JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Johns Hopkins University, Dept. of Biostatistics Working Papers

3-30-2007

TRENDS IN PARTICULATE MATTER AND MORTALITY: AN APPROACH TO THE ASSESSMENT OF UNMEASURED CONFOUNDING

Holly Janes
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, hjanes@jhsph.edu

Francesca Dominici
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, fdominic@jhsph.edu

Scott Zeger
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, szeger@jhsph.edu

Suggested Citation

Janes, Holly; Dominici, Francesca; and Zeger, Scott, "TRENDS IN PARTICULATE MATTER AND MORTALITY: AN APPROACH TO THE ASSESSMENT OF UNMEASURED CONFOUNDING" (March 2007). *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working Paper 104.
http://biostats.bepress.com/jhubiostat/paper104

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors

Trends in Particulate Matter and Mortality: An Approach to the Assessment of Unmeasured Confounding

Holly Janes, Francesca Dominici, Scott Zeger

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, U.S.A.

Abstract

We propose a method for diagnosing confounding bias under a model which links a spatially and temporally varying exposure and health outcome. We decompose the association into orthogonal components, corresponding to distinct spatial and temporal scales of variation. If the model fully controls for confounding, the exposure effect estimates should be equal at the different temporal and spatial scales. We show that the overall exposure effect estimate is a weighted average of the scale-specific exposure effect estimates.

We use this approach to estimate the association between monthly averages of fine particles ($PM_{2.5}$) over the preceding 12 months and monthly mortality rates in 113 U.S. counties from 2000-2002. We decompose the association between $PM_{2.5}$ and mortality into two components: 1) the association between "national trends" in $PM_{2.5}$ and mortality; and 2) the association between "local trends," defined as county-specific deviations from national trends. This second component provides evidence as to

COBRA
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

whether counties having steeper declines in $PM_{2.5}$ also have steeper declines in mortality relative to their national trends.

We find that the exposure effect estimates are different at these two spatio-temporal scales, which raises concerns about confounding bias. We believe that the association between trends in $PM_{2.5}$ and mortality at the national scale is more likely to be confounded than is the association between trends in $PM_{2.5}$ and mortality at the local scale. If the association at the national scale is set aside, there is little evidence of an association between 12-month exposure to $PM_{2.5}$ and mortality.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

In environmental epidemiology we often conduct observational studies in which exposures to environmental agents cannot be controlled by the investigator. Inference about the health effects of the exposures is generally drawn from a statistical model that controls for potential confounders by including these factors as covariates. Confounding bias caused by omitting important confounders from the regression model is the most common threat to the validity of the exposure effect estimates.[1–7]

This paper illustrates an approach to diagnosing confounding bias under a causal model linking an environmental exposure and health outcome, estimated using spatio-temporal data. To test the model, we decompose the association between the exposure and health outcome into orthogonal components, corresponding to distinct scales of spatial and temporal variation. If the model adequately controls for confounding, then the exposure effect estimates should be similar at the different spatial and temporal scales. We show that the overall exposure effect estimate is a weighted average of the scale-specific exposure effect estimates. Differences among the scale-specific estimates indicates confounding by omitted covariates.

We illustrate our approach in a study of the mortality effect of 12-month exposure to fine particulate matter ($PM_{2.5}$). We develop a log-linear regression model for multi-site time-series data to estimate the association between month-to-month variation in mortality rates and month-to-month variation in average $PM_{2.5}$ over the preceding year in 113 U.S. counties and for the period 2000-2002. We decompose the association between $PM_{2.5}$ and mortality into two components: 1) the association between "national trends" in $PM_{2.5}$ and mortality; and 2) the association between county-specific deviations from the national

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

trend, that is, between "local trends." This second component provides evidence as to whether counties having steeper declines in $PM_{2.5}$ also have steeper declines in mortality with respect to their national trends.

If monthly mortality rates are caused by average $PM_{2.5}$ concentration in the previous year, the associations between the national and local trends should be the same, absent confounding and measurement error. Our proposed approach allows us to assess the validity of this causal hypothesis.

We hypothesize that the association between the national trends in $PM_{2.5}$ and mortality is likely to be confounded by slowly time-varying factors, such as changes in industrial activities and the economy, improving health care, and large scale weather events.[8–11] Our approach can be used to focus on the component of association that is least likely to be confounded, the association between the local trends.

The statistical framework proposed in this paper draws from both cohort studies of long-term exposure[12–15] and multi-site time series studies of short-term exposure.[16–23] As in cohort studies, we focus on long-term average exposure (averaged over the previous year). As in time-series studies, we estimate associations between temporal changes in exposure and outcome within counties, to guard against bias due to county-specific characteristics that do not vary with time.

Methods

We construct mortality counts (Y_t^c) and number of people at risk (N_t^c) for each county c and month t for six strata (two sexes and three age groups: 65-74 years; 75-84 years; and $>$ 85 years), using Medicare enrollment files. Our study population includes 8.2 million

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

Medicare enrollees living on average six miles from an EPA $PM_{2.5}$ monitor.

The locations of the 113 US counties included in the study are shown in Figure 1. The counties are categorized into seven geographic regions. The regions are based on our previous national multi-site time-series studies of PM_{10} and mortality and of $PM_{2.5}$ and hospital admissions.[16,24] These counties have nearly complete $PM_{2.5}$ data (no gaps larger than three weeks) for the period over which exposure was averaged, 1999 to 2002.

Estimating county-specific annual average $PM_{2.5}$: For each county and each month, we calculate the average level of $PM_{2.5}$ over the preceding year (denoted by PM_t^c) as follows. First, we estimate the smooth trend in $PM_{2.5}$ using a linear regression model with outcome monitor-specific daily $PM_{2.5}$ level, and as predictor a natural cubic spline of time with 16 degrees of freedom. Second, for each month, we calculate the average $PM_{2.5}$ over the previous year using the fitted values from the regression model described above. This modelled $PM_{2.5}$ allows us to impute small gaps of missing data when calculating annual averages. For counties with multiple monitors, we use the one with the most complete data, that is, the one with the smallest maximum and average gap in observations and with the longest observation period. We use data from a single monitor rather than from all the monitors within a county because averaging ambient $PM_{2.5}$ concentrations across monitors that are online for varying periods of time might induce spurious trends.

Measurement error in county-specific annual average $PM_{2.5}$ trends: To investigate whether the observed variation in PM_t^c trends across counties represents true between-county variability in long-term exposure, rather than differences between monitors within a county ("measurement error"), we perform the following analysis. First, for each monitor, we linearly regress PM_t^c on t and estimate the slope. Here, we use all monitors with at least 80% of the data available and with no gaps longer than 1 month. Second, we

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

fit a one-way random effects model to the monitor-specific estimated slopes and calculate: 1) the variability of the slopes within county (measurement error) (σ_w^2); 2) the variability of the slopes between counties (σ_b^2); and 3) the intraclass correlation coefficient, $\rho = \sigma_b^2/(\sigma_b^2 + \sigma_w^2)$.

Analysis of variance of county-specific annual average $PM_{2.5}$ trends: To quantify the variability of PM_t^c in space and time, we conduct the following analysis of variance. We fit a linear model with PM_t^c as the dependent variable, and with the following predictors: 1) county-specific intercepts (the spatial dimension); 2) a natural cubic spline of month with 16 degrees of freedom (the time dimension); and 3) an interaction between the county-specific indicators and the smooth function of time (the space-by-time interaction).

Causal model for annual average $PM_{2.5}$ and mortality: Within each age-sex stratum, we consider the following causal model for the health effects of air pollution:

$$\log E(Y_t^c) = \log N_t^c + \delta_0^c + \delta_1 PM_t^c. \tag{1}$$

The parameters δ_0^c are county-specific intercepts, which are included in the model to control for unmeasured county-specific characteristics that do not vary with time. The parameter δ_1 denotes the association between month-to-month variation in PM_t^c and month-to-month variation in mortality.

Estimates from model (1) are likely to be confounded by factors that cause trends in $PM_{2.5}$ and mortality. Examples of such confounders are policy changes affecting the economy, industrial activity, and health care and large scale weather events.[8–11] A popular approach to controlling for unmeasured temporal confounding at the national level is to add to the model a smooth function of time:

$$\log E(Y_t^c) = \log N_t^c + \beta_0^c + \beta_1 PM_t^c + s(t; d), \tag{2}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

where $s(t;d)$ is a smooth function of time modelled using a natural cubic spline with d degrees of freedom. We emphasize that this model controls for temporal trends at the national level, since $s(t;d)$ is common to all counties. The parameters β_0^c are county-specific intercepts. This model is equivalent to the following:

$$\log E(Y_t^c) = \log N_t^c + \eta_0^c + \eta_1 \widehat{PM}_t + \eta_2 (PM_t^c - \widehat{PM}_t) + s^*(t; d-1). \tag{3}$$

The term $\widehat{PM}_t$ denotes the national trend in annual average $PM_{2.5}$, calculated as the fitted values of a linear regression model having PM_t^c as dependent variable (for all counties) and a natural cubic spline of time with d degrees of freedom ($s(t;d)$) as predictor. The term $s^*(t;d-1)$ is a smooth function of time modelled using a natural cubic with $d-1$ degrees of freedom, orthogonal to $\widehat{PM}_t$ and PM_t^c.

Models (2) and (3) yield the same predicted values. The only difference between the two models is in parametrization: model (3) takes the smooth function $s(t;d)$ in model (2), which is represented by a set of d basis functions, and breaks it into: 1) $\widehat{PM}_t$, which is a linear combination of the d basis functions; and 2) the remaining smooth function, $s^*(t;d-1)$. The parameters η_2 in model (3) and β_1 in model (2) are exactly the same.

Model (3) allows us to estimate the association between $PM_{2.5}$ and mortality trends at two different scales: national and local. The parameter η_1 denotes the association between month-to-month variation in the national trend in $PM_{2.5}$, $\widehat{PM}_t$, and month-to-month variation in the national trend in mortality rates. The parameter η_2 denotes the association between month-to-month variation in county-specific deviations in PM_t^c from the national trend, and month-to-month variation in county-specific Y_t^c from the national trend. In other words, η_2 provides evidence as to whether counties having steeper declines in PM_t^c also have steeper declines in mortality relative to the national trend.

CORRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

If model (1) describes the causal link between annual average $PM_{2.5}$ and mortality, then the estimates of η_1 and η_2 in model (3) should be equal, absent confounding and measurement error. Therefore, a comparison of $\hat{\eta}_1$ of $\hat{\eta}_2$ provides important evidence on the causal hypothesis formulated in model (1).

In model (3), the term $\widehat{PM}_t$ controls for the national trend in annual average $PM_{2.5}$, and $s^*(t; d-1)$ controls for the remaining national trend in mortality. This implies that the effect of $\widehat{PM}_t$ (η_1), which represents the association between trends in $PM_{2.5}$ and mortality at the national scale, is potentially confounded by time-varying factors such as changes in the economy and health care. We focus on η_2, the association between trends in $PM_{2.5}$ and mortality at the local scale, because we believe that this exposure effect is less likely to be confounded. In order to bias the estimation of η_2, a confounder must cause county-specific deviations in PM_t^c and mortality from their national trends. An example of such a factor is "health consciousness," a characteristic of counties that relates to their aggressiveness in implementing national air pollution regulatory standards and in improving health care.

It can be shown that the $PM_{2.5}$-mortality association as measured by model (1) is a composite of two pieces of information:

$$\widehat{\delta}_1 \approx w\,\widehat{\eta}_1 + (1-w)\,\widehat{\eta}_2, \tag{4}$$

where $\hat{\eta}_1$ and $\hat{\eta}_2$ are the estimated coefficients of the national and local $PM_{2.5}$ trends from model (3), $w = (1/V_1)/(1/V_1 + 1/V_2)$, and V_1 and V_2 are the statistical variances of $\widehat{\eta}_1$ and $\widehat{\eta}_2$. That is, $\widehat{\delta}_1$ is a weighted average of the association between the national $PM_{2.5}$ and mortality trends and the association between the local $PM_{2.5}$ and mortality trends.

We also consider a pooled model that combines information across age-sex strata and

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

allows for stratum- and region-specific smooth functions of time:

$$\log E(Y_t^c) = \log N_t^c + \alpha_0^{cs} + \alpha_1 PM_t^c + s^{rs}(t;d), \tag{5}$$

where α_0^{cs} are county and age-sex stratum-specific intercepts, $s^{rs}(t;d)$ is a stratum- and region-specific smooth function of time modelled using a natural cubic spline with d degrees of freedom, and α_1 is the $PM_{2.5}$ effect common to all age-sex strata. When $d = 0$, model (5) is an age-sex stratum pooled version of model (1), and α_1 is the association between $PM_{2.5}$ and mortality without control for trends. When $d > 0$, model (5) is a pooled version of model (2), or equivalently of model (3). The parameter α_1 is the association between month-to-month deviations in $PM_{2.5}$ and mortality from their respective stratum- and region-specific trends, i.e., the association between local trends.

In all log-linear models, we use a negative binomial variance model,[25]

$$Var(Y_t^c) = E(Y_t^c)\left(1 + E(Y_t^c)/\phi\right).$$

We fit the models by iterating between fitting the log-linear model for fixed ϕ, and estimating ϕ using a method of moments estimator.[26]

We report results for all models when $d = 16$ degrees of freedom are used to model the national trend over 3 years.

Sensitivity analyses: We assess the sensitivity of the results to different choices of d, from $d = 0$ to $d = 32$. We vary d on the $\log_2$ scale so as to maintain the same knots as d increases. We also calculate robust standard errors,[27] which account for residual autocorrelation in monthly mortality rates. Robust and model-based standard errors are similar, and hence we report only the results using model-based standard errors. We also explore the sensitivity of our results to the time period over which $PM_{2.5}$ is averaged. We fit the same models, using average $PM_{2.5}$ over the previous two years as exposure.

COBRA
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Results

In the measurement error analysis of $PM_{2.5}$ trends, we find that 80% of the total variability in monitor-specific trends is attributable to variability among counties.

Table 1 summarizes the results of the analysis of variance of PM_t^c. We find that 91% of the total variance in PM_t^c can be attributed to the space component, and 5% to the space-by-time component. Note that the space-by-time variance of PM_t^c, which provides the main source of information for estimating η_2 in model (3), is larger than the variance due to the time component, and accounts for 57% of the temporal variance.

Figure 2A displays regional and national linear trends in annual average $PM_{2.5}$ concentrations. We estimate these trends by linearly regressing PM_t^c on t. Figure 2B shows regional and national trends in log mortality rates. These trends are estimated by log-linearly regressing Y_t^c on t with offset $\log N_t^c$. The log-linear models are fit separately for each age-sex stratum, and the fitted values are averaged across strata. Annual average $PM_{2.5}$ concentrations are decreasing over time in all regions except in the Northeast and Central regions. Mortality rates are decreasing in all regions. This information is used to estimate the association between the national trends in $PM_{2.5}$ and mortality in model (3).

Figure 3A shows how county-specific linear trends in PM_t^c deviate from the national linear trend. County-specific PM_t^c trends are calculated by linearly regressing PM_t^c on t. The deviations are the differences between these county-specific trends and the national trend. The deviations are centered at zero in order to draw attention to the trends, rather than to the levels. Figure 3B shows how county-specific linear log mortality rate trends deviate from the national linear trend. For each county and age-sex stratum, we calculate the trend in the log mortality rate by log-linearly regressing Y_t^c on t with offset $\log N_t^c$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

Deviations are the differences between the county- and stratum-specific trends and the national stratum-specific trend. The deviations are centered at zero and averaged across age-sex strata. Three counties with very different trends– Los Angeles county (CA), Peoria county (IL), and De Kalb county (GA)– are identified. This plot examines whether counties in which $PM_{2.5}$ is decreasing faster than the national trend also have mortality rates decreasing faster than the national trend. In LA county, for example, $PM_{2.5}$ is increasing relative to the national trend, but mortality is decreasing relative to the national trend. Observe the substantial variability in the county-specific deviations from the national trend. This information is used to estimate the association between local trends in $PM_{2.5}$ and mortality in model (3).

Figure 4 shows a scatterplot of the slopes estimated by linearly regressing PM_t^c on t versus the slopes estimated by log-linearly regressing Y_t^c on t with offset $\log N_t^c$. The mortality rate slopes are averaged across age-sex strata. Los Angeles county (CA), Peoria county (IL), and De Kalb county (GA) are again highlighted. The median $PM_{2.5}$ slope is -0.048 (interquartile range [IQR] = 0.056), which corresponds to an average decrease of 0.58 μg/m^3 $PM_{2.5}$ concentration per year ($12 \times 0.048 = 0.58$). The median log mortality rate slope is -2.112 $\times 10^{-3}$ (IQR = 1.904 $\times 10^{-3}$), which corresponds to a 2.50% decrease in the mortality rate each year on average ($e^{12 \times -2.112 \times 10^{-3}} = 0.9750$). We evaluate the association between the $PM_{2.5}$ slopes and the mortality slopes using a weighted linear regression model, where the weights are the inverse variances of the mortality slope estimates. The regression line is superimposed. There is no evidence of a positive association between the rates of change in $PM_{2.5}$ and log mortality rates (slope estimate = -0.001; 95% CI = -0.006 to 0.003).

Table 2 displays the results of models (1) and (3), separately for each age-sex stratum.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

We report results for model (3) when $d = 16$, but note that any $d \geq 8$ provides qualitatively similar results. The first column contains estimates of δ_1 from model (1), and the second and third columns show estimates of η_1 and η_2 from model (3). As expected from Figure 2, we find a strong evidence of an association between national trends in $PM_{2.5}$ and mortality (second column). However, there is no evidence of an association between local trends in any of the strata (third column). This is consistent with the data displayed in Figure 3 and the exploratory analysis shown in Figure 4.

The first column of Table 2 contains results from model (1). These estimates quantify the association between annual average $PM_{2.5}$ and mortality without control for temporal confounding. In each age-sex stratum, $\widehat{\delta_1}$ lies between $\widehat{\eta}_1$ (second column) and $\widehat{\eta}_2$ (third column). This follows from the weighted average result, equation (4). Observe that the positive association between $PM_{2.5}$ and mortality estimated based on model (1) (δ_1) is a combination of a very strong positive association between national trends (η_1) and a null association between local trends (η_2). The large difference between these two effects (η_1 and η_2) suggests that they should not be combined in a weighted average. In the fourth column of Table 2, we show the weight that is given to the national trend component, $\frac{1/V_1}{1/V_1+1/V_2}$. We find that the national trend component accounts for about 40% of the information contained in δ_1.

Figure 5 shows estimates of the association between annual average $PM_{2.5}$ and mortality based on the pooled model (5), as a function of the degrees of freedom allowed in each stratum- and region-specific trend term per year. When $d = 0$, we estimate the association without control for temporal confounding. We estimate that a 1 $\mu g/m^3$ increase in $PM_{2.5}$ is associated with an 0.86% increase in mortality (95% CI = 0.64% to 1.09%). This corresponds to an 8.96% increase in mortality for each 10 $\mu g/m^3$ increase in $PM_{2.5}$,

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

which is remarkably similar to the $PM_{2.5}$ effect estimated in previous cohort studies.[12–14] However, as $d > 0$ (that is, as we start to control for smooth trends in $PM_{2.5}$ and mortality), the evidence changes. For $d \geq 8$ we find no evidence of an association between local trends in $PM_{2.5}$ and mortality.

Figure 5 also displays the results of model (5) separately for each year. Again if there is a causal association between exposure and outcome, the estimated association should be similar in different subsets of the data. When $d = 0$, the three year-specific $PM_{2.5}$ effects are very different, but all statistically significant. The change in mortality associated with a 1 $\mu g/m^3$ increase in $PM_{2.5}$ ranges from a 4.02% decrease in 2001 (95% CI = 3.25% to 4.79%) to a 5.30% increase in 2002 (95% CI = 4.41% to 6.19%). As d increases, the three year-specific estimates become more similar, and settle around a null effect.

We explore the sensitivity of our results to the time period over which $PM_{2.5}$ concentrations are averaged, by using $PM_{2.5}$ averaged over the previous two years as exposure (and using mortality data for 2001 and 2002). The results of the age-sex stratum-specific models are shown in Table 3. For model (3), using now just two years (24 months) of mortality data, we report results when $d = 8$ degrees of freedom are used to model the national trend. Results are qualitatively similar for all $d \geq 4$. The results shown in Table 3 are qualitatively similar to those in Table 2. We find an association between national trends for most strata, but no association between local trends.

Discussion

This paper illustrates an approach to the assessment of confounding bias in observational studies where environmental exposures and health outcomes vary in time and

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

space. We introduce a causal model for the association between monthly variations in annual average $PM_{2.5}$ and mortality rates. We show how this association can be decomposed into two components: the association between national trends in $PM_{2.5}$ and mortality, and the association between local trends in $PM_{2.5}$ and mortality. We find a very large association at the national scale, and no evidence of association at the local scale. We believe that the national trend component is more likely to be confounded than the local trend component. If we set aside the association between national trends, we are left with no evidence of an effect of $PM_{2.5}$ on mortality.

Chay, Dobkin, and Greenstone[9] estimated the association between trends in air pollution and adult mortality in the US using an instrumental-variables approach. Following the Clean Air Act of 1970, counties were designated as "attainment" or "non-attainment" according to their levels of total suspended particulates (TSP). These authors compared changes in TSP levels and mortality rates across attainment and non-attainment counties. They found that, while non-attainment status was associated with large reductions in TSP in the years 1971-1972, non-attainment status was not significantly associated with reductions in adult or elderly mortality.

In another recent paper, Laden and colleagues[28] used extended follow-up data from the Harvard Six Cities Study[14] to examine trends in average $PM_{2.5}$ and mortality rates in six U.S. cities. They partitioned time into two periods, 1974-1989 and 1990-1998. Controlling for average $PM_{2.5}$ in the first time period, they found that a reduction in average $PM_{2.5}$ in the second period was associated with a reduction in the mortality rate.

In our analysis we define long-term exposure as average $PM_{2.5}$ over the preceding year. National monitoring data for $PM_{2.5}$ started in 1999 and therefore we do not have data to estimate exposures for longer time periods. Our sensitivity analysis suggests that, when a

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

different exposure averaging period is used, results do not change qualitatively. Determining the appropriate long-term $PM_{2.5}$ exposure measure is an important scientific question that deserves further research.

Our analysis focuses on 113 counties with relatively complete $PM_{2.5}$ data over the study period, and uses data from the best single monitor for each county. We conducted the same analysis using a larger set of 250 US counties (meeting less strict $PM_{2.5}$ measurement criteria) and using as exposure the annual average $PM_{2.5}$ concentration averaged across all monitors in each county. This produced very similar results.

In these data, we estimate that 20% of the total variability in $PM_{2.5}$ trends is within-county variability (measurement error). Using a regression calibration correction,[29] we estimate that our $PM_{2.5}$ local trends coefficient is attenuated by 20% (1 - 0.80, where 0.80 is the intraclass correlation). In contrast, we assume that the national trend in $PM_{2.5}$ is estimated without error, since it is based on data from 113 counties. We conclude that the attenuation of the local trends coefficient is not enough to explain the discrepancy between the effects of the local and national $PM_{2.5}$ trends.

Our study, as with most air pollution studies, is potentially affected by various sources of bias. This bias comes from three sources. First, we use county-level exposure to represent individual-level exposure. Previous studies have shown that this tends to bias exposure effects towards the null.[30,31] The second source of bias is due to the lack of information on area-level time-varying confounders that affect both $PM_{2.5}$ and mortality trends. We control for such factors by including a smooth function of time in the regression models. The third source of bias is due to the lack of adjustment for individual-level covariates beyond age and sex. However, previous cohort studies have found the air pollution-mortality association to be robust to the adjustment for both time-varying and

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

time-invariant individual-level confounders.[32]

Our proposed methods can be used more generally to diagnose unmeasured confounding in observational studies where the exposure and outcome vary in time and space. We decompose the exposure variable into orthogonal components and allow each component to have a unique effect on the outcome. If there is a causal link between exposure and outcome, then the exposure components must affect the outcome equally, assuming there is no confounding or covariate measurement error. Therefore, differences in these scale-specific effects are a useful diagnostic tool for assessing confounding and its magnitude. If the exposure effects differ, we suggest focusing on the exposure effects that are thought least likely to be confounded. *A priori* knowledge about the potential confounders can guide the partitioning: the least confounded exposure effects are those corresponding to scales of variation at which the confounders are approximately constant.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

References

1. Greenland S, Morgenstern H. Confounding in health research. *Annual Review of Public Health.* 2001;22:189-212.

2. Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosomatic Medicine.* 2004;66:868-75.

3. Dominici F, McDermott A, Hastie T. Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association.* 2004;99:938-48.

4. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society, Series A.* 2006;169:179-203.

5. Burnett R, Ma R, Jerrett M, et al . The spatial association between community air pollution and mortality: A new method analyzing correlated geographic data. *Environmental Health Perspectives.* 2001;109:375-80.

6. Jerrett M, Burnett RT, Willis A, et al . Spatial analysis of the air pollution-mortality relationship in the context of ecologic confounders. *Journal of Toxicology and Environmental Health A.* 2003;66:1735-77.

7. Touloumi G, Samoli E, Pipikou M, Le Tertre A, Atkinson R, Katsouyanni K. Seasonal confounding in air pollution and health time-series studies: Effect on air pollution effect estimates. *Statistics in Medicine.* 2006;25:4164-78.

8. Curriero FC, Heiner KS, Samet JM, et al . Temperature and mortality in 11 cities of the eastern United States. *American Journal of Epidemiology.* 2002;155:80-7.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

9. Chay K, Dobkin C, Greenstone M. The Clean Air Act of 1970 and adult mortality. *Journal of Risk and Uncertainty.* 2003;27:279-300.

10. Chay K, Greenstone M. Air quality, infant mortality, and the Clean Air Act of 1970. Tech. Rep. Working Paper No. 04-08MIT Department of Economics 2003.

11. Greenstone M. The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufacturers. *Journal of Political Economy.* 2002;110:1175-219.

12. Pope CA, Thun MJ, Namboodiri MM, et al . Particulate air pollution as a predictor of mortality in a prospective study of US adults. *American Journal of Respiratory and Critical Care Medicine.* 1995;151:669-74.

13. Pope CA, Burnett RT, Thun MJ, et al . Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association.* 2002;287:1132-41.

14. Dockery DW, Pope CA, Xu X, et al . An association between air pollution and mortality in six US cities. *New England Journal of Medicine.* 1993;329:1753-9.

15. Jerrett M, Burnett RT, Ma R, et al . Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology.* 2005;16:791-801.

16. Dominici F, Peng RD, Bell ML, et al . Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association.* 2006;295:1-9.

17. Bell ML, Dominici F, Samet J. Time series studies of particulate matter. *Annual Review of Public Health.* 2004;25:247-80.

COBRA
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

18. Samet J.M., Dominici F., Curriero F.C., et al . Fine particulate air pollution and mortality in 20 US cities, 1987 - 1994. *New England Journal of Medicine.* 2000;343:1742-49.

19. Samet J. M., Zeger S. L., Dominici F., et al . *The National Morbidity, Mortality, and Air Pollution Study, Part I: Methods and Methodological Issues.* Cambridge MA: Health Effects Institute 2000. Available at: http://pubs.healtheffects.org/view.php?id=228.

20. Samet J. M., Zeger S. L., Dominici F., et al . *The National Morbidity, Mortality, and Air Pollution Study, Part II: Morbidity and Mortality from Air Pollution in the United States.* Cambridge MA: Health Effects Institute 2000. Available at: http://pubs.healtheffects.org/view.php?id=118.

21. 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. *Environmental Health Perspectives.* 2001;109:349–353.

22. Samoli E, Touloumi G, Zanobetti A, et al . Investigating the dose-response relation between air pollution and total mortality in the APHEA-2 multicity project. *Occupational and Environmental Medicine.* 2003;60:977–982.

23. 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;12:521-531.

24. Peng RD, Dominici F, Pastor-Barriuso R, et al . Seasonal analyses of air pollution and mortality in 100 US cities. *American Journal of Epidemiology.* 2005;161:585-94.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

25. McCullagh P, Nelder JA. *Generalized Linear Models.* Boca Raton, FL: Chapman and Hall 1989.

26. Agresti A. *Categorical Data Analysis.* Hoboken, NJ: John Wiley and Sons, Inc. 2002.

27. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986;73:13-22.

28. Laden F, Schwartz J, Speizer FE, et al . Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities Study. *American Journal of Respiratory and Critical Care Medicine.* 2006;173:667-72.

29. Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models.* Boca Raton, FL: Chapman and Hall 1995.

30. Zeger SL, Thomas D, Dominici F, et al . Measurement error in time-series studies of air pollution: Concepts and consequences. *Environmental Health Perspectives.* 1999;108:419-26.

31. Dominici F, Zeger SL, Samet JM. A measurement error model for time-series studies of air pollution and mortality. *Biostatistics.* 1999;1:157-75.

32. Krewski D., Burnett R.T., Goldberg M.S., et al . *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Part II: Sensitivity Analyses.* Cambridge MA: Health Effects Institute 2000.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

Table 1: Variability in PM_t^c in space, time, and space-by-time dimensions. This is based on a linear model with dependent variable PM_t^c and independent variables: 1) county-specific intercepts (space dimension); 2) a smooth function of time modelled as a natural cubic spline of month with 16 degrees of freedom (time dimension); and 3) an interaction between the county-specific indicators and the smooth function of time (space-by-time interaction). The first column shows the percent of the total variance of PM_t^c attributable to each of the three components, and the second column shows the percent of the total temporal variation in PM_t^c attributable to the "time" and the "space-by-time" components.

	% Variance	% of Temporal Variability
Space	90.90	—
Time	3.90	42.92
Space x Time	5.19	57.08
Residual	< 0.01	—
Total	100.00	—

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Table 2: Point estimates and 95% confidence intervals (CIs) for long-term effects of $PM_{2.5}$ on mortality, by age-sex stratum. The percent change in the mortality rate per 1 $\mu g/m^3$ increase in $PM_{2.5}$ is shown. The first three columns summarize the effects of PM_t^c (from model (1)), $\widehat{PM}_t$ (from model (3)), and $PM_t^c - \widehat{PM}_t$ (from model (3)), respectively. Estimates in the first column are approximately a weighted average of estimates in the second and third columns, according to the weighted average result (equation (4)). The fourth column shows the weight that is given to the national trend component.

Age (years)	Sex	% Change in mortality rate per $\mu g/m^3$ increase in $PM_{2.5}$ (δ_1) Point estimate (95% CI)	% Change in mortality rate per $\mu g/m^3$ increase in $PM_{2.5}$ national trend (η_1) Point estimate (95% CI)	% Change in mortality rate per $\mu g/m^3$ increase in $PM_{2.5}$ local trends (η_2) Point estimate (95% CI)	% Information from national trend ($\frac{1/V_1}{1/V_1+1/V_2}$)
65-74	Men	1.48 (0.93 to 2.03)	3.55 (2.77 to 4.34)	0.04 (-0.58 to 0.67)	40.66
	Women	0.83 (0.24 to 1.43)	1.97 (1.12 to 2.83)	-0.03 (-0.71 to 0.66)	40.15
75-84	Men	0.85 (0.34 to 1.35)	2.48 (1.83 to 3.14)	-0.34 (-0.87 to 0.19)	40.87
	Women	0.77 (0.28 to 1.27)	2.29 (1.66 to 2.93)	-0.31 (-0.82 to 0.21)	40.77
85+	Men	0.70 (0.03 to 1.38)	1.38 (0.52 to 2.26)	< 0.01 (-0.71 to 0.73)	41.26
	Women	0.59 (0.05 to 1.12)	1.65 (1.01 to 2.29)	-0.22 (-0.74 to 0.31)	41.19

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

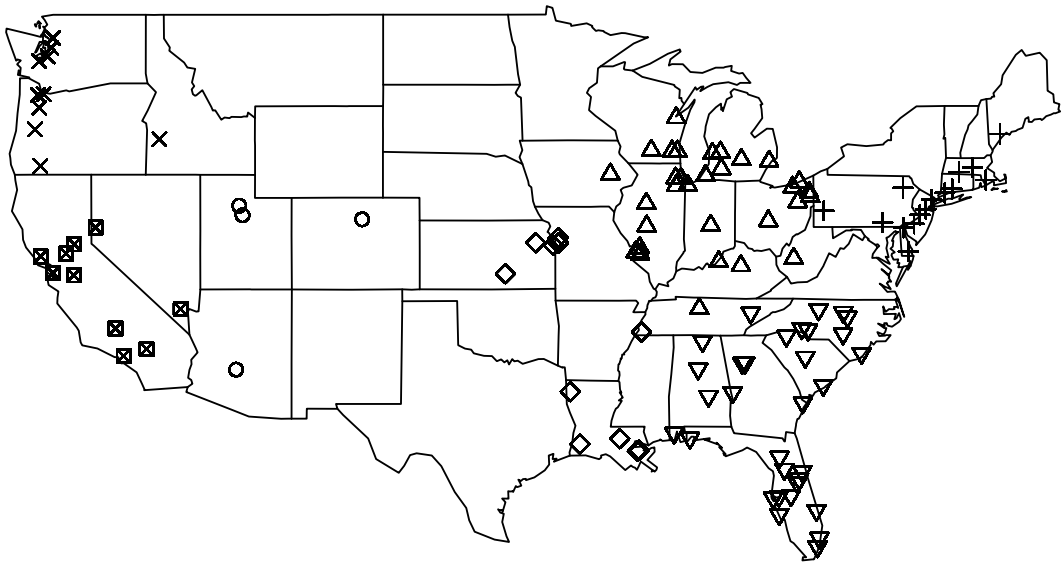
Table 3: Point estimates and 95% CIs for long-term effects of $PM_{2.5}$ on mortality, by age-sex stratum, using $PM_{2.5}$ concentrations averaged over the previous two years as exposure. The percent change in the mortality rate per 1 $\mu g/m^3$ increase in $PM_{2.5}$ is shown. The first 3 columns summarize the effects of PM_t^c (from model (1)), $\widehat{PM}_t$ (from model (3)), and $PM_t^c - \widehat{PM}_t$ (from model (3)), respectively. Estimates in the first column are approximately a weighted average of estimates in the second and third columns, according to the weighted average result (equation (4)). The fourth column shows the weight that is given to the national trend component.

Age (years)	Sex	% Change in mortality rate per $\mu g/m^3$ increase in $PM_{2.5}$ (δ_1) Point estimate (95% CI)	% Change in mortality rate per $\mu g/m^3$ increase in $PM_{2.5}$ national trend (η_1) Point estimate (95% CI)	% Change in mortality rate per $\mu g/m^3$ increase in $PM_{2.5}$ local trends (η_2) Point estimate (95% CI)	% Information from national trend ($\frac{1/V_1}{1/V_1+1/V_2}$)
65-74	Men	0.74 (-0.48 to 1.97)	4.48 (2.57 to 6.43)	-1.25 (-2.61 to 0.14)	36.25
	Women	0.24 (-1.06 to 1.57)	1.48 (-0.57 to 3.58)	-0.40 (-1.90 to 1.12)	35.51
75-84	Men	0.51 (-0.61 to 1.64)	2.87 (1.27 to 4.49)	-0.73 (-1.89 to 0.45)	36.14
	Women	0.83 (-0.24 to 1.90)	2.85 (1.31 to 4.41)	-0.11 (-1.23 to 1.03)	35.86
85+	Men	-0.70 (-2.15 to 0.76)	0.18 (-1.84 to 2.23)	-1.37 (-2.87 to 0.15)	36.25
	Women	-0.34 (-1.48 to 0.82)	2.17 (0.63 to 3.73)	-1.54 (-2.65 to -0.40)	36.23

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

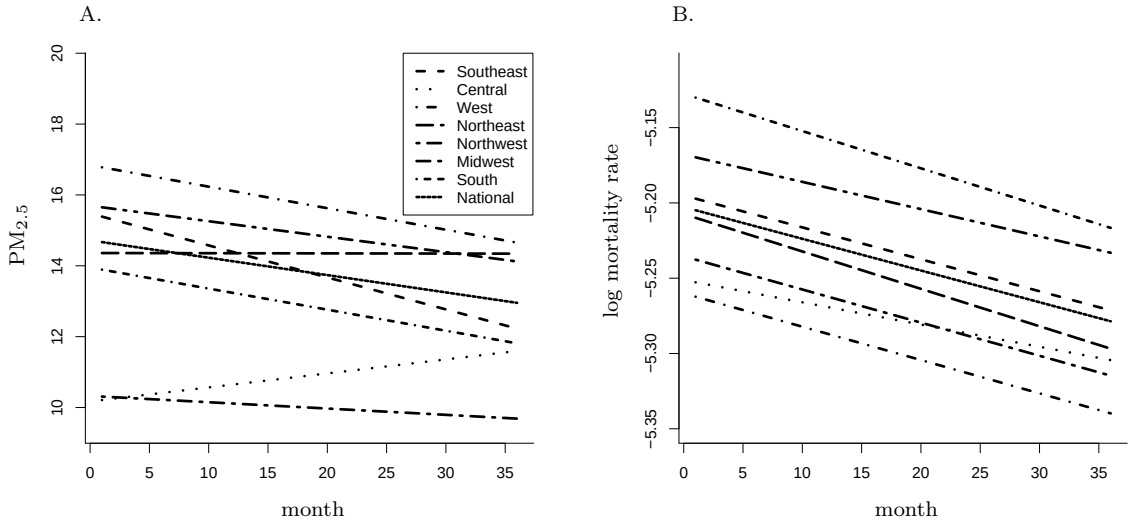
Figure 1: The location of the 113 counties used in the analysis. Each region is plotted using a different symbol.



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

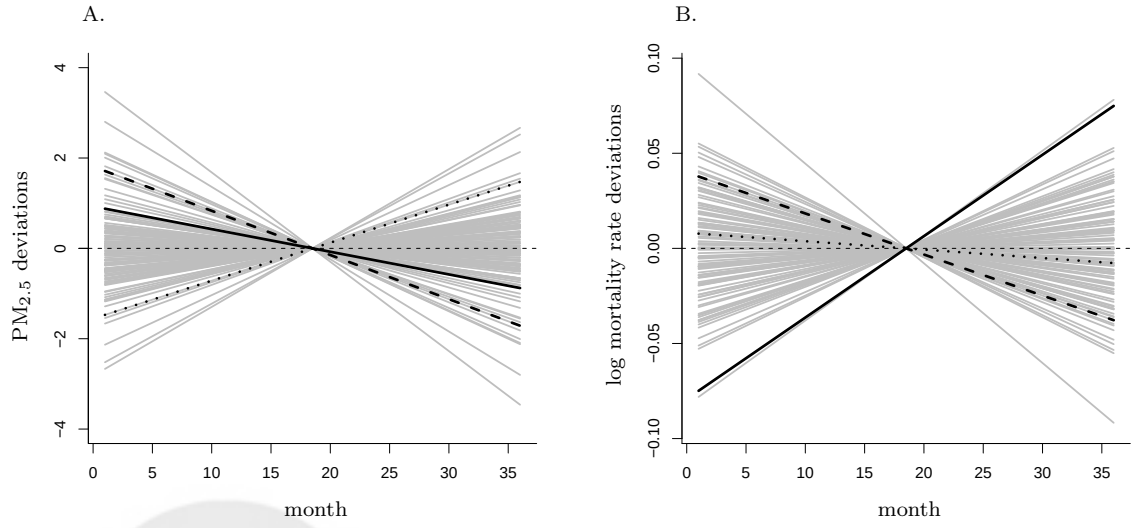
Figure 2: Regional and national linear trends in (A) PM_t^c and (B) log mortality rates. Trends in PM_t^c are calculated based on linear models, and log mortality rate trends are calculated using log-linear models. These mortality trends are then averaged across age-sex strata.



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

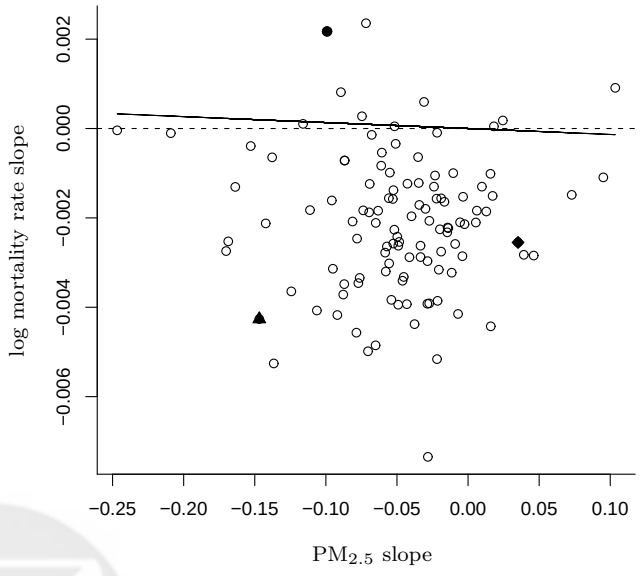
Figure 3: County-specific deviations in (A) linear PM_t^c trends and (B) linear log mortality rate trends from their respective national linear trends. The mortality deviations are averages of age-sex stratum-specific deviations from their respective national trends. Three counties Los Angeles, CA (dotted line), De Kalb County, GA (dashed line), and Peoria County, IL (solid line) counties, are highlighted.



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104

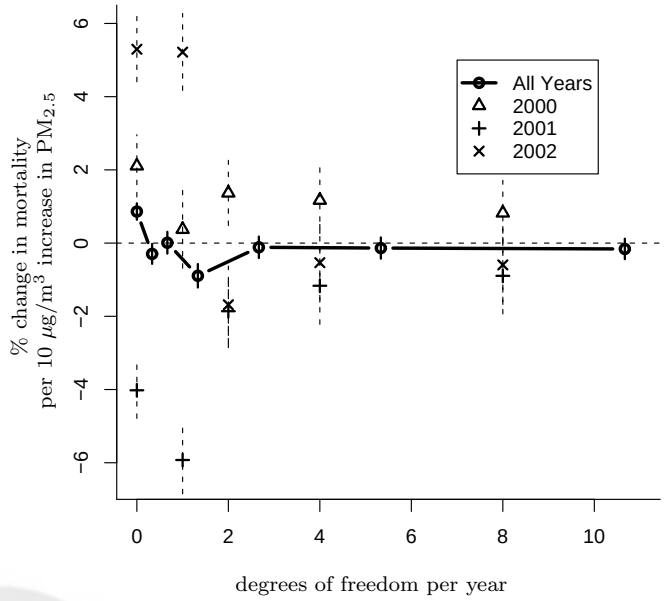
Figure 4: County-specific linear rates of change in PM_t^c versus county- and stratum-specific linear rates of change in mortality. The mortality trends are averaged across age-sex strata. A weighted linear regression model is overlaid, where the weights are the inverse variances of the mortality slope estimates. Three counties, Los Angeles, CA (diamond), De Kalb County, GA (triangle), and Peoria County, IL (circle), are highlighted.



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Figure 5: The percent increase in the mortality rate associated with a 1 μg/m^3 increase in $PM_{2.5}$ based on model (5), as a function of the degrees of freedom per year. Confidence intervals are superimposed. Estimates are also shown separately for each year.



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/jhubiostat/paper104