plausible.

Confidence intervals in Figure 3 were produced via parametric bootstrap. Using the estimated β_1, β_2, and θ values from model C1, 100 new mortality time series were generated according to the Poisson model [Equation 3]. Model C1 with θ fixed at 60 $\mu g/m^3$ was then fit to each of the generated mortality time series. At each PM_{10} concentration, the 95% confidence interval corresponds to the 97.5 and 2.5 percentiles of the 100 estimated C1 concentration–response functions.

Next, the dual change-point models C2 and UC2 were fit to the Cook County data
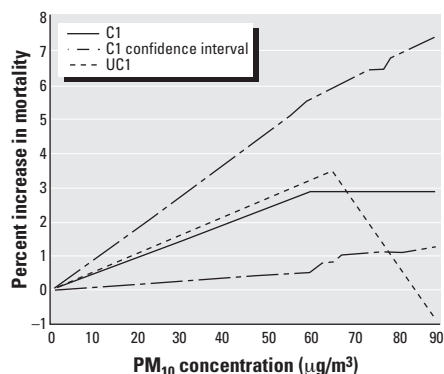
with θ_1 fixed at 25 $\mu g/m^3$ and θ_2 fixed at 50 $\mu g/m^3$. The location of these change points is similar to the knot location used by Daniels et al. (2000). The use of two change points allows extra flexibility in the C2 and UC2 concentration–response functions compared with C1 and UC1. The check described above was performed on the constrained iterative estimation procedure for model C2 estimation. Once again the PM_{10} parameter estimates from the two methods agreed. Figure 4 contains a plot of the estimated C2 function for Cook County and a 95% pointwise confidence interval. The confidence intervals, as for model C1, were produced via parametric bootstrap. Model C2 says that PM_{10} has no effect on mortality until the first change point of 25 $\mu g/m^3$ (an actual threshold), has an increasing effect for concentrations between 25 and 50 $\mu g/m^3$, and has no incremental effect for concentrations beyond 50 $\mu g/m^3$. If an actual concentration–response function were shaped like the estimated UC2 function (dotted), it would say that the effect of PM_{10} on mortality is beneficial at some PM_{10} levels and that higher concentrations of PM_{10} can be less harmful than lower concentrations of PM_{10}. This is not biologically plausible.

The above analyses were repeated using 4 instead of 8 degrees of freedom per year for the smooth function of time. Fewer degrees of freedom for the smooth function of time correspond to less aggressive control of time-related confounders such as seasonality. The results of using 4 degrees of freedom were similar to those obtained using 8 degrees of freedom: The shape of the estimated concentration–response functions were similar.

The shape of the estimated C1 and C2 concentration–response functions could have public health implications. The estimated C1 concentration–response function suggests that a PM_{10} standard would be inefficient unless it were set below the change point of 60 $\mu g/m^3$. The reason for this is that according to the estimated C1 concentration–response function, PM_{10} does not have an incremental effect

on mortality beyond 60 $\mu g/m^3$. Similarly, the estimated C2 concentration–response function suggests that a PM_{10} standard will reduce PM_{10}-induced mortality only if it reduces PM_{10} concentrations in the range between 25 and 50 $\mu g/m^3$.

The estimated UC1 and UC2 concentration–response functions were both biologically implausible. This is not an uncommon occurrence. In many studies that examine the shape of the concentration–response function, the estimated concentration–response functions are decreasing in certain air pollution ranges and/or indicate that air pollution at certain concentrations can be beneficial to health (Daniels et al. 2000; Dominici et al. 2002a; Moolgavkar and Luebeck 1996; Smith et al. 2000b). One might ask whether I am overinterpreting the negative slopes of the unconstrained functions UC1 and UC2: Are the negatives slopes actually statistically different from zero? Although the question of whether the negative slopes are statistically significant is not of particular concern, the reason for using constrained concentration–response functions is to preclude concentration–response functions that are believed implausible, irrespective of statistical significance.

Table 1 contains p-values for change in deviance tests. The tests compare the fitted concentration–response functions with a null model of no PM_{10} concentration–response function and a linear PM_{10} concentration–response function. The linear concentration–response function fit to the Cook County data has a positive slope, so the constrained and unconstrained versions of the linear concentration–response function are the same. The "Test" column gives the forms of the concentration-response functions being compared. "Null" indicates a model that does not contain a PM_{10} concentration–response function. "Linear" indicates a model that contains a linear PM_{10} concentration–response function. The "p-value constrained" column contains the p-values of the change in deviance test for the constrained forms of the concentration–response functions (C1 and C2). The "p-value unconstrained" column contains the p-values for the unconstrained forms of the concentration–response functions (UC1 and UC2). The p-values would not be small enough to reject the linear
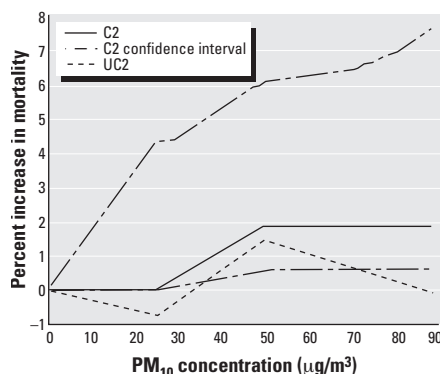

Figure 3. Plot of the estimated C1 function with 95% pointwise confidence interval for Cook County. The estimated UC1 function is included for comparison.

Figure 4. Plot of the estimated C2 function with 95% pointwise confidence interval for Cook County. The estimated UC2 function is included for comparison.

Table 1. Comparing the fitted concentration–response functions using the change in deviance test.

| | p-Value | |
Test	Constrained	Unconstrained
Linear versus null	0.08	0.08
C1, UC1 versus null	0.18	0.05
C1, UC1 versus linear	0.40	0.10
C2, UC2 versus null	0.10	0.07
C2, UC2 versus linear	0.21	0.14

concentration–response function at conventional significance levels. Of the two constrained concentration–response functions, model C2 gave the best fit to the data, as measured by the likelihood value. Because models C1 and C2 use the same number of parameters (C1 uses a parameter for estimation of the change-point location), this comparison suggests that model C2 is the preferred constrained concentration–response function.

It is possible that a biologically implausible fitted concentration–response function will show a statistically significant improvement over a linear concentration–response function that is nondecreasing (biologically plausible). For example, if we fit an unconstrained single change-point piecewise linear concentration–response function to Cook County with the change point fixed at 65 $\mu g/m^3$, this would result in the biologically implausible concentration–response function of Figure 3 (dotted line), which shows mortality decreasing markedly for concentrations > 65 $\mu g/m^3$. The *p*-value for the test of whether this concentration–response function is an improvement over a linear concentration–response function is 0.03, suggesting that the biologically implausible concentration–response function is an improvement over a biologically plausible linear concentration–response function. However, rejecting the linear concentration–response function in favor of a biologically implausible nonlinear concentration–response function raises problems of interpretation that could be avoided by the use of biologically plausible nonlinear concentration–response functions. These potential anomalies illustrate the importance of having sufficiently flexible parameterizations of biologically plausible concentration–response functions.

Discussion and Conclusion

In this article I introduce a method for fitting PM_{10} mortality concentration–response functions that constrains the concentration–response functions to be biologically plausible, i.e., a nondecreasing function of the PM_{10} concentration. Simulations showed an increase in statistical estimation precision that results from constraining the concentration–response function estimates to be biologically plausible. Biases in the constrained

concentration–response function are small and are more than compensated for by a decrease in estimation variance. This was the case whether or not PM_{10} was assumed to have an effect on mortality.

The use of constrained concentration–response functions in contrast to unconstrained concentration–response functions was illustrated using data from Cook County. The estimated unconstrained concentration–response functions were biologically implausible. It is difficult to understand and apply unconstrained concentration–response functions because increases in PM_{10} could imply decreases in mortality, or PM_{10} could be beneficial to health. The use of biologically plausible concentration–response functions avoids these problems of interpretation without introducing material bias. However, it is important to keep in mind that a concentration–response function not biologically plausible may indicate a more serious problem, such as inadequate adjustment for covariates. In such instances further investigation into the form of the model being used may be required.

The constrained concentration–response functions fit to the Cook County data did not provide evidence against a linear concentration–response function, and similar results have been reported in previous studies (Dominici et al. 2002a; Schwartz and Zanobetti 2000). Smith et al. (2000a) did not find statistically significant evidence in favor of nonlinear concentration–response functions. However, they question whether testing a nonlinear versus a linear concentration–response function is the appropriate formulation. Schwartz and Zanobetti (2000) also present an argument as to why the concentration–response relationship may look like a logistic curve. Other studies have shown evidence in favor of nonlinear particulate air pollution concentration–response functions (Daniels et al. 2000; Smith et al. 2000b).

The fact that for the Cook County data the constrained concentration–response functions were not statistically different from a linear concentration–response function should not affect the desirability of using constrained concentration–response functions. Constrained concentration–response functions are more accurate than unconstrained concentration–response functions, have low bias, and are biologically plausible.

REFERENCES

Daniels MJ, Dominici F, Samet JM, Zeger SL. 2000. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am J Epidemiol 152:397–406.

Dominici F, Daniels M, Zeger SL, Samet JM. 2002a. Air pollution and mortality: estimating regional and national dose-response relationships. J Am Stat Assoc 97:100–111.

Dominici F, McDermott A, Zeger SL, Samet JM. 2002b. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 156:193–203.

Dominici F, Samet JM, Zeger SL. 2000. Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. J R Stat Soc Ser A 163:263–302.

EarthInfo Environmental Databases. 2003. Available: http://earthinfo.com/databases/databases.htm [accessed 30 January 2004].

Green PJ, Silverman BW. 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. London:Chapman and Hall.

Hastie TJ, Tibshirani RJ. 1990. Generalized Additive Models. London:Chapman and Hall.

Health Effects Institute. 2003. Special Report: Revised Analyses of Time-Series Studies of Air Pollution and Health. Cambridge, MA:Health Effects Institute.

HEI Perspectives. 2002. Understanding the Health Effects of Components of the Particulate Matter Mix: Progress and Next Steps. Cambridge, MA:Health Effects Institute.

Ito K, Kinney PL, Thurston GD. 1995. Variations in PM-10 concentrations within two metropolitan areas and their implications for health effects analyses. Inhal Toxicol 7:735–745.

McCullagh P, Nelder JA. 1989. Generalized Linear Models. 2nd ed. London:Chapman and Hall.

Moolgavkar SH, Luebeck EG. 1996. A critical review of the evidence on particulate air pollution and mortality. Epidemiology 7:420–428.

Pickle LW, Mungiole M, Jones GK, White A. Atlas of United States Mortality. Hyattsville, MD:National Center for Health Statistics, 1996.

Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. 2000. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. N Engl J Med 343:1742–1749.

Schwartz J. 2000. Assessing confounding, effect modification, and thresholds in the association between ambient particles and daily deaths. Environ Health Perspect 108:563–568.

Schwartz J, Ballester F, Saez M, Perez-Hoyos S, Bellido J, Cambra K, et al. 2001. The concentration–response relation between air pollution and daily deaths. Environ Health Perspect 109:1001–1006.

Schwartz J, Zanobetti A. 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 11:666–672.

Smith RL, Davis JM, Sacks J, Speckman P, Styer P. 2000a. Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama. Environmetrics 11:719–743.

Smith RL, Davis JM, Speckman P. 1999. Human health effects of environmental pollution in the atmosphere. In: Statistics for the Environment 4: Statistical Aspects of Health and the Environment (Barnett V, Stein A, Turkman F, eds). Chichester, UK:John Wiley and Sons, 91–116.

Smith RL, Spitzner D, Kim Y, Fuentes M. 2000b. Threshold dependence of mortality effects for fine and coarse particles in Phoenix, Arizona. J Air Waste Manag Assoc 50:1367–1379.

U.S. Environmental Protection Agency. 2004. AirData: Access to Air Pollution Data. Available: http://www.epa.gov/air/data/aqsdb.html [accessed 30 January 2004].