A Multi-Site Time Series Study of Hospital Admissions and Fine Particles:
A Case-Study for National Public Health Surveillance

Francesca Dominici

(fdominic@jhsph.edu)

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

EPA Workshop October 17 2007

Sponsored by the EPA, CDC Center of Excellence, and NIEHS

A NATIONAL SYSTEM FOR TRACKING POPULATION HEALTH

- **Multiple government databases contain massive amounts of information on the environmental, social, and economic factors that determine health**
- **Research on population health could be rapidly advanced by:**
  - **integrating these existing databases**
  - **bringing to bear new statistical models that would describe major threats and their causes**
- **These integrated databases and new analysis tools would create a national system for population health research**

Air pollution and health: Fundamental questions

- Is there a risk at current levels?
- How can we estimate it?
- How big is the risk?
- What causes it?

Health Effects Fine Particles: Objectives

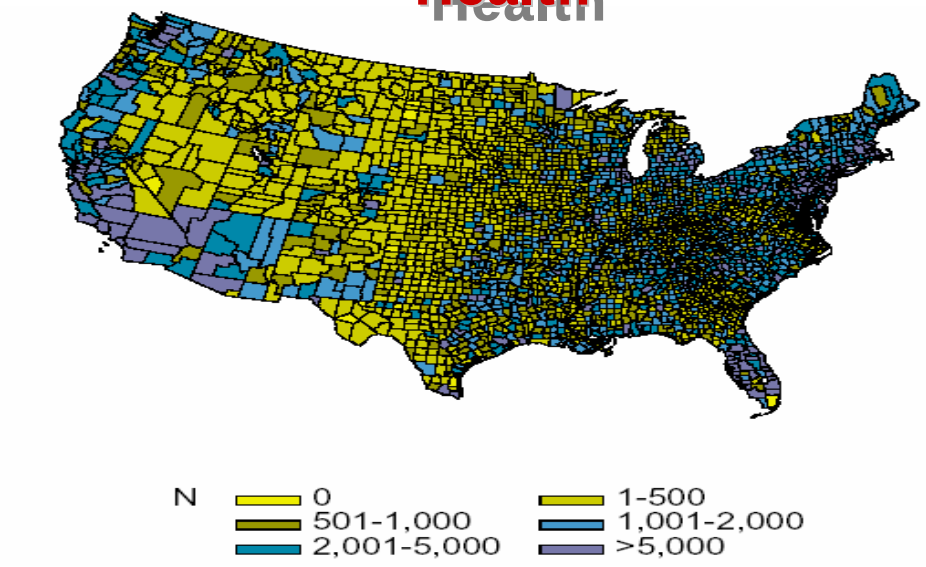
1. assemble a national database of time series data for the period 1999-2005 on hospital admissions rates for cardiovascular and respiratory diseases, fine particulates, and weather for 204 US counties
2. develop state-of-the-art statistical methods
3. estimate maps of relative risks of hospital admissions associated with short-term changes in fine particles
4. illustrate how integration and analysis of national databases can lead to a national health monitoring system

Integrating National Data Sources

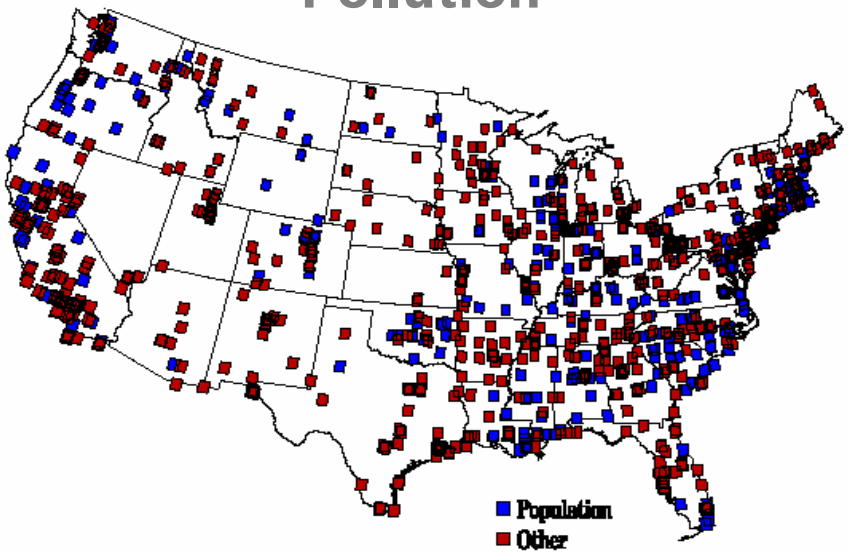
- **NCHF:** 48 million identification numbers
- **MCBS:** subset of 15,000 Medicare participants with additional information on risk factors
- **AIRS:** air pollution monitoring network
- **NOAA:** weather monitoring network
- **US Census:** location characteristics

Integrating national data bases

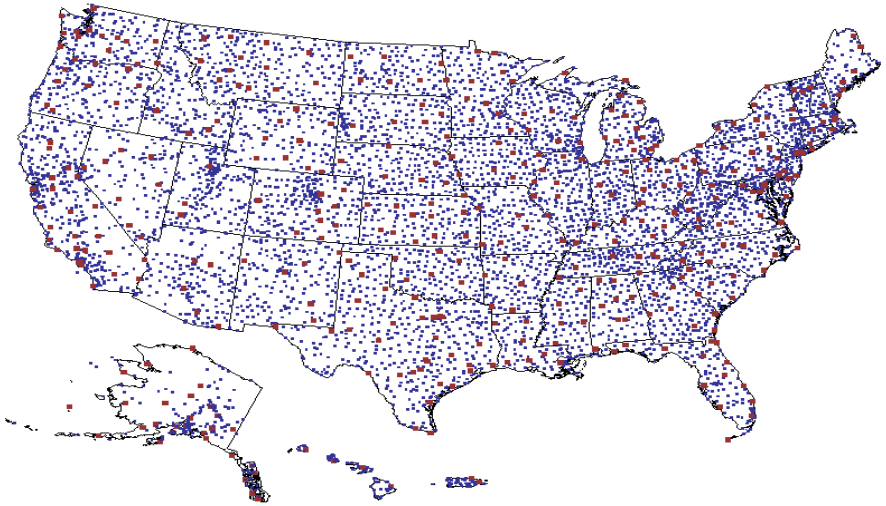
Health



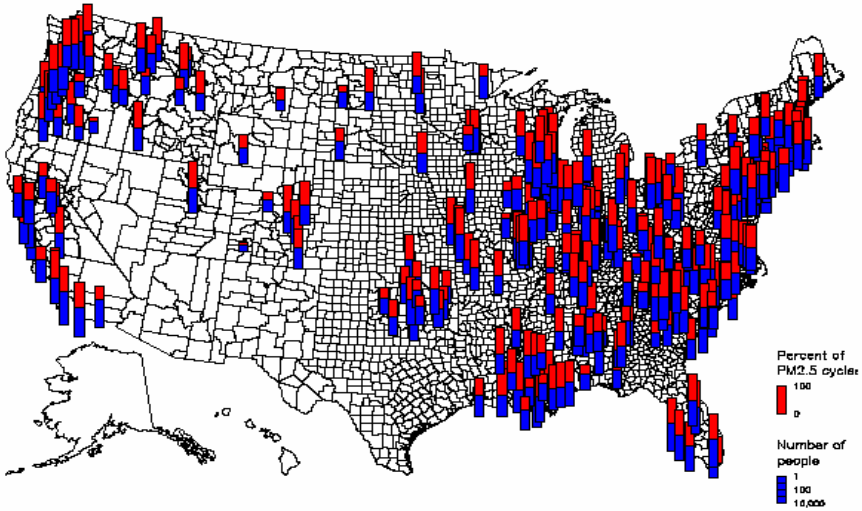
Pollution

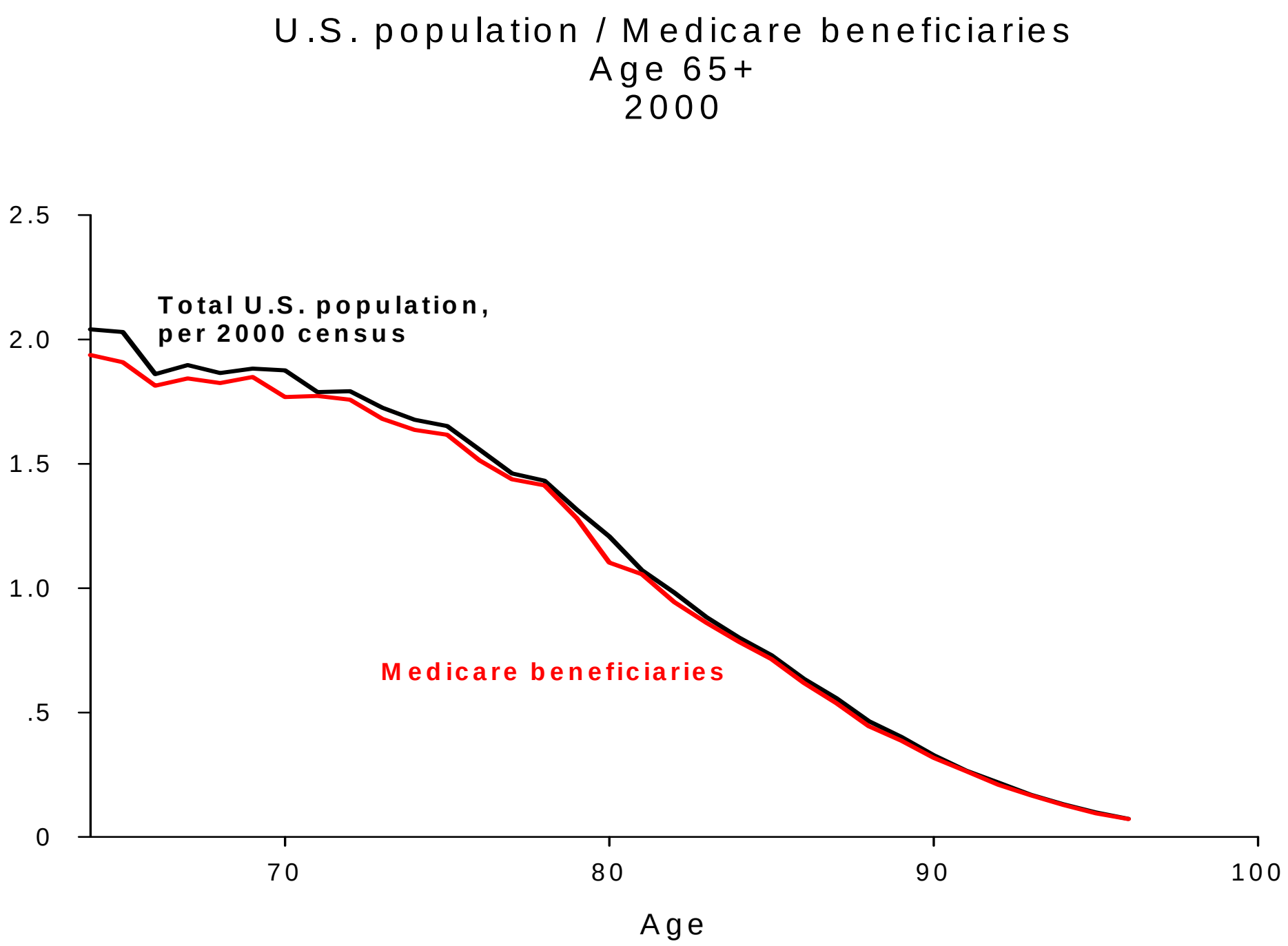


Weather



204 counties with matched data

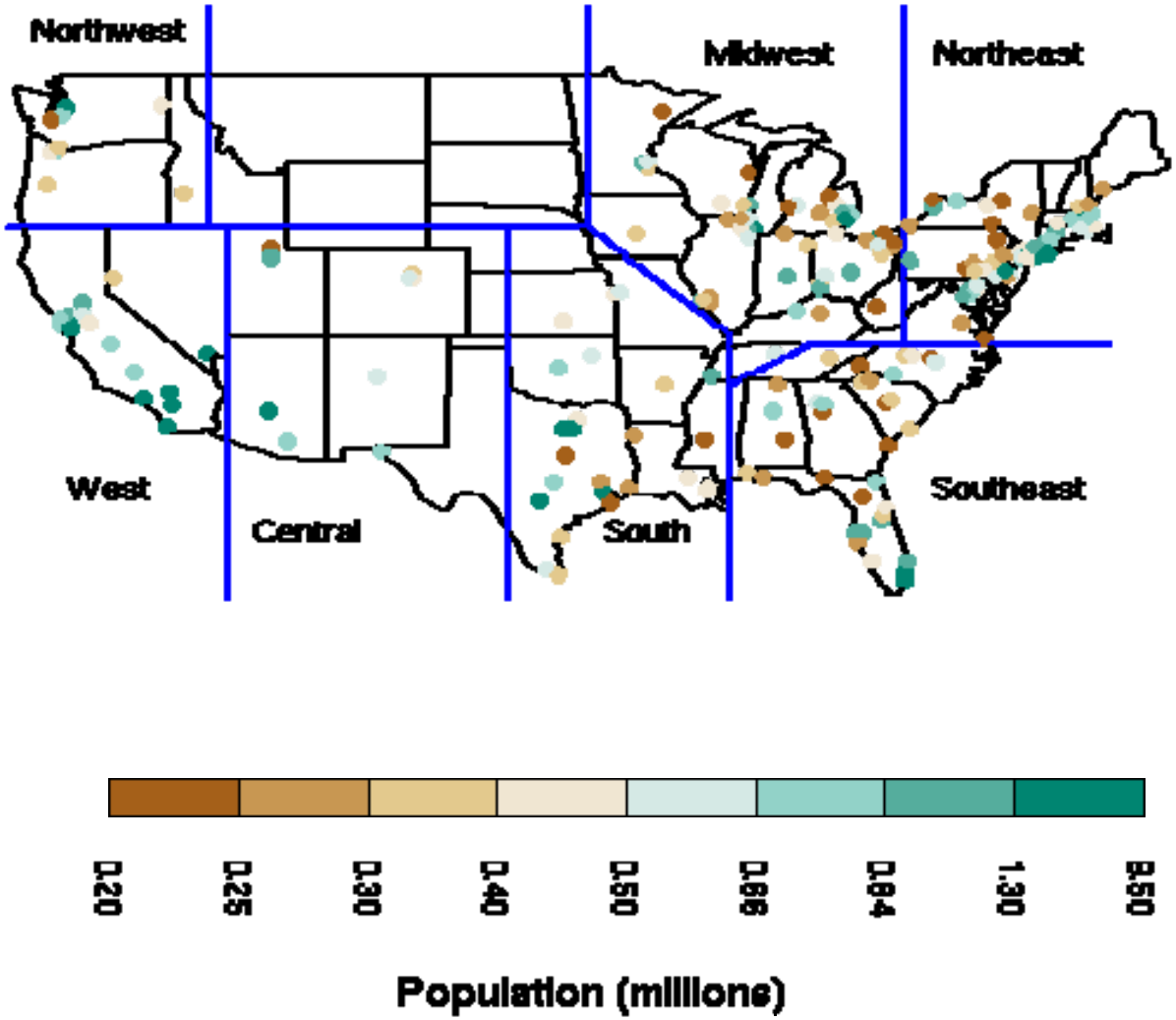



U.S. population / Medicare beneficiaries
Age 65+
2000
2.5
2.0
1.5
1.0
.5
0
Total U.S. population,
per 2000 census
Medicare beneficiaries
70
80
90
100
Age

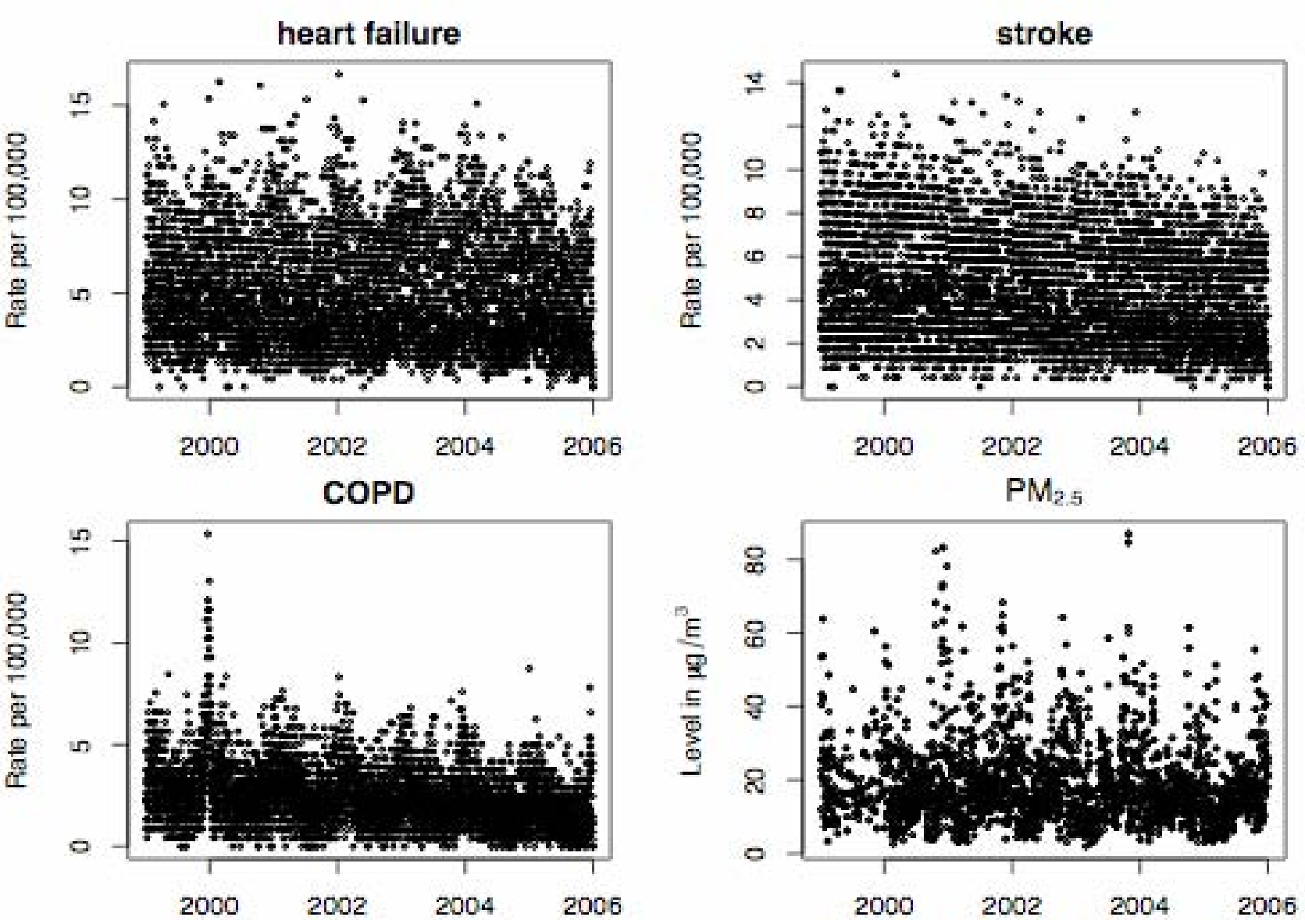
National Medicare Cohort (1999—2005)

- 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

MCAPS study population: 204 counties with populations larger than 200,000 (11.5 million people)



Daily time series of hospitalization rates and $PM_{2.5}$ levels in Los Angeles county (1999-2005)



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 and 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

Challenges

- For any given city, we try to estimate a small pollution effect relative to confounding effects of trend, season and weather
- Strong role of other time-dependent factors
- High correlation between non linear predictors
- Sensitivity of findings to model specifications

JAMA®
The Journal of the American Medical Association

$PM_{2.5}$ → **Hospital Admissions**

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

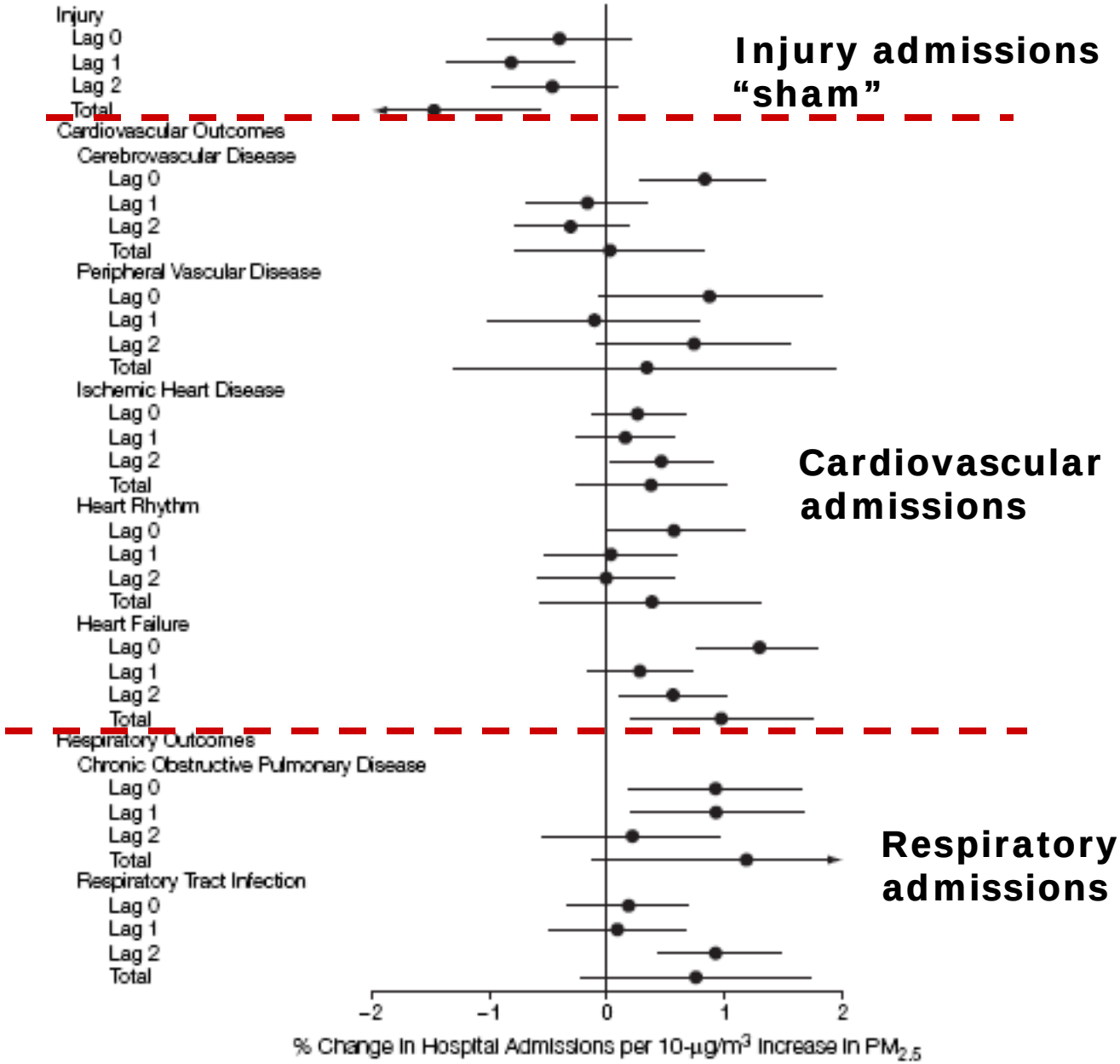
Context Evidence on the health risks associated with short-term exposure to fine particles (particulate matter ≤2.5 μm in aerodynamic diameter [$PM_{2.5}$]) is limited. Results from the new national monitoring network for $PM_{2.5}$ make possible systematic research on health risks at national and regional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

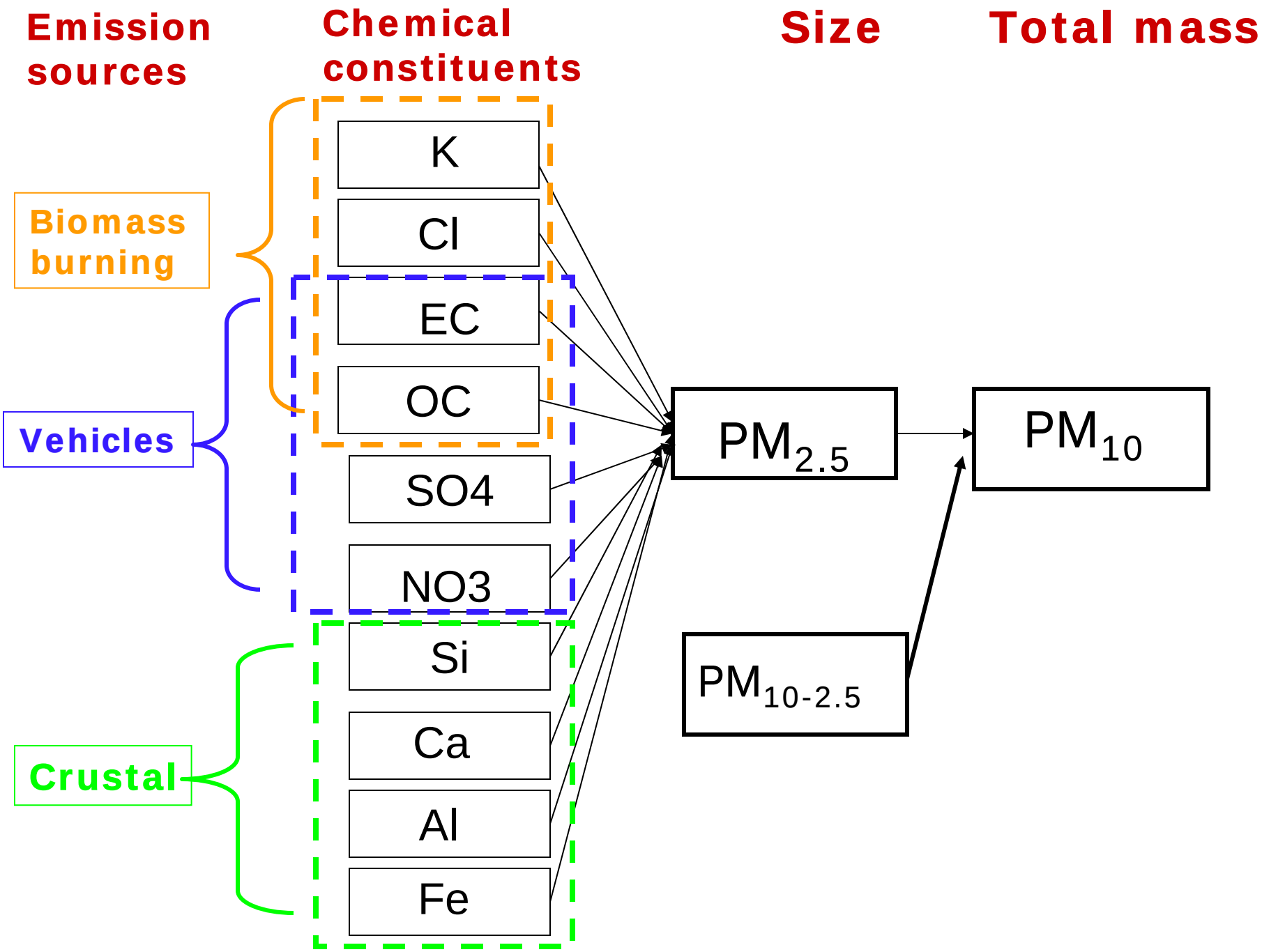
Figure 2. Percentage Change in Hospitalization Rate by Cause per 10-μg/m³ Increase in $PM_{2.5}$ on Average Across 204 US Counties



New Scientific Questions

What are the mechanisms of PM toxicity?

- **Size?**
- **Chemical components?**
- **Sources?**


Emission sources
Chemical constituents
Size
Total mass
Biomass burning
Vehicles
Crustal
K
Cl
EC
OC
SO4
NO3
Si
Ca
Al
Fe
PM2.5
PM10-2.5
PM10

Air pollution and health: Questions and (some) answers

- **Is there a risk?**
  - Multi-site time series studies such as NMMAPS (1987—2000) provide strong evidence of short-term association between air pollution and mortality
  - Preliminary results from Medicare data (1999—2002) indicate that current air pollution levels still affect health
- **How can we estimate it?**
  - National datasets are powerful resources for assessing the health effects of air pollution
  - Statistical models that can integrate information across space and time
  - National average estimates for the effect of PM are robust to various model formulations and statistical methods

Reproducible research

- We want to reproduce previous findings
  - "Did you do what you said you did?"
- Test assumptions, robustness of findings; check methodology
  - "Is what you did any good?"
- Implement and test new methodology
  - "I can do it better!"

NMMAPSdata R Package - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://ihapss.biostat.jhsph.edu/data/NMMAPS/R/

New York Times | Slashdot | Yahoo! | Yahoo! Calendar | Roger Peng's Page of ... | JHSPH Biostatistics | Roger Peng's Other Ho... | TinyURL! | BugMeNot.com

Stumble! All | I like it! | Not-for-me | ... | Menu

iHAPSS
Internet-based Health & Air Pollution Surveillance System

NMMAPSdata R Package

Publications | Software | Data | Rweb | iHAPSS

NMMAPSdata R Package

Current version: 0.3-4

The NMMAPSdata R package contains daily mortality, air pollution, and weather data originally assembled as part of the National Mortality, Morbidity, and Air Pollution Study (NMMAPS).

There is a technical report available which contains a brief overview of the package and contains examples of multi-city time series analysis of air pollution and mortality.

- The files simple.R, seasonal.R, and tdlm.R referenced in the report contain example code and functions for reproducing NMMAPS analyses.

Database summary information

- Time frame: January 1, 1987 -- December 31, 2000
- Causes of death: Total non-accidental, CVD, respiratory, pneumonia, COPD, accidental
  - Age categories: < 65, 65--74, >= 75
- Pollutants: PM_{10}, $PM_{2.5}$, CO, O_3, SO_2, NO_2
- Weather: Temperature, dewpoint temperature, relative humidity
- Number of Cities: 108

More detailed information about the database can be found on the IHAPSS website at http://www.ihapss.jhsph.edu/.

Package requirements

- R version 1.9.0 or higher.
- bzip2 compression capability. Most people will *not* have to worry about this since R comes with bzip2 compression capability by default. However, on some Unix-like systems it is possible that the version of R was compiled without it. NMMAPSdata will give an error when the package is loaded if bzip2 capability is not present.
- Approximately 380MB of disk space to store the package.

For **Unix**, **Linux**, and **Mac OS X** users, there is a source package available.

Done

rpeng's Ho | Trash | Start Her

[Inbox for rpeng@jhsph.edu - Mozilla Thunderbird | NMMAPSdata R Package - Mozilla Firefox | 1:12 PM

seasonal-techreport.pdf

File Edit Document View Window Help

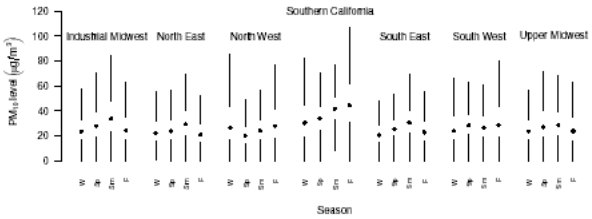


Figure 2: Boxplots of regionally averaged daily PM_{10} levels (in $\mu g/m^3$) by season.

	Winter	Spring	Summer	Fall
PM_{10} only (100 cities)	$0.15_{(-0.08,0.39)}$	$0.14_{(-0.14,0.42)}$	$0.36_{(0.11,0.61)}$	$0.14_{(-0.06,0.34)}$
with SO_2 (79 cities)	$0.24_{(-0.12,0.60)}$	$0.05_{(-0.44,0.55)}$	$0.47_{(-0.04,0.97)}$	$0.15_{(-0.21,0.51)}$
with O_3 (72 cities)	$0.21_{(-0.05,0.47)}$	$0.21_{(-0.08,0.51)}$	$0.32_{(0.04,0.59)}$	$0.01_{(-0.28,0.29)}$
with NO_2 (68 cities)	$0.18_{(-0.15,0.51)}$	$0.15_{(-0.22,0.51)}$	$0.34_{(-0.04,0.72)}$	$0.16_{(-0.18,0.51)}$

Table 3: Seasonal PM_{10} estimates in two-pollutant models.

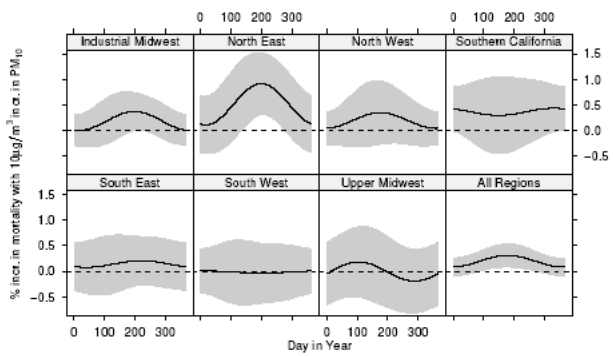


Figure 3: National and regional seasonal curves estimated from the sine/cosine model for lag 1 PM_{10}. Gray regions indicate pointwise 95% posterior intervals.

18 of 20 8.5 x 11 in

Zooey

```
x <- 1:ndays
B <- periodicBasis(x, 1, ndays, intercept = TRUE, ortho = FALSE)

bc <- extractBetaCov(r, pollutant = "pm10")
initTLNise()
g <- tlnise(bc$beta, bc$cov, regionInd, prnt = FALSE, labelY = names(r),
            intercept = FALSE, seed = 54321)
pooledRegion <- matrix(g$gamma[,1], byrow = TRUE, ncol = ncol(regionInd),
                       nrow = nrow(g$theta))
pooledRegionSD <- matrix(g$gamma[,2], byrow = TRUE, ncol = ncol(regionInd),
                         nrow = nrow(g$theta))
b <- B %*% pooledRegion
zero <- matrix(0, nrow = ndays, ncol = 7 * 3)
curveRegion <- lapply(1:ncol(pooledRegion), function(i) {
    B <- zero
    ## This is weird
    B[, seq(i, 7 * 3, 7)] <- periodicBasis(x, 1, ndays, intercept = TRUE, ortho
= FALSE)
    s <- sqrt(diag(B %*% g$Dgamma %*% t(B)))
    cbind(b[,i], b[,i] - 2*s, b[,i] + 2*s)
})

## Overall estimate
initTLNise()
g <- tlnise(bc$beta, bc$cov, rep(1, NROW(bc$beta)), seed = 12345, prnt = FALSE)
V <- diag(B %*% g$Dgamma %*% t(B))
b <- B %*% g$gamma[,1]
curve <- cbind(b, b - 2*sqrt(V), b + 2*sqrt(V))

## data(regions)
regionFullNames <- c("Industrial Midwest", "Northeast", "Northwest",
                     "Southern California", "Southeast", "Southwest",
                     "Upper Midwest")
## regionFullNames <- with(regions, description[match(regionNames, value)])

## Plot
Y <- rbind(do.call("rbind", curveRegion), curve) * 1000
conf <- Y[,2:3]
rng <- range(Y) + c(-1, 1) * .05
y <- Y[,1]
nRegions <- length(curveRegion) + 1
x <- rep(1:ndays, nRegions)
g <- ordered(rep(c(regionFullNames, "All Regions"), each = ndays),
             levels = c(regionFullNames, "All Regions"))

## trellis.device(dev = x11, height = 5, width = 8, color = FALSE)
p <- xyplot(y ~ x | g, as.table = TRUE, subscripts = TRUE, type = "l",
            panel = function(x, y, subscripts, ...) {
                llines(x, Y[subscripts, 2], lty = 3, lwd = 2)
                llines(x, Y[subscripts, 3], lty = 3, lwd = 2)
                panel.xyplot(x, y, ...)
                panel.abline(h = 0, lty = 2)
            },
            strip = strip.custom(bg = 0),
            ylim = rng, layout = c(4, 2),
            ylab = list(expression(paste("% increase in mortality with ",
                "10-", mu, "g/", m^3," increase in ", PM[10])), cex = 0.9),
            xlab = "Day in year",
            scales = list(alternating = 1, tck = c(0.8, 0))
            )
print(p)
:
```

Zooey

seasonal-techreport.pdf

11:52 AM

Discussion

- Linking national databases and developing statistical methods that can properly analyze these them, are essential steps for a successful national public health tracking system
- Because of the small risks to be detected and the large number of potential confounders, single-site studies are generally swamped by statistical error
- A national system, that routinely analyze data from multiple locations in a systematic fashion, is a very promising approach for tracking population health

Explosion of Information
e.g.
large databases
on population health and exposure
to potentially toxic agents

Expertise in:

- **Integration of complex databases**
- **Statistical Methods**
- **Reproducibility**

More confusion

More knowledge and
Better health risk assessment

Acknowledgments

- **Our team:**
  - R. Peng
  - S. Zeger
  - J. Samet
  - A. McDermott
  - M. Bell
  - L. Pham
- **Our sponsors:**
  - EPA
  - JHU CDC Center of Excellence
  - NIEHS