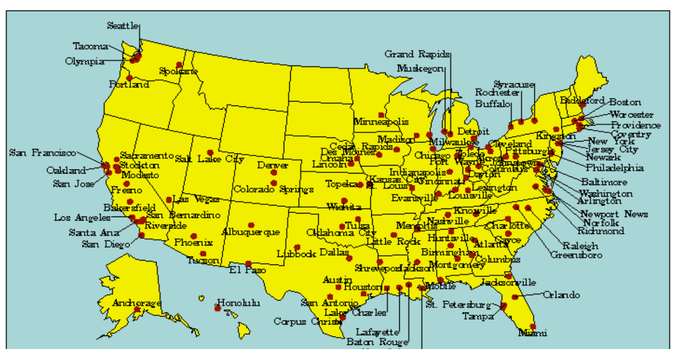
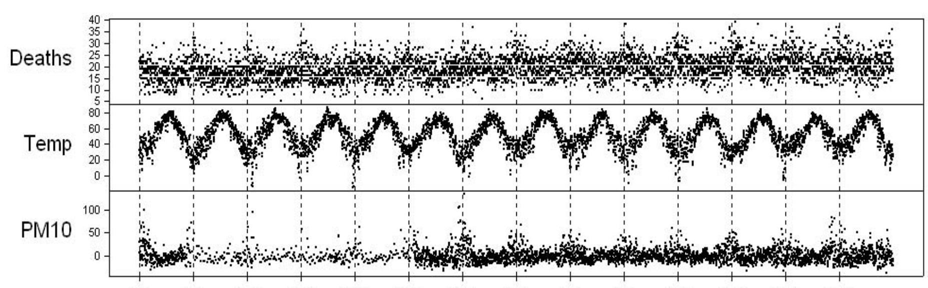
Lecture 6
NMMAPS Case study: Hierarchical Models for Estimating Health Effects of Air Pollution

National
Morbidity
Mortality
Air
Pollution
Study

1987—2000





Now the issue is getting another look as EPA faces a December 2005 deadline for revisiting its $PM_{2.5}$ standard. EPA scientists, after reviewing piles of new data implicating $PM_{2.5}$ in health effects, have proposed tightening the 1997 standard to further reduce ambient concentrations of fine particles. Some scientists and industry groups remain skeptical, noting that researchers still haven't pinned down what makes particles dangerous—whether it's mainly size, and that the tiniest particles are most potent; or chemistry, such as metal content; or some combination of the two. Despite 8 years and some $400 million in research, finding out exactly how fine particles do their dirty work has proved frustratingly elusive, researchers say. "We've gotten glimpses, but we don't yet have enough systematic coverage of the problem," says epidemiologist Jon Samet of Johns Hopkins University in Baltimore, Maryland.

Urgent need to gather scientific evidence on the health effects of PM2.5 on a national scale

National Morbidity, Mortality, and Air Pollution Study (NMMAPS), 1987—2000

- 108 urban communities
- Cause-specific mortality data from NCHS
  - all-cause (non-accidental), CVD, respiratory, COPD, pneumonia, accidental
- Weather from NWS
  - 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

National Medicare Cohort (1999—2002)

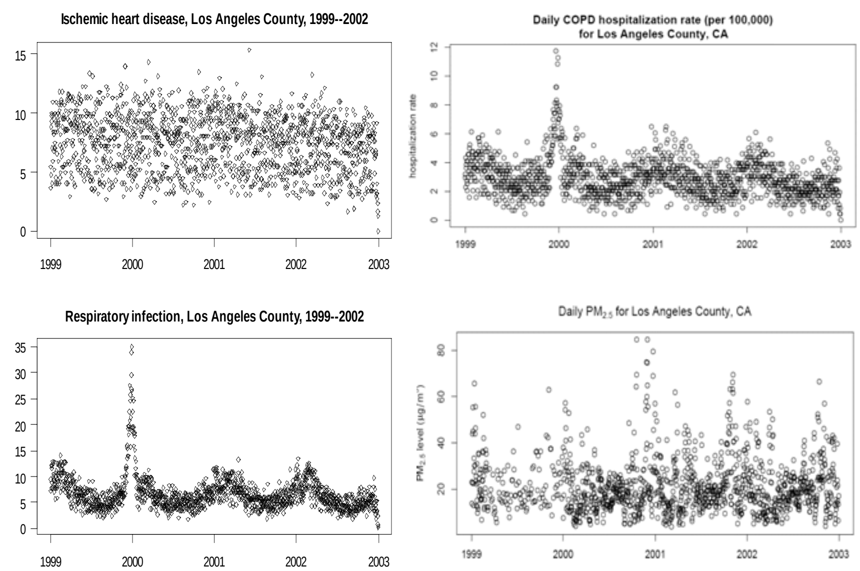
- National study of fine particles ($PM_{2.5}$) and hospital admissions in Medicare
- Data include:
  - Billing claims (NCHF) for everyone over 65 enrolled in Medicare (~48 million people),
    - date of service
    - treatment, disease (ICD 9), costs
    - age, gender, and race
    - place of residence (ZIP code/county)
  - Approximately 204 counties linked to the air pollution monitoring
  - Study population includes 11.5 million Medicare enrollees living on average 5.9 miles from a monitor.

Health Outcomes

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

- cerebrovascular disease
- peripheral disease
- ischemic heart disease
- heart rhythm
- heart failure
- chronic obstructive pulmonary disease
- respiratory infection
- injuries (as a sham outcome)

Daily time series of hospitalization rates and PM2.5 levels in Los Angeles county (1999-2002)



Exposure and Effect Modifiers

- **Daily PM2.5 ambient levels for the period 1999-2002 for each of the 204 locations**
- **To explore effect modification of risk by PM2.5 sulfate composition, we gathered PM2.5 speciation data from 2000 to 2002**
- **Sulfate concentrations in the PM2.5 mass were available for at least one entire year for 100 of the 204 cities included in the study. Most cities have measurements every six days, with some variation**

Multi-site time series studies

- **Compare day-to-day variations in hospital admission rates with day-to-day variations in pollution levels within the same community**
- **Avoid problem of unmeasured differences among populations**
- **Key confounders**
  - ➢ **Seasonal effects of infectious diseases**
  - ➢ **Weather**

Statistical Methods

- **Within city. Semi-parametric regressions for estimating associations between day-to-day variations in air pollution and mortality controlling for confounding factors**
- **Across cities. Hierarchical Models for estimating:**
  - **– national-average relative rate**
  - **– regional-average relative rate**
  - **– exploring heterogeneity of air pollution effects across the country**

Pooling log-relative rates across counties

- **To produce a national average relative rate we used Bayesian hierarchical models**
- **We combine relative rates 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 above but separately within each of the seven regions.**

Pooling

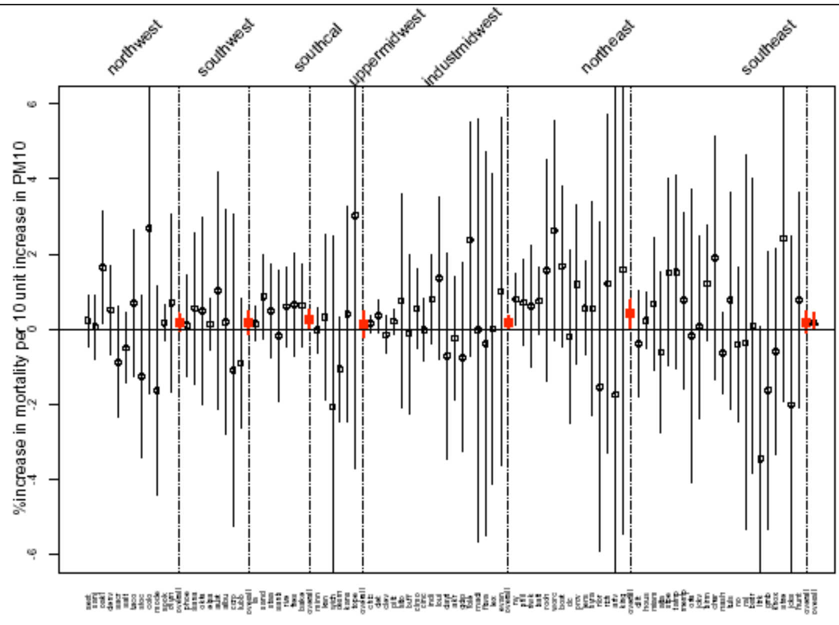
City-specific relative rates are pooled across cities to:

1. **estimate a *national-average* air pollution effect on mortality;**
2. **explore geographical patterns of variation of air pollution effects across the country**

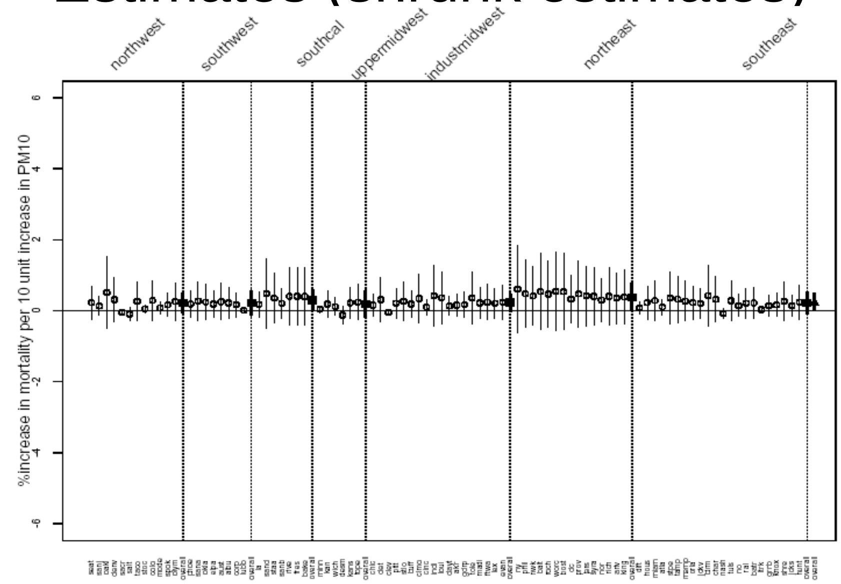
Pooling

- Implement the old idea of ***borrowing strength across studies***
- **Estimate heterogeneity and its uncertainty**
- **Estimate a national-average effect which takes into account heterogeneity**

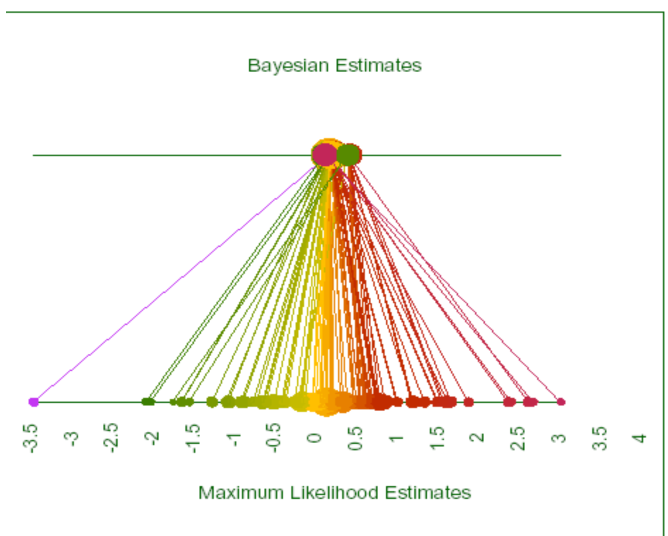
City-specific and regional estimates



City-specific Bayesian Estimates (shrunk estimates)



Shrinkage



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

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)

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

Exploring Effect Modification

- **To explore effect modification of air pollution risks by location-specific characteristics, we fitted a weighted linear regression where the dependent variable is the location-specific relative rate estimate and the independent variable is the location-specific characteristic**
- **We consider average sulfate concentration in the PM mass, ozone, and temperature as predictors across the period 2000-2002.**

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

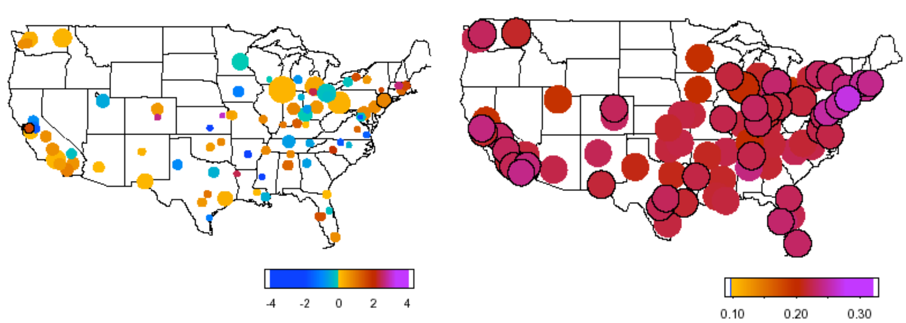
Maximum likelihood and Bayesian estimates of air pollution effects

Use only city-specific information

Borrow strength across cities

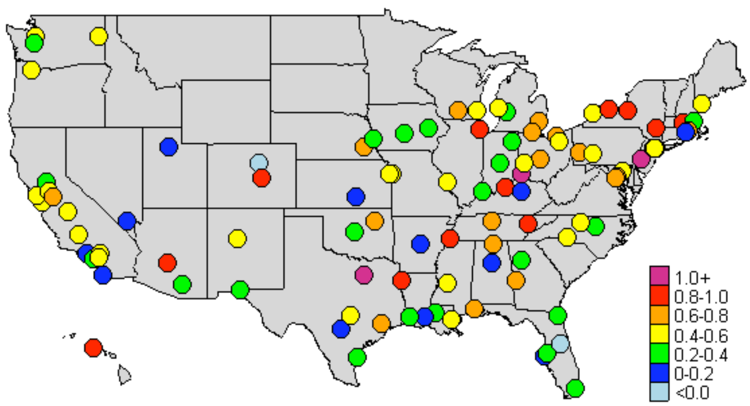
MLE

BAYESIAN ESTIMATES

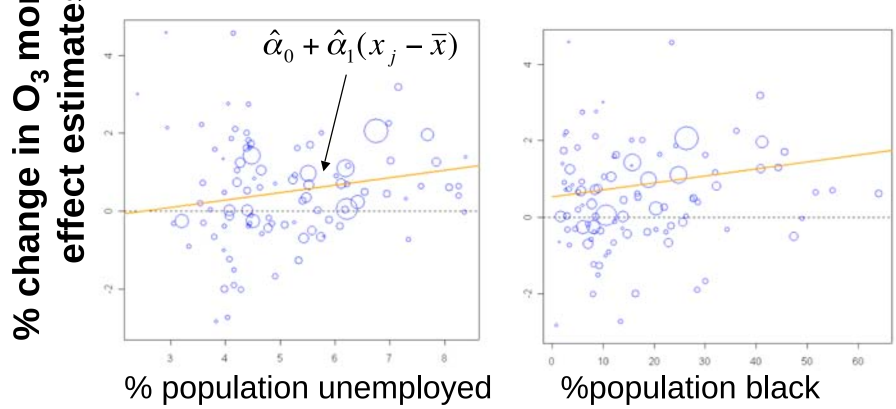


Dominici, McDermott, Zeger, Samet EHP 2003

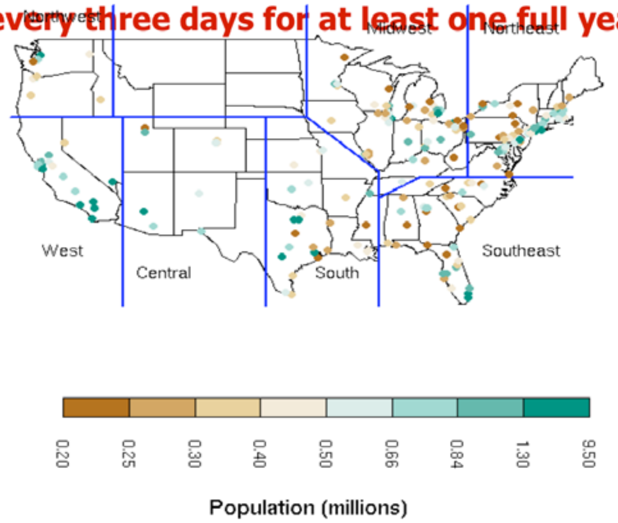
Percent Increase in mortality per 10 ppb increase in the previous week's daily O_3 (Bayesian community-specific estimates, Bell JAMA 2004)



Communities' effect estimates vs. unemployment and race: The size of the circle corresponds to the inverse of the standard error of the community's ıaximum likelihood estimate.. The orange line reflects sults from the second stage analysis (Bell, AJE 2007).



Study population: 204 counties with populations larger than 200,000 and with data collected once every three days for at least one full year



A three-stage normal normal model

city　　region

$$y_j^r = \theta_j^r + \varepsilon_j^r, j = 1,..,J^r, r = 1,..,R$$

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

$$\theta_j^r = \theta^r + \xi_j^r$$

$$\xi_j^r \sim N(0,\tau_1^2)$$ Variance within regions

$$\theta^r = \theta + \delta^r$$

$$\delta^r \sim N(0,\tau_2^2)$$ Variance across regions

Three stage model

y_j^r Estimated relative rate for city j in region r

θ_j^r True relative rate for city j in region r

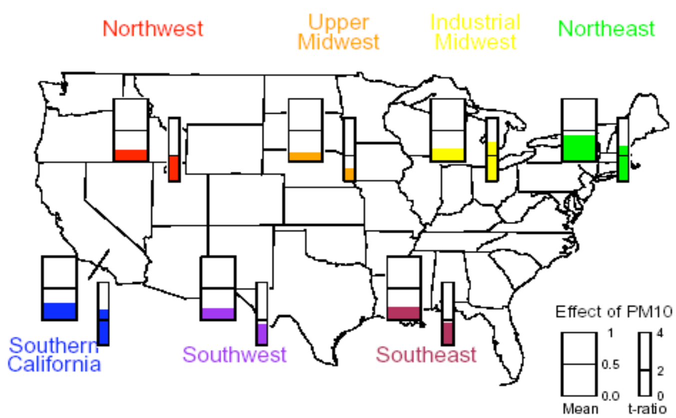
θ^r True region-average relative rate

θ True national-average relative rate

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

Within city — Across cities Within region — Across regions

Regional map of air pollution effects

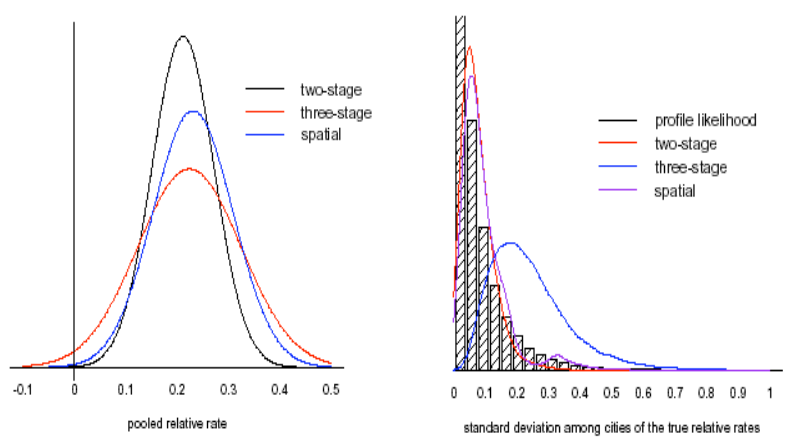


Partition of the United States used in the 1996 Review of the NAAQS

Three Models

- “**Three stage**”- as in previous slide
- “**Two stage**”- ignore region effects; assume cities have exchangeable random effects
- **Two stage with “spatial” correlation -**city random effects have isotropic exponentially decaying autocorrelation function

Comparison between heterogeneity models



Results Stratified by Cause of Death

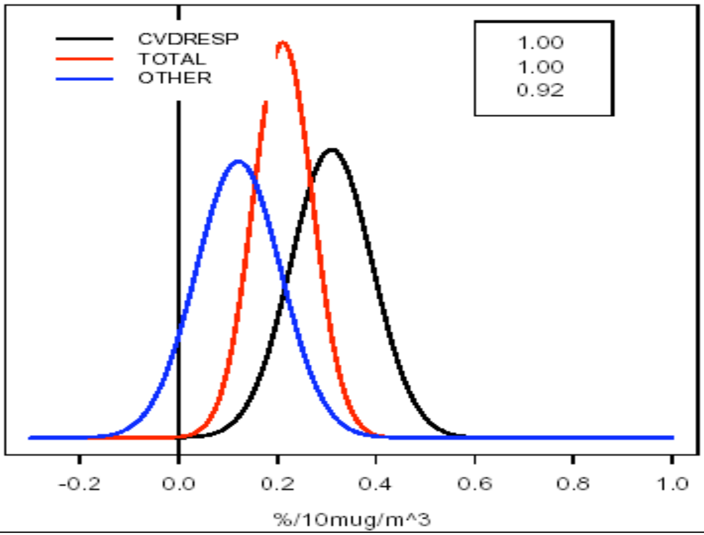
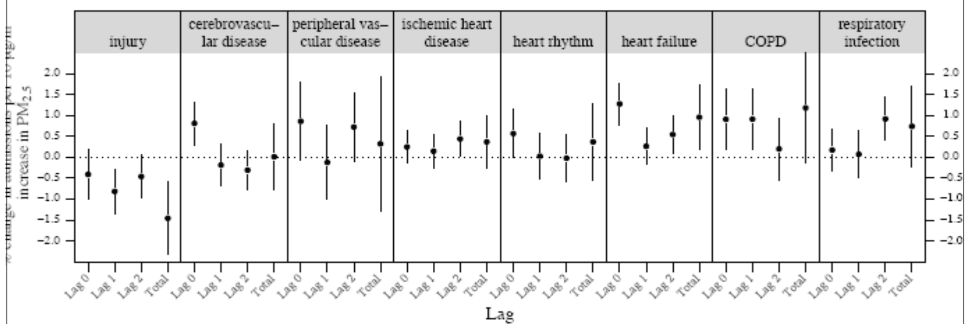


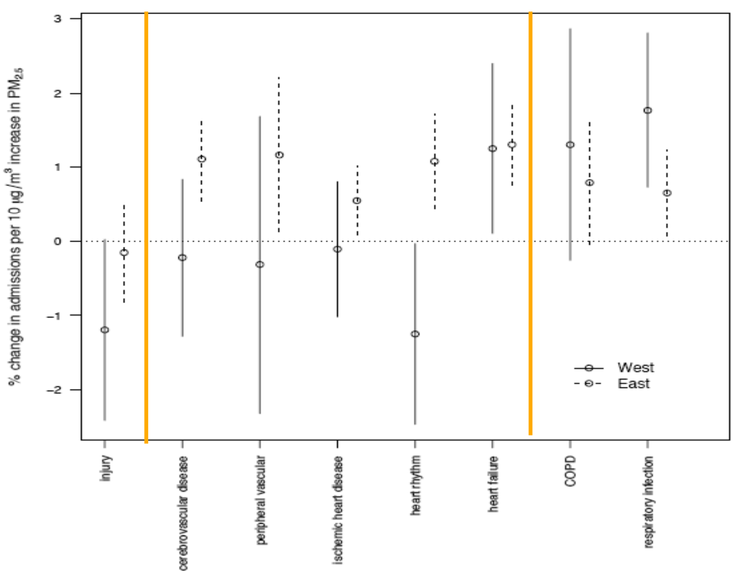
Table 1: Mean and interquartile range among counties of hospitalization rates (number of cases per 100,000 people) for each outcome for the period 1999-2002

Variable	ICD-9	Mean (IQR)
Ischemic heart disease	410-414 and 429	8.3 (7.1,9.4)
Heart rhythm	426-427	3.8 (3.3,4.2)
Heart failure	428	5.7 (4.7,6.6)
Cerebrovascular disease	430-438	5.5 (4.8,6)
Peripheral vascular	440-448	1.7 (1.5,1.9)
Respiratory infection	464-466 and 488-487	5.5 (4.7,6.2)
COPD	490-492	2.6 (2.1,3.2)
Accident	800-849	4.2 (3.7,4.5)
$PM_{2.5}$ Levels ($\mu g/m^3$)		13.4(11.3,15.2)
Days with $PM_{2.5}$		817(434,1295)

Point estimates and 95% posterior intervals (PI) of the percent change in admission rates per 10 units increase in PM2.5 concentration on average across the 204 counties (national average relative rates) for single lag (lags 0,1, and 2 days) and distributed lag models for to 2 days (Total) for all outcomes.



Point Estimates and 95% PI of the percent change in admission rates per 10 units increase in PM2.5 concentration for the East and West regions for all outcomes.



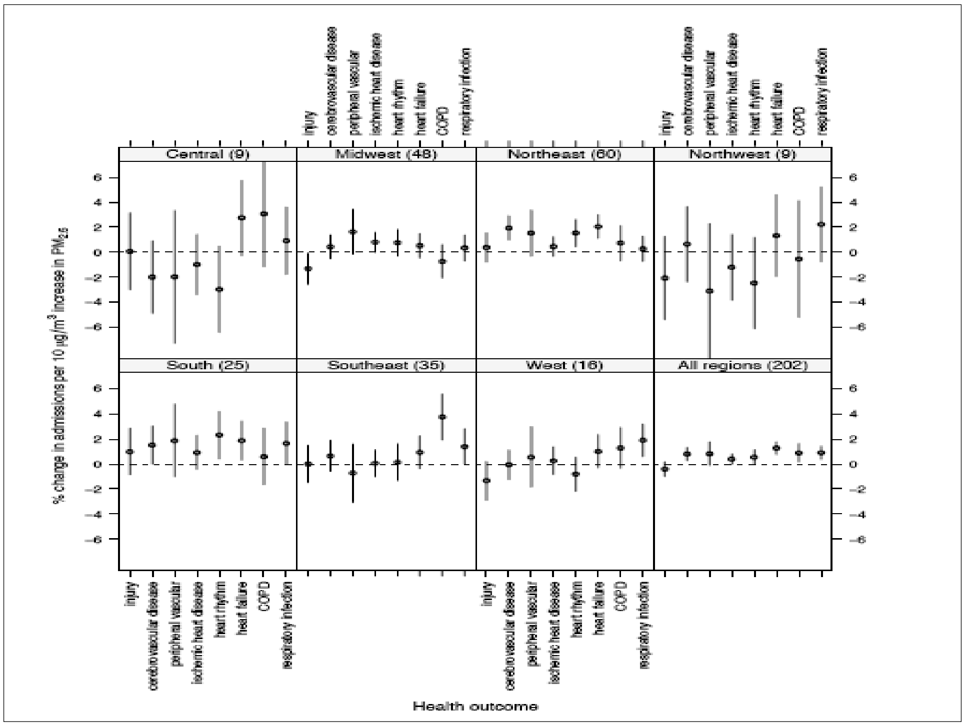
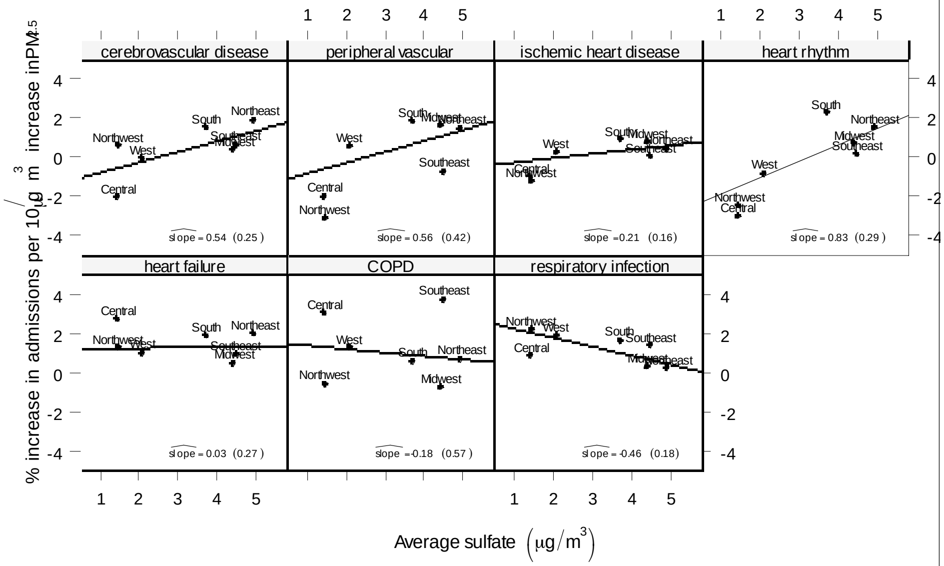


Table 3: average sulfate concentration in the $PM_{2.5}$ mass and inter-quartile range across cities for the period 2000-2002 for each geographical region. The last column denotes the number of counties (100 total) within each region where sulfate concentrations measurements are available.

Region	*Fine Sulfate* ($\mu g / m^3$) *(IQR)*	*# counties*
East	4.49 (3.74 – 5.17)	82
South	3.70 (3.50 – 4.99)	12
Southeast	4.48 (3.96 – 4.99)	18
Northeast	4.92 (4.14 – 5.76)	27
Midwest	4.40 (3.70 – 5.18)	25
West	1.79 (1.34 – 1.72)	18
Central	1.42 (1.34 – 1.58)	6
West	2.08 (1.32 – 2.86)	10
Northwest	1.45 (1.42 – 1.49)	2

Can we explain regional heterogeneity of risks by average Sulfate concentration?



NEWS

Evidence Mounts That Tiny Particles Can Kill

A)lthough many questions remain about how fine particles kill people, the NMMAPS study shows there's no mistaking that PM is the culprit…?

Results: National Averages

- We found evidence of a positive association between day-to-day variation in concentration and hospital admissions for all outcomes, except injuries, for at least one exposure lag
  - The largest effect was found at lag 0 for most of the cardiovascular outcomes
  - For respiratory outcomes, we found that the largest effects occurred at lags 0 and 1 for COPD and at lag 2 for respiratory infections
- We did not find any positive association for injuries or for other external causes or when using lag -1 as the exposure indicator
- The main results were robust to the number of degrees of freedom used to adjust for temporal confounding and to the adjustment for weather

Results: regional heterogeneity

- For the two groups of outcomes (cardiovascular and respiratory), the estimated relative rates have very distinct regional patterns
- For cardiovascular diseases, all estimates in the East US were positive while estimates in the West US were close to zero
- For respiratory diseases, we found positive effects in all US with slightly larger effects in the West US

Results: effect modification

- *Sulfate*: ***We found evidence of effect modification of the relative rates by average sulfate concentration with positive slopes for the cardiovascular outcomes (except heart failure) and negative slopes for the two respiratory outcomes.***

Findings

- NMMAPS has provided at least four important findings about air pollution and mortality

1. There is evidence of an association between acute exposure to particulate air pollution and mortality
2. This association is strongest for cardiovascular and respiratory mortality
3. The association is strongest in the Northeast region of the USA
4. The exposure-response relationship is linear

A Big Challenge

- Doing research in a controversial political context can lead to a process which can be highly non scientific
- Expect to face consultants who use “quasi-scientific” arguments that create confusion about findings

Criticisms

- **Heterogeneity**: in presence of heterogeneity of air pollution effects across the country, the national-average estimate is un-meaningful
- **Adjustment for confounders**: the associations are spurious and are the results of inadequate adjustments for confounders
- **Other Pollutants**: associations are not due to PM but to other pollutants and extreme weather

Heterogeneity

Is it appropriate to pool?

What are the data saying about heterogeneity?

- Chi-squared tests of homogeneity are always accepted (need to have 30% smaller standard errors to reject the null)
- Profile likelihood has a peak at zero
- Bayesian approach: marginal posterior distribution of the between-city standard deviation indicates that heterogeneity is very small

Why do a joint analysis of all the cities?

- Individual cities can be selected to show one point or another
- Results from individual cities are swamped by statistical error
- There is no reason to expect that two neighboring cities with similar sources of particles would have qualitative different relative risks

What are the public policy implications?

- A national estimate of the air pollution effect provides evidence on the amount of hazard from exposure to air pollution
- EPA needs a single number for the entire country

Other Pollutants

Have PM studies adequately separated the effects of PM, weather, and co-pollutant?

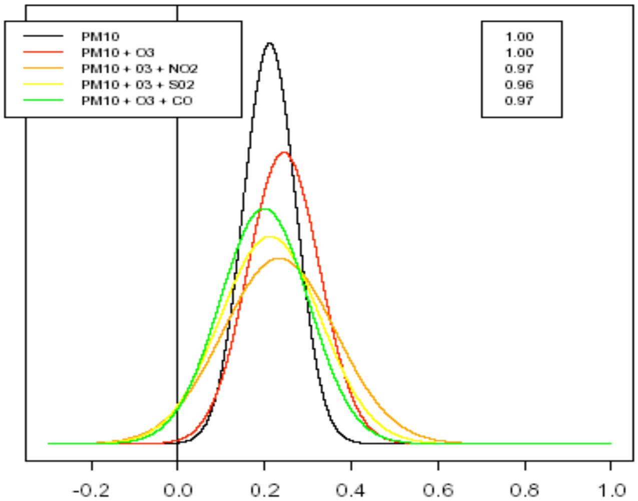
Other pollutants

- This is a complicate matter since many of the same mechanisms are postulated to underlie the effects of different pollutants

A simpler question is:

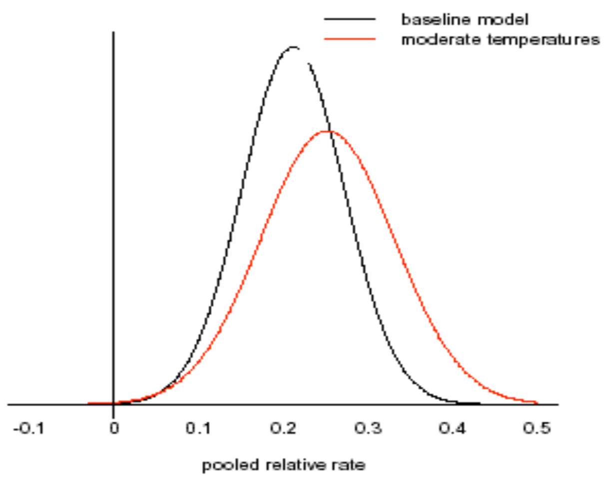
- Does the effect of PM on mortality sensitive to the adjustment for weather, seasonality and other pollutants?

Sensitivity of the pooled effect to the inclusion of other pollutants in the model



Posterior distributions of the pooled PM effects under 5 multi pollutant models

Sensitivity of the pooled effect to adjustment for weather



Findings

- Pooled estimates of the PM effects on mortality are robust to:
- Adjustment for confounding factors
- Inclusion of other pollutant in the models
- Exclusion of days with more extreme temperatures

Discussion

- To disentangle the effects of particulate matter from the effects of the other pollutants is difficult
- Very limited data is currently available on PM composition to better characterize the risk
- Multi site analyses provide a robust approach for exploring confounding and effect modifications to other pollutant and weather