

Multi-site Time Series Analysis

Long-term Exposure

**SAMSI Spatial Epidemiology
Fall 2009**

Howard Chang
hhchang@jhsph.edu

Multi-site Time Series Analysis

Data

Exposure varies both spatially and temporally.

Health outcomes are only available as aggregated counts at various locations.

Approach

- First estimate the health effects the best we can at each location:
 - Measurement error and confounder control.
- Then pool information across locations via Bayesian hierarchical modeling:
 - Borrow information across locations.
 - Examine effect modification due to location-specific characteristics.

Cohort Study (Long-Term Exposure)

The American Cancer Society Cohort

Population

Include approx 550,000 adults in 151 US metropolitan area (subset of 1.2 millions from all US from the original study initiated in 1982).

Each individuals completed a questionnaire with information on age, weight, family history of cancer, use of medication/vitamins, diet, alcohol/tobacco use, ... etc

Outcome

Person time: periodic follow-up to link participates to the Nation Death Index.

Exposure

Mean concentration of air pollutant for the entire study period was obtained by averaging monitoring measurements for each metropolitan area.

Cohort Study (Long-Term Exposure)

The American Cancer Society Cohort

Statistical Model

Stage I

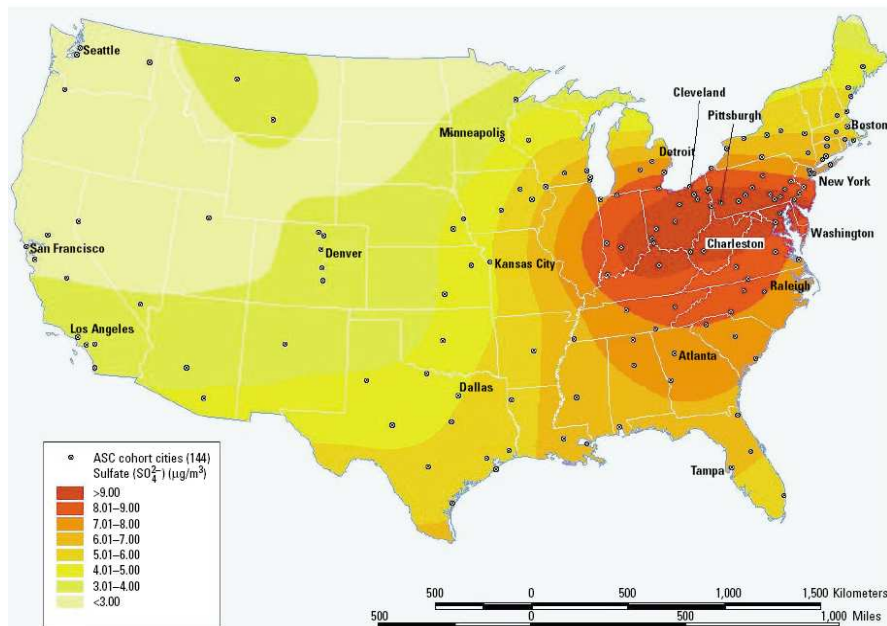
- Survival data are modeled on the *individual level* with and indicators for each metropolitan area using Cox-proportional hazards model stratified by 1-year category.
- Obtain estimates of community-specific log relative risks $\beta(s)$ at locations $s = 1, 2, \dots, S$ and its covariance matrix V .

Stage II

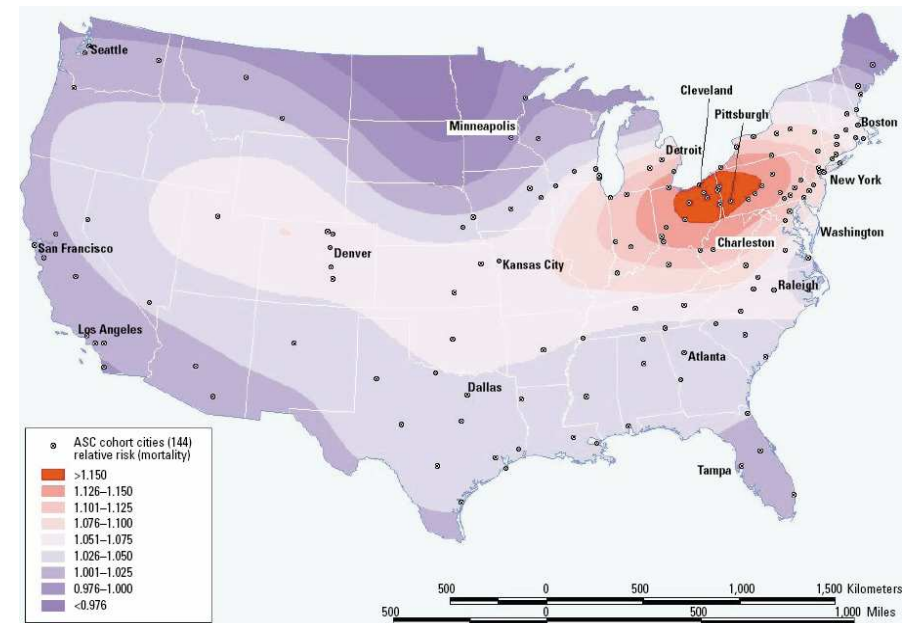
- Fit a linear random effect model with $\beta(s)$ and community-level long-term exposure to sulfate.
- Include a spatial random effect estimated via LOESS smoother

ACS Cohort Example

Mortality



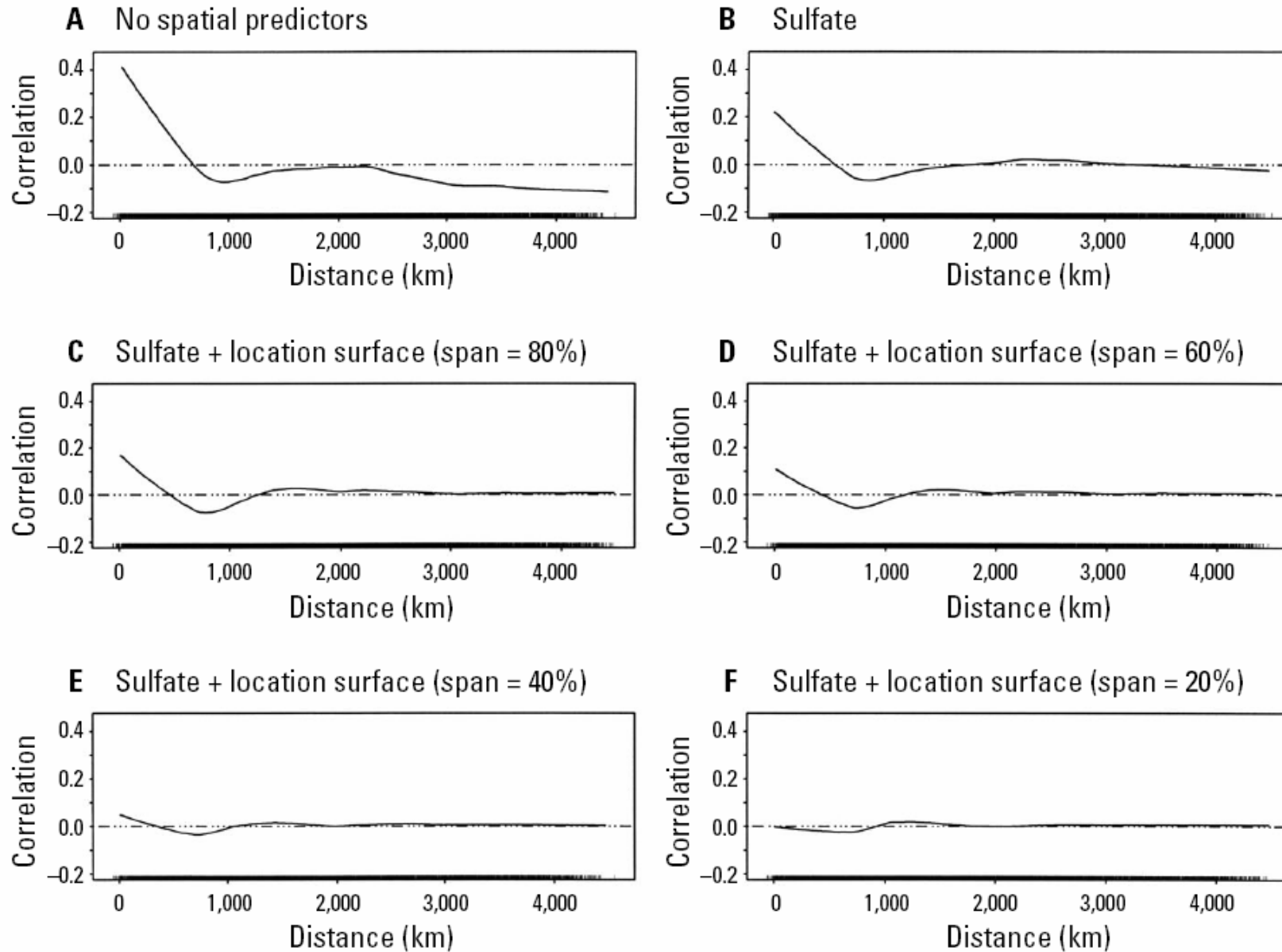
Sulfate



Burnett R, Ma R, Jerrett M, et al. The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data. *Environ Health Perspect.* 2001; 109(suppl 3):375-380.

ACS Spatial RE Example

ACF of Residuals



ACS Sulfate Pooled Relative Risks

Model type (model number)	Span (%)	Regression coefficient (β) (standard error)	Relative risk ^a (95% confidence interval)
Cox (1)	NA	0.0118 (0.00177)	1.051 (1.036, 1.066)
Random-effects Cox (2)	NA	0.0125 (0.00252)	1.055 (1.033, 1.077)
Spatial (3)	NA	NA	NA
Spatial (4)	NA	0.0127 (0.00252)	1.055 (1.033, 1.077)
Spatial (5)	80	0.0106 (0.00277)	1.046 (1.022, 1.070)
Spatial (6)	60	0.0093 (0.00261)	1.040 (1.018, 1.062)
Spatial (7)	40	0.0085 (0.00245)	1.036 (1.016, 1.058)
Spatial (8)	20	0.0081 (0.00219)	1.035 (1.016, 1.053)

Cohort Design

Including spatial random effects in cohort study for the health effects of long-term exposure is essential to control for *unmeasured (ecological) confounders* that are risk factors for mortality and are correlated with pollution in space. [Similar to *temporal confounder control* in time series analysis]

Also positive correlation between individuals in health outcome in the same community due to unmeasured risk factors will result in an under-estimation of the standard error of the health effect coefficients.

Cohort versus Time Series

The magnitude of associations reported by cohort studies are typically larger than those from time series analysis.

Time series analysis captures the *number of excess* death attributed to short-term increase in air pollution; cohort study directly measures the *year of life lost* due to air pollution.

Study design issue	Time-series studies	Cohort studies
Outcome	Counts	Person-time
Exposure variance	Temporal	Spatial
Time from exposure to outcome	Short (days or weeks)	Cumulative (years, lifetime)
Duration of exposure considered	Short term	Can be long term, in the past, etc.

Künzli N, Medina S, Kaiser R, Quénel P, Horak, F Jr., and Studnicka M. Assessment of Deaths Attributable to Air Pollution: Should We Use Risk Estimates based on Time Series or on Cohort Studies? American Journal of Epidemiology 2001 11 1060-1055

Time Series and Long-term Exposure

With continuing monitoring of air pollution and health data, we can examine long-term exposure in a time series framework.

Define long-term exposure as: *previous 12-month* average exposure

Consider the following model:

$$h^c(a, t) = h^c(a) \exp(P M_t^c \beta)$$

$h(\cdot)$ = location-specific baseline hazard function

c = location

a = age group (1-year per group above 65)

t = month index between 2000~2006

Partitioning of Evidence

With the assumption that death is independent and constant hazard over one-year age interval, we can use a log-linear model:

$$\log E(Y_{at}^c) = \log(T_{at}^c) + \log(h^c(a)) + PM_t^c \beta$$

This estimates the overall association between monthly risk of death and $PM_{2.5}$ concentration of the previous year in county c .

- Model $\log(h^c(a))$ using thin-plate splines of age with 3 df and location-specific intercepts. This accounts for location-age baseline hazard. This avoids cross-section confounding.

Partitioning of Evidence

We can re-write the previous model as:

$$\log E(Y_{at}^c) = \log(N_{at}^c) + \log(h^c(a)) + (PM_t^c - \overline{PM_t})\beta_1 + \overline{PM_t}\beta_2$$

β_2 = How PM2.5 associates with mortality *nationally*

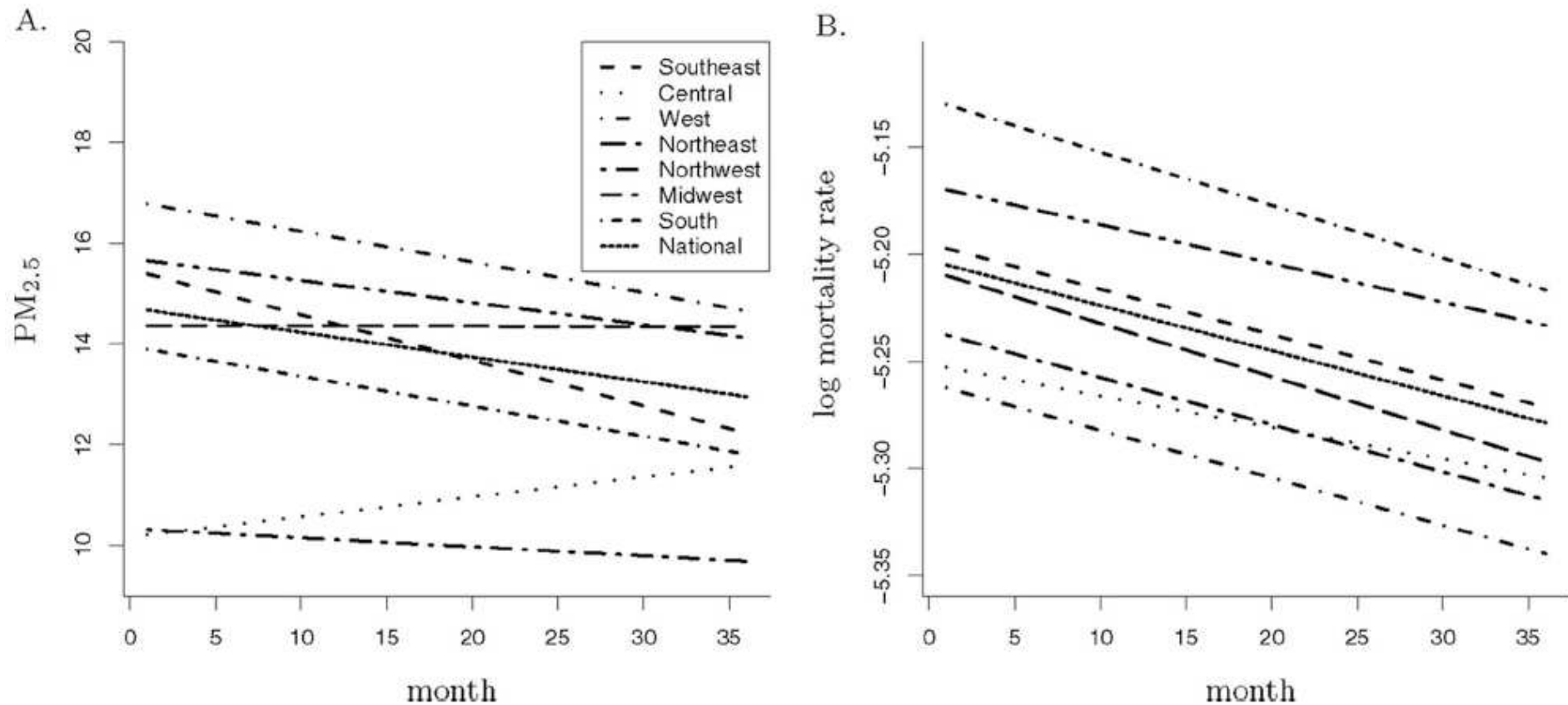
This estimate can be confounded by variables that vary slowly in time. Examples:

- trends in smoking prevalence
- health care access

Greven S, Dominici F and Zeger S (2009). Spatio-Temporal Approach for Estimating Chronic Effects of Air Pollution. *Johns Hopkins University, Department of Biostatistics Working Papers, Paper 190.*

Partitioning of Evidence

Note that both mortality and PM2.5 decreased over time



Janes H, Dominici F, Zeger SL. Trends in Particulate Matter and Mortality in 113 U.S. Counties, 2000-2002: Evidence on the Long Term Effects of Air Pollution. *Epidemiology* 18 416-423.

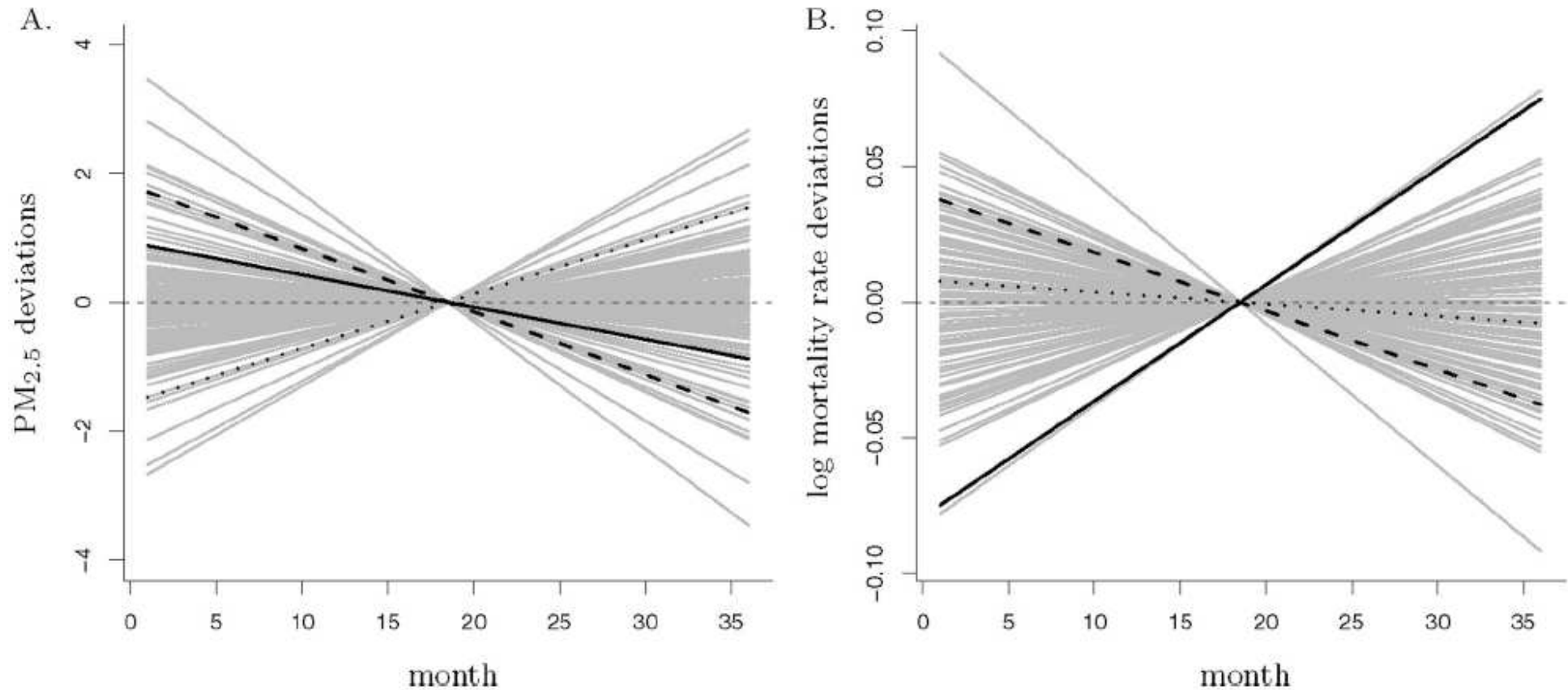
Partitioning of Evidence

$$\log E(Y_{at}^c) = \log(N_{at}^c) + \log(h^c(a)) + (PM_t^c - \overline{PM_t})\beta_1 + \overline{PM_t}\beta_2$$

β_1 = association between faster decline in mortality and faster decline in PM_{2.5} level at county c .

Can be confounded by variables that vary similarly with mortality and PM_{2.5}. For example, people resided in locations with faster decreases in PM_{2.5} level relative to the national level, were also becoming more health conscious.

Partitioning of Evidence



Janes H, Dominici F, Zeger SL. Trends in Particulate Matter and Mortality in 113 U.S. Counties, 2000-2002: Evidence on the Long Term Effects of Air Pollution. *Epidemiology* 18 416-423.

Partitioning of Evidence

