Re-Analyses of the NMMAPS Data Base:
Shape of the Exposure-Response Relationship and Mortality Displacement

Dominici F., Daniels M., McDermott A., Zeger S.L., Samet J.M.

Abstract

We summarize the results of three re-analyses of the NMMAPS data for PM_{10} and mortality: 1) estimation of the shape of the exposure-response relationship and location of any threshold, 20 largest US cities (1987-1994); 2) estimation of regional and national average exposure-response curves, 88 largest US cities (1987-1994); and 3) frequency and time domain log linear regression analyses to assess the extent of mortality displacement, Philadelphia (1974-1987). These analyses did not show any substantial differences with respect to previous analyses.

Acknowledgments

Research described in this article was partially supported by a contract and grant from Health Effects Institute (HEI), an organization jointly funded by the Environmental Protection Agency (EPA R824835) and automotive manufacturers. The contents of this article do not necessarily reflect the views and policies of HEI, nor do they necessarily reflect the views and policies of EPA, or motor vehicles or engine manufacturers. Funding for Francesca Dominici was provided by a grant from the Health Effects Institute (Walter A. Rosenblith

New Investigator Award). Research described within was also supported by a grant from the National Institute of Environmental Health Sciences for the Johns Hopkins Center in Urban Environmental Health (P30ES03819-12).

1 Introduction

Recent work by Dominici et al. (2002c) and Ramsay et al. (2002) called for caution in the use of the S-Plus software `gam` for fitting Generalized Additive Models (GAM) (Hastie and Tibshirani, 1990) to estimate relative rates of mortality/ morbidity in time series studies of air pollution and health. Dominici et al. (2002c) recently reported that when the data to which the GAM are applied have the following two characteristics 1) small estimated regression coefficients relative to the effects of confounders; and 2) confounding factors are modeled using at least two non-parametric smooth functions, the defaults in the S-PLUS software (Version 3.4) `gam` do not assure convergence of its iterative estimation procedure. Biased estimates of the regression coefficients and their standard errors can result. In parallel, Ramsay et al. (2002) point out that the S-Plus function `gam` uses a computational approximation which, in the presence of correlation between the air pollution variable and the non-linear functions included in the model, can underestimate the standard errors of the relative rates (see also Chambers and Hastie (1992) pages 303-304 and commentaries by Lumley and Sheppard (2003) and Samet et al. (2003). Standard errors are underestimated even if more stringent convergence parameters are used.

Recently, some progress has been made to overcome the limitations of the Splus software `gam`. First, default convergence parameters of the S-Plus function `gam` (Version 6.1) have made more stringent with a change from 10^{-3}, the default, to 10^{-7}. Second, a revised version of the S-Plus `gam` which calculates asymptotically exact standard errors of the relative rate estimate has been released (see Dominici et al. (2002b) for reference and `http://biosun01.biostat.jhsph.edu/~ fdominic/research.html` for software) .

In this short communication, we present findings of new analyses of three NMMAPS databases: 1) estimation of a national average dose-response curve for the largest 20 US cities (Daniels et al., 2000); 2) estimation of a national average dose-response curve for the largest 88 cities (Dominici et al., 2002a); and 3) frequency and time domain log-linear regression to assess the extent of mortality displacement (Zeger et al., 1999). All other re-analyses of the National Morbidity Mortality Air Pollution Study are summarized in a separate report to the Health Effects Institute submitted in October 2002 (Dominici et al., 2002d).

2 Methods

Estimation of dose-response. In the original NMMAPS analyses (Daniels et al., 2000; Dominici et al., 2002a), we extended the NMMAPS city-specific regression model (Dominici et al., 2000) by assuming that the expected value of mortality is a natural cubic spline of air pollution adjusted by several confounders. The confounders included smooth functions of time and temperature, originally modelled as smoothing splines (Kelsall et al., 1997; Dominici et al., 2000). Within each city, the shape of the exposure response curve was estimated by using the `gam` software with default convergence parameters.

Here we re-estimate the shape of the exposure-response curve using the same approach as in Daniels et al. (2000) and Dominici et al. (2002a), but modelling the smooth functions of time and temperature as natural cubic splines and using Generalized Linear Models

(McCullagh and Nelder, 1989) and the S-Plus function `glm` .

More specifically, to estimate the shape of the exposure-response, we model the logarithm of the expected value of daily mortality as a smooth function of air pollution after adjusting for other confounders:

$$E(Y_t) = \exp\{S(X_t, \text{knots} = \boldsymbol{\nu}) + \text{confounders}\} \tag{1}$$

where X_t and Y_t are the air pollution and mortality time series, and $S(\cdot, \cdot)$ is a natural cubic spline with k knots at locations $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$. In addition, to examine whether the health effects of air pollution are negligible below some level, a linear threshold model can be used:

$$E(Y_t) = \exp\{\theta(X_{t-\ell} - h)^+ + \text{confounders}\} \tag{2}$$

where ($x^+ = x$ if $x \geq 0$ and $x^+ = 0$ if $x < 0$) and h is an unknown change-point that is estimated from the data. In both modelling approaches, the term "confounders" represents time-varying covariates (as, for example, long-term trends and seasonality in the mortality time-series, and weather variables), which may bias the particle-mortality association. These confounders are modeled as natural cubic splines (Table 1).

Mortality Displacement. The original work by Zeger et al. (1999) developed and illustrated an approach for estimating the association between air pollution and mortality from time series data that is resistant to short-term harvesting. The method was based on the concept that harvesting alone creates associations only at the shorter-time scales. It used frequency-domain log-linear regression (FDLLR) (Kelsall et al., 1999) to decompose the information about the pollution-mortality association into distinct time scales and then created harvesting-resistant estimates by excluding the short-term information which is affected by harvesting. This method was applied to Total Suspended Particles (TSP) and mortality counts from Philadelphia for 1974-1988 and showed that the TSP-mortality association in Philadelphia was inconsistent with the harvesting-only hypothesis and that the harvesting-resistant estimates of the TSP relative risk were actually larger, not smaller, than the ordinary estimates. As the software for fitting FDLLR used `gam` with default convergence parameters, re-analyses are now necessary.

Here we re-analyze the Philadelphia data using two methods: 1) applying the FDLLR software but with stringent convergence parameters (10^{-15}) for the S-Plus function `gam`; and 2) developing an alternative method based on the time-domain analogue of FDLLR which uses natural cubic splines and GLM. The time-domain log linear regression approach (TDLLR) is based on a Fourier decomposition of air pollution time series into a set of independent exposure variables, each representing a different time scale. These variables are then used as predictors in a Poisson regression model to estimate a separate relative rate of mortality on each exposure time-scale while controlling for time and temperature. More specifically, let X_t^c be the air pollution time series, and Y_t^c be the mortality time series in location c. We first decompose the air pollution series X_t^c into distinct component series X_{kt}^c, one for each distinct time scale k, and then we calculate the association between Y_t^c, without decomposition, and each of the time scale components X_{kt}^c. The decomposition is obtained by applying the discrete Fourier transform to the X_t^c series (Bloomfield, 1976). Specifically, we assume:

$$E(Y_t^c) = \exp\{\sum_k \beta_k^c X_{kt}^c + \text{confounders}\} \tag{3}$$

where β_k^c (the parameters of interest), denotes the log relative rate of daily mortality for each 10 unit increase in the air pollution level in location c on a time scale k. The TDLLR and its application to four NMMAPS cities are described in detail elsewhere (Dominici et al., 2002d).

3 Results

Exposure Response. Differences in the shape of the exposure-response curves were not observed when comparing the new with the original analyses. Figure 1 shows exposure-response curves for total (TOTAL) mortality, cardiovascular and respiratory (CVDRESP) mortality, and other causes (OTHER) mortality, for the 20 largest US cities, 1987-1994. Figure 2 shows regional PM_{10}-mortality exposure-response curves for TOTAL, for each region, and the national average, for the 88 largest US cities, 1987-1994.

The national average exposure-response curves for both the 20 and 88 cities are linear, and the posterior probability of zero knots is approximately 1. At the regional level, the data from cities in several regions (Northwest, Upper Midwest, and Southeast) indicate some modest departure from a linear model. In particular, the Northwest and Upper Midwest regions show a levelling off (saturation effect) at higher levels of PM. However, the uncertainty boundaries for these regions indicate compatibility of the data with a linear relationship, and we cannot explain why these regions might have other than a linear exposure-response curve.

Figure 3 shows posterior probabilities for a threshold for the effect of PM_{10} on the cause-specific mortality groupings, for the 20 largest US cities, 1987-1994. The posterior distributions on the location of the threshold are skewed to the right for TOTAL and CVDRESP, and skewed to the left for OTHER. The category of OTHER had the highest, most-probable threshold, at $65\mu/m^3$. However, for CVDRESP, the data give more support to a low value of h (as $0, 5, 10$ $\mu g/m^3$), lower than TOTAL. The threshold, if any, for CVDRESP may be lower than for TOTAL.

Mortality Displacement. Figure 4 shows the frequency domain estimate of the mortality relative rate associated with air pollution as a function of frequency. The horizontal axes denote the Fourier frequencies (bottom) and the time scale in days (top) at which the association is measured. The dotted curves denote the estimated relative rates, plus or minus two estimated standard errors, respectively, at each frequency. The time-scale estimates (points connected by line segments) are plotted on top of the frequency domain results (continuous curve). Time scales and frequency domain results are similar, indicating little sensitivity of the results to the modelling assumptions. In addition, consistent with results for the four NMMAPS cities (Dominici et al., 2002d), relative rate estimates at longer time-scales are larger than relative rate estimates at shorter time-scales, indicating a pattern of air pollution effects at different time scales that is inconsistent with the "harvesting only" hypothesis.

4 Discussion

We have presented NMMAPS re-analyses of the shape of the exposure-response curves for the 20 and 88 cities, and frequency and time domain log-linear regression (FDLLR and TDLLR) to assess mortality displacement.

The exposure-response analyses were performed by using the same modelling approaches described in previous publications, but replacing the smoothing spline with natural cubic splines with the same number of degrees of freedom for the adjustment of time-varying confounders. On average, the shape of the exposure-response curve is linear, confirming results reported in the original analyses.

The analyses of mortality displacement were performed by using FDLLR software with `gam` with stringent convergence parameters and by developing and applying an alternative time domain approach (TDLLR) (Dominici et al., 2002d), which used natural cubic splines and GLM software. As reported in previous analyses, both methods produce patterns of relative rates estimates at different time scales that are inconsistent with the hypothesis of mortality displacement.

References

Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons.

Chambers, J. and Hastie, T. (1992). *Statistical Models in S*. Chapman and Hall, London.

Daniels, M., Dominici, F., Samet, J. M., and Zeger, S. L. (2000). "Estimating PM10-Mortality Dose-Response Curves and Threshold Levels: An Analysis of Daily Time-Series for the 20 Largest US Cities." *American Journal of Epidemiology*, 152, 397–412.

Dominici, F., Daniels, M., Zeger, S. L., and Samet, J. M. (2002a). "Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships." *Journal of the American Statistical Association*, 97, 100–111.

Dominici, F., McDermott, A., and Hastie, T. (2002b). "Semiparametric Regression in Time Series Analyses of Air Pollution and Health." *Technical Report, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins.*

Dominici, F., McDermott, A., Zeger, S. L., and Samet, J. M. (2002c). "0n Generalized Additive Models in Time Series Studies of Air Pollution and Health." *American Journal of Epidemiology*, 156, 3, 1–11.

— (2002d). "Airborne Particulate Matter and Mortality: Time-Scale Effects in Four US Cities." *American Journal of Epidemiology (under revision).*

Dominici, F., Samet, J. M., and Zeger, S. L. (2000). "Combining Evidence on Air pollution and Daily Mortality from the Twenty Largest US cities: A Hierarchical Modeling Strategy (with discussion)." *Royal Statistical Society, Series A, with discussion*, 163, 263–302.

Green, P. J. (1995). "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination." *Biometrika*, 82, 711–732.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall, New York.

Kelsall, J., Samet, J. M., and Zeger, S. L. (1997). "Air Pollution, and Mortality in Philadelphia, 1974-1988." *American Journal of Epidemiology*, 146, 750–762.

Kelsall, J., Zeger, S. L., and Samet, J. M. (1999). "Frequency Domain Log-Linear Models; Air pollution and Mortality." *The Royal Statistical Society, Series C*, 48, 331–344.

Lumley, T. and Sheppard, L. (2003). "Time Series Analyses of Air Pollution and Health: Straining at Gnats and Swallowing Camels?" *Epidemiology*, 14, ?–?

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (Second Edition)*. New York: Chapman & Hall.

Ramsay, T., Burnett, R., and Krewski, D. (2002). "The effect of concurvity in generalized additive models linking mortality and ambient air pollution." *Epidemiology (to appear)*.

Samet, J., Dominici, F., McDermott, A., and S.L., Z. (2003). "New Problems for an Old Design: Time Series Analyses of Air Pollution and Health." *Epidemiology*, 14, ?–?

Zeger, S. L., Dominici, F., and Samet, J. M. (1999). "Harvesting-resistant estimates of pollution effects on mortality." *Epidemiology*, 89, 171–175.

Table Legend

1. Potential confounding factors in the estimation of the city-specific relative rates associated with particulate air pollution levels, and the rationale for their inclusion in the model.

Predictors	Primary reasons for inclusion
Indicator variables for the three age groups	Allow different baseline mortality rate within each age group
Indicator variables for the day of the week	Allow different baseline mortality rate within each day of the week
Natural cubic splines of time with 7 degrees of freedom per year	To adjust for long term trend and seasonality
Natural cubic splines of temperature with 6 degrees of freedom	To control for the known effects of weather on mortality
Natural cubic splines of dewpoint with 3 degrees of freedom	To control for the known effects of humidity on mortality
Separate natural cubic splines of time (2 df for year) for each age group contrast	To separately adjust for seasonality within each age group

Figures Legend

1. Mortality-particulate matter < 10 μm in aerodynamic diameter (PM_{10}) dose-response curves for total (TOTAL) mortality, cardiovascular, and respiratory (CVDRESP) mortality, and other causes (OTHERS) mortality, 20 largest US cities, 1987-1994. The exposure-response curves for the mean lag, current day, and previous day PM_{10} are denoted by solid lines, squared points, and triangle points, respectively.

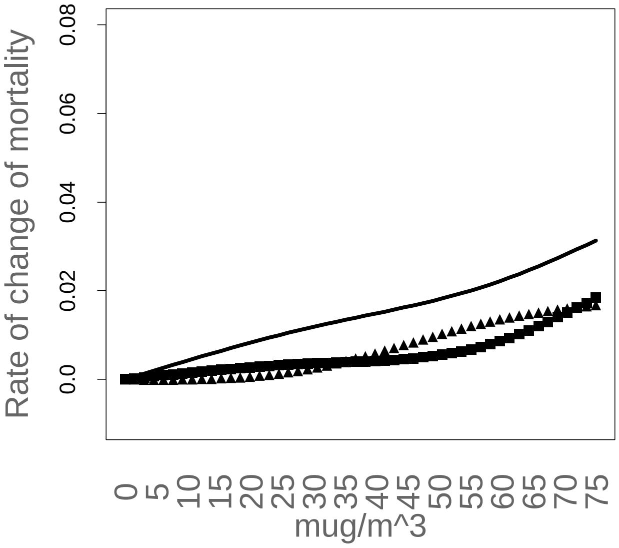
2. Regional and national-average PM_{10}-mortality exposure-response curves. Solid black curves are obtained by fitting the spline model with the Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) and allowing for an unknown number and location of knots. The curves with the empty dots are obtained by setting one knot at $40\mu_g/m^3$ and fitting the spline model with a Gibbs sampler. The linear curves are obtained by fitting the hierarchical linear model with a Gibbs sampler without borrowing strength across regions. The shaded area denotes the 95% confidence bands for the curve with a fixed knot.

3. Posterior probabilities of the thresholds for each cause-specific mortality and for the mean lag particulate matter $< 10\mu$ m in aerodynamic diameter (PM_{10}), 20 largest US cities, 1987-1994. TOTAL, total mortality; CVDRESP, cardiovascular mortality and respiratory mortality; OTHERS, other mortality.

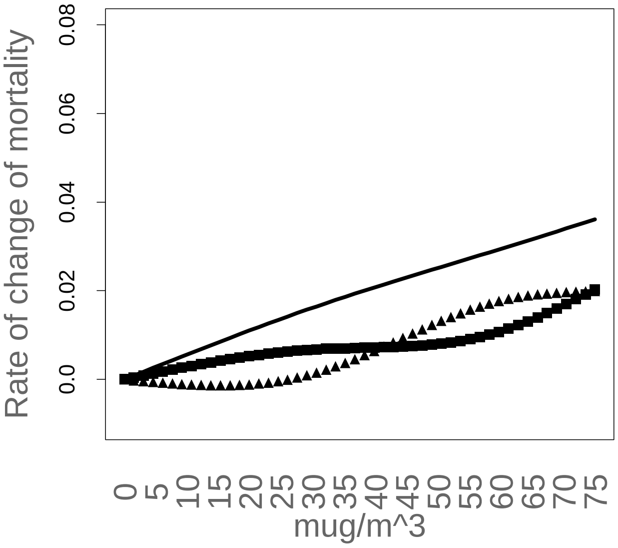
4. Philadelphia database 1974-1988: comparison between frequency domain (continuous curve) and time scale estimates (points connected by line segments), showing the log-relative rates of Total mortality by frequency and frequency grouping. The dotted lines

show plus and minus two standard errors for the frequency domain estimates and the bars represent plus and minus two standard errors for the time-scale estimates.

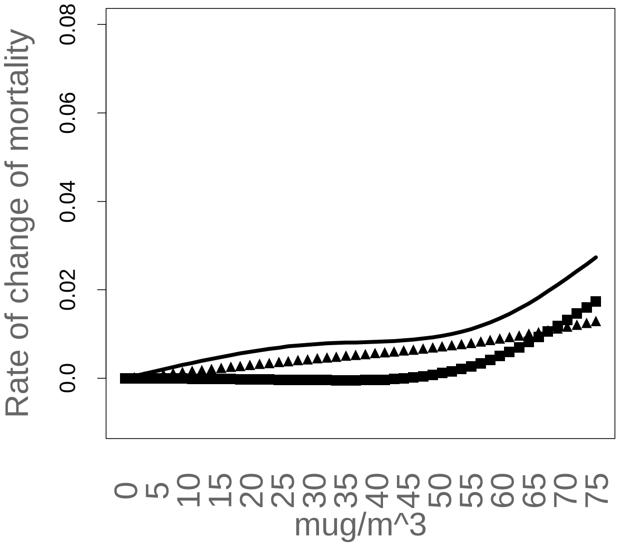
TOTAL

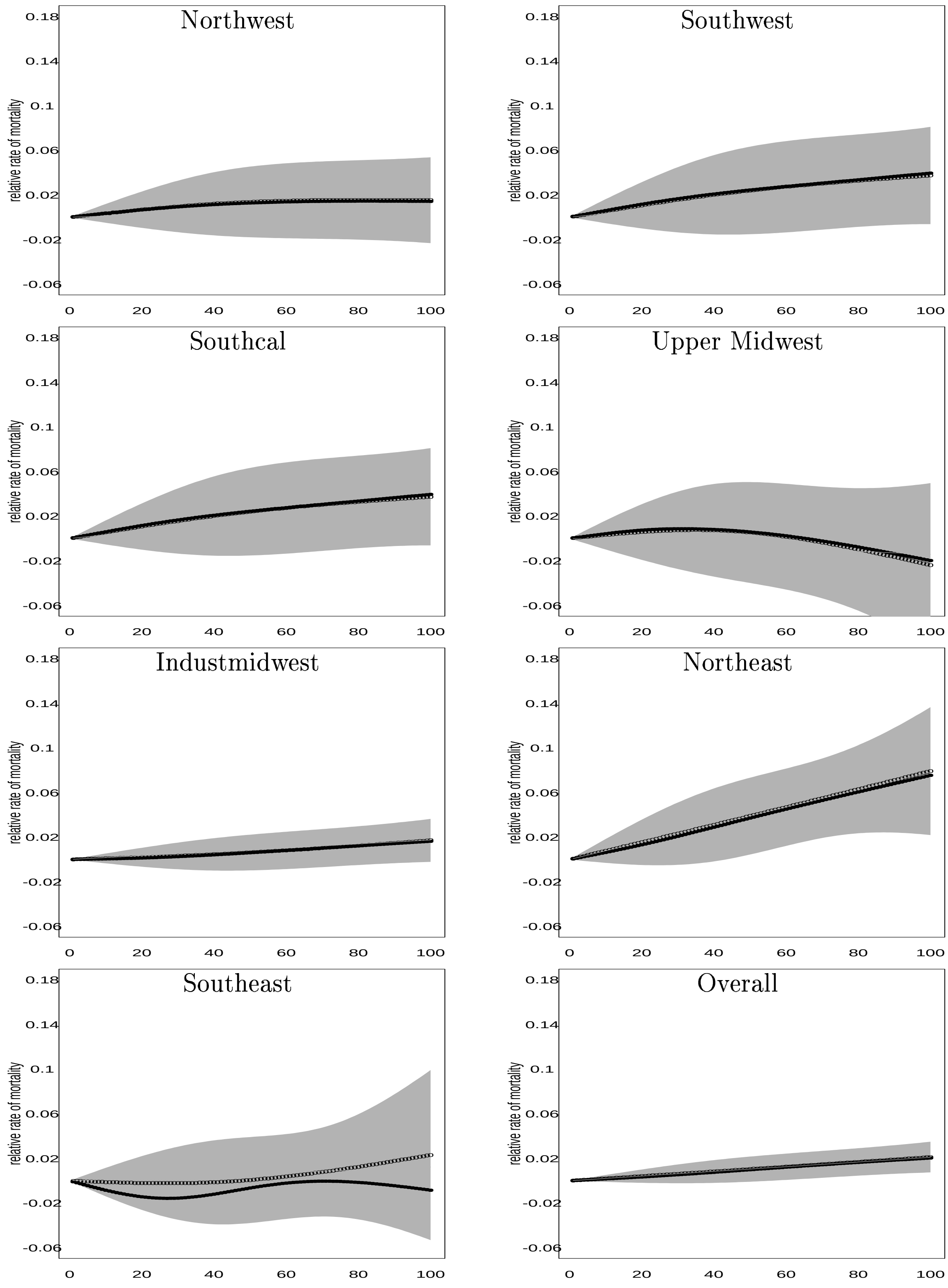


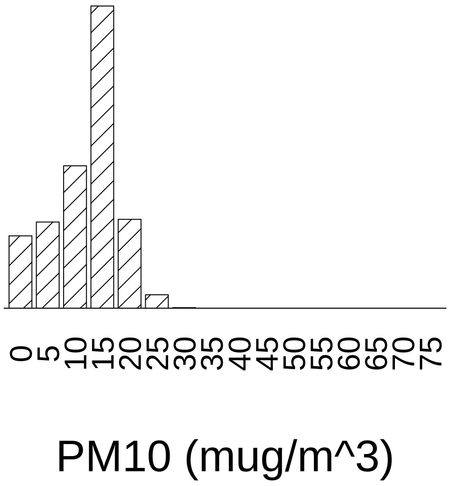
CVDRESP



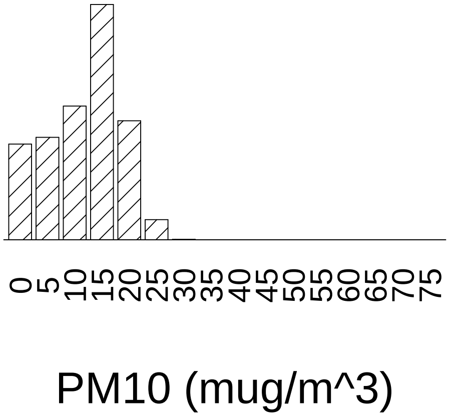
OTHER




Northwest
Southwest
Southcal
Upper Midwest
Industmidwest
Northeast
Southeast
Overall
relative rate of mortality
0.18
0.14
0.1
0.06
0.02
-0.02
-0.06
0
20
40
60
80
100



TOTAL



CVDRESP

OTHER

Philadelphia (1974-1987)