Hierarchical Models for Estimating the Health Effects of Air Pollution

Roger D. Peng, PhD

Department of Biostatistics

Johns Hopkins Bloomberg School of Public Health

2009-07-03

Good

Bad

Ugly

What are the challenges in studying air pollution and health?

- Estimating small (but important) health effects in the presence of much stronger signals
- Results inform substantial policy decisions, affect many stakeholders
  - EPA regulations can cost billions of dollars
- Complex statistical methods are needed and subjected to intense scrutiny

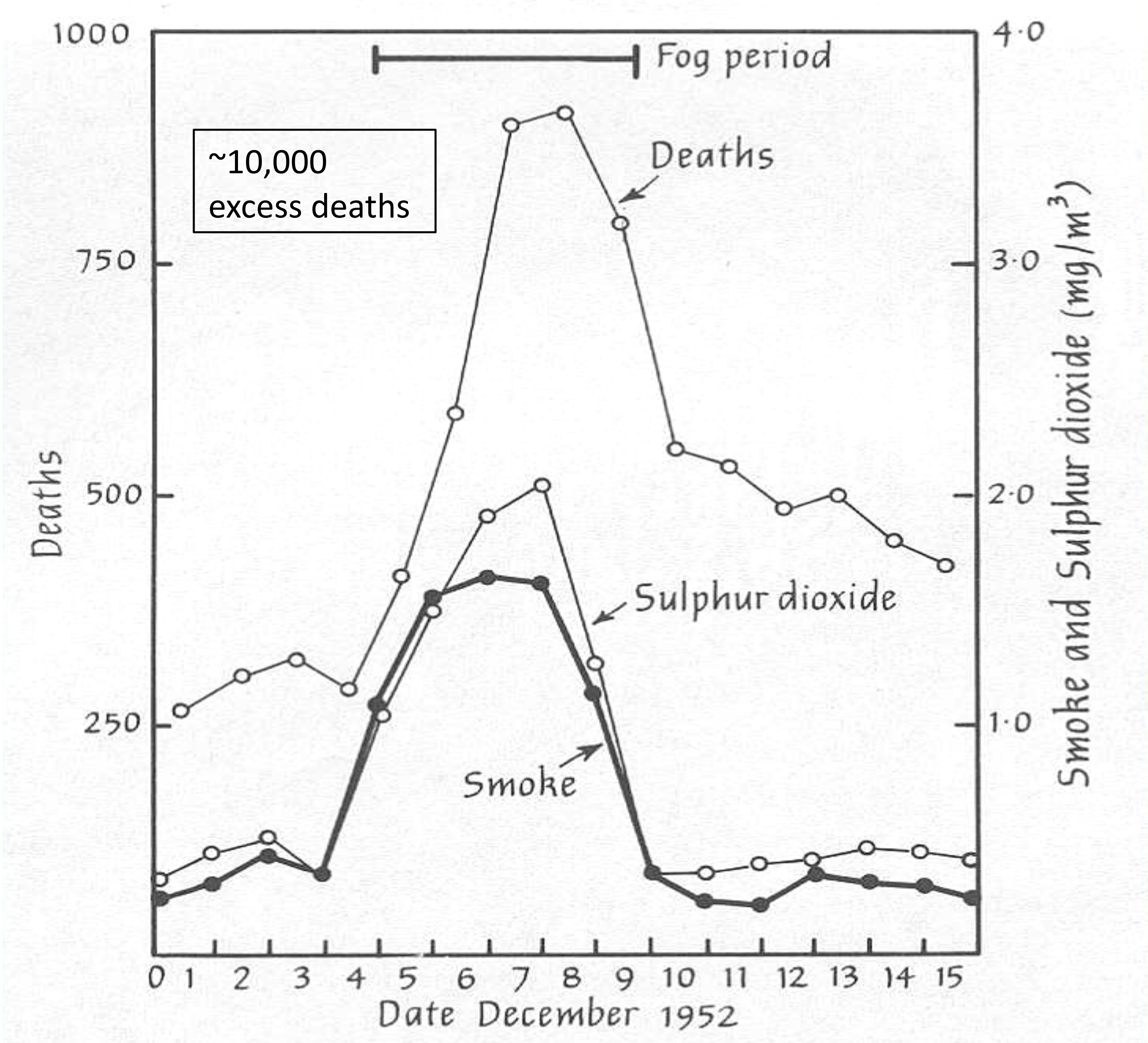
Types of Population-level Air Pollution Studies

Time series

- Examine large populations (cities, counties)
- Estimate short-term, acute effects

Cross-sectional

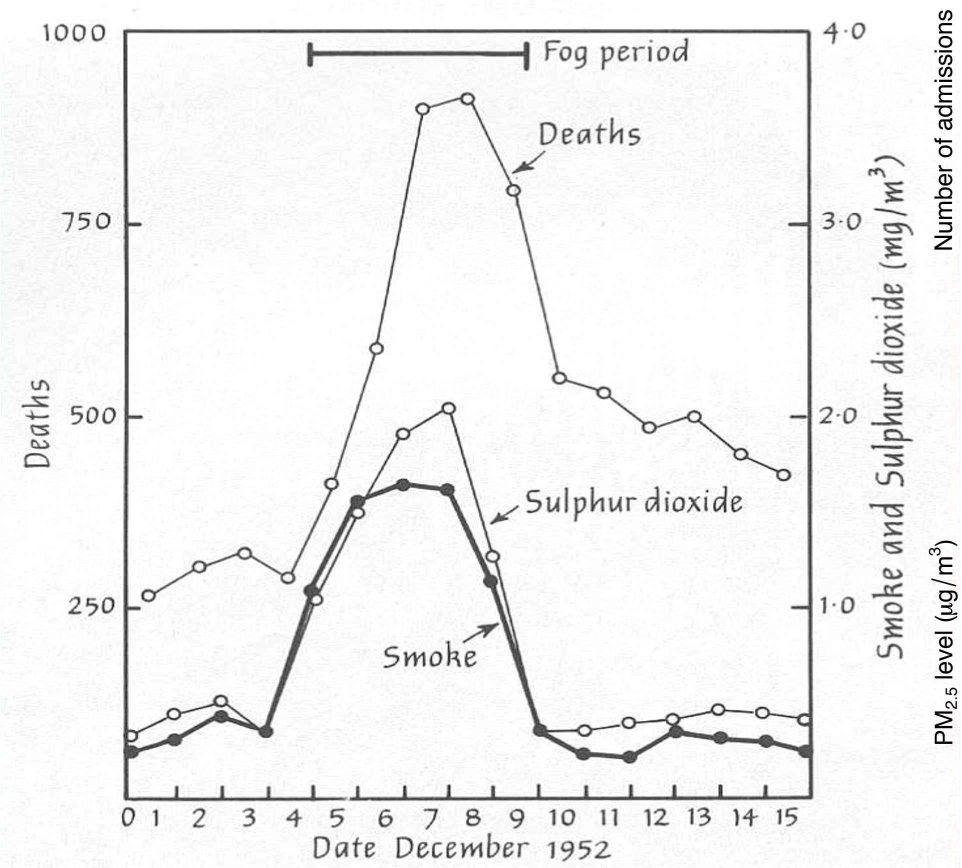
- Examine individual people
- Estimate long-term, chronic effects
- Better assessment of effect of lifetime exposure



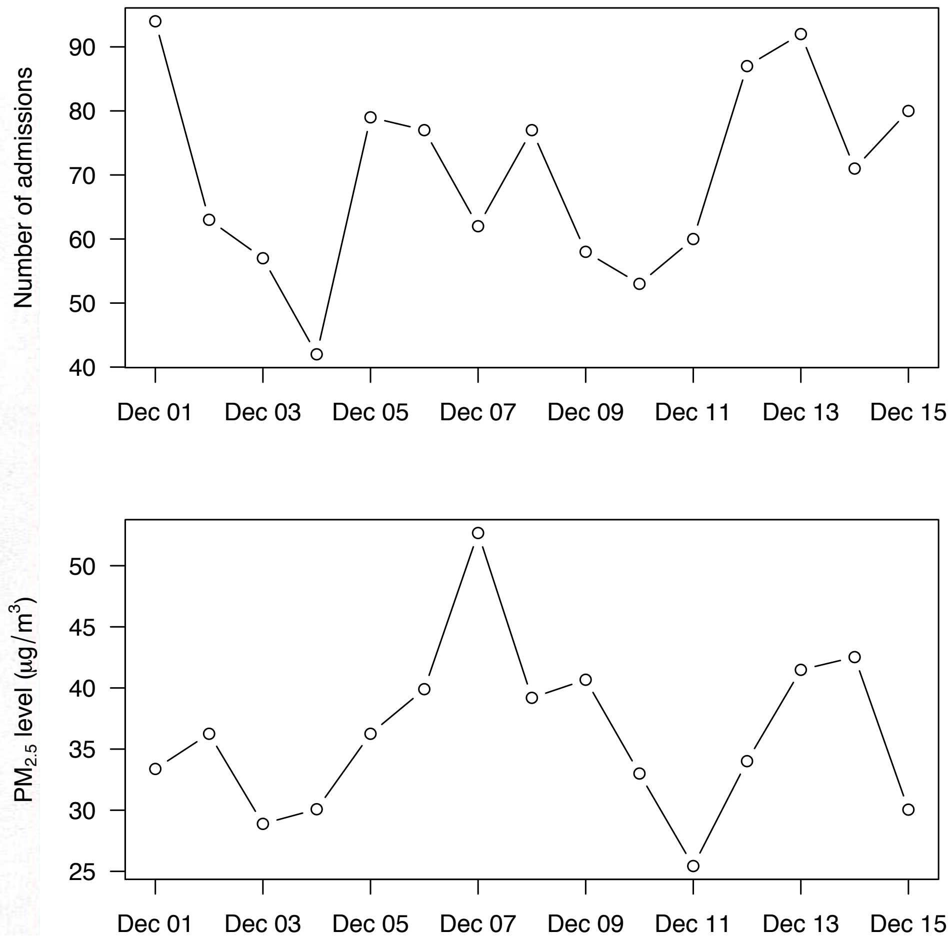
~10,000
excess deaths
Fog period
Deaths
Sulphur dioxide
Smoke
1000
750
500
250
4·0
3·0
2·0
1·0
Deaths
Smoke and Sulphur dioxide (mg/m³)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Date December 1952

Air pollution and health: Then and now

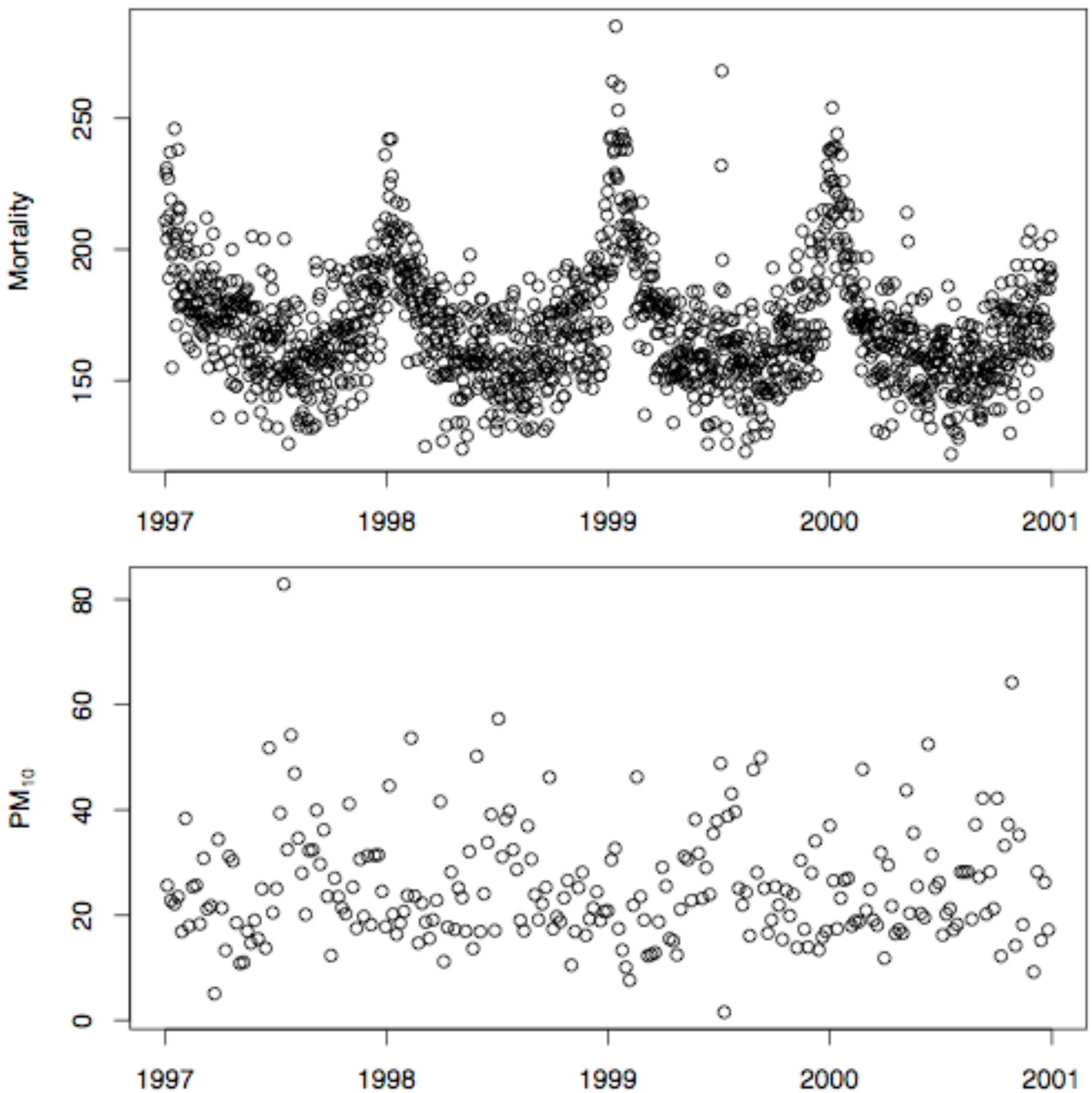
London, December, 1952



Hospital admissions and $PM_{2.5}$ in Chicago, December 2005

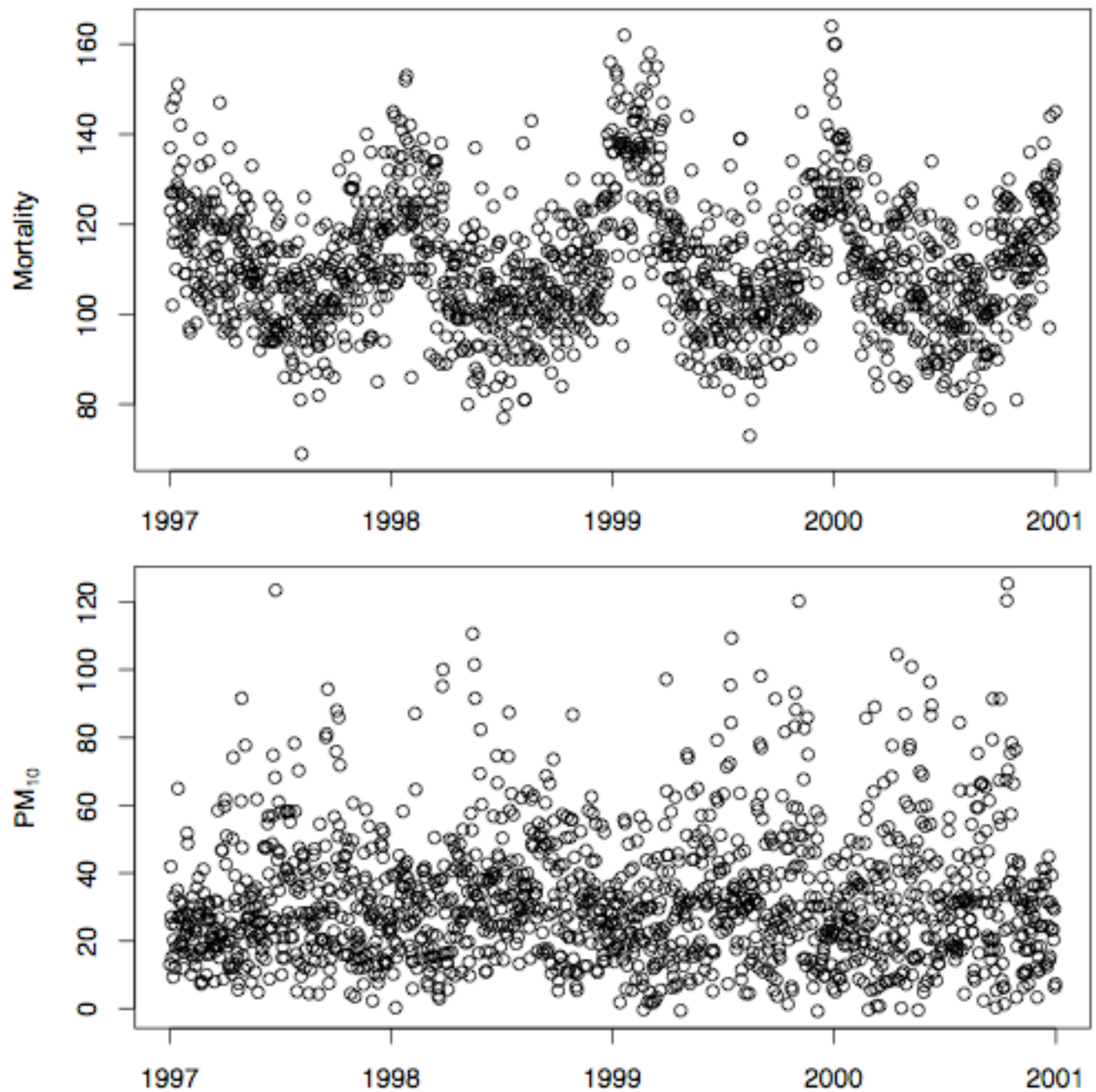


New York



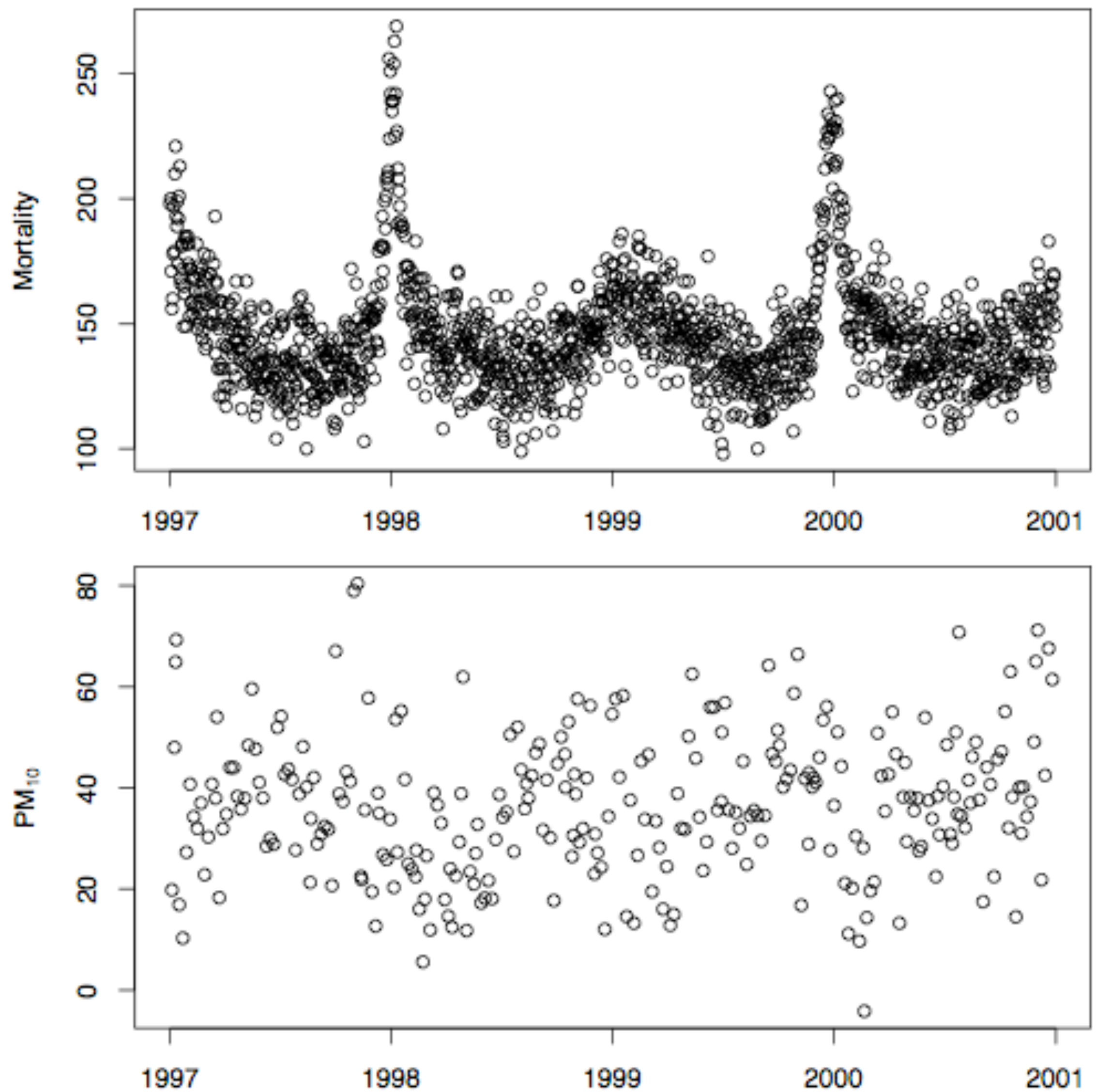
Mortality
250
200
150
1997
1998
1999
2000
2001
PM_{10}
80
60
40
20
0
1997
1998
1999
2000
2001

Chicago

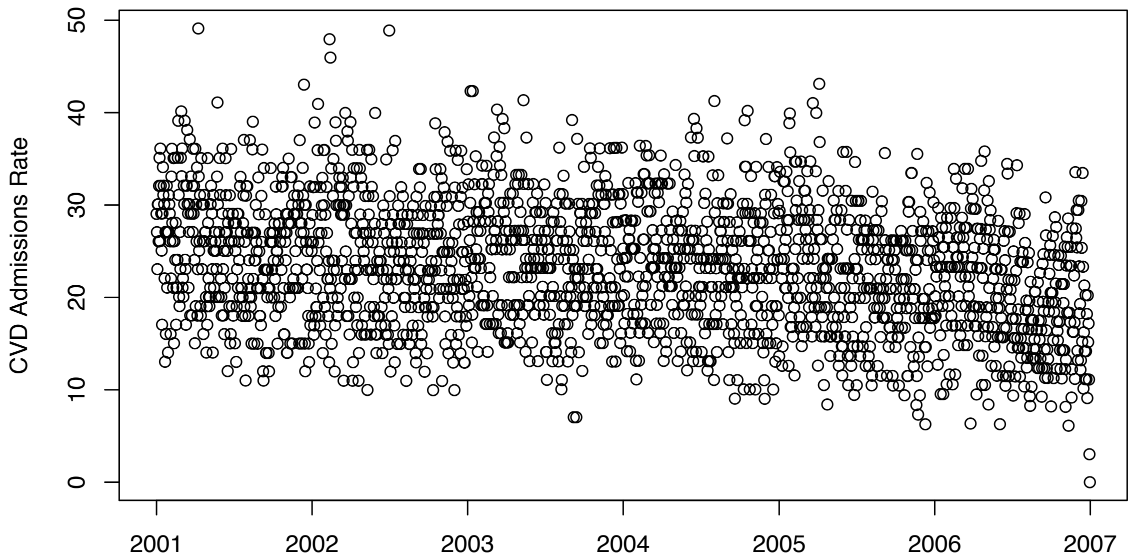


Mortality
160
140
120
100
80
1997
1998
1999
2000
2001
PM_{10}
120
100
80
60
40
20
0
1997
1998
1999
2000
2001

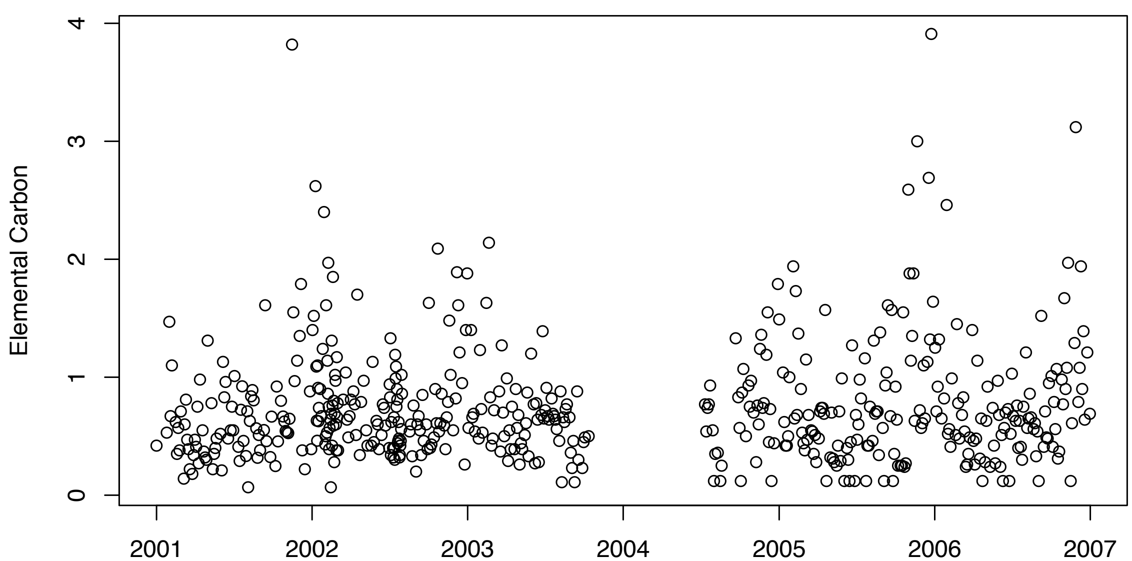
Los Angeles


Mortality
250
200
150
100
1997
1998
1999
2000
2001
PM_{10}
80
60
40
20
0
1997
1998
1999
2000
2001

Baltimore



CVD Admissions Rate
0
10
20
30
40
50
2001
2002
2003
2004
2005
2006
2007



Elemental Carbon
0
1
2
3
4
2001
2002
2003
2004
2005
2006
2007

Time Series Regression Model

$$Y_t = \beta\, x_t + \textit{other stuff}$$

Mortality

Risk

Pollution

City-specific Model

Semiparametric model

$$\begin{aligned}
Y_t^c &\sim \text{Poisson}(\mu_t^c) \\
\log \mu_t^c &= \beta^c\, x_{t-\ell}^c + \text{DOW}_t + \text{AgeCat} \\
&\quad + s(\text{temp}_t;\ df_1) + s(\text{temp}_{t,1-3};\ df_2) \\
&\quad + s(\text{dew pt}_t;\ df_3) + s(\text{dew pt}_{t,1-3};\ df_4) \\
&\quad + s(t;\ df_5) + s(t;\ df_6) \times \text{AgeCat}
\end{aligned}$$

City-specific Model

Semiparametric model

Pollutant series
(PM_{10} or $PM_{2.5}$)

$$
\begin{aligned}
Y_t^c &\sim \text{Poisson}(\mu_t^c) \\
\log \mu_t^c &= \boxed{\beta^c\, x_{t-\ell}^c} + \text{DOW}_t + \text{AgeCat} \\
&\quad + s(\text{temp}_t;\, df_1) + s(\text{temp}_{t,1-3};\, df_2) \\
&\quad + s(\text{dew pt}_t;\, df_3) + s(\text{dew pt}_{t,1-3};\, df_4) \\
&\quad + s(t;\, df_5) + s(t;\, df_6) \times \text{AgeCat}
\end{aligned}
$$

City-specific Model

Semiparametric model

Weather

$$
\begin{aligned}
Y_t^c &\sim \text{Poisson}(\mu_t^c) \\
\log \mu_t^c &= \beta^c\, x_{t-\ell}^c + \text{DOW}_t + \text{AgeCat} \\
&\quad + \boxed{s(\text{temp}_t;\, df_1) + s(\text{temp}_{t,1-3};\, df_2)} \\
&\quad + \boxed{s(\text{dew pt}_t;\, df_3) + s(\text{dew pt}_{t,1-3};\, df_4)} \\
&\quad + s(t;\, df_5) + s(t;\, df_6) \times \text{AgeCat}
\end{aligned}
$$

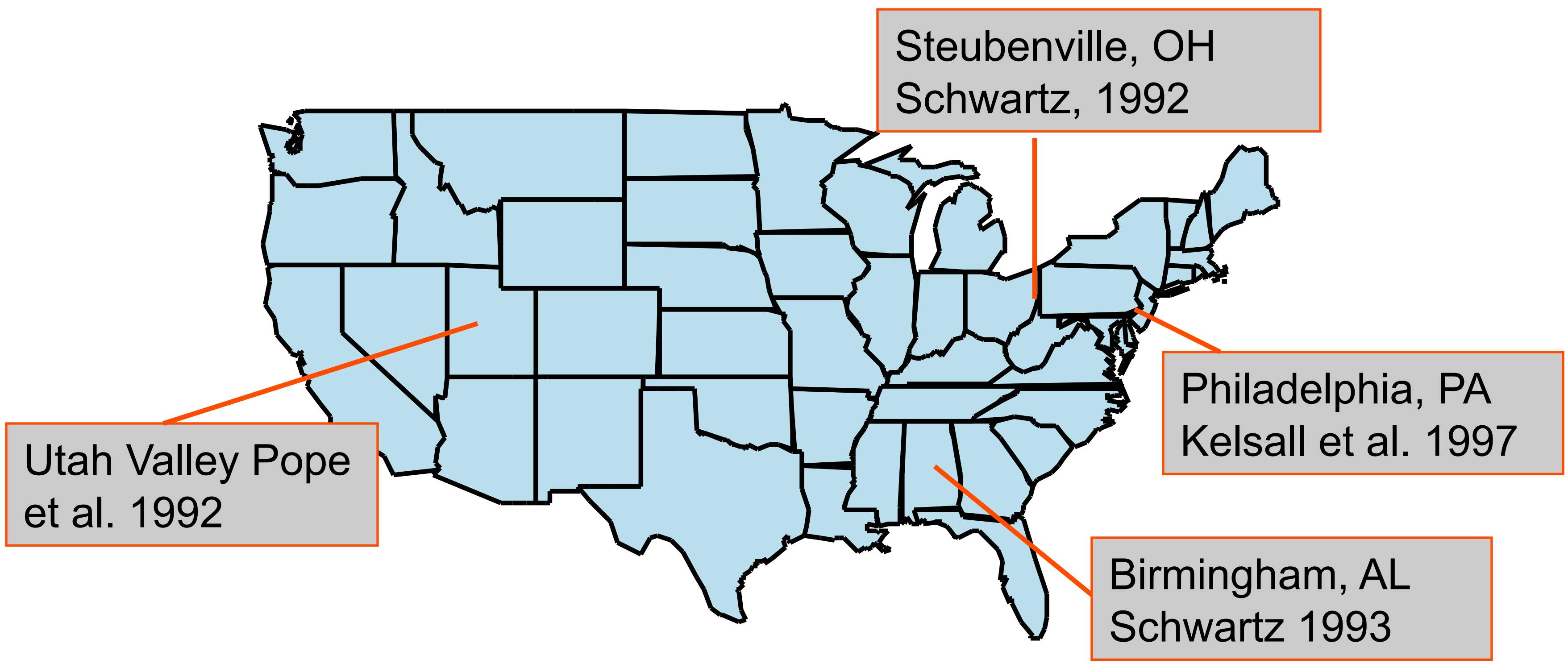
City-specific Model

Semiparametric model

$$
\begin{aligned}
Y_t^c &\sim \text{Poisson}(\mu_t^c) \\
\log \mu_t^c &= \beta^c x_{t-\ell}^c + \text{DOW}_t + \text{AgeCat} \\
&\quad + s(\text{temp}_t;\, df_1) + s(\text{temp}_{t,1-3};\, df_2) \\
&\quad + s(\text{dew pt}_t;\, df_3) + s(\text{dew pt}_{t,1-3};\, df_4) \\
&\quad + \boxed{s(t;\, df_5) + s(t;\, df_6) \times \text{AgeCat}}
\end{aligned}
$$

Seasonal and long-term trends

Single-city Time Series Studies in the U.S.



National

Morbidity

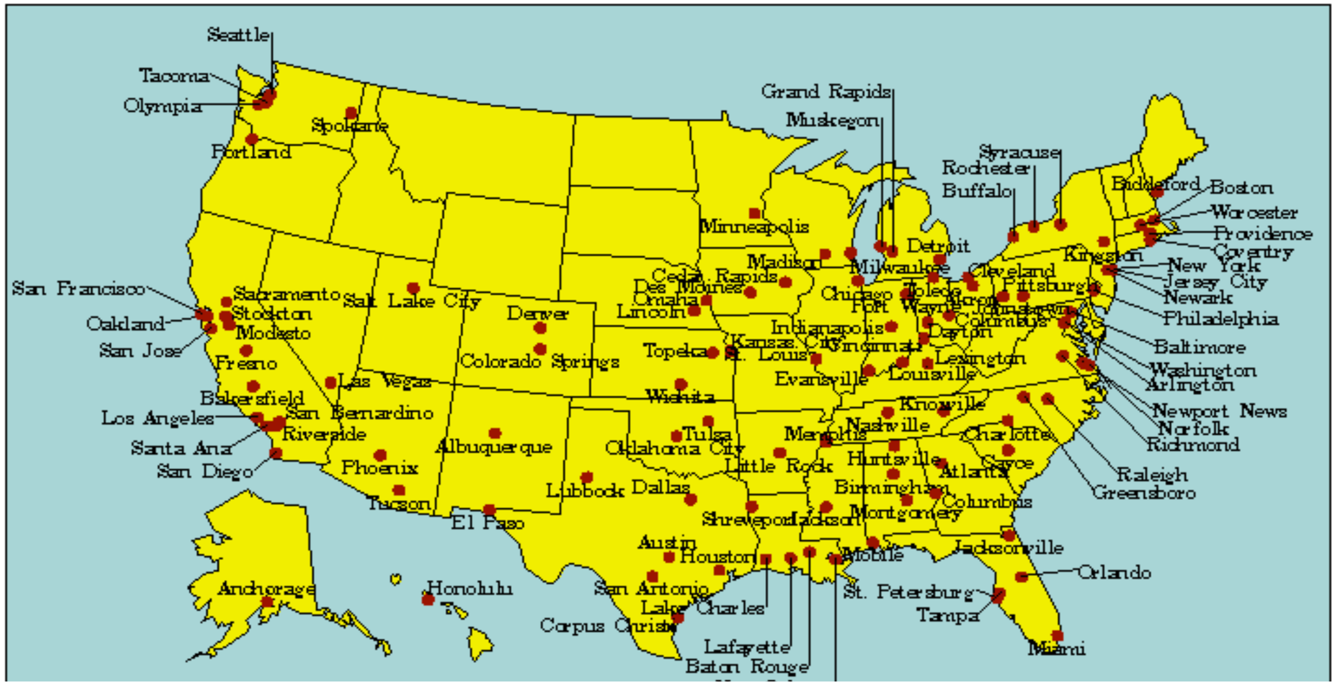
Mortality

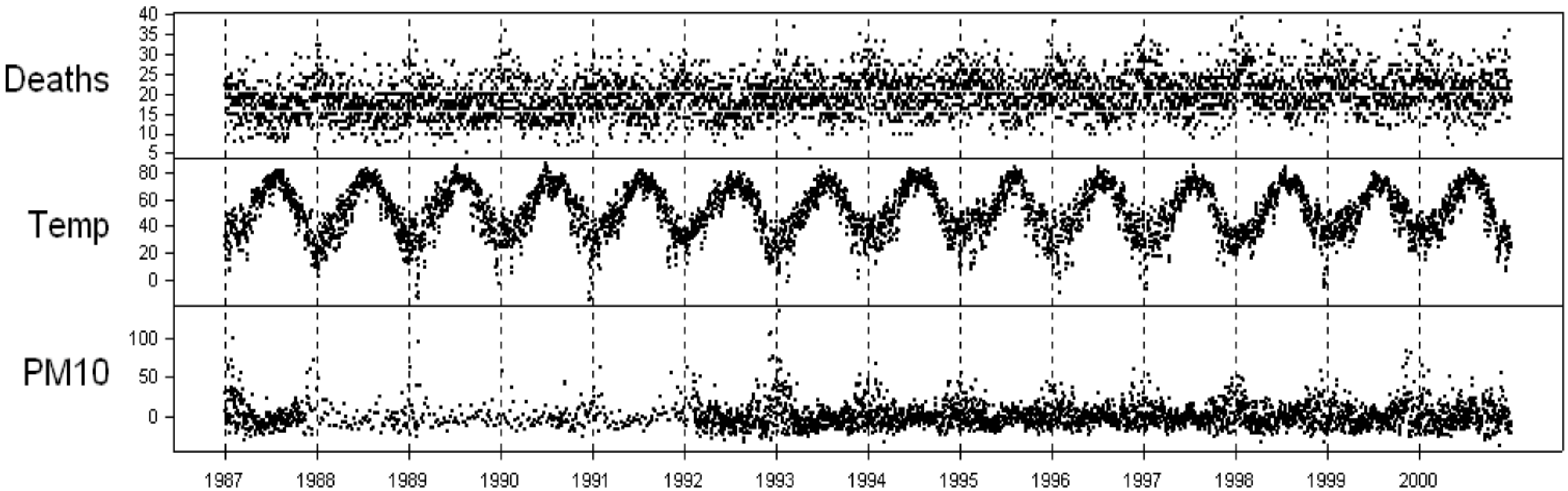
Air

Pollution

Study

1987—2000

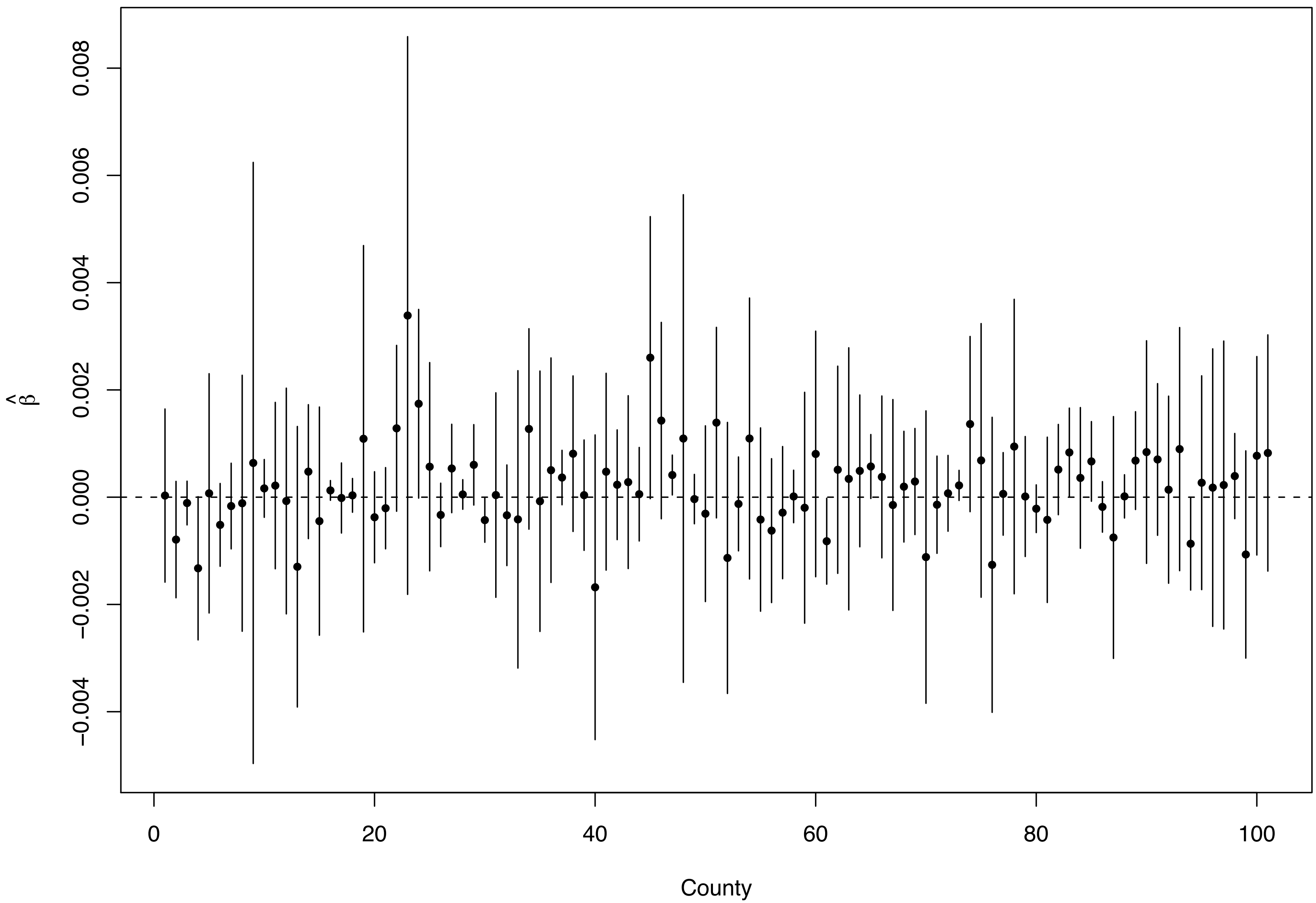




National Morbidity, Mortality, and Air Pollution Study (NMMAPS), 1987—2005

- 108 urban communities
- Cause-specific mortality data from NCHS
  - all-cause (non-accidental), CVD, respiratory, COPD, pneumonia, accidental
- Weather from NOAA
  - Temperature, dew point, relative humidity
- Air pollution data from the EPA
  - PM_{10}, $PM_{2.5}$, O_3, NO_2, SO_2, CO
- U.S. Census 1990, 2000

NMMAPS City-specific Risk Estimates for Mortality and PM_{10}



Why a Joint Analysis of All Cities?

- Individual cities can be selected to show one point or another (publication bias)
- Uniform application of methodology
- Results from individual cities are swamped by statistical noise (remember we're estimating small effects)
- There is no reason to expect that two neighboring cities with similar sources of particles would have qualitatively different relative risks
- "People are people" regardless of where they live

Pooling

- Implement the old idea of borrowing strength across studies
- Estimate heterogeneity between studies
- Estimate a national average effect which takes into account heterogeneity as well as statistical uncertainty

Public Policy Implications

- A national estimate of the air pollution effect provides evidence on the amount of hazard from exposure to air pollution
- Having a single number quantifying the risk is useful for EPA which has to set *national* standards for air pollutants

National Medicare Cohort Air Pollution Study (MCAPS), 1999—2006

- Billing claims for ~48 million adults 65 and older enrolled in Medicare
  - Date of service
  - Treatment, disease (ICD-9), costs
  - Age, gender, race
  - Place of residence (ZIP, county)
- Approximately 200 counties linked with air pollution and weather data

MCAPS Health Outcomes

Daily counts of county-wide hospital admissions for a primary diagnosis:

- Cardiovascular
  - cereberovascular disease
  - peripheral vascular disease
  - ischemic heart disease
  - heart rhythm
  - heart failure
- Respiratory
  - chronic obstructive pulmonary disease
  - respiratory infection

$PM_{2.5}$ → Hospital Admissions

JAMA®
The Journal of the American Medical Association

ORIGINAL CONTRIBUTION

Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases

Francesca Dominici, PhD
Roger D. Peng, PhD
Michelle L. Bell, PhD
Luu Pham, MS
Aidan McDermott, PhD
Scott L. Zeger, PhD
Jonathan M. Samet, MD

Context Evidence on the health risks associated with short-term exposure to fine particles (particulate matter ≤2.5 μm in aerodynamic diameter [$PM_{2.5}$]) is limited. Results from the new national monitoring network for $PM_{2.5}$ make possible systematic research on health risks at national and regional scales.

Objectives To estimate risks of cardiovascular and respiratory hospital admissions associated with short-term exposure to $PM_{2.5}$ for Medicare enrollees and to explore heterogeneity of the variation of risks across regions.

Design, Setting, and Participants A national database comprising daily time-series data daily for 1999 through 2002 on hospital admission rates (constructed from

March 8 2005

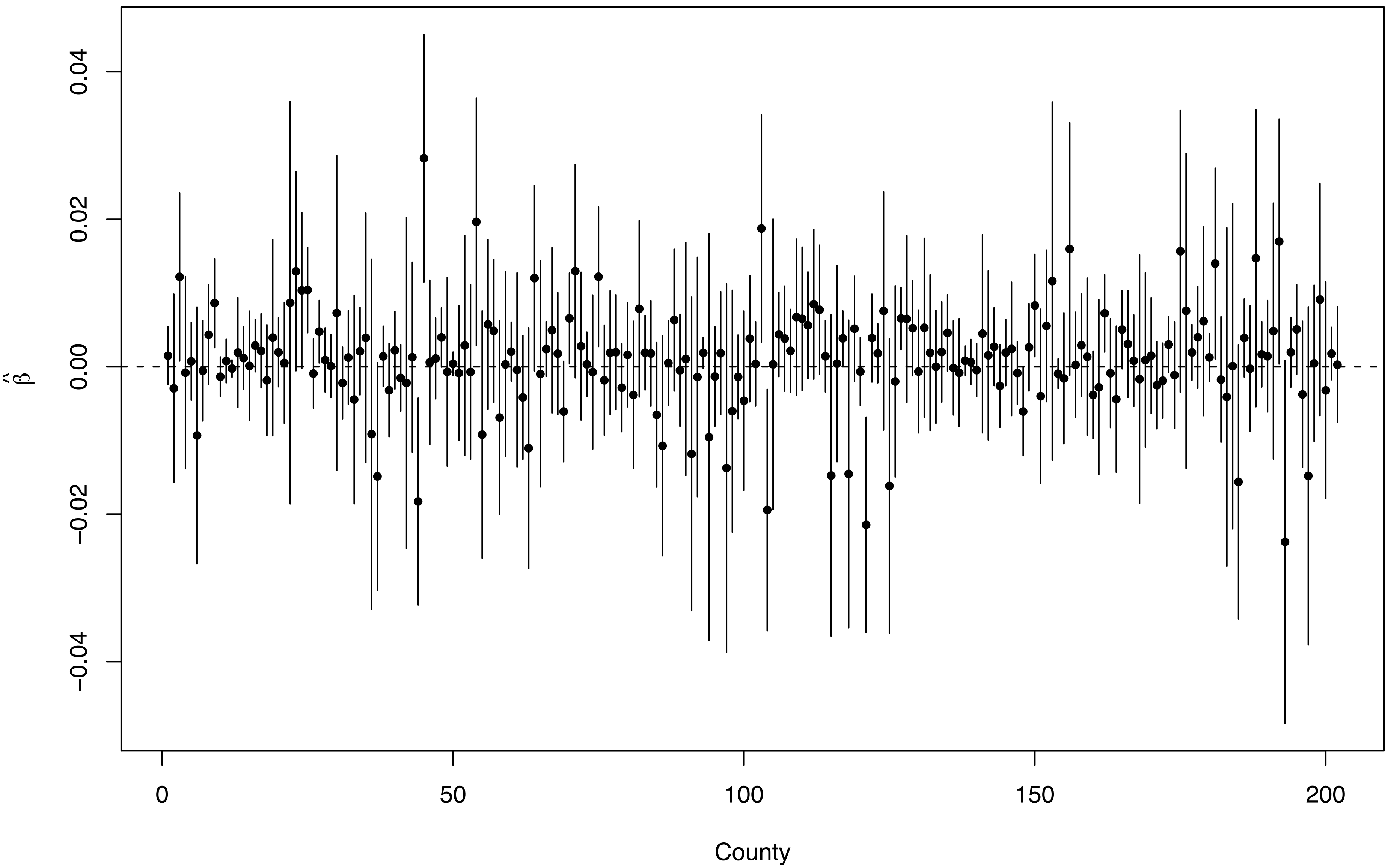
Methods for Multi-site Time Series Studies

Within city: Semi-parametric regressions for estimating associations between day-to-day variations in air pollution and mortality, controlling for confounding factors

Across cities: Bayesian hierarchical models for estimating:

- national-average relative risk
- exploring heterogeneity of air pollution effects across the country

County-specific Maximum Likelihood Estimates ($PM_{2.5}$ and heart failure)



Pooling Log-relative Risks Across Counties

- To produce a national average relative rate we used Bayesian hierarchical models
- We combine (log) relative risks across counties accounting for within-county statistical error and for between-county variability of the “true” relative rates (also called “heterogeneity”)
- To produce regional estimates we used the same two-stage hierarchical model described below but separately within each region

Two stage model

y_j Estimated relative rate for city j

θ_j True relative rate for city j

θ True national-average relative rate

$$y_j = \theta + (y_j - \theta_j) + (\theta_j - \theta)$$

Within city Across cities

Statistical variation/noise Heterogeneity

A Two-stage normal normal model

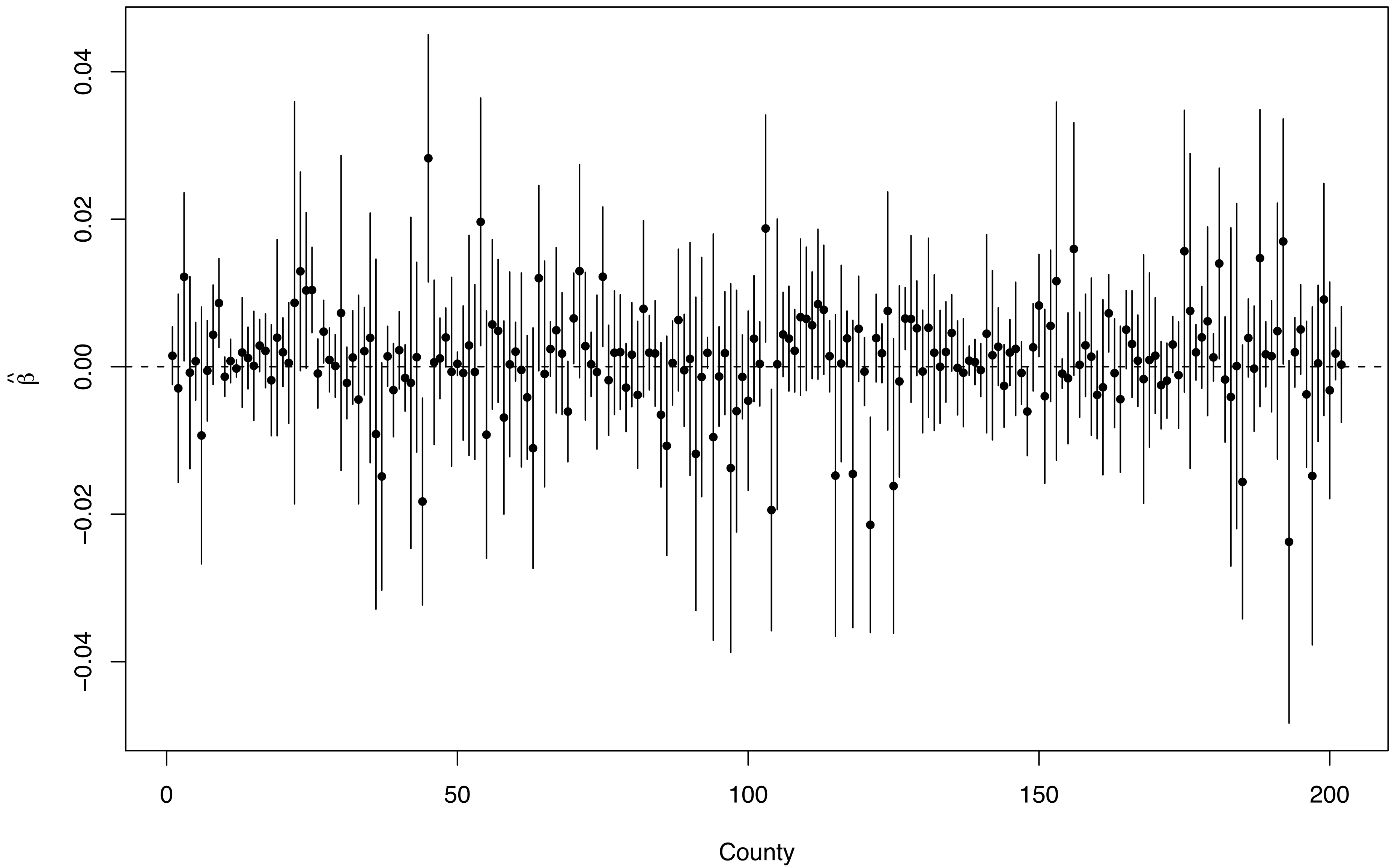
$$y_j = \theta_j + \varepsilon_j; j = 1,..,J$$

$$\varepsilon_j \sim N(0,\sigma_j^2)$$ Statistical variance (known)

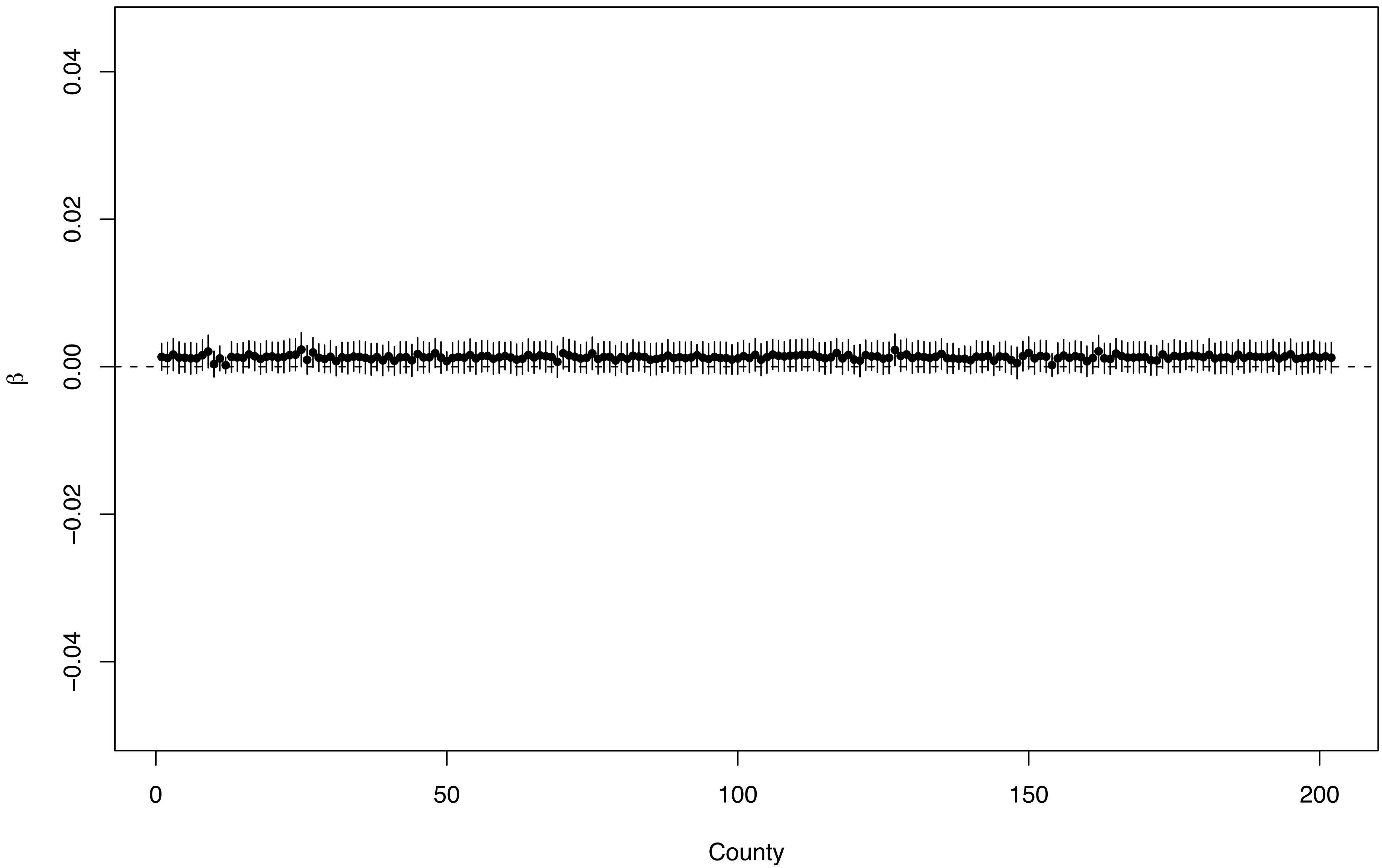
$$\theta_j = \theta + N(0,\tau^2)$$

Between cities variance (unknown)

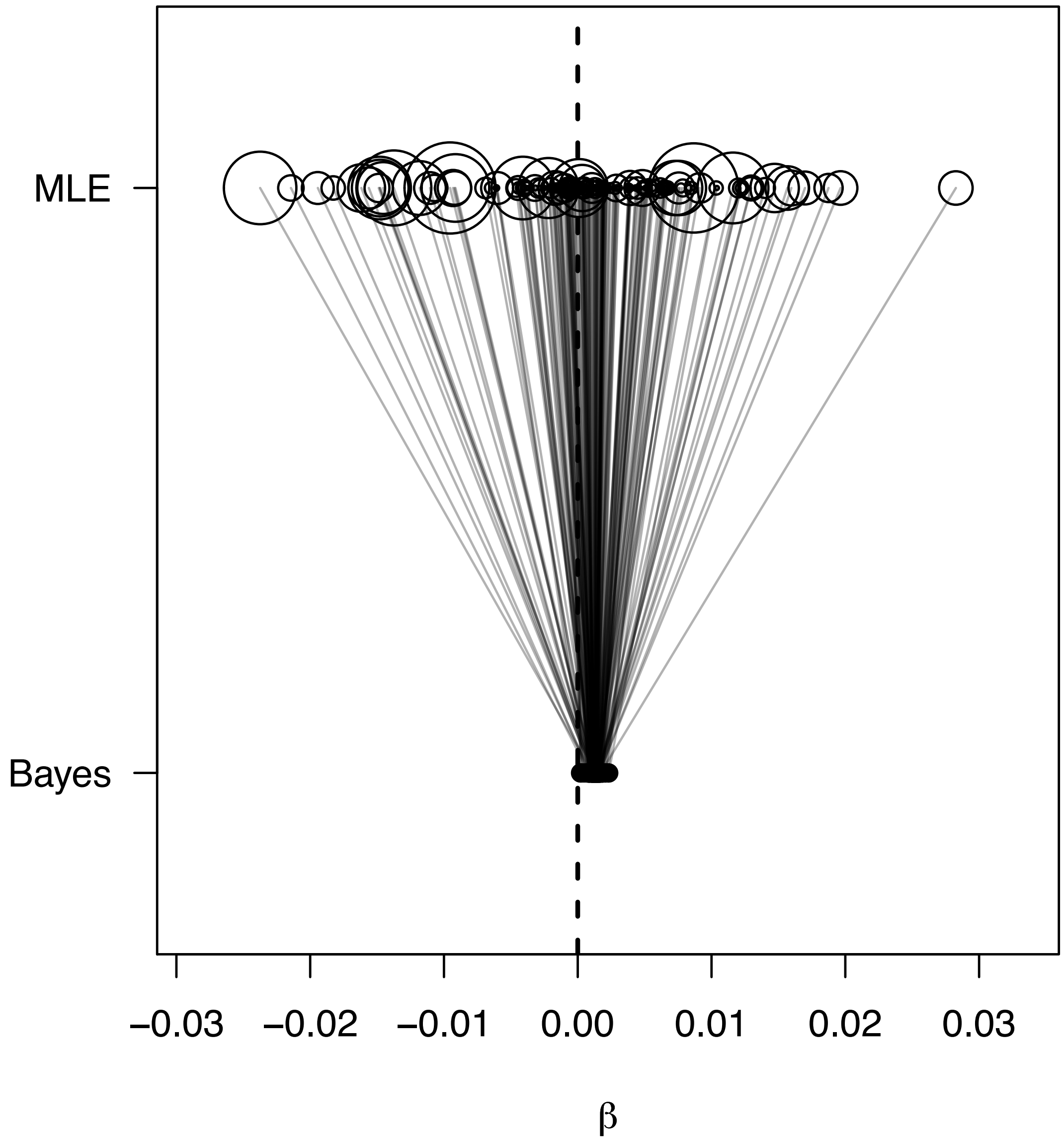
County-specific Maximum Likelihood Estimates ($PM_{2.5}$ and heart failure)



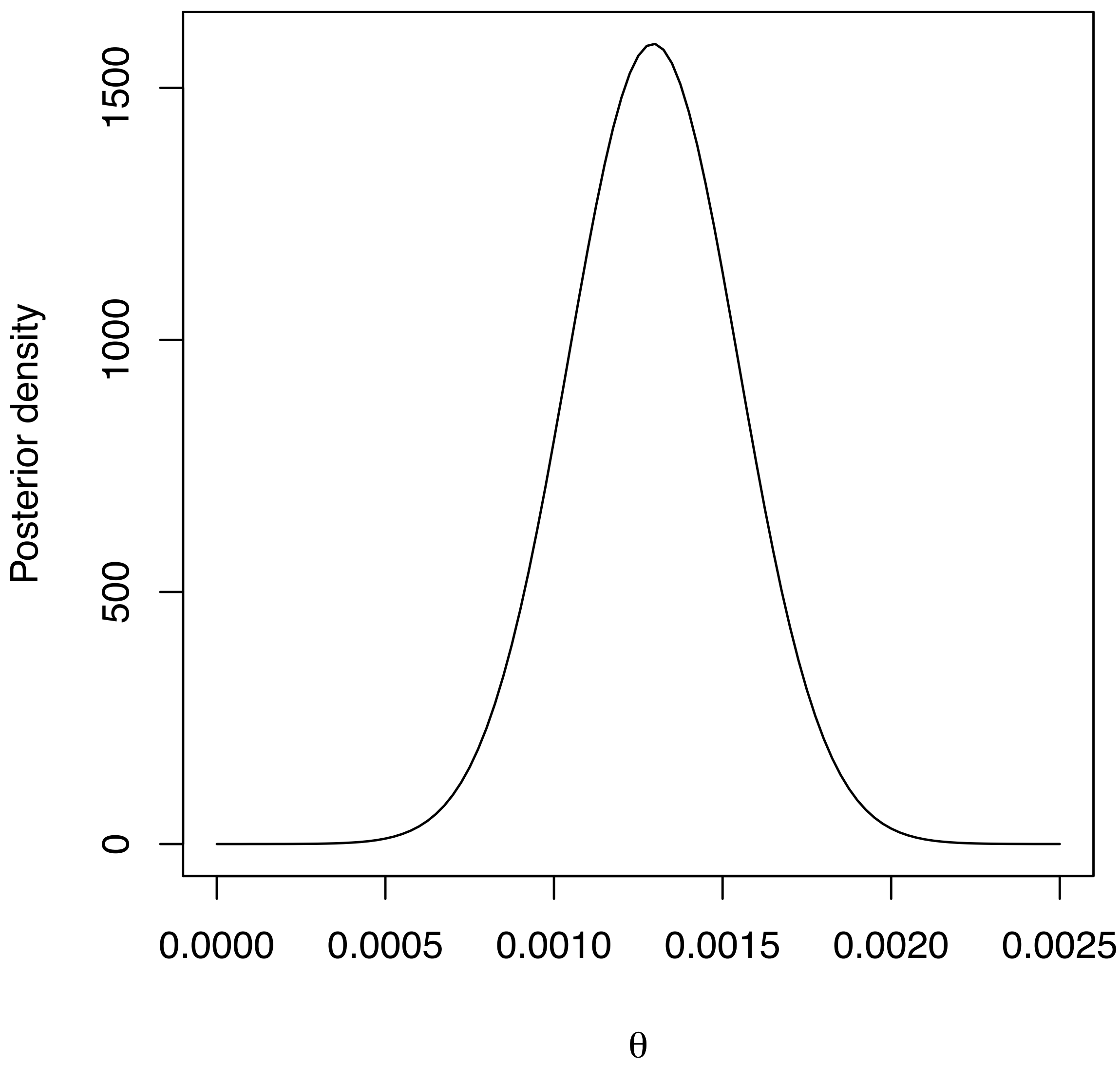
County-specific Bayesian estimates (shrunken)



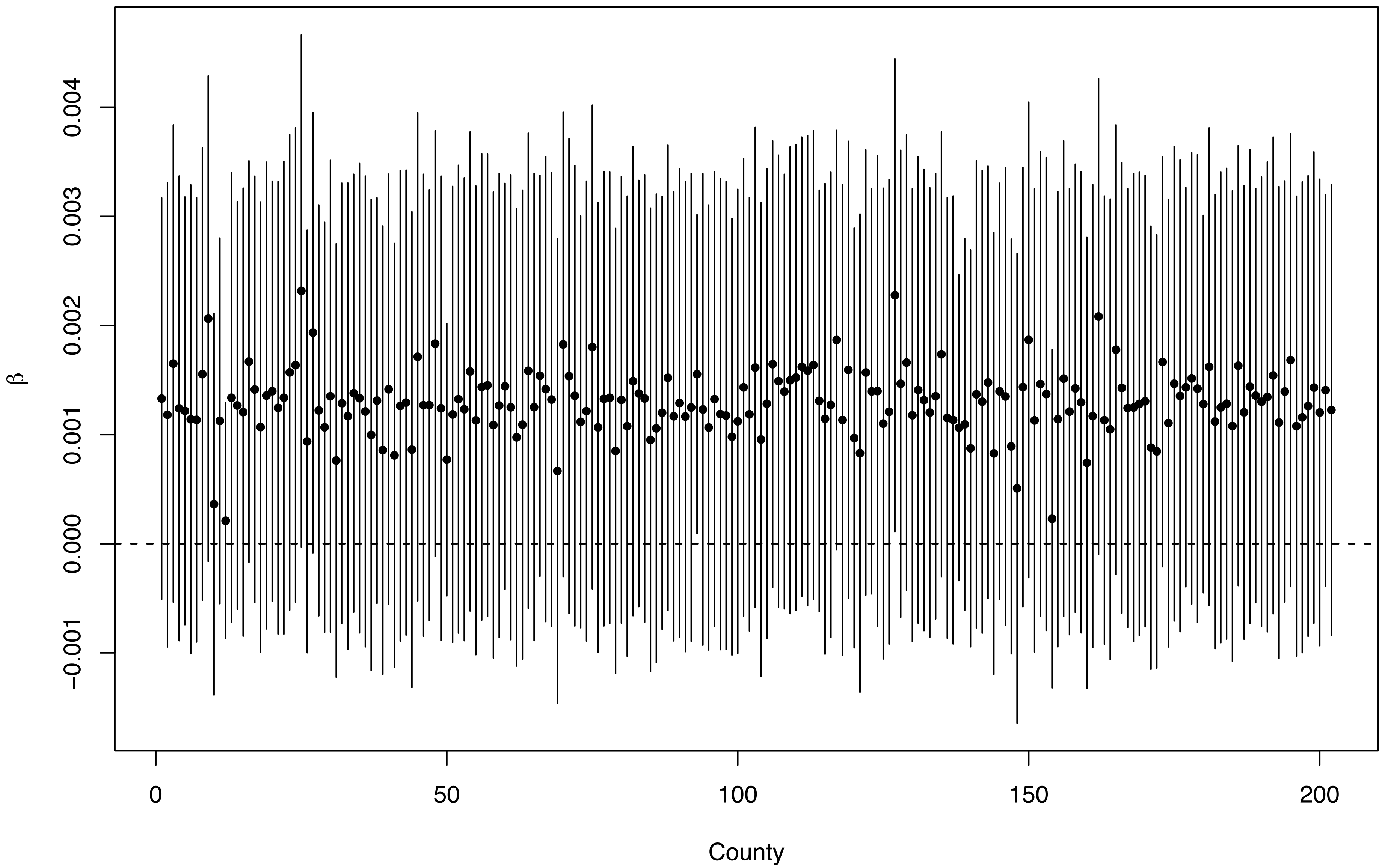
Shrinkage!



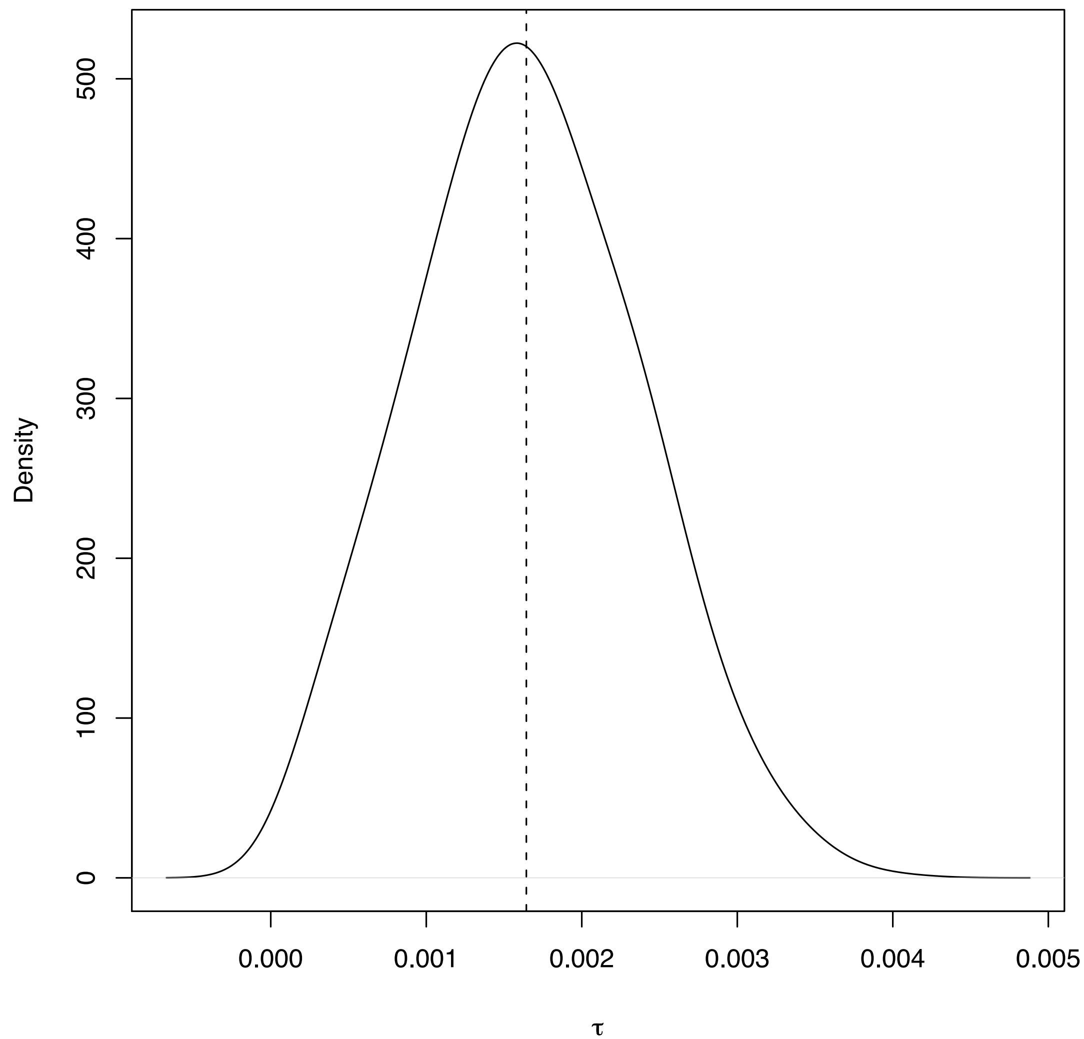
National Average Estimate (Posterior Distribution)



County-specific Bayesian estimates (shrunken)



Heterogeneity Parameter (Posterior Distribution)



Exploring Effect Modification

- To explore effect modification of air pollution risks by location-specific characteristics, we can include a covariate in the second level of the model
- Alternatively, we can fit a weighted linear regression where the dependent variable is the location-specific (log) relative risk estimate and the independent variable is the location-specific characteristic

A Two-stage normal normal model with level-2 covariate

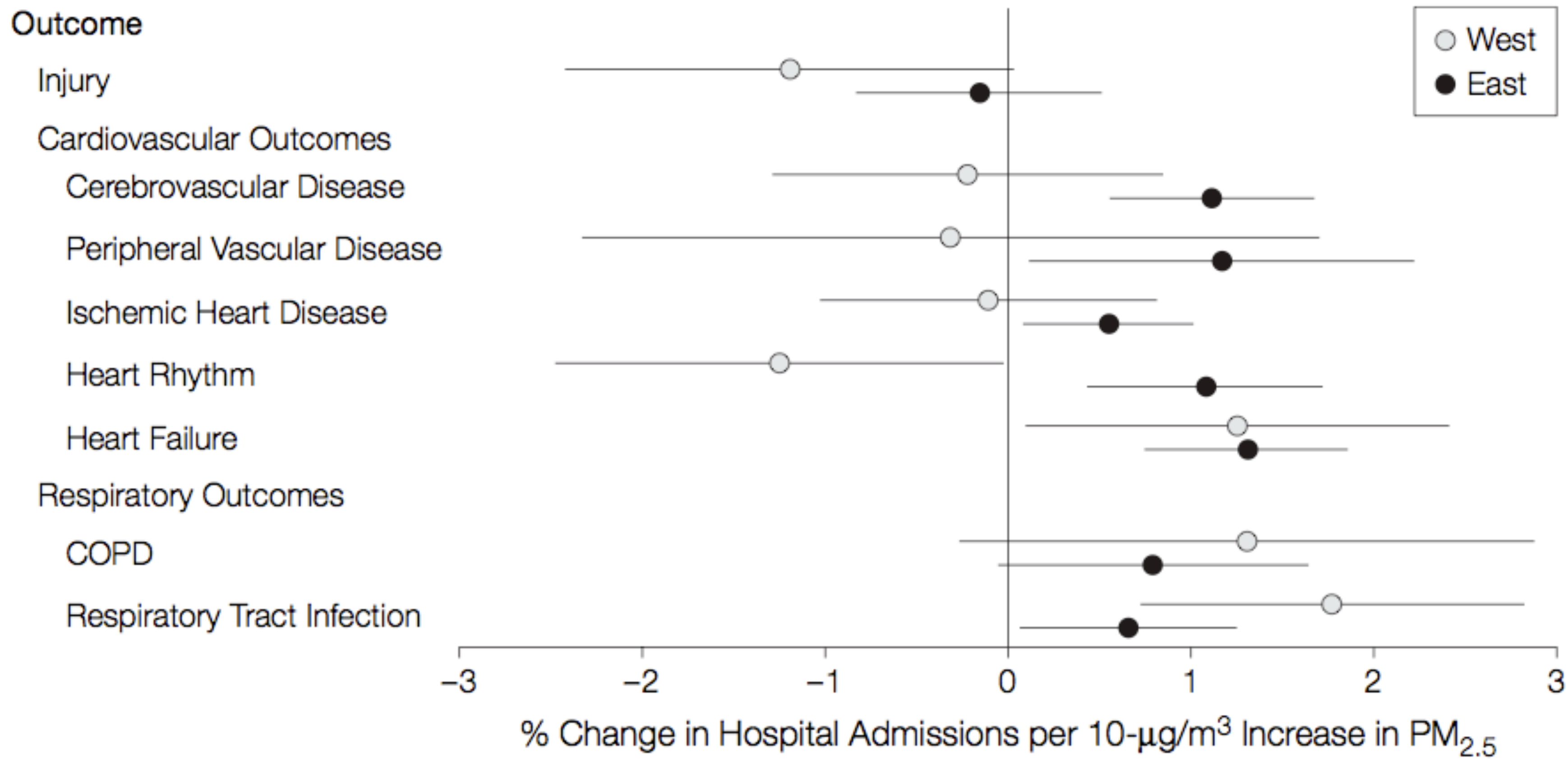
$$y_j = \theta_j + \varepsilon_j; j = 1,..,J$$

$$\varepsilon_j \sim N(0,\sigma_j^2)$$ Statistical variance

$$\theta_j = \alpha_0 + \alpha_1(x_j - \bar{x}) + N(0,\tau^2)$$

Effect modifier

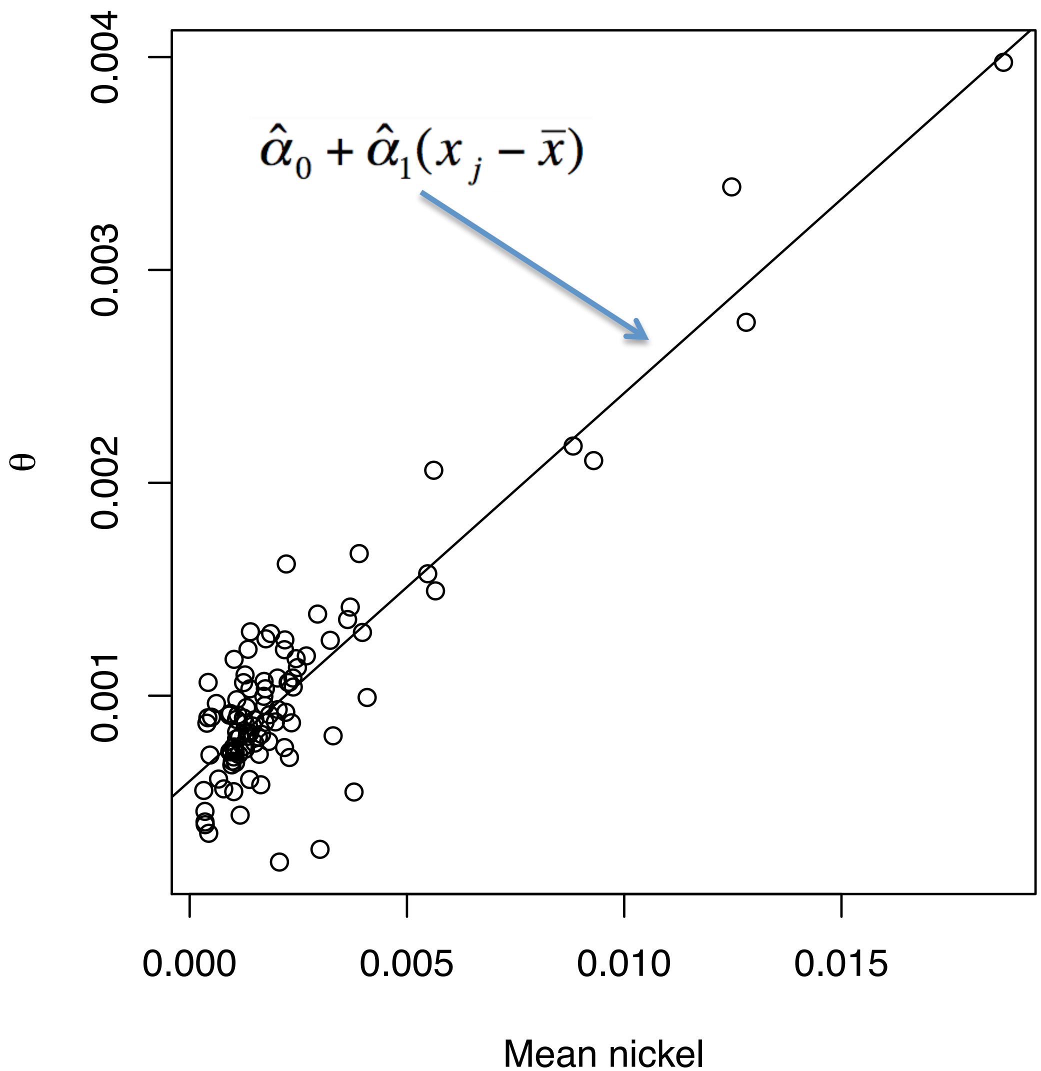
Figure 4. Percentage Change in Hospitalization Rate by Cause per 10-μg/m³ Increase in $PM_{2.5}$ for the US Eastern and Western Regions for all Outcomes



Point estimates and 95% posterior intervals of the percentage change in admission rates per 10 μg/m³. $PM_{2.5}$ indicates particulate matter of less than or equal to 2.5 μm in aerodynamic diameter; COPD, chronic obstructive pulmonary disease.

Dominici, et al. *JAMA*, 2006

Effect Modification by Long-term Nickel Levels



A two-stage normal normal model with spatially correlated random effects

$$y_j = \theta_j + \varepsilon_j$$

$$i = 1,\ldots,n_j, j = 1,..,J$$

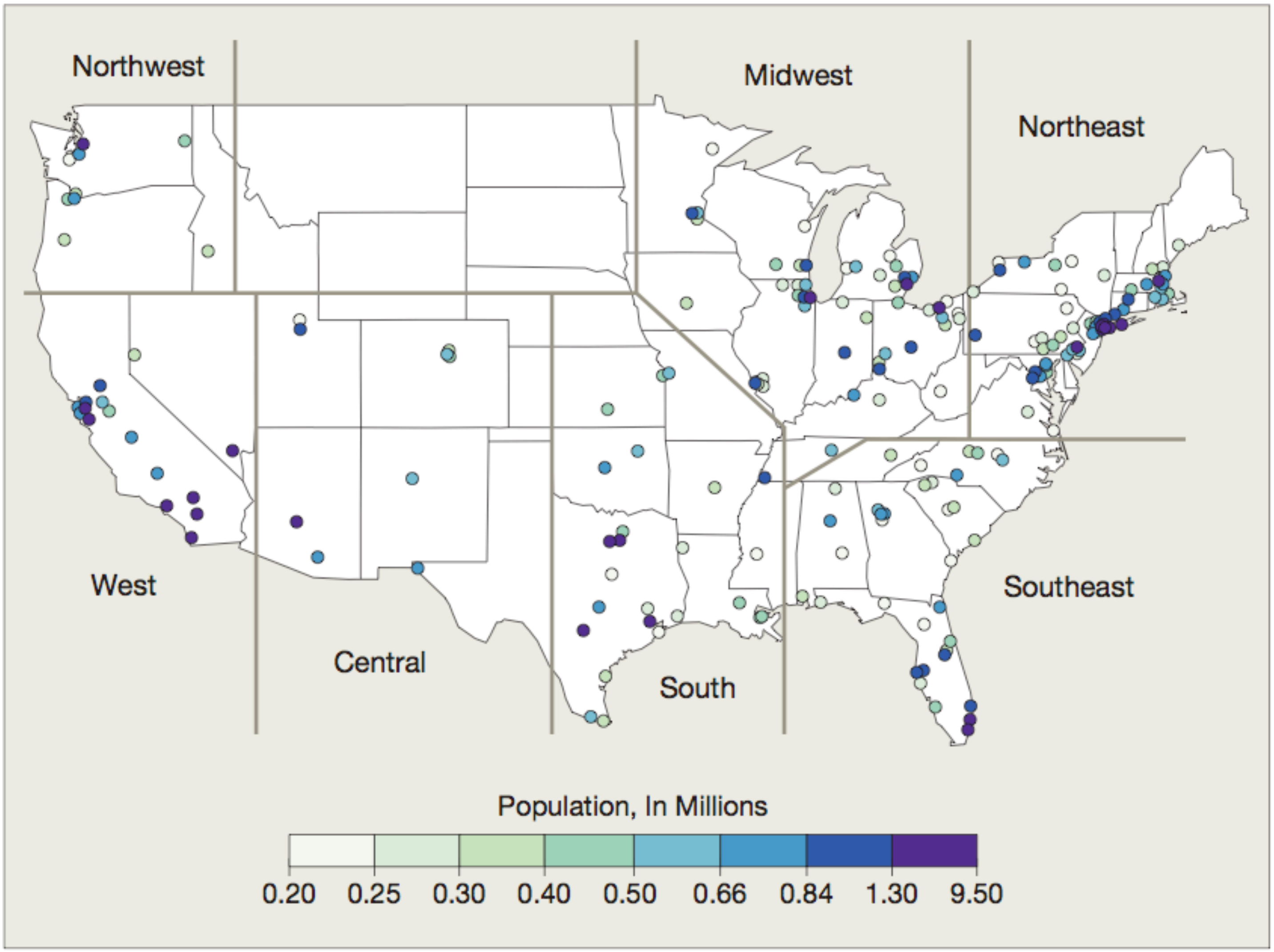
$$\varepsilon_j \sim N(0,\sigma_j^2)$$

$$\theta_j = \theta + N(0,\tau^2)$$

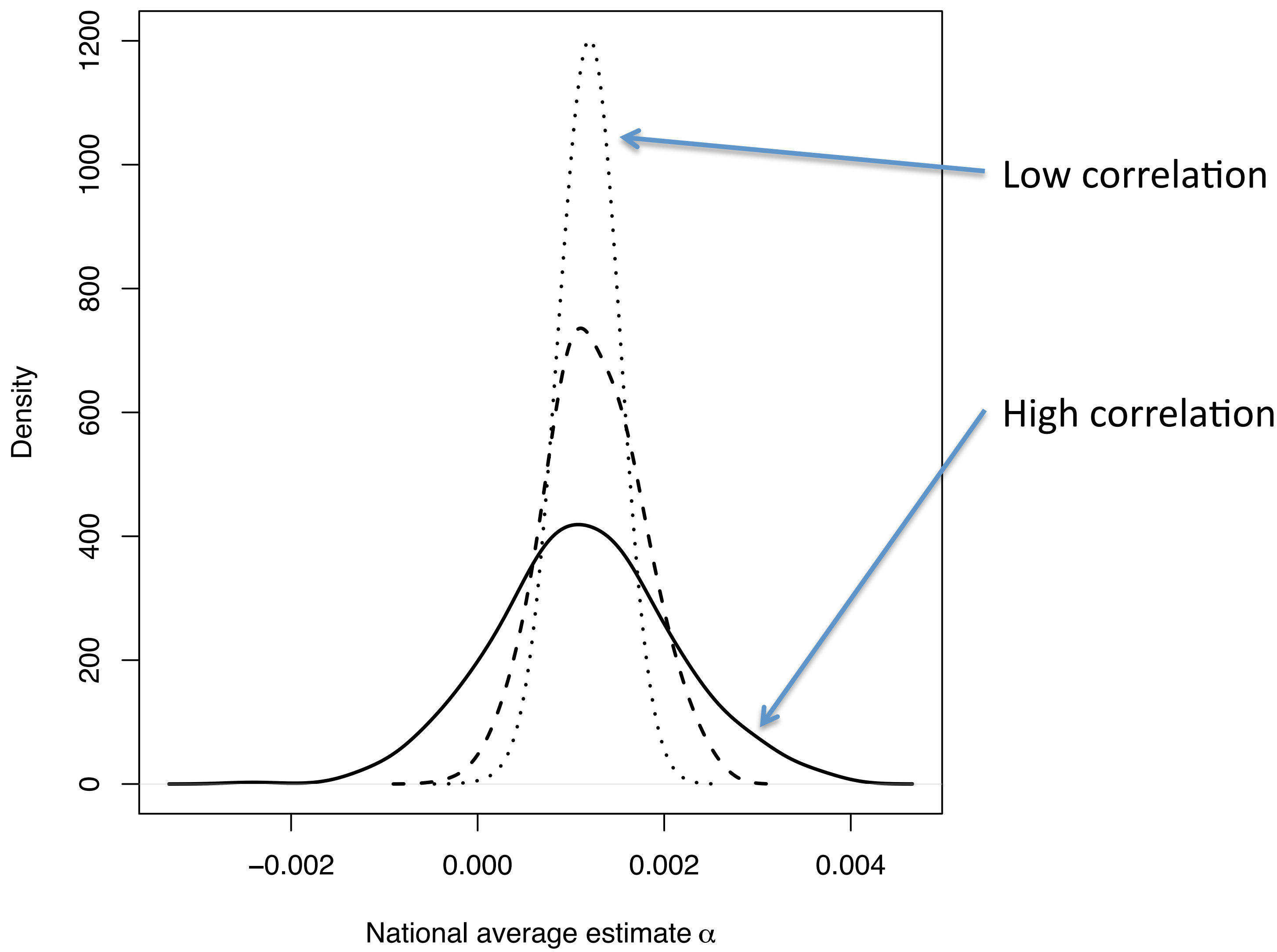
$$cor(\theta_j,\theta_k) = \exp(-\phi \times d(j,k))$$

Cities that are closer to each other will have more similar relative rates

Spatial Distribution of MCAPS Counties



The Effect of Modeling Spatial Correlation of Risks



Scientific Story Thus Far...

- There is strong evidence of an association between day-to-day variation in PM and day-to-day variation in mortality/morbidity
- There appears to be heterogeneity in the risks across locations, particularly for hospital admissions outcome
- For the two groups of outcomes (cardiovascular and respiratory), the estimated relative rates have very distinct regional patterns
- PM chemical component levels may explain some heterogeneity, but more work is needed

Scientific Story Thus Far...

- Individual city-specific analyses give highly variable results due to substantial noise in estimation
- Multi-city studies using hierarchical models provide much more precise risk estimates, both nationally and at a city-specific level
- Hierarchical models allow us to quantify the heterogeneity across locations
- Understanding and explaining the heterogeneity in risk is a *major scientific goal for the future*