



COMMENTARY

On the Use of Generalized Additive Models in Time-Series Studies of Air Pollution and Health



Francesca Dominici¹, Aidan McDermott¹, Scott L. Zeger¹, and Jonathan M. Samet²

¹ Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

² Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

Received for publication May 6, 2002; accepted for publication June 25, 2002.

The widely used generalized additive models (GAM) method is a flexible and effective technique for conducting nonlinear regression analysis in time-series studies of the health effects of air pollution. When the data to which the GAM are being applied have two characteristics—1) the estimated regression coefficients are small and 2) there exist confounding factors that are modeled using at least two nonparametric smooth functions—the default settings in the `gam` function of the S-Plus software package (version 3.4) do not assure convergence of its iterative estimation procedure and can provide biased estimates of regression coefficients and standard errors. This phenomenon has occurred in time-series analyses of contemporary data on air pollution and mortality. To evaluate the impact of default implementation of the `gam` software on published analyses, the authors reanalyzed data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) using three different methods: 1) Poisson regression with parametric nonlinear adjustments for confounding factors; 2) GAM with default convergence parameters; and 3) GAM with more stringent convergence parameters than the default settings. The authors found that pooled NMMAPS estimates were very similar under the first and third methods but were biased upward under the second method. *Am J Epidemiol* 2002;156:000–00.

air pollution; algorithms; backfitting; generalized additive models; models, statistical; time series

Abbreviations: GAM, generalized additive model(s); GLM, generalized linear model(s); NMMAPS, National Morbidity, Mortality, and Air Pollution Study; PM₁₀, particulate matter <10 μm in diameter.

Since they were originally described, generalized additive models (GAM) (1, 2) have been effectively applied in a variety of research areas. A MEDLINE search using the term “generalized additive models” revealed 128 articles published in the last 8 years in the fields of genetics, epidemiology, molecular biology, and medicine. In time-series studies of air pollution and mortality, GAM has been the

most widely applied method, because it allows for nonparametric adjustments for nonlinear confounding effects of seasonality, trends, and weather variables (3–10). It is a more flexible approach than fully parametric alternatives (11–13). GAM has been widely used in many time-series analyses, including those of data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) (6, 7, 14,

Reprint requests to Dr. Francesca Dominici, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205-3179 (e-mail: fdominic@jhsph.edu).

15). The NMMAPS addresses the relation between daily mortality counts and air pollution in the 90 largest cities in the United States.

GAM extends traditional generalized linear models (GLM) (16) by replacing linear predictors of the form $\eta = \sum_j \beta_j x_j$ with $\eta = \sum_j f_j(x_j)$, where $f_j(x_j)$ are unspecified nonparametric functions. Methods for estimating f_j include smoothing splines (17) or LOESS smoothers (18–20). GLM with regression splines (to which we refer here as the fully parametric alternative of the GAM with nonparametric smoothers) commonly define f_j to be regression splines, such as natural cubic splines or B-splines with a prespecified number of knots at known locations (21, 22).

Estimation in GAM is based on a combination of the local scoring algorithm (1) and the backfitting algorithm (23, 24). The local scoring algorithm is a generalization of the Fisher scoring procedure for finding maximum likelihood estimates in GLM (16, 25). The backfitting algorithm is suitable for fitting any additive model, and in GAM it is used within the local scoring iteration when several smooth functions are included in the model.

Unlike linear regression models, which are fitted by using weighted least squares and have an exact solution, the estimation procedure for a GAM (or a GLM) requires iterative approximations in order to find the optimal estimates. More specifically, the backfitting procedure in GAM with smoothing splines maximizes a penalized log likelihood defined as $l_p(\eta, \mathbf{y}) = l(\eta, \mathbf{y}) + P$, where \mathbf{y} is the vector of the observations, $l(\eta, \mathbf{y})$ is the likelihood function of the linear predictor η , and P is a quadratic penalty term used to account for smoothness (2, 24). This is equivalent to maximizing a posterior distribution in a Bayesian analysis using a prior that favors a smoother relation. From a frequentist standpoint, this may result in improved mean squared error, but it may also result in bias, the extent of which depends on P (26). Asymptotic bias and variance properties of backfitting estimators have been explored by Opsomer and Ruppert (27) and Opsomer (28).

Convergence of the local scoring algorithm is controlled by two user-defined parameters: 1) ϵ , which controls the convergence precision, and 2) M , which controls the maximum number of iterations allowed. Convergence of the backfitting algorithm is controlled by two additional user-defined parameters: 1) ϵ_{bfp} , which controls the convergence precision, and 2) M_{bfp} , which controls the maximum number of iterations to be used in backfitting.

In this paper, we discuss the use of GAM for estimating relative rates of mortality/morbidity associated with exposure to air pollution in time-series analyses of air pollution and health. Our simulation studies show that when the relative rates to be estimated are small and at least two nonparametric smooth functions are included in the model (smoothing splines or LOESS smoothers), as is often done in time-series studies of air pollution and mortality, the default convergence parameters in the S-Plus statistical function `gam` may be too lax to assure convergence of the backfitting algorithm and may lead to biased estimates.

Below we provide details on a simulation study to evaluate the impact of default implementation of the `gam` function in S-Plus software on published analyses. We next reanalyze

the NMMAPS data using three different methods: 1) Poisson regression with natural cubic splines to achieve nonlinear adjustments for confounding factors; 2) GAM with smoothing splines and default convergence parameters; and 3) GAM with smoothing splines and more stringent convergence parameters than the default settings. We then discuss advantages and disadvantages of using GAM as compared with GLM. Implementation of the GAM estimation procedure for air pollution time-series data is detailed in the Appendix.

SIMULATION STUDY FOR CITY-SPECIFIC ANALYSES

To assess the effects of convergence parameters on city-specific relative rate estimates in time-series studies of air pollution and health, we conducted a simulation study. First, we fitted the following GAM with default convergence parameters to the Pittsburgh, Pennsylvania, NMMAPS database (1987–1994):

$$Y \sim \text{Poisson}(\mu_t)$$

$$\log \mu_t = \alpha + \beta \text{PM}_{10t} + s_1(\text{temperature}, 6) + s_2(\text{time}, 7/\text{year}) + \eta I_{\text{dow}}$$

where Y_t denotes the daily number of deaths among people older than 75 years, β denotes the log relative rate of mortality associated with a 10-unit increase in particulate matter less than 10 μm in diameter (PM_{10}), $s_1(\text{temperature}, 6)$ and $s_2(\text{time}, 7/\text{year})$ are smooth functions (smoothing splines) of temperature and calendar time designed to control for trend, seasonality, and weather, and I_{dow} are indicator variables for day of the week. This is a simplified version of the GAM originally used in the published NMMAPS analyses (4, 7), with only two smooth functions and no overdispersion. In the simulation, the true β was set to a 0.51 percent increase in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{10} , the value estimated from the actual Pittsburgh data.

We then simulated 1,000 mortality time series from a Poisson distribution with mean equal to the predictive values (μ_t) from model 1, also shown in figure 1 (top left). Each of the simulated data sets was analyzed using three methods: 1) `gam` with smoothing splines `s` and default parameters (`gam + s + default`); 2) `gam` with smoothing splines `s` and more stringent convergence parameters (`gam + s`); and 3) Poisson regression with natural cubic splines `ns`, with equally spaced knots and more stringent convergence parameters (`glm + ns`). The default settings in version 3.4 of the S-Plus software package (Insightful Corporation, Seattle, Washington) and our suggested, more stringent parameter values for the S-Plus function `gam` convergence control are summarized in table 1.

In the top left panel of figure 2, we plot the 1,000 estimates of β obtained under `glm + ns` (x-axis) versus the 1,000 estimates of β obtained under `gam + s + default` (y-axis). The horizontal and vertical lines are placed at the true values (0.51). The distribution of `glm + ns` estimates is centered at the true value, while the distribution of `gam + s + default` estimates is displaced upwards. In the top right panel of figure 2, we plot the 1,000 estimates of β obtained under `glm + ns` (x-axis) versus the 1,000 estimates of β obtained under `gam + s` with more stringent convergence criteria (y-axis). Here the two sets of estimates are much

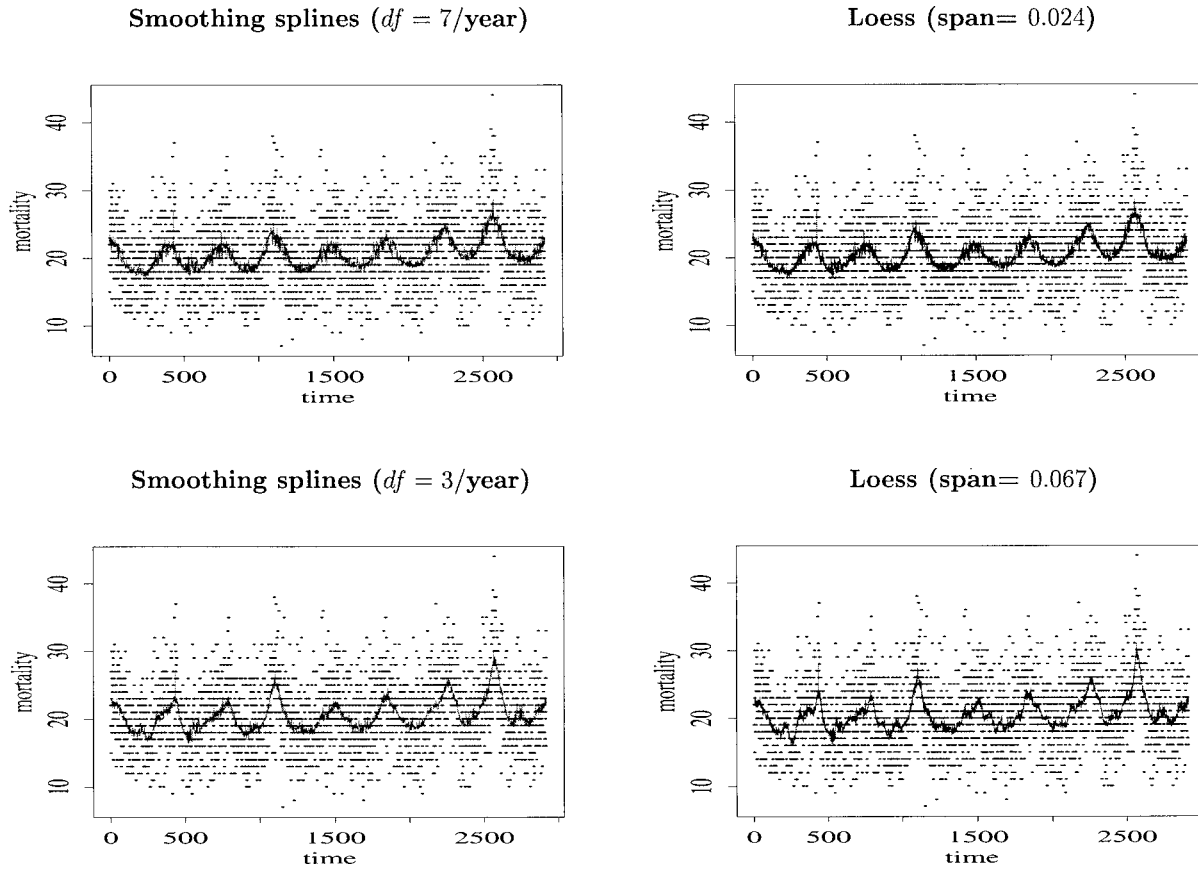


FIGURE 1. Fitted values for daily mortality (number of deaths) under generalized additive models with smoothing splines (left) and LOESS smoothers (right). Black dots are the observed mortality counts for the Pittsburgh, Pennsylvania, database (1987–1994) of the National Morbidity, Mortality, and Air Pollution Study. Time is given in days.

more similar and are closer to the true value, suggesting that GAMs with more stringent convergence parameters provide relative rate estimates closer to the true rate.

We then repeated this simulation study, replacing smoothing splines with LOESS smoothers. We fitted the following GAM with LOESS smoothers with default convergence parameters to the Pittsburgh database (1987–1994):

$$\log \mu_t = \alpha + \beta PM_{10t} + lo(\text{temperature}, 0.024) + lo(\text{time}, 0.4) + \eta I_{dow} \quad (2)$$

In model 2, we chose the spans in the LOESS smoothers to obtain a correlation of 0.99 between fitted values of model 1 and model 2. The estimated β was a 0.53 percent increase in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{10} . We then simulated 1,000 mortality time series from a Poisson distribution with mean equal to the predictive values (μ_t) from model 2, also shown in figure 1 (top right). Each of the simulated data sets was analyzed using three methods: 1) `gam` with LOESS smoother `lo` and default parameters (`gam + lo + default`); 2) `gam` with LOESS smoother `lo` and more strin-

TABLE 1. Default settings and more stringent convergence parameters in the `gam` function of the S-Plus* statistical software package (version 3.4)

Parameter	<code>gam + s + default</code>	<code>gam + s</code>	<code>glm + ns</code>
ϵ	10^{-3}	10^{-15}	10^{-15}
M	10	1,000	1,000
ϵ_{bf}	10^{-3}	10^{-15}	10^{-15}
M_{bf}	10	1,000	1,000

* Insightful Corporation, Seattle, Washington.

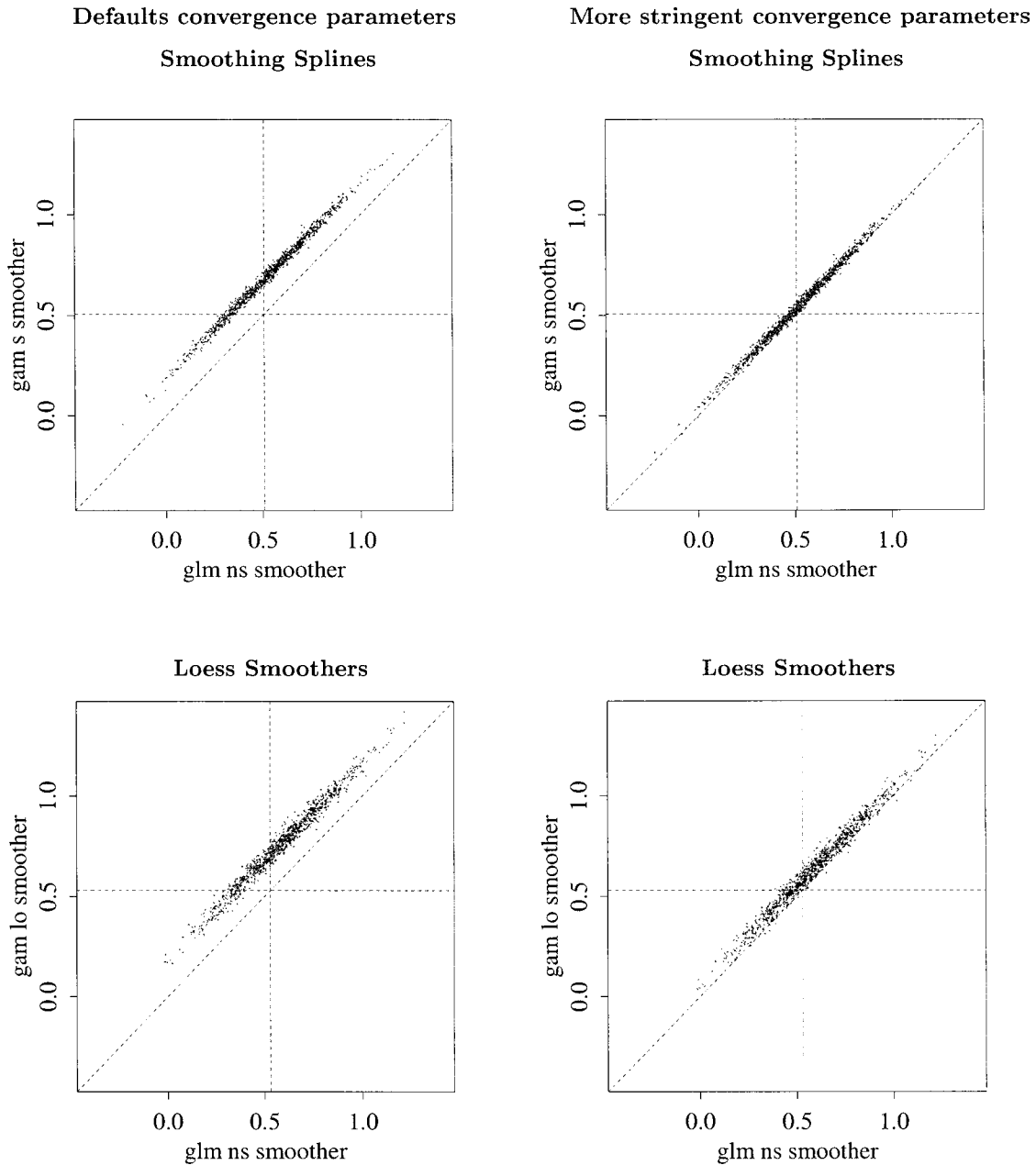


FIGURE 2. Top: estimates of β obtained by fitting a Poisson regression model with parametric nonlinear adjustments for confounding factors $glm + ns$ versus estimates obtained by fitting generalized additive models with smoothing splines $gam + s$. Vertical and horizontal lines indicate true values. Note that small bias still remains (see table 2). Bottom: estimates of β obtained by fitting a Poisson regression model with parametric nonlinear adjustments for confounding factors $glm + ns$ versus estimates obtained by fitting generalized additive models with LOESS smoothers $gam + lo$. Vertical and horizontal lines indicate true values. Note that small bias still remains (see table 2).

gent convergence parameters ($gam + lo$); and 3) Poisson regression with natural cubic splines ns ($glm + ns$) and more stringent convergence parameters.

In the bottom panels of figure 2, we plot the 1,000 estimates of β obtained under $glm + ns$ (x-axis) versus the 1,000 estimates of β obtained under $gam + lo$ + default and versus the 1,000 estimates of β obtained under $gam + lo$. The horizontal and vertical lines are placed at the true

values (0.50). Smoothing splines and LOESS smoothers show similar bias in the relative rate estimates.

To further investigate the relations among bias, size of the true coefficient, and amount of control from trend and seasonality in models 1 and 2, we repeated the simulation study by lowering the number of degrees of freedom in the smoothing splines from 7 to 3 per year or, equivalently, by increasing the span in the LOESS smoother from 0.024 to

TABLE 2. Percent bias of relative rate estimates under generalized additive models with smoothing splines and LOESS smoothers for default and more stringent convergence parameters and for different degrees of control for trend and seasonality

Model	True log relative rate	Percent bias		
		gam + s + default	gam + s	glm + ns
$s_1(\text{time}, 7/\text{year}) + s_2(\text{temperature}, 6)$	0.51	36	7	2
$s_1(\text{time}, 3/\text{year}) + s_2(\text{temperature}, 6)$	0.73	22	17	0
		gam + lo + default	gam + lo	glm + ns
$l_0(\text{time}, 0.024) + l_0(\text{temperature}, 0.4)$	0.53	42	18	9
$l_0(\text{time}, 0.067) + l_0(\text{temperature}, 0.4)$	0.73	24	18	1

0.067. The fitted values of models 1 and 2 are shown in figure 1 (bottom left and right). Note that with less control for trend and seasonality, the estimated β for Pittsburgh becomes larger (i.e., 0.73 percent vs. 0.51 percent). Table 2 summarizes the results of our simulations. The first column indicates the nonparametric smoother included in GAM models 1 and 2; the second column represents the true log relative rate (β); and the third column summarizes the percent bias of $\text{gam} + s + \text{default}$, $\text{gam} + s$, and $\text{glm} + \text{ns}$ with respect to the true β . Even when the data are generated from a GAM with smoothing splines or LOESS smoothers, $\text{gam} + s + \text{default}$ and $\text{gam} + \text{lo} + \text{default}$ produce a bias of 36–42 percent. The bias is not negligible (7–18 percent) when more stringent convergence parameters are used and when a smaller number of degrees of freedom (or a larger span in the LOESS smoother) is used in the smooth function of time to adjust for trend and seasonality.

To investigate whether the degree of bias depends on the size of the true relative rate of mortality (β), we repeated the simulation study generating 1,000 mortality time series from model 1, with β 's equal to 0.05 percent, 0.1 percent, 0.5 percent, 0.7 percent, and 1 percent increases in mortality per 10-unit increase in PM_{10} , respectively. For each sampled mortality time series, we estimate the relative rate of mortality using $\text{gam} + s + \text{default}$. Figure 3 (left) plots the relative bias versus the true β values. The relative bias is defined as $(\hat{\beta} - \beta)/\beta$, where $\hat{\beta}$ is the average of the 1,000 $\text{gam} + s + \text{default}$ estimates. The bias decreases as the size of the true coefficient increases, indicating that, on a percentage scale, estimates of smaller effects are affected more by inappropriate control in the convergence criteria.

We also performed additional simulation studies to assess the relations between bias and degree of concurvity in the GAM. The degree of concurvity (29) can be estimated by calculating the correlation between the daily time series of PM_{10} (x_{1t}) and the fitted values (\hat{x}_{1t}) from the linear model $x_{1t} = s_1(\text{temperature}, 6) + s_2(\text{time}, 7/\text{year}) + \epsilon$, where ϵ is distributed as $N(0, \sigma^2)$. For the Pittsburgh data, such concurvity is equal to 0.60. We then generate 1,000 mortality time series from model 1, replacing x_{1t} with $\sim x_{1t}$, where $\sim x_{1t} = \hat{x}_{1t} + N(0, \sigma^2)$, and we chose σ^2 so that the correlations between $\sim x_{1t}$ and \hat{x}_{1t} are equal to 0, 0.3, 0.6, 0.8, and 0.9. Figure 3 (right) plots relative bias versus concurvity. Here we take β

to be the average of the 1,000 $\text{gam} + s + \text{default}$ estimates and β to be a 0.51 percent increase in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{10} . As expected, bias increases as concurvity increases.

NMMAPS REANALYSES

We next reanalyzed the NMMAPS data by calculating city-specific relative rate estimates under three methods: 1) $\text{gam} + s + \text{default}$; 2) $\text{gam} + s$; and 3) $\text{glm} + \text{ns}$. Figure 4 (top left) shows the 90 city-specific estimates at lag 1 obtained with $\text{glm} + \text{ns}$ (x-axis) versus the published NMMAPS estimates (13, 15) obtained with $\text{gam} + s + \text{defaults}$ (y-axis). Figure 4 (top right) shows the 90 city-specific estimates at lag 1 obtained with $\text{glm} + \text{ns}$ (x-axis) versus the 90 city-specific estimates at lag 1 obtained with $\text{gam} + s$ and more strict convergence parameters. The black square is plotted at the pooled estimate across the 90 cities; its size corresponds to ± 2 standard errors of the pooled estimates.

Visual inspection of the plot indicates substantial agreement between the two sets of city-specific estimates. Cities whose estimates are unchanged fall exactly on the $y = x$ line. The original pooled estimate under $\text{gam} + s + \text{default}$ (a 0.41 percent increase in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{10} (posterior standard error 0.05)) is larger than the pooled estimate under $\text{glm} + \text{ns}$ (a 0.21 percent increase in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{10} (posterior standard error 0.06)). The pooled estimate under $\text{gam} + s$ with the strict convergence criterion is similar to the pooled estimate under $\text{glm} + \text{ns}$ (a 0.27 percent increase in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{10} (posterior standard error 0.05)). If we use 3 degrees of freedom per year in the smooth function of time under $\text{glm} + \text{ns}$, we obtain a pooled estimate of 0.31 percent (posterior standard error 0.05).

Figure 4 (bottom left) shows the standard errors of 90 city-specific estimates at lag 1 obtained with $\text{glm} + \text{ns}$ (x-axis) versus $\text{gam} + s + \text{defaults}$ (y-axis). Figure 4 (bottom right) shows the standard errors of 90 city-specific estimates at lag 1 obtained with $\text{glm} + \text{ns}$ (x-axis) versus $\text{gam} + s$ (y-axis). GAM consistently gives smaller estimated standard errors than $\text{glm} + \text{ns}$. The estimated standard errors in the two models are roughly proportional to each other. This is consistent with recent work by Ramsay et al. (29), who

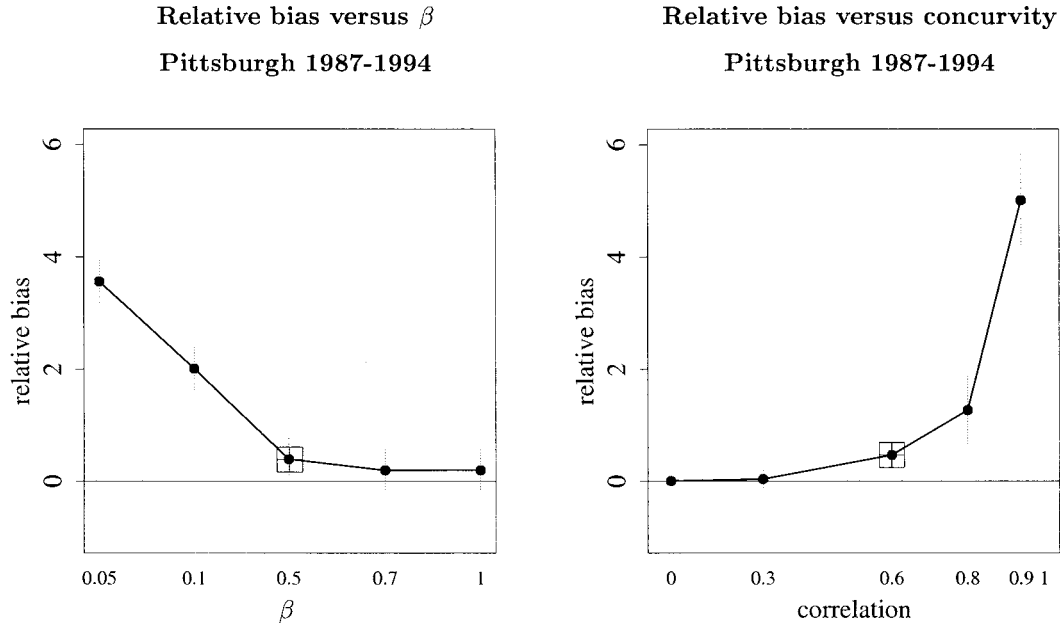


FIGURE 3. Right: relative bias in the log relative rate $\hat{\beta}$ versus the true β . The relative bias is measured as $(\hat{\beta} - \beta)/\beta$, where $\hat{\beta}$ is obtained by averaging the `gam + s + default` estimates across the 1,000 iterations. The square is plotted at $\beta = 0.51$ percent increase in mortality per $10\text{-}\mu\text{g}/\text{m}^3$ increase in particulate matter less than μm in diameter—for example, the estimate for the Pittsburgh, Pennsylvania, database (1987–1994) of the National Morbidity, Mortality, and Air Pollution Study—using `glm + ns`. Left: relative bias versus different values for concavity. The relative bias is measured as $(\hat{\beta} - \beta)/\beta$, where $\hat{\beta}$ is obtained by averaging the `gam + s + default` 1,000 estimates and β is set to be 0.51 percent. The square is plotted at the observed concavity for Pittsburgh (0.6).

showed that inability of the GAM to detect concavity can lead to underestimation of the standard errors of relative rate estimates. Standard errors are underestimated even if more stringent convergence parameters are used (figure 4, bottom right).

Figure 5 compares pooled estimates across 90 cities under the three regression methods. The 90 city-specific estimates are pooled across cities under 1) a fixed-effect model, 2) a random-effect model with a moment estimator of the across-city variance (30), and 3) a Bayesian two-stage model with a noninformative prior of the across-city variance and a Markov chain Monte Carlo estimation procedure (31), as in the papers by Dominici et al. (7, 13). Relative to `glm + ns`, we observe that 1) `gam + s + default` gives a larger pooled estimate and 2) `gam + s` with restricted convergence criteria gives a similar but slightly larger pooled estimate. The pooled estimates show little sensitivity to the specific pooling procedure used. The Markov chain Monte Carlo method provides slightly more conservative confidence intervals.

SIMULATION STUDY FOR POOLED ANALYSES

To characterize precisely the extent of bias in the pooled NMMAPS results, we performed a second simulation study. We first estimated the 90 city-specific relative rates of mortality at lag 1 from the NMMAPS database by using `glm + ns` with restricted convergence parameters and with the same set of confounding factors used in the NMMAPS anal-

yses (7, 13, 15). More specifically, we replaced all of the smoothing splines in the NMMAPS GAM model (7, 13) with natural cubic splines that had the same number of degrees of freedom. The spline knots were equally spaced at appropriate quantiles of the distribution of each covariate.

For each city, we obtained the maximum likelihood estimate of the linear predictor $\mu_i^c = \exp\{(\hat{\beta}^c \text{PM}_{10i} + \sum_j \text{ns}_j^c(x_{ji}, \lambda_j))\}$ using a fully parametric version of the GAM of Dominici et al. (7, 13). We then pooled the 90 city-specific estimates ($\hat{\beta}^c$) using a two-stage hierarchical normal model with a noninformative prior on the across-city variance. The pooled estimate under this model is 0.21 percent per $10\text{-}\mu\text{g}/\text{m}^3$ increase in PM_{10} (posterior standard error 0.06), as reported above.

For each city, we then generated 100 mortality time series Y_i^c from an overdispersed Poisson distribution with mean equal to the city-specific fitted values μ_i^c . We analyzed each city-specific simulated data set with `gam + s + default` and `glm + ns` to obtain two sets of 90×100 relative rate estimates and their standard errors. We then pooled the city-specific estimates across cities under a random-effect model (30) using an estimate of the across-city variance obtained from the Markov chain Monte Carlo procedure.

Figure 6 shows histograms for the 100 pooled estimates obtained under `glm + ns` (left) and under `gam + s + default` (right). The vertical lines are placed at the true value—that is, at the pooled estimate from the `glm + ns`. First, the distribution of the 100 pooled estimates under `glm + ns` is centered at the true pooled value (pooled estimate

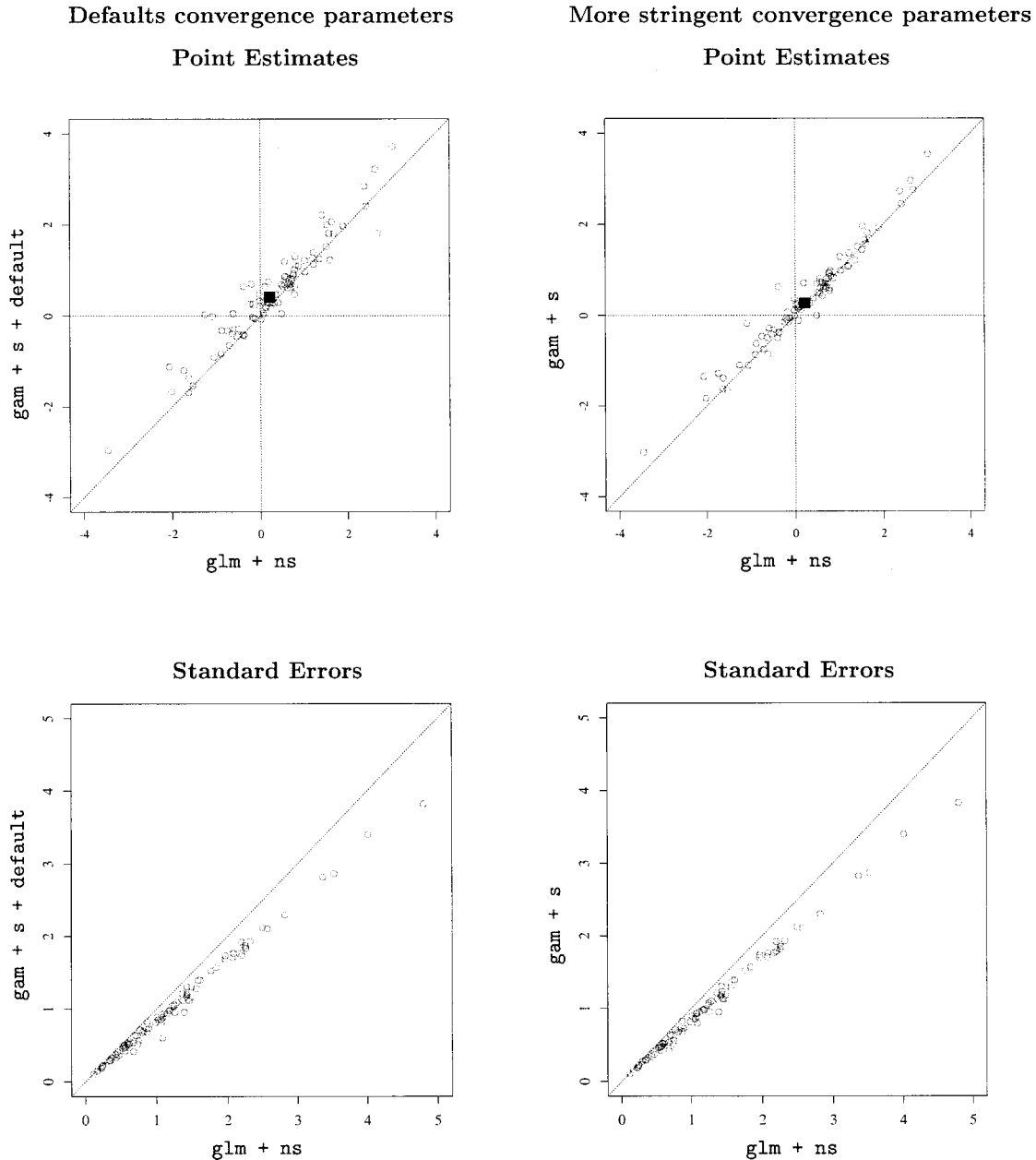


FIGURE 4. Relative rate estimates (top) and standard errors (bottom) for mortality across 90 US cities in the National Morbidity, Mortality, and Air Pollution Study. Shown are estimates obtained by fitting a Poisson regression model with parametric nonlinear adjustments for confounding factors $glm + ns$ versus estimates obtained by fitting generalized additive models with smoothing splines $gam + s + default$ (left) and by using more restrictive convergence criteria $gam + s$ (right). The black square is plotted at the pooled estimate across the 90 cities; its size corresponds to ± 2 standard errors of the pooled estimates.

from $glm + ns$), indicating that the $glm + ns$ model provides unbiased estimates not only of the city-specific effects but also of the pooled effects. Second, as in the simulation study for the city-specific analyses, the 100 pooled estimates under $gam + s$ are all larger than the true value, showing a bias similar to that indicated by the simulation study for city-specific analyses.

DISCUSSION

In this analysis, we examined the effects of using default convergence criteria options in S-Plus gam software on time-series studies of daily mortality counts and air pollution. Our simulations using time-series data for Pittsburgh (1987–1994) and a GAM model with two smooth functions (smoothing splines or LOESS) suggested that GAMs with

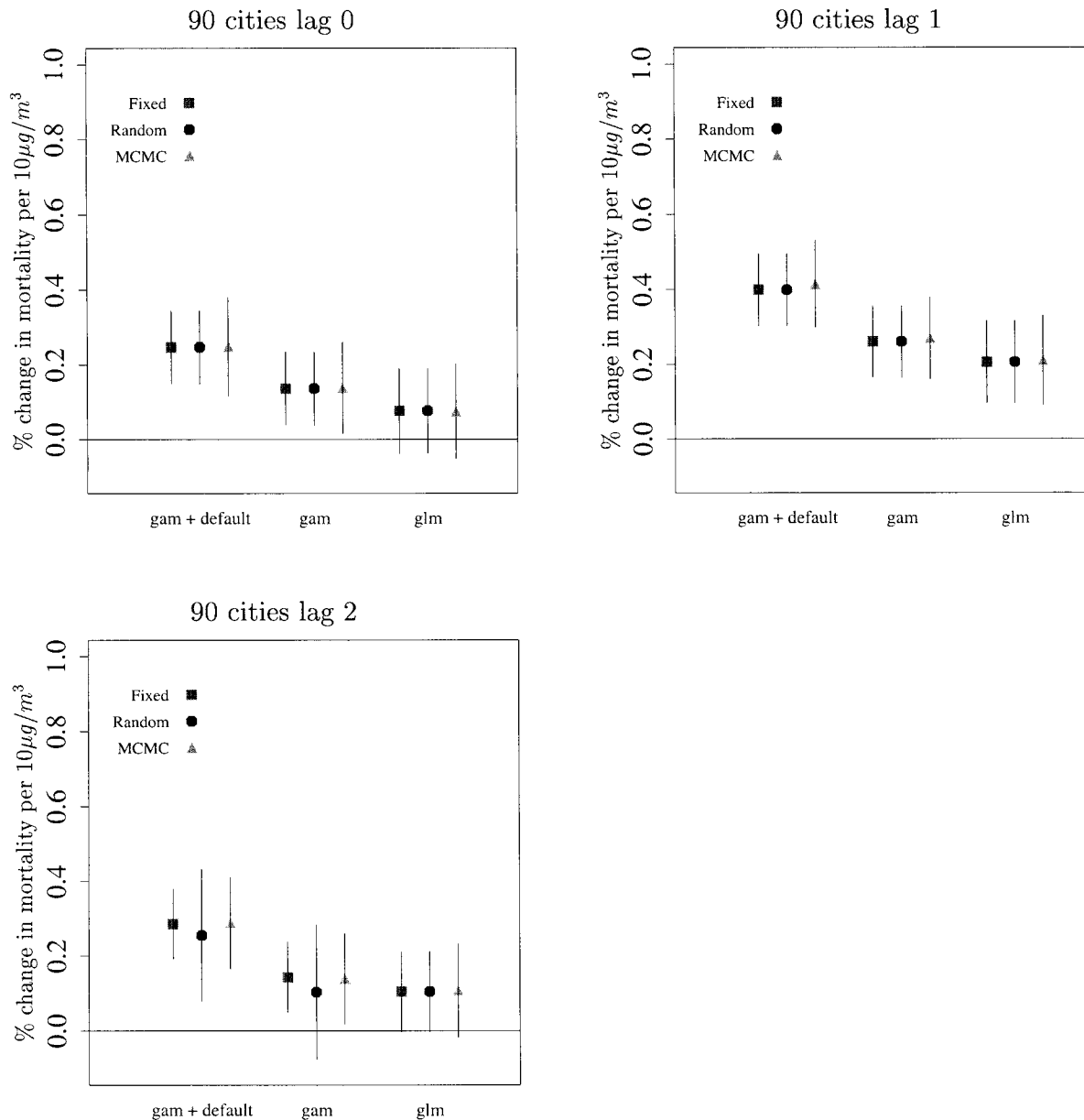


FIGURE 5. Pooled relative rate estimates for mortality across 90 US cities in the National Morbidity, Mortality, and Air Pollution Study under three models: 1) $\text{gam} + \text{s} + \text{default}$, 2) $\text{gam} + \text{s}$, and 3) $\text{glm} + \text{ns}$. The 90 city-specific estimates obtained under each model were pooled across cities under 1) a fixed-effect model, 2) a random-effect model (30) with a moment estimator of the across-city variance, and 3) a Bayesian two-stage model with a noninformative prior for the across-city variance and a Markov chain Monte Carlo estimation procedure (31), as in the paper by Dominici et al. (7).

default convergence criteria overestimate the true parameter by 36–42 percent and that such bias is larger when 1) LOESS smoothers are used instead of smoothing splines, 2) the magnitude of the regression coefficients is small (e.g., <0.001), and 3) concurvity among the time-series covariates (32)—the nonparametric analog of multicollinearity—is greater.

We observed that for time-series studies of air pollution and mortality, the bias is dependent on the nonparametric adjustment for confounding factors such as trend, season-

ality, and weather. A smaller number of degrees of freedom in the smoothing spline or a larger span in LOESS leads to larger estimated relative rates, smaller concurvity, and therefore less bias from lack of convergence. For example, in our simulations for Pittsburgh, we used 7 degrees of freedom per year. If we change the number of degrees of freedom to 3, the bias drops from 36 percent to 22 percent. However, if we compare the relative bias for 3 versus 7 degrees of freedom for a fixed particle effect and concurvity, we find greater bias with fewer degrees of freedom.

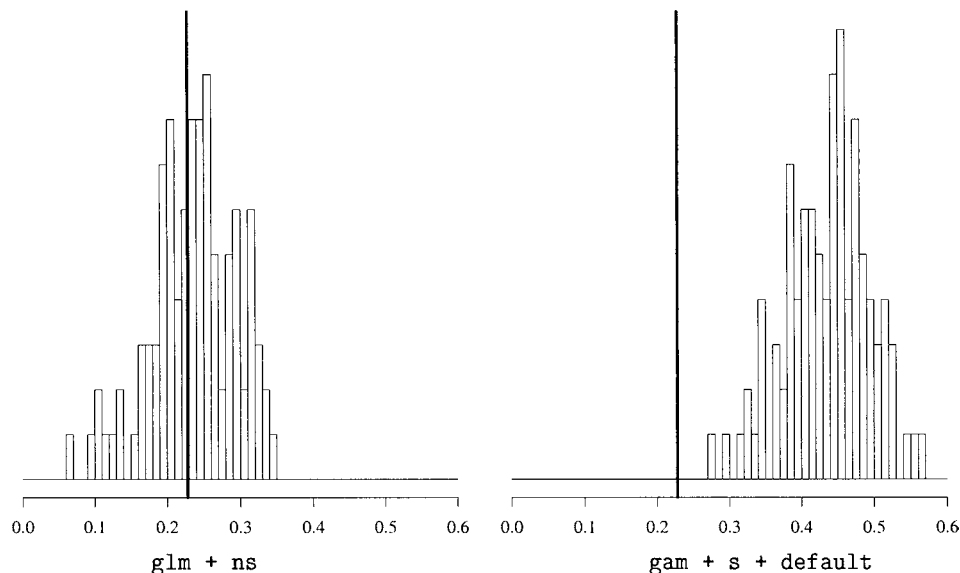


FIGURE 6. Percent change in mortality per 10- $\mu\text{g}/\text{m}^3$ increase in particulate matter less than 10 μm in diameter for 100 pooled estimates from a simulation study using `glm + ns` (left) and `gam + s + default` (right). Solid vertical lines indicate the true parameter values.

In simulation studies using NMMAPS data for the 90 largest US cities and using city-specific regression models with smooth functions (smoothing splines) to adjust for nonlinear confounders as described elsewhere (13, 15), we found that pooled relative rate estimates obtained under GLMs with natural cubic splines and iteratively reweighted least squares better detect true relative rates than GAMs with smoothing splines and default convergence parameters for the local scoring and the backfitting algorithm.

Although GAM with nonparametric smoothers provides a more flexible approach for adjusting for nonlinear confounders compared with fully parametric alternatives in time-series studies of air pollution and health, we have found that the use and implementation of GAMs requires extreme caution. Specifically: 1) default convergence parameters need to be modified; 2) GAM optimizes a penalized likelihood that can itself lead to increased bias in pollution effect estimates in exchange for decreased variance; and 3) even when the convergence of the backfitting algorithm is guaranteed, failure of the GAM to detect concavity can also lead to underestimation of the standard error of relative rate estimates (29). The NMMAPS reanalysis empirically confirmed the theoretical results of Ramsey et al. (29). It showed that the degree of bias in the standard errors is proportional to the size of the standard errors and that this underestimation remains even when more stringent convergence parameters are used.

Imposing stricter convergence criteria altered the NMMAPS results quantitatively but not qualitatively. In the reanalysis, the pooled estimate across 90 cities at lag 1 moved from a 0.41 percent (posterior standard error 0.05) increase in total mortality for a 10-unit increase in PM_{10} to a 0.27 percent (posterior standard error 0.05) increase. When

GLMs with natural cubic splines were used, the pooled estimate was 0.21 percent (posterior standard error 0.06). In every analysis, however, there was strong evidence for a positive association between acute exposure to PM_{10} and death, even with very conservative adjustments for trend, seasonality, and weather. The differences among these pooled time-series estimates are also small relative to the difference between the effects of acute exposures obtained from the time-series studies (like NMMAPS) and the effects of chronic exposures estimated from cohort studies (33–36).

The findings of this analysis in no way diminish the utility of GAM and other nonparametric regression techniques in epidemiologic or other research. A significant advantage of nonparametric regression is that epidemiologists need not rely on difficult-to-verify assumptions about the functional form of the dependence of the outcome on individual risk factors or confounders. To guard against the problems identified here, convergence criteria must be made substantially more stringent by users and/or distributors of statistical software.

Use of default settings was standard practice in environmental epidemiology until we started investigating this issue. Researchers can guard against such problems by regularly assessing the sensitivity of their findings to the convergence criteria used. An excellent overview of software reliability for S-Plus, SAS, and SPSS is provided by McCulloch (37, 38).

As a result of the simulation studies described here, the default parameters in the `gam` function of S-Plus, version 6.1, have already been revised, with somewhat more stringent convergence parameters ($\epsilon = 10^{-7}$, $M = 30$, $\epsilon_{bf} = 10^{-7}$, $M_{bf} = 30$) (39). This problem is likely to be shared by other software packages.

ACKNOWLEDGMENTS

This research was partially supported by the Health Effects Institute (Boston, Massachusetts), an organization jointly funded by the Environmental Protection Agency (grant EPA R824835) and automotive manufacturers. Funding for Dr. Francesca Dominici was provided by a grant (the Walter A. Rosenblith New Investigator Award) from the Health Effects Institute. Funding was also provided by the Johns Hopkins Center in Urban Environmental Health (grant 5P30ES03819-12).

The authors thank Professor Trevor Hastie and Drs. Rick Burnett, Tim Ramsay, Thomas Lumley, Lianne Sheppard, David Smith, Rafael Irizarry, and Giovanni Parmigiani for their comments.

REFERENCES

- Hastie T, Tibshirani R. Generalized additive models. *Stat Sci* 1986;1:297-318.
- Hastie TJ, Tibshirani RJ. Generalized additive models. New York, NY: Chapman and Hall, Inc, 1990.
- Schwartz J. Nonparametric smoothing in the analysis of air pollution and respiratory illness. *Can J Stat* 1994;22:471-88.
- Kelsall JE, Samet JM, Zeger SL, et al. Air pollution and mortality in Philadelphia, 1974-1988. *Am J Epidemiol* 1997;146:750-62.
- Schwartz J. Air pollution and hospital admissions for heart disease in eight US counties. *Epidemiology* 1999;10:17-22.
- Samet JM, Dominici F, Curriero F, et al. Fine particulate air pollution and mortality in 20 U.S. cities: 1987-1994 (with discussion). *N Engl J Med* 2000;343:1742-57.
- Dominici F, Samet JM, Zeger SL. Combining evidence on air pollution and daily mortality from the twenty largest US cities: a hierarchical modeling strategy (with discussion). *J R Stat Soc Ser A* 2000;163:263-302.
- Katsouyanni K, Toulomi G, Samoli E, et al. Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology* 2001;12:521-31.
- Moolgavkar S. Air pollution and hospital admissions for diseases of the circulatory system in three U.S. metropolitan areas. *J Air Waste Manag Assoc* 2000;50:1199-206.
- Schwartz J. Assessing confounding, effect modification, and thresholds in the associations between ambient particles and daily deaths. *Environ Health Perspect* 2000;108:563-8.
- Samoli E, Schwartz J, Wojtyniak B, et al. Investigating regional differences in short-term effects of air pollution on daily mortality in the APHEA project: a sensitivity analysis for controlling long-term trends and seasonality. *Environ Health Perspect* 2002;109:349-53.
- Moolgavkar S. Air pollution and daily mortality in three U.S. counties. *Environ Health Perspect* 2002;108:777-84.
- Dominici F, Daniels M, Zeger SL, et al. Air pollution and mortality: estimating regional and national dose-response relationships. *J Am Stat Assoc* 2002;97:100-11.
- Samet JM, Zeger SL, Dominici F, et al. The National Morbidity, Mortality, and Air Pollution Study. Part I: methods and methodological issues. (Report no. 94.I). Cambridge, MA: Health Effects Institute, 1999.
- Samet JM, Zeger SL, Dominici F, et al. The National Morbidity, Mortality, and Air Pollution Study. Part II: morbidity and mortality from air pollution in the United States. (Report no. 94.II). Cambridge, MA: Health Effects Institute, 2000.
- McCullagh P, Nelder JA. Generalized linear models. 2nd ed. New York, NY: Chapman and Hall, Inc, 1989.
- Green PJ, Silverman BW. Nonparametric regression and generalized linear models: a roughness penalty approach. London, United Kingdom: Chapman and Hall Ltd, 1994.
- Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *J Am Stat Assoc* 1988;83:596-610.
- Cleveland R, Cleveland W, McRee JE. Seasonal-trend decomposition procedure based on LOESS. *J Offic Stat* 1990;6:3-73.
- Chambers J, Hastie T. Statistical models in S. London, United Kingdom: Chapman and Hall Ltd, 1992.
- de Boor C. A practical guide to splines. New York, NY: Springer-Verlag New York, 1978.
- Cheney W, Kincaid D. Numerical mathematics and computing. Pacific Grove, CA: Brooks/Cole, 1999.
- Friedman JH, Stuetzle W. Projection pursuit regression. *J Am Stat Assoc* 1981;76:817-23.
- Buja A, Hastie T, Tibshirani R. Linear smoothers and additive models. *Ann Stat* 1989;17:453-5.
- Nelder JA, Wedderburn RW. Generalized linear models. *J R Stat Soc Ser A* 1972;135:370-84.
- Hastie T, Ikeda D, Tibshirani R. Bayesian backfitting (with discussion). *Stat Sci* 2000;15:193-223.
- Opsomer J, Ruppert D. Fitting a bivariate additive model by local polynomial regression. *J Am Stat Assoc* 1997;25:186-211.
- Opsomer J. Asymptotic properties of backfitting estimators. *J Multivar Anal* 2000;73:166-79.
- Ramsay T, Burnett R, Krewski D. The effect of concurrency in generalized additive models linking mortality and ambient air pollution. *Epidemiology* (in press).
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
- Gilks WR, Richardson S, Spiegelhalter DJ, eds. Markov chain Monte Carlo in practice. London, United Kingdom: Chapman and Hall Ltd, 1996.
- Donnell DJ, Buja A, Stuetzle W. Analysis of additive dependencies and concurrencies using smallest additive principal components. *Ann Stat* 1994;22:1635-73.
- Dockery D, Pope CA, Xu X, et al. An association between air pollution and mortality in six U.S. cities. *N Engl J Med* 1993;329:1753-9.
- Pope CA, Thun M, Namboodiri M, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med* 1995;151:669-74.
- Krewski D, Burnett RT, Goldberg MS, et al. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Cambridge, MA: Health Effects Institute, 2000.
- Pope CA, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002;287:1132-41.
- McCullough B. Assessing the reliability of statistical software: part I. *Am Stat* 1998;52:358-66.
- McCullough B. Assessing the reliability of statistical software: part II. *Am Stat* 1999;53:149-59.
- Insightful Corporation. How to tighten default convergence criteria for gam and glm. Seattle, WA: Insightful Corporation, 2000. (<http://www.insightful.com/support/faqdetail.asp?FAQID=139andIsArchive=0>).
- Stone CJ. Additive regression and other nonparametric models. *Ann Stat* 1985;13:689-705.

41. Stone CJ. The dimensionality reduction principle for generalized additive models. *Ann Stat* 1986;14:590–606.

APPENDIX

Generalized Additive Models and the Local Scoring Algorithm

Here we briefly review the local scoring algorithm used to fit a generalized additive model (GAM) and its implementation for air pollution time-series data. The local scoring algorithm (2, 24) is analogous to the use of iteratively reweighted least squares (16, 25) for solving likelihood and nonlinear regression equations. At each iteration, an adjusted dependent variable is formed and an additive regression model is fitted. The estimation procedure of the additive regression model depends on the number and nature of the smooth functions included in the GAM. When the smooth functions are parametric (e.g., regression splines), the additive regression model is fitted using weighted least squares, and the GAM is equivalent to a generalized linear model. When the smooth functions in the linear predictors are nonparametric (smoothing splines or LOESS), the additive regression model is fitted using the backfitting algorithm (40, 41). The backfitting algorithm cycles through the variables and estimates each coordinate function by smoothing partial residuals. If a single smooth function is included in the model, the backfitting algorithm provides a closed-form solution of the parameter estimates.

In most time-series analyses of air pollution and health that have used GAM, the models include several nonparametric smooth functions of calendar time and weather variables. More specifically, let y_t be the daily air pollution counts and let x_{1t} , x_{2t} , and x_{3t} be the daily time series of air pollution, calendar time, and temperature, respectively. A typical model assumes

$$y_t \sim \text{Poisson}(\mu_t) \\ \log \mu_t = \alpha + \beta x_{1t} + s_2(x_{2t}, \lambda_2) + s_3(x_{3t}, \lambda_3) = \eta_t, \quad (3)$$

where β denotes the relative rate of mortality associated with a 10- $\mu\text{g}/\text{m}^3$ change in air pollution levels and $s_2(x_{2t})$ and $s_3(x_{3t})$ are nonparametric smooth functions with degrees of freedom λ_2 and λ_3 modeled as smoothing splines or LOESS functions.

The local scoring algorithm for model 3 consists of the following steps:

- Initialize: $\alpha = \beta = s_2 = s_3 = 0$ and $m = 0$, where $s_j = (s_j(x_{jt}), \dots, s_j(x_{jT}))$, $j = 2, 3$.

- Iterate: $m = m + 1 < M$ (outer loop).

1. Form the adjusted dependent variable:

$$z = \eta^{m-1} + (y - \mu^{m-1})\partial\eta/\partial\mu^{m-1},$$

where $\eta^{m-1} = \alpha^{m-1} + \beta^{m-1}\mathbf{x} + s_2^{m-1} + s_3^{m-1} = \log(\mu^{m-1})$ and $\mathbf{z} = (z_1, \dots, z_T)$, $\mathbf{x} = (x_1, \dots, x_T)$ (similarly for η, μ, \mathbf{y}).

2. Form the weights $\mathbf{w} = (\partial\mu/\partial\eta^{m-1})^2 V^{-1}$, where V is the variance of \mathbf{y} at μ^{m-1} .

3. Fit an additive model to \mathbf{z} using the backfitting algorithm with weights \mathbf{w} and estimate $\alpha^m, \beta^m, s_2^m, s_3^m$, and η^m , as follows:

- 3.1. Cycle $j = 1, 2, 3$ (inner loop).

- 3.2. Calculate residuals by removing the estimated functions or covariate effects of all of the other variables:

$$r_{t1} = y_t - s_2^{m-1} - s_3^{m-1} \\ r_{t2} = y_t - \alpha^{m-1} - \beta^{m-1}\mathbf{x} - s_3^{m-1} \\ r_{t3} = y_t - \alpha^{m-1} - \beta^{m-1}\mathbf{x} - s_2^{m-1}, t = 1, \dots, T$$

- 3.3. Estimate the s_j^m by smoothing the residuals with respect to the next covariate:

$$\hat{s}_2^m(x_{1t}) = \text{smooth}(r_{2t}|x_{2t}) \\ \hat{s}_3^m(x_{1t}) = \text{smooth}(r_{3t}|x_{3t})$$

The term “smooth($r_j|x_{jt}$)” denotes a smoothing of the data (r_j, x_{jt}) at the point x_{jt} . The parameter estimates (α^m, β^m) are obtained by fitting weighted least squares on the data (r_1, x_1).

- 3.4. Compute the backfitting convergence criterion:

$$\text{RSS}^m = \frac{1}{T} \sum_{t=1}^T \left(y_t - \alpha^m - \beta^m x_{1t} - \sum_{j=2}^3 s_j^m(x_{jt}) \right)^2$$

- 3.5. Stop when $|\text{RSS}^m - \text{RSS}^{m-1}| < \epsilon_{bf}$.

- Compute the local scoring convergence criterion:

$$\Delta^m = |(D^{m-1} - D^m)/(D^{m-1} + \epsilon)|,$$

where $D^{m-1} = D(\mathbf{y}; \eta^{m-1})$ is the deviance for a fitted model η^m (2).

- Stop for $\Delta^m < \epsilon$.