Biostatistics (2000), **1**, 3, *pp*. 279–292
Printed in Great Britain

Generalized additive distributed lag models: quantifying mortality displacement

A. ZANOBETTI*

Department of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA, 02115, USA and Dipartimento di Statistica, Universita di Firenze, Firenze, Italy

M. P. WAND, J. SCHWARTZ

Harvard School of Public Health, Boston, MA, USA

L. M. RYAN

Dana Farber Cancer Institute, Boston, MA, USA

SUMMARY

There are a number of applied settings where a response is measured repeatedly over time, and the impact of a stimulus at one time is distributed over several subsequent response measures. In the motivating application the stimulus is an air pollutant such as airborne particulate matter and the response is mortality. However, several other variables (e.g. daily temperature) impact the response in a possibly non-linear fashion. To quantify the effect of the stimulus in the presence of covariate data we combine two established regression techniques: *generalized additive models* and *distributed lag models*. Generalized additive models extend multiple linear regression by allowing for continuous covariates to be modeled as smooth, but otherwise unspecified, functions. Distributed lag models aim to relate the outcome variable to lagged values of a time-dependent predictor in a parsimonious fashion. The resultant, which we call *generalized additive distributed lag models*, are seen to effectively quantify the so-called 'mortality displacement effect' in environmental epidemiology, as illustrated through air pollution/mortality data from Milan, Italy.

Keywords: Environmental epidemiology; Non-parametric regression; Smoothing; Time series.

1. INTRODUCTION

In the early 1900s a number of studies reported associations between air pollution and daily deaths (Schwartz and Dockery, 1992; Schwartz, 1991; Fairley, 1990). Those findings led to a large number of followup studies (Pope *et al.*, 1995a). Most of these studies (Katsouyanni *et al.*, 1997; Anderson *et al.*, 1997; Schwartz, 1994b; Schwartz and Dockery, 1992) examined the association with pollution on the mean of several days. Presumably, the impact of a unit change in pollution appears with some latency, and persists for some time after exposure. That is, the effect in an individual is distributed over time. While most studies found multi-day averages of pollution are better predictors of daily death counts than a single

*To whom correspondence should be addressed

day's exposure, only one has systematically examined the shape of the distributed lag (Schwartz, 2000b). One method of modeling such behavior is through a linear regression model with lags of the stimulus as covariates. But since the number of lags can be relatively high, and the exposure is serially correlated, ordinary linear regression is susceptible to collinearity-related problems. A remedy is the *distributed lag model* (DLM) first proposed by Almon (1965) which restricts the coefficients to being a low-degree polynomial in the lags. Recently, parametric distributed lag models have been used to determine the association over time between air pollution and mortality (Pope and Schwartz, 1996; Schwartz, 2000b). An extension of this idea, which aims for more flexibility than that afforded by polynomial functions, is to subject the coefficients to non-parametric smoothing. This approach has been investigated, through the use of spline functions, by Corradi (1977).

These papers have focused on DLMs examining the effect of air pollution exposure over a few days on daily deaths. If one wishes to examine effects at substantially longer lags, a smoothing approach is even more advantageous. We will motivate such an application below.

While polynomial and, in particular, spline-based DLMs are flexible enough to model the effect of the stimulus of interest, it is usually advantageous to incorporate other covariate information. For example, in air pollution studies weather and seasonality have confounding effects on mortality. To allow their entry into a DLM in a flexible fashion we combine the model with the *generalized additive model* popularized by Hastie and Tibshirani (1990). Generalized additive models were introduced into time series epidemiology studies in 1993 (Schwartz, 1993, 1994a), and have since become the standard approach to this covariate control problem (Schwartz, 1994b; Hoek *et al.*, 1997; Kelsall *et al.*, 1997; Bremner *et al.*, 1999; Dominici *et al.*, 2000). Generalized additive models are an extension of generalized linear models that allow for the modeling of nonlinear effects through the use of smoothing. We call the result a *generalized additive DLM*.

The following section provides some background on the motivational application: the 'mortality displacement issue' from environmental epidemiology. While the association between particulate matter and mortality or morbidity is generally accepted, controversy remains about the importance of the associations. If it is due solely to the deaths of frail individuals, which are brought forward by only a brief period of time, the public health implications of the associations are smaller. Recently some papers have addressed this issue, using different approaches to investigate the 'mortality displacement effect' or 'harvesting effect' (Zeger *et al.*, 1999; Schwartz, 2000a).

Zeger *et al.* (1999) used frequency domain log–linear regression and estimated a mortality–air-pollution association that is resistant to short-term harvesting. Schwartz (2000a) used filtering in the time domain to look at how the mortality–air-pollution association varies as the time scale of the exposure varies, using the STL algorithm (Cleveland *et al.*, 1990) to decompose the time series. In addition Spix (Spix *et al.*, 1993, Spix, *Daily time-series of mortality counts: estimating the harvesting effect*. (submitted for publication)) has addressed a related issue. If the pool of susceptible individuals varies due to other risk factors, such as cold weather or influenza, the apparent risk due to pollution may change. She estimated the size of the risk pool change using moving averages of mortality and examined an interaction with air pollution.

This paper takes an alternative, and in some ways more direct, approach to estimating the effect of air pollution on daily deaths net of any short-term rebounds due to mortality displacement, as well as any longer-term lagged effect.

In Section 2 we describe the mortality displacement issue in more detail to motivate the model. In Section 3.1 we describe the data and develop models geared towards quantification of the mortality displacement. Analysis of air pollution data from Milan, Italy, and a new quantification of the mortality displacement is given in Section 4.

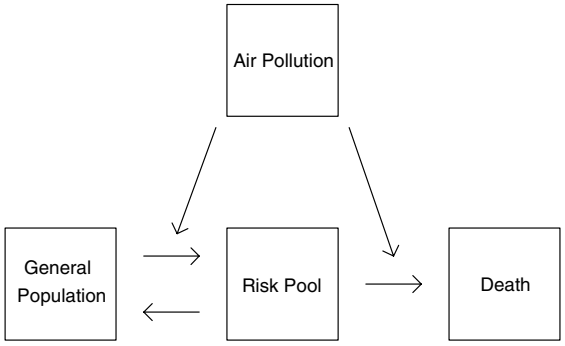

Fig. 1. The mechanics that lead to the mortality displacement.

2. MOTIVATION: THE 'MORTALITY DISPLACEMENT'

As noted, a number of analyses have estimated the relative risk of mortality per a given change in air pollution. The public health significance of short-term effects of air pollution varies depending on the extent of time the deaths are being advanced.

This question interacts directly with the distributed lag mentioned above. In an individual, the risk of air pollution exposure is presumably positive or zero at all lags. This need not be true in a population. Assume that at any given time there is a pool of people at elevated risk of dying because of acute or chronic conditions. Then an air pollution episode could increase the risk for those individuals, increasing the death rate out of the pool, and decreasing the pool size. On subsequent days, the number of deaths observed will be less than expected, inducing a negative association with air pollution at those lags. The finite nature of the risk pool therefore creates the possibility of this negative association, with the result that many of the early deaths may only be advanced by a short period. This is sometimes referred to as mortality displacement and this pathway is illustrated in the right side of Figure 1, where air pollution effects the rate of transition out of the risk pool to death. While previous approaches have sought to net out these negative effects in the course of estimating the pollution impact (Zeger *et al.*, 1999; Schwartz, 2000a), our approach seeks to directly estimate any negative association at longer lags.

In reality the relationship of air pollution to the risk pool is more complex. Zelikoff *et al.* (1999), for example, have shown that exposure to airborne particles substantially exacerbates pneumonia in rats. Similar results have been reported for influenza infections in mice (Clarke *et al.*, 1997). Toxicological data also suggest that effects of exposure may be seen over several subsequent days. For example, Clarke *et al.* (1999) reported changes in tidal volume of rats immediately following exposure to concentrated ambient air particles, and increases in inflammatory markers approximately 36 hours following exposure. Gold and coworkers (Gold *et al.*, 2000) have seen changes in electrocardiograms within hours of exposure to airborne particles. However, the strongest associations of air pollution with defibrillator discharges was with a two-day lag (Peters *et al.*, 2000). Hence air pollution could shift the distribution of intensity of many illnesses to the right, increasing the number of persons whose illness is now life threatening. This may occur with multiple lags depending on the mechanism and the individual. Such impact is indicated on the left side of Figure 1, where air pollution potentially increases the rate of transition into the risk pool, or the rate of recovery out of it. Moreover, the effect of air pollution on increasing recruitment into the risk pool may occur with different lags than the effect on increasing the transition to death.

Figure 2 shows the type of lag structure that arises when there is harvesting. The horizontal axis is the lag number and the vertical axis shows the coefficients β_ℓ. Each β_ℓ has the interpretation: effect of

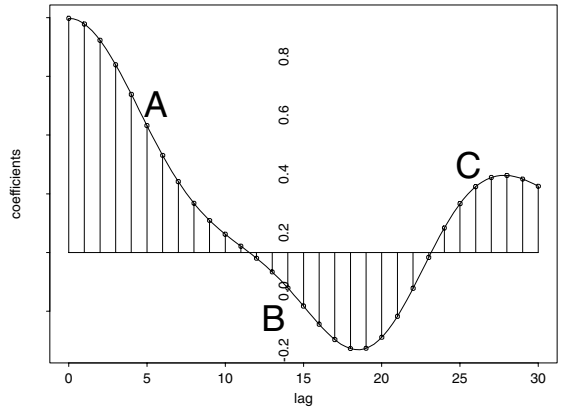

Fig. 2. Lag structure corresponding to the mortality displacement effect.

pollution level ℓ days ago on mortality today. (This is made mathematically precise in the models given in Section 3.3.)

A is the sum of the positive coefficients for low lags and represents the fact that pollution levels in the past few days have a positive effect on mortality. This is what most previous studies have reported. However, the negative coefficients in B mean that pollution levels a longer period ago have a negative effect. This 'rebound effect' is due to the depletion of the risk pool, normally made up of elderly and sick people who have had their death brought forward a few days or weeks by episodes of high pollution.

Most discussions of mortality displacement end by stating that A overestimates the public health significance of pollution, since it is really $A + B$ (where B is negative) that represents deaths brought forward by a noticeable amount of time. However, if the increased risk to individuals of an air pollution episode die out slowly over time, or if there is increased recruitment into the risk pool due to air pollution, and this occurs more slowly than the effects on mortality illustrated in period A, then we may also have a period C, subsequent to the harvesting period B, where the regression coefficients are once more positive. The cumulative effect is then $A + B + C$. To examine this possibility we need to allow flexibility for a relatively long and nonlinear distributed lag. However, the total explanatory power of air pollution on daily mortality is quite modest, making it likely that high degree polynomials will be estimated poorly. The desire for parsimony in a flexible model is also a motivating factor for this paper.

3. DATA AND MODELS

3.1. Description of the data

Data from a study in Milan, Italy, are available to study the mortality displacement. These are daily time series of mortality, meteorological variables, and air pollution. The Milan study was conducted in the central urban area of the city for the ten-year period 1980–89. The mean population of the municipality of Milan, during the study period, was 1.5 million inhabitants. The mortality data are daily counts derived from death certificates and consist of residents of Milan who died in the city from natural causes (International Classification of Diseases, ICD, ninth revision: 1-799). The mean daily deaths in the 10 years is 32. For this illustration we use total suspended particulate (`TSP`) as our pollution measure. Meteorological data are expressed by mean temperature and relative humidity. Further details may be found in Vigotti *et al.* (1996) and in Rossi *et al.* (1999).

Table 1. *Summary statistics of Milan mortality data*

	Average	Max	Min	S.dev
Mean temperature	14	32	−6	8
Relative humidity (%)	62.0	99.7	0	17
TSP	14.2	529	4	81
Total mortality (ICD-9: 1-799)	32	66	10	7.7

TSP was measured by the β attenuation method and was available only in two stations and the mean was recorded if both were present. The correlations between monitors was high (0.89–0.91). After the daily mean was computed 19% of the values for TSP were missing.

Although these are the only data analysed in this paper, there are many types of data with similar features that could benefit from the model and fitting procedure that we develop to analyze the Milan data. The essential features are: a response, a stimulus whose effect on the response is distributed across several time lags and measurements on possible confounders.

3.2. Imputation of missing data

In the time series of the air pollutant, missing values are present because the stations were not always working during the 10 years. Some of our proposed models involve 45 lags of air pollution, so it is important to impute the missing data. Since air pollution varies seasonally, and also with weather parameters, a generalized additive model was fit estimating air pollution as smooth functions of these variables. Weather fronts usually persist for just a few days, and strongly influence pollution levels. This usually induces autocorrelation in pollution values. The partial autocorrelation function of the deviance residuals of our model showed a high value at lag 1. Therefore the one-day-lagged pollutant was introduced in the model to correct for an autocorrelation of order one and improve model fit. The final model is:

$$\texttt{air.pollutant}_t = \theta\,\texttt{air.pollutant}_{t-1} + f_1(\texttt{min.temp}_t) + f_2(\texttt{rel.humid}_t) + f_3(t) + \texttt{error}_t$$

This model explained about 70% of the variability of the dependent variable. If more than one consecutive day's pollution values were missing, the model fitting was iterated.

3.3. Smooth DLMs

Let $(x_1, y_1), \ldots, (x_T, y_T)$ denote a data set that is ordered with respect to time. In the motivating example, x_t represents a daily pollution measure and y_t represents a mortality count for the same day. Additional covariates may be measured, but we ignore these for now for the purposes of developing our approach. The *unconstrained* distributed lag model of order q is

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + \varepsilon_t, \qquad t = q+1, \ldots, T \tag{1}$$

where the ε_t are independent random variables with mean zero and constant variance (Almon, 1965). The idea here is the outcome y_t at time t may depend on x_t measured not only on the current day, but also on previous days. Note that the overall impact of a unit change in exposure over the next q days is given by $\sum_{\ell=0}^{q} \beta_\ell$.

Because of collinearity-related problems it is common to constrain the β_ℓ to be a simple function of the lag number. This constraint can be generalized to the polynomial DLM of order (q, p) (PDLM(q, p))

where the β_ℓ are constrained to be a pth degree polynomial function of ℓ. It is straightforward to formulate least squares procedures for fitting such a model: see, for example, Davidson and MacKinnon (1993, pp. 674–676).

The PDLM(q, p) imposes the restriction that β_ℓ be a polynomial function of ℓ:

$$\beta_\ell = \sum_{j=0}^{p} \tau_j \ell^j, \quad \ell = 0, \ldots, q. \tag{2}$$

As with ordinary regression, polynomial models are not suitable for modeling more localized structure (see, e.g. Hastie and Tibshirani, 1994). We propose to overcome this by extending (2) to

$$\beta_\ell = \sum_{j=0}^{p} \tau_j \ell^j + \sum_{k=1}^{K} \nu_k (\ell - \kappa_k)_+^p \tag{3}$$

where $\kappa_1, \ldots, \kappa_K$ is a set of K distinct numbers between 0 and q. Note that β_ℓ is a piecewise pth degree polynomial in ℓ, with join points, or *knots* at the κ_k.

Equation (3) is commonly referred to as a *regression spline* function of ℓ. However, estimation of the coefficients via ordinary least squares will normally lead to a β_ℓ function that is too wiggly. This problem can be overcome by restricting the size of the ν_k. This can be achieved through the addition of a penalty term $\lambda \sum_{k=1}^{K} \nu_k^2$ to the least squares criterion. This strategy has become known as *penalized spline* (or P-spline) smoothing (e.g. Eilers and Marx, 1996).

The full prescription for estimation of the coefficients is:

$$\text{minimize} \sum_{t=q+1}^{T} (y_t - \alpha - \beta_0 x_t - \cdots - \beta_q x_{t-q})^2 + \lambda^2 \sum_{k=1}^{K} \nu_k^2, \quad \text{where } \beta_\ell \text{ is as given in (3).}$$

The solution is easiest to express in matrix notation. Let $\mathbf{X}$ be the $(T - q) \times (q + 2)$ design matrix corresponding to (1),

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0^p & (0-\kappa_1)_+^p & \cdots & (0-\kappa_K)_+^p \\ 0 & 1 & 1 & \cdots & 1^p & (1-\kappa_1)_+^p & \cdots & (1-\kappa_K)_+^p \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & q & \cdots & q^p & (q-\kappa_1)_+^p & \cdots & (q-\kappa_K)_+^p \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \mathbf{0}_{(p+2)\times(p+2)} & \mathbf{0}_{(p+2)\times K} \\ \mathbf{0}_{K\times(p+2)} & \mathbf{I}_{K\times K} \end{bmatrix}.$$

If $\boldsymbol{\theta} = [\alpha, \tau_0, \tau_1, \ldots, \tau_p, \nu_1, \ldots, \nu_K]^\mathsf{T}$ then our fitting criterion is

$$\|\mathbf{y} - \mathbf{X}\mathbf{T}\boldsymbol{\theta}\|^2 + \lambda^2 \boldsymbol{\theta}^\mathsf{T} \mathbf{D} \boldsymbol{\theta}$$

where $\|\mathbf{v}\| = \sqrt{\mathbf{v}^\mathsf{T}\mathbf{v}}$. The fitted $\boldsymbol{\beta}$ values are then

$$\widehat{\boldsymbol{\beta}}_\lambda = \mathbf{T}(\mathbf{T}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{T} + \lambda^2 \mathbf{D})^{-1} \mathbf{T}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{y}, \tag{4}$$

which we call the *penalized spline distributed lag estimate* of the β_ℓ.

3.4. Extension to generalized additive DLMs

In our main motivating data set we have

$$(x_t, y_t) = (\texttt{mortality}_t, \texttt{TSP}_t), \quad t = 1, \ldots, 3652$$

i.e. daily mortality and level of `TSP` for the city of Milan, Italy, for 3652 consecutive days. While a DLM as described by (1) and (4) might help describe the short-term effect of `TSP` levels on daily mortality, it would be weakened by the fact that it does not account for confounding due to, for example, meteorological variables such as temperature and relative humidity. Also, the fact that mortality is a count variable means that a distribution suited to count data, such as the Poisson, is needed.

These considerations lead to what we call a *generalized additive DLM*. The general form of such a model is

$$g\{E(y_t)\} = \alpha + \boldsymbol{\gamma}^\mathsf{T}\mathbf{z}_t + \sum_{j=1}^{d} f_j(s_{jt}) + \sum_{\ell=0}^{q} \beta_\ell x_{t-\ell}, \quad t = q+1, \ldots, T \tag{5}$$

where g is a link function, $\mathbf{z}_t$ is a vector of variables modeled linearly (these are often dummy variables), s_{jt} is the jth variable modeled as smooth function.

Estimation can be performed via maximum penalized likelihood, according to the distributional assumption for y_t. In this paper we concentrate on the case where y_t is from a one-parameter exponential family (e.g. Poisson, binomial, gamma). However, to account for over-dispersion this can be extended to quasi-likelihood models (see, e.g., Hastie and Tibshirani, 1990).

The model can be fit directly using penalized splines for estimation of the f_j. The starting point is the following X-matrix:

$$\mathbf{X} = [\mathbf{1}|\mathbf{X}_{\text{linear}}|\mathbf{X}_{\text{smooth}}|\mathbf{X}_{\text{lag}}]$$

where

$$\mathbf{X}_{\text{linear}} = \begin{bmatrix} \mathbf{z}_{q+1}^\mathsf{T} \\ \vdots \\ \mathbf{z}_T^\mathsf{T} \end{bmatrix}, \quad \mathbf{X}_{\text{smooth}} = [\mathbf{X}_{\text{smooth},1}|\cdots|\mathbf{X}_{\text{smooth},d}],$$

$$\mathbf{X}_{\text{smooth},j} = \begin{bmatrix} s_{j,q+1} & \cdots & s_{j,q+1}^p & (s_{j,q+1}-\kappa_{j1})_+^p & \cdots & (s_{j,q+1}-\kappa_{jK_j})_+^p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{jT} & \cdots & s_{jT}^p & (s_{jT}-\kappa_{j1})_+^p & \cdots & (s_{jT}-\kappa_{jK_j})_+^p \end{bmatrix}$$

and

$$\mathbf{X}_{\text{lag}} = \begin{bmatrix} x_{q+1} & x_q & \cdots & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ x_T & x_{T-1} & \cdots & x_{T-q} \end{bmatrix} \mathbf{U}, \quad \mathbf{U} = \begin{bmatrix} 1 & 0 & \cdots & 0^p & (0-\kappa_1)_+^p & \cdots & (0-\kappa_K)_+^p \\ 1 & 1 & \cdots & 1^p & (1-\kappa_1)_+^p & \cdots & (1-\kappa_K)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & q & \cdots & q^p & (q-\kappa_1)_+^p & \cdots & (q-\kappa_K)_+^p \end{bmatrix}. \tag{6}$$

The last K_j columns of $\mathbf{X}_{\text{smooth},j}$ correspond to knots in a regression spline model for the s_{jt} and, analogously to the smooth DLM (3), the corresponding coefficients need to be penalized in the fitting process to obtain smooth estimates of the f_j. This approach to smoothing is described in Eilers and Marx (1996) and Marx and Eilers (1998).

In the case of normal errors the model can be fit using the ridge regression

$$\widehat{\boldsymbol{\theta}} \equiv (\mathbf{X}^\mathsf{T}\mathbf{X} + \boldsymbol{\Lambda})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$$

where $\mathbf{y} = [y_{q+1}, \ldots, y_T]^\mathsf{T}$ and

$$\boldsymbol{\Lambda} = \text{diag}(0, \mathbf{0}, \mathbf{0}_p, \lambda_1^{2p}\mathbf{1}_{K_1}, \ldots, \mathbf{0}_p, \lambda_d^{2p}\mathbf{1}_{K_d}, 0, \mathbf{0}_p, \lambda^2).$$

Here $\mathbf{0}$ is a vector of zeroes of length the same as that of $\boldsymbol{\gamma}$, $\mathbf{0}_p$ is a vector of zeroes of length p and $\mathbf{1}_K$ is a vector of ones of length K. The λ_j, $j = 1, \ldots, d$, represent smoothing parameters for each of the j smooth components.

The estimated coefficients in $\widehat{\boldsymbol{\theta}}$ directly give estimates of α and $\boldsymbol{\gamma}$. It is also straightforward to form estimates of the f_j (for arbitrary abscissae) and the β_ℓ from $\widehat{\boldsymbol{\theta}}$. We adopt the convention

$$\sum_{t=q+1}^{T} \gamma_j z_{jt} = \sum_{t=q+1}^{T} f_j(s_{jt}) = 0 \tag{7}$$

so that the functions are vertically centered about zero and hence are interpretable as departures from the mean.

In the generalized situation iteratively reweighted least squares, with weights proportional to the current estimated variance of y, is required. The most challenging part of the algorithm is the generalized ridge regression. In the applications the size of the matrix to be inverted is approximately 100×100, so speed and numerical stability issues need to be factored in. We found that the QR-decomposition-based approach described, for example, in the Appendix of Hastie (1996) was suitable for meeting this challenge.

A noteworthy point about the fitting procedure is that the use of low-rank smoothers such as penalized splines circumvents the need for *backfitting*, the usual method for fitting additive models. This reduces the computational burden quite significantly, as pointed out by Hastie (1996) and Marx and Eilers (1998).

The covariance matrix of the estimated coefficients (not including the constant term) is approximately

$$\operatorname{cov}(\widehat{\boldsymbol{\theta}}_r) \simeq (\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}\mathbf{W}\widetilde{\mathbf{X}}_{\mathrm{r}} + \boldsymbol{\Lambda}_{\mathrm{r}})^{-1}\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}\operatorname{cov}(\mathbf{y})\widetilde{\mathbf{X}}_{\mathrm{r}}(\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}\mathbf{W}\widetilde{\mathbf{X}}_{\mathrm{r}} + \boldsymbol{\Lambda}_{\mathrm{r}})^{-1}$$

where $\mathbf{W} = \operatorname{diag}(\mathbf{w})$ contains the weights at convergence. If the y_t are (conditionally) uncorrelated and belong to a one-parameter exponential family then $\operatorname{cov}(\mathbf{y})$ can be estimated by $\mathbf{W}$.

It is straightforward to obtain estimates of the standard errors of the $\hat{f}_j$ and $\hat{\beta}_\ell$. For example,

$$\widehat{\mathrm{SE}}(\hat{\beta}_\ell) = \sqrt{\ell\text{th diagonal entry of } \{\mathbf{U}\operatorname{cov}(\widehat{\boldsymbol{\theta}}_r)_{\mathrm{lag}}\mathbf{U}^{\mathsf{T}}\}}$$

where $\operatorname{cov}(\widehat{\boldsymbol{\theta}}_r)_{\mathrm{lag}}$ is the block of $\operatorname{cov}(\widehat{\boldsymbol{\theta}}_r)$ corresponding to distributed lag terms.

3.5. Selection of the smoothing parameters

In the normal errors case the fitted values of the additive DLM may be written as

$$\widehat{\mathbf{y}}_{\boldsymbol{\Lambda}} = \mathbf{G}_{\boldsymbol{\Lambda}}\mathbf{y} \quad \text{where} \quad \mathbf{G}_{\boldsymbol{\Lambda}} = \tfrac{1}{T-q}\mathbf{1}\mathbf{1}^{\mathsf{T}} + \widetilde{\mathbf{X}}_{\mathrm{r}}(\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}\widetilde{\mathbf{X}}_{\mathrm{r}} + \boldsymbol{\Lambda}_{\mathrm{r}})^{-1}\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}.$$

The generalized cross validation (GCV) criterion for selection of $\boldsymbol{\Lambda}$ is

$$\mathrm{GCV}(\boldsymbol{\Lambda}) = \frac{(T-q)\|\mathbf{I} - \mathbf{G}_{\boldsymbol{\Lambda}}\mathbf{y}\|^2}{\{\operatorname{tr}(\mathbf{I} - \mathbf{G}_{\boldsymbol{\Lambda}})\}^2}. \tag{8}$$

Extension to the generalized context involves replacement of $\mathbf{G}_{\boldsymbol{\Lambda}}$ by

$$\mathbf{G}_{\boldsymbol{\Lambda}}^{\mathbf{w}} = \tfrac{1}{T-q}\mathbf{1}\mathbf{1}^{\mathsf{T}} + \widetilde{\mathbf{X}}_{\mathrm{r}}(\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}\mathbf{W}\widetilde{\mathbf{X}}_{\mathrm{r}} + \boldsymbol{\Lambda}_{\mathrm{r}})^{-1}\widetilde{\mathbf{X}}_{\mathrm{r}}^{\mathsf{T}}\mathbf{W}. \tag{9}$$

where $\mathbf{W} = \operatorname{diag}(\mathbf{w})$ contains the weights at convergence. (e.g., Hastie and Tibshirani, 1990, p. 159).

In principle, this can be minimized over $(\lambda_1, \ldots, \lambda_p, \lambda)$ to select each of these smoothing parameters. But there are some practical problems with such an approach (see, e.g., Hastie and Tibshirani, 1990, pp. 159–160). Nevertheless, we can still use (8) to select *one* of the smoothing parameters if the remaining ones are fixed by other means. This strategy is used in our analysis of the Milan mortality data, described in the next section.

4. Analysis of Milan mortality data

We analyzed the Milan mortality data using a Poisson additive distributed model, as described in Section 3.4, with mortality as the response variable. Our model included smooth functions of temperature and relative humidity as well as day number. The motivation for this seasonal control is that unmeasured covariates, such as diet, often have seasonal patterns which may be confounded with the seasonal patterns of weather and pollution variables. In contrast, shorter-term variations in, for example, diet and smoking are unlikely to correlate with short-term fluctuations in pollution. Hence removing the long-wavelength components of the mortality data will avoid this potential for confounding. We choose a smoothing parameter that reduced the residuals of the mortality regression to white noise. We also included dummy variables for days of the week, indicators for holidays and influenza epidemics. Further details are provided in Rossi (Rossi *et al.*, 1999).

During the ten-year period in Milan, there was also a strong effect due to extreme temperatures. It was difficult to model the effect of these extreme days well without using excessive degrees of freedom in the smooth function of temperature thereby fitting noise at more moderate temperature levels. To correct for these we added dummy variables for (1) mean temperature between 27°C and 28°C, (2) mean temperature between 28°C and 29°C, and (3) mean temperature above 29°C, in addition to the smooth functions of temperature and humidity. It should be noted that, for daily time series data, individual factors such as cigarette smoking or a diet are not confounders since these do not vary significantly over short time intervals, and long time intervals are removed by the smooth function of time.

We fitted two models. One was a traditional model that examined the effect of `TSP` on the same day. The other was our DLM that examined the effect of `TSP` distributed over the same day and the previous 45 days.

The model was fit using penalized splines for smooth terms as described in Section 3.4. We chose smoothing parameters to use the same number of degree of freedom as in the paper of Rossi (Rossi *et al.*, 1999), so that the differences would be attributable to our use of a DLM. That study used 3.2 degrees of freedom for temperature (plus three dummy variables for extreme days) and 3.7 for relative humidity. We decided that 8 knots would be sufficient for estimating those smooth functions. We used 118 knots (one for each change of month) for the smooth function of time since this has much more fine detail. The results were not sensitive to changes of the number of knots. The smoothing parameter for this latter component was chosen to make the deviance residuals appear to be 'white', via visual inspection of the sample autocorrelation function. The smoothing parameter for time corresponded to $\lambda = 220$ with 32.9 degrees of freedom.

The smooth functions are shown in Figure 3.

The curves for mean temperature and relative humidity each show an approximate U-shaped relationship between these variables and mortality. This reflects the fact that people are more vulnerable at extreme values of these factors.

The analysis of the residuals of the model showed no pattern remaining.

We then used GCV to choose the smoothing parameter for the distributed lag component. While it might be considered desirable to choose all of the smoothing parameters jointly via GCV, this would involve a very challenging optimization problem, without any guarantee of providing reasonable answer: see the discussion on page 161 of Hastie and Tibshirani (1990). In addition, we felt it appropriate to use the same number of degrees of freedom for covariate as the original paper.

To assess whether the results were sensitive to variation of the smoothing parameter λ, Table 2 shows the estimate of the `TSP` effect for different values of λ. The values of $\hat{\beta}_{\text{tot}}$, $\hat{\beta}_{\text{tot}} = \sum_{\ell=0}^{45} \hat{\beta}_\ell$, do not change significantly. The corresponding plot (not shown) presents a pattern going from very wiggly for the first two values of λ, to a almost quadratic shape for $\lambda = 10^8$.

The values of the smoothing parameter chosen via GCV is $\lambda = 8.2 \times 10^6$, which correspond to the

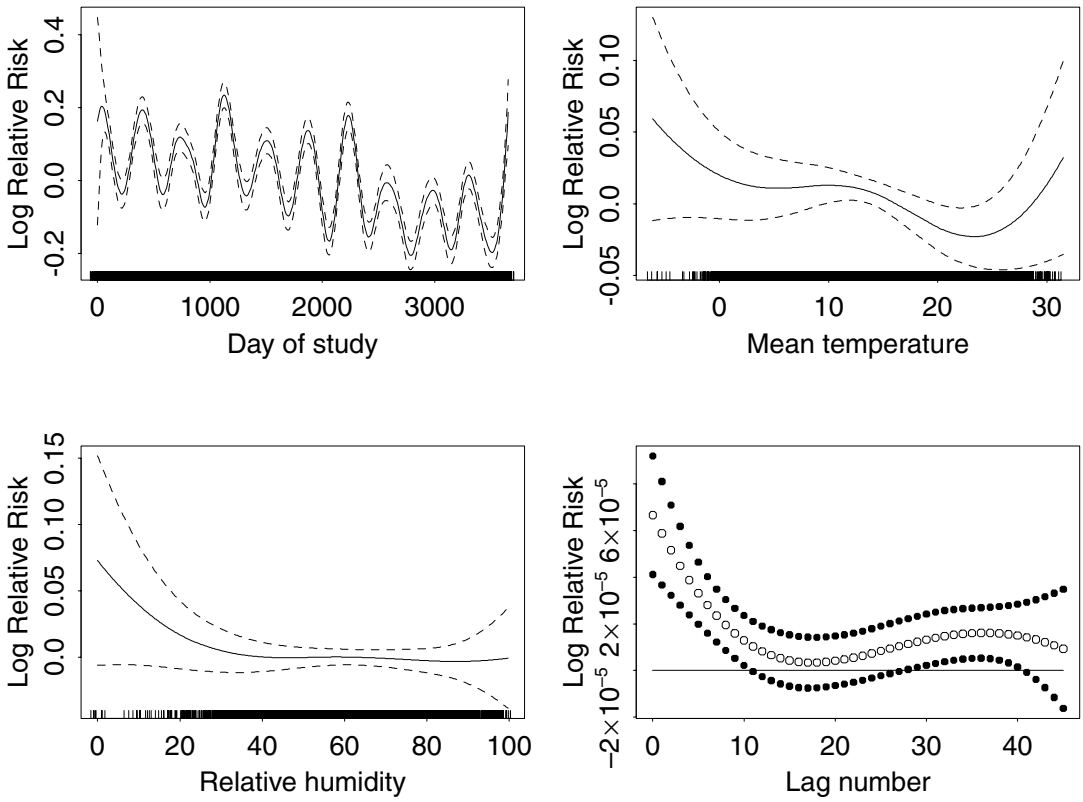

Fig. 3. Smooth curve components of generalized additive DLM fit to the Milan mortality data. The dashed lines regions (or dots in the fourth panel) correspond to plus and minus twice the estimated pointwise standard deviation of each estimate.

Table 2. `TSP` *estimates for some smoothing parameters*

λ	$\hat{\beta}_{\text{tot}}$	SE
10^4	0.000783	0.000163
10^5	0.000771	0.000162
10^6	0.000758	0.000161
8.2×10^6	0.000749	0.000161
10^7	0.000749	0.000161
10^8	0.000753	0.000161

Table 3. *Relative risk estimates and approximate 95% confidence intervals*

Generalized DLM	Linear `TSP` lag 0
1.067 (1.038, 1.096)	1.022 (1.014, 1.031)

Table 4. *Results: decomposition of the* `TSP` *effect*

	$\hat{\beta}$	SE	RR	CI 5%	CI 95%
A	0.000424	0.000106	1.037	1.019	1.056
B	0.000018	0.000027	1.002	0.997	1.006
C	0.000307	0.000165	1.027	0.998	1.056
$A+B+C$	0.000749	0.000161	1.067	1.038	1.096

last panel in Figure 3. The value of $\hat{\beta}_{\text{tot}}$ was also not sensitive to varying the degrees of freedom for temperature and relative humidity. We were unable to substantially vary the degrees of freedom for time with these data without either allowing significant positive autocorrelation at short lags or inducing major negative autocorrelation at negative lags.

This plot is very interesting as the resulting DLM shows the basic features postulated in Figure 2. A unit impulse of air pollution on one day produces an immediate effect that decays over the succeeding week; we did not observe a negative effect due to the mortality displacement, but between lags 14 and 20 the values are almost zero. This is followed by a period of several weeks where small positive effects persist, and suggests a long-term effect.

Table 3 shows the estimated relative risk (RR) for the interquartile range (IQR) of `TSP` ($RR = \exp(\beta \times IQR(\texttt{TSP}))$), and approximate 95% confidence interval (CI), based on the generalized additive DLM. For comparison, we also give the estimates for the lag 0 of `TSP`.

We can then decompose the `TSP` effect in the three parts A, B and C as in Figure 2; the results are shown in Table 4, where the relative risks are computed for the interquartile range of `TSP`.

A corresponds to the sum of the first positive parameters and describes the positive effect of `TSP` in the first 15 days, with a $RR = 1.037$ (CI 1.019–1.056). Even if not negative, B should represents the depletion of the risk pool. The sum of the parameter between lag 16 and lag 20 is almost zero. C is the sum of the last 25 days and the relative risk of this period is higher ($RR = 1.03$, CI 0.998–1.056). The positive parameters could be due to either an increase of the people at risk or a longer-term effect of air pollution on individuals or both. The sum of all days ($A + B + C$) represents the total cumulative effect in the long term with $RR = 1.067$ (CI 1.038–1.096).

As sensitivity analysis we applied parametric DLMs of degree two, three and four to the data, obtaining similar results.

5. Discussion

In this work we studied the mortality displacement with the aim to estimate the effect of air pollution adjusting for this phenomena. Our approach was to directly fit the 'rebound effect' of negative correlation between exposure and outcome due to the presumed depletion of the risk pool. We did this using a DLM in order to study the distribution of the effect of air pollution over time. To allow for some flexibility, we estimated this model with a smoothing function, and precisely with a penalized spline.

We found that allowing up to 45 days of delay to capture the potential rebound effect of mortality dis-

placement, there was no evidence of a reduction of the effect size estimate for TSP. Rather, the estimated overall effect was increased compared with results using just a one- or two-day mean. This does not mean that there is no short-term depletion of the risk pool. Rather, it indicates that any such depletion is counter–balanced by a combination of continuing risk out for lags up to 45 days and possible increased recruitment into the risk pool at other time lags. Lags greater than 45 days are difficult to examine because the seasonal control has filtered out fluctuations in daily deaths with period much longer than 45 days. These results are consistent with the results published by Zeger *et al.* (1999) and Schwartz (2000a), who also reported higher effect sizes in their harvesting resistant analyses. The report of similar patterns in three different cities using three different methodologies strongly indicates that the effects of particulate air pollution are not restricted to moving forward deaths by a few days or weeks, and that current published time series results may be under estimating the effect of airborne particles. This conclusion is strengthened by the lack of sensitivity of the estimated overall effect of particles to the choice of smoothing parameter for the DLM shown in Table 2 or to weather control.

The use of one-day average for TSP appears to underestimate the relative risk. The estimated relative risk based on the generalized additive DLM suggests that TSP may be responsible for approximately 7% of the deaths in Milan and that they are brought forward by a nontrivial amount. For comparison breast or prostate cancer are responsible for substantially smaller numbers of deaths.

The Milan data has previously been analyzed using autoregressive Poisson models and later (Rossi *et al.*, 1999) with the generalized additive model. The estimated effect was a 3.3% increase in total mortality for a 100 μg/m^3 change in TSP. This result for total mortality appear quite reasonable with respect to our results. One of the main differences between the present work and the previous result is the fact that we had to estimate the missing values for TSP to be able to estimate the DLM, and in Milan we had 19% of missing values.

Although Milan did not experience any air pollution episodes, the DLM shows an interesting shape which describe a dip and also a persistent, long-term effect of the air pollutant. This could be explained by some toxicological studies. These have shown that exposure of animals to combustion particles can produce inflammatory damage in the lung (Li *et al.*, 1996; Gilmour *et al.*, 1996; Pritchard, 1996; Costa and Dreher, 1997). This suggests that exposure over time intervals of weeks may have some additional cumulative effect that is not captured in the current short-term regression analyses. Prospective cohort studies of particulate air pollution and daily deaths (Dockery *et al.*, 1993; Pope *et al.*, 1995b) have reported substantially larger effects of long-term exposure to 10 μg/m^3 of fine particles than are indicated by the daily time series studies, suggesting that the difference may represent an effect of chronic exposure.

ACKNOWLEDGEMENT

This work was supported in part by NCI grant CA48061 (L.M.R.), EPA grant R824757 (M.P.W.) and EPA grant R827353 (A.Z. and J.S.).

REFERENCES

ALMON, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica* **33**, 178–196.

ANDERSON, H. R., SPIX, C., MEDINA, S., SCHOUTEN, J. P., CASTELLSAGUE, J., ROSSI, G., ZMIROU, D., TOULOUMI, G., WOJTYNIAK, B., PONKA, A., BACHAROVA, L., SCHWARTZ, J. AND KATSOUYANNI, K. (1997). Air pollution and daily admissions for chronic obstructive pulmonary disease in 6 European cities: results from the APHEA project. *European Respiratory Journal* **10**, 1064–1071.

BREMNER, S. A., ANDERSON, H. R., ATKINSON, R. W., MCMICHAEL, A. S., STRACHAN, D. P., BLAND, J. M.

AND BOWER, J. S. (1999). Short term association between outdoor air pollution and mortality in London 1992–4. *Occupational and Environmental Medicine* **56**, 237–244.

CLARKE, R. W., HEMENWAY, D. R., FRANK, R., *et al.* (1997). Particle associated Sulfate exposure enhances murine influenza mortality. *Americal Journal of Respiratory and Critical Care Medicine* **A245**.

CLARKE, R. W., CATALANO, P. J., KOUTRAKIS, P., MURTHY, G. G., SIOUTAS, C., PAULAUSKIS, J., COULL, B., FERGUSON, S. AND GODLESKI, J. J. (1999). Urban Air Particulate Inhalation Alters Pulmonary Function and Induces Pulmonary Inflammation in a Rodent Model of Chronic Bronchitis. *Inhal. Toxicol.* **11**, 637–656.

CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E. AND TERPENING, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics* **6**, 3–73.

CORRADI, C. (1977). Smooth distributed lag estimators and smoothing spline functions in Hilbert spaces. *Journal of Econometrics* **5**, 211–220.

COSTA, D. AND DREHER, K. (1997). Bioavailable transition metals in particulate matter mediate cardiopulmonary injury in health and compromised animal models. *Journal of Environmental Health Perspective* **105**, 1053–1060.

DAVIDSON, R. AND MACKINNON, J. G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

DOCKERY, D. W., POPE, C. A., XU, X., SPENGLER, J. D., WARE, J. H., FAY, M. E., FERRIS, B. G. JR. AND SPEIZER, F. E. (1993). An association between air pollution and mortality in six US cities. *New England Journal of Medicine* **329**, 1753–1759.

DOMINICI, F., SAMET, J. AND ZEGER, S. L. (2000). Combining evidence on air pollution and daily mortality from the largest 20 US cities: a hierarchical modeling strategy. *Journal of the Royal Statistical Society*, Series A, in press.

EILERS, P. H. C. AND MARX, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Sciences* **89**, 89–121.

FAIRLEY, D. (1990). The relationship of daily mortality to suspended aprticulates in Santa Clara County, 1980–1986. *Environmental Health Perspective* **89**, 159–168.

GILMOUR, P. S., BROWN, D. M., LINDSAY, T. G., BESWICH, P. H., MACNEE, W. AND DONADSON, K. (1996). Adverse health effects of PM10 particles: involvement of iron in generation of hydroxy radical. *Occupational and Environmental Medicine* **53**, 817–822.

GOLD, D. R., LITONJUA, A., SCHWARTZ, J., VERRIER, M., MILSTEIN, R., LARSON, A., LOVETT, E. AND VERRIER, R. (2000). Ambient Pollution and Heart Rate Variability. *Circulation*, in press.

HASTIE, T. J. (1996). Pseudosplines. *Journal of the Royal Statistical Society*, Series B **58**, 379–396.

HASTIE, T. J. AND TIBSHIRANI, R. J. (1990). *Generalized additive models.* New York: Chapman and Hall.

HASTIE, T. J. AND TIBSHIRANI, R. J. (1994). Discussion of the paper by Royston and Altman. *Applied Statistics* **43**, 460.

HOEK, G., SCHWARTZ, J., GROOT, B. AND EILERS, P. (1997). Effects of ambient particulate matter and ozone on daily mortality in Rotterdam, The Netherlands. *J. Arch. Environ. Health* **52**, 455–463.

KATSOUYANNI, K., TOULOUMI, G., SPIX, C., SCHWARTZ, J., BALDUCCI, F., MEDINA, S., ROSSI, G., WOJTYNIAK, D., SUNYER, J., BACHAROVA, L., SCHOUTEN, J. P., PONTAK, A. AND ANDERSON, H. R. (1997). Short term effects of ambient sulphur dioxide and particulate mater on mortality in 12 European cities: results from time series data from the APHEA project. *British Medical Journal* **314**, 1658–1663.

KELSALL, J. E., SAMET, J. M., ZEGER, S. L. AND XU, J. (1997). Air Pollution and Mortality in Philadelphia: 1974–1988. *Americal Journal of Epidemiology* **146**, 750–762.

LI, X. Y., GILMOUR, P. S., DONALDSON, K. AND MACNEE, W. (1996). Free radical activity and pro inflammatory effects of particulate air pollution (PM10) *in vivo* and *in vitro*. *Thorax* **51**, 1216–1222.

MARX, B. D. AND EILERS, P. H. C. (1998). Direct generalized additive modeling with penalized likelihood. *Comparative Statistical Data Analysis* **28**, 193–209.

PETERS, A., LIU, E., VERRIER, R. L., SCHWARTZ, J., GOLD, D. R., MITTLEMAN, M., BALIFF, J., OH, A., ALLEN, G., MONAHAN, K. AND DOCKERY, D. (2000). Air pollution and incidences of cardiac arrhythmia. *Epidemiology* **11**, 11–17.

POPE, C. A., DOCKERY D. W. AND SCHWARTZ, J. (1995a). Review of Epidemiologic Evidence of Health Effects of Particulate Air Pollution. *Inhal. Toxicol.* **7**, 1–18.

POPE, C. A., THUN, M. J., NAAMBOODIRI, M. M., DOCKERY, D. W., EVANS, J. S., SPEIZER, F. E. AND HEATH, C. W. JR. (1995b). Particulate air pollution as a predictor of mortality in a prospective study of US Adults. *American Journal of Respiratory and Critical Care Medicine* **151**, 669–674.

POPE, C. A. AND SCHWARTZ, J. (1996). Time series for the analysis of pulmonary health data. *American Journal of Respiratory and Critical Care Medicine* **154**, S229–S233.

PRITCHARD, R. J., GHIO, A. J., LEHMANN, J. R., WINSETT, D. W., TEPPER, J. S., PARK, P., GILMOUR, M. I., DREHER, K. L. AND COSTA, D. L. (1996). Oxidant generation and lung injury after particulate air pollutant exposure increase with the concentrations of associated metals. *Inhal. Toxicol.* **8**, 457–477.

ROSSI, G., VIGOTTI, M. A., ZANOBETTI, A., REPETTO, F., GIANNELLE, V. AND SCHWARTZ, J. (1999). Air pollution and cause specific mortality in Milan, Italy, 1980-1989. *Archiv Environmental Health* **54**, 158–164.

SCHWARTZ, J. (1991). Particulate air pollution and daily mortality in Detroit. *Environmental Research* **56**, 204–213.

SCHWARTZ, J. AND DOCKERY, D. W. (1992). Increased mortality in Philadelphia associated with daily air pollution concentrations. *American Review of Respiratory Diseases* **145**, 600–604.

SCHWARTZ, J. (1993). Air pollution and mortality in Birmingham, AL. *American Journal of Epidemiology* **137**, 1136–1147.

SCHWARTZ, J. (1994a). Air pollution and daily mortality: a Review and Meta Analysis. *Environmental Research* **64**, 36–52.

SCHWARTZ, J. (1994b). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *The Canadian Journal of Statistic* **22**, 471–487.

SCHWARTZ, J. (2000a). Harvesting and long-term exposure effects in the relationship between air pollution and mortality. *American Journal of Epidemiology* **151**, 440–448.

SCHWARTZ, J. (2000b). The distributed lag between air pollution and daily deaths. *Epidemiology* **11**, 320–326.

SPIX, C., HEINRICH, J., DOCKERY, D. W., SCWARTZ, J., VOELKSCH, G., SCWINKOWSKI, K., COELLEN, C. AND WICHMANN, H. E. (1993). Air pollution and daily mortality in Erfutrt, East Germany, 1980–89. *Journal of Environmental Health Perspective* **101**, 518–526.

VIGOTTI, M. A., ROSSI, G., BISANTI, L., ZANOBETTI, A. AND SCHWARTZ, J. (1996). Short term effect of urban air pollution on respiratory health in Milan, Italy, 1980–89. *Journal of Epidemiology and Community Health* **50**, S71–S75.

ZELIKOFF, J. T., NADZIEJKO, C., FANG, T., GORDON, C., PREMDASS, C. AND COHEN, M. D. (1999). Short term, low-dose inhalation of ambient particulate matter exacerbates ongoing pneumoncoccal infections in *Streptococcus pneumoniae*-infected rats. In PHALEN, R. F. and BELL, Y. M. (eds), *Proceedings of the Third Colloquium on Particulate Air Pollution and Human Health*, Air Pollution Health Effects Laboratory, University of California, Irvine, CA, Vol. 8, pp. 94–101.

ZEGER, S. L., DOMINICI, F. AND SAMET, J. (1999). Harvesting-resistant Estimates of air pollution effects on mortality. *Epidemiology* **10**, 171–175.

[*Received October 8, 1999*; *revised February 24, 2000*; *accepted for publication February 25, 2000*]