How Risky is Breathing?
Statistical Methods in Air Pollution Risk Estimation

Francesca Dominici
Department of Biostatistics
Johns Hopkins University
Bloomberg School of Public Health

Sponsored by the EPA and the NIEHS
Joint Work with

- Michelle Bell
- Aidan McDermott
- Roger Peng
- Jonathan Samet
- Scott Zeger
From crisis to questions

• We began with crisis---Meuse Valley 1932, Donora 1948, London 1952 and have moved to questions:
  – Are there adverse effects of today’s air pollution?
  – How large are these risks?
  – What is the cost-benefit ratio for control?
London by day: The 1952 Fog
APS Power Plant, Four Corners Area
Steel Mill, Steubenville, Ohio
How do epidemiologist/biostatisticians estimate the health effects of air pollution

- **Goal**: To assess health effects of ambient air pollution
- **Health effects**: acute or chronic
- **Exposure**: generally from ambient monitors
- **Confounding**: can never be ruled out in observational studies
Some early single-city time series studies in the U.S.

- Utah Valley, KY: Pope et al., 1992
- Philadelphia, PA: Kelsall et al., 1997
- Birmingham, AL: Schwartz, 1993
- Steubenville, OH: Schwartz, 1992
Multi-site time series studies

The National Morbidity Mortality Air Pollution Study (NMMAPS, 1987-2000)

The National Medicare Air Pollution Study (MCAPS, 1999-2005)
Data and Methods
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
  - $\text{PM}_{10}$, $\text{PM}_{2.5}$, $\text{O}_3$, $\text{NO}_2$, $\text{SO}_2$, $\text{CO}$
- U.S. Census 1990, 2000
The National Medicare Cohort Study, 1999-2005 (MCAPS)

• Medicare data include:
  – Billing claims 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 PM$_{2.5}$ monitoring network
• Study population includes **11.5 million** Medicare enrollees living on average **5.9 miles** from a PM$_{2.5}$ monitor
Daily time series of hospitalization rates and PM$_{2.5}$ levels in Los Angeles county (1999-2002)

- **Ischemic heart disease, Los Angeles**
  - Hospitalizations per 100
  - 1999-2003

- **Respiratory infection, Los Angeles**
  - Hospitalizations per 100
  - 1999-2003

- **Daily COPD hospitalization rate (per 100,000)**
  - for Los Angeles County, CA
  - 1999-2003

- **Daily PM$_{2.5}$ for Los Angeles County, CA**
  - 1999-2003
MCAPS study population: 204 counties with populations larger than 200,000 (11.5 million people)
Multi-site time series models of air pollution and mortality

- **Stage 1 (within city):** Poisson regressions for estimating short-term association between air pollution and mortality, controlling for time-varying confounders
- **Stage 2 (between cities):** Hierarchical model for pooling information across neighboring cities and obtaining a national average effect
Methods for 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
How strong is the evidence?
Ozone and Short-term Mortality in 95 US Urban Communities, 1987-2000

Michelle L. Bell, PhD
Aidan McDermott, PhD
Scott L. Zeger, PhD
Jonathan M. Samet, MD
Francesca Dominici, PhD

Context  Ozone has been associated with various adverse health effects, including increased rates of hospital admissions and exacerbation of respiratory illnesses. Although numerous time-series studies have estimated associations between day-to-day variation in ozone levels and mortality counts, results have been inconclusive.

Objective  To investigate whether short-term (daily and weekly) exposure to ambient ozone is associated with mortality in the United States.
Is there an association between ozone and mortality?

**Figure 1.** Percentage Change in Daily Mortality for a 10-ppb Increase in Ozone for Total and Cardiovascular Mortality, for Single-Lag and Distributed-Lag Models

The single-lag model reflects the percentage increase in mortality for a 10-ppb increase in ozone on a single day. The distributed-lag model reflects the percentage change in mortality for a 10-ppb increase in ozone during the previous week. Error bars indicate 95% posterior intervals.

Bell, Samet, McDermott, Zeger, Dominici JAMA 2004
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 \( \leq 2.5 \) \( \mu \text{m} \) 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
Figure 2. Percentage Change in Hospitalization Rate by Cause per 10-μg/m³ Increase in PM$_{2.5}$ on Average Across 204 US Counties

- Sham outcome
- Cardiovascular outcomes
- Respiratory outcomes
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

<table>
<thead>
<tr>
<th>Outcome</th>
<th>West</th>
<th>East</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebrovascular Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peripheral Vascular Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemic Heart Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Rhythm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Failure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory Tract Infection</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

% Change in Hospital Admissions per 10-μg/m³ Increase in PM$_{2.5}$

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.
Results

• We found evidence of an association between Ozone and daily mortality on a national scale

• We found evidence of an association between PM$_{2.5}$ and daily hospitalizations on a national scale
  – We found evidence of spatial heterogeneity in the effect of PM$_{2.5}$ on risk for hospitalization.
  – There are known differences in the composition of PM$_{2.5}$, including a greater sulfate component in the East and a greater nitrate component on the West that might explain these differences
New Challenges

What are the mechanisms of PM toxicity? Size? Chemical components? Sources?
Health Effects of size and chemical constituents of Particulate Matter

- Sulfate
- Nitrate
- Organic carbon
- PM$_{10-2.5}$
- PM$_{2.5}$
- O$_3$

Health
Health Effects of PM$_{2.5}$ Speciation data

1. We have constructed a database of 52 PM$_{2.5}$ chemical constituents for 204 U.S. counties for 2000 to 2005

2. We have identified a subset of PM$_{2.5}$ components that substantially contribute and/or co-vary with daily PM$_{2.5}$ concentrations (Bell et al EHP 2006)
Legend

Sulfate Events

- 0_Year
  - 0.6195 - 2.0364
  - 2.0365 - 3.3865
  - 3.3866 - 4.3633
  - 4.3634 - 5.1136
  - 5.1137 - 6.8292

Sulfate PM$_{2.5}$ average for 203 U.S. counties, 2000-2005 (ng/m$^3$)
PM$_{2.5}$ chemical composition

- OC: 33%
- SO$_4^{2-}$: 30%
- NO$_3^-$: 14%
- NH$_4^+$: 12%
- EC: 5%
- Si: 1%
- Na$^+$: 1%
- Other: 4%
Statistical challenges in estimating health effects of PM chemical constituents

- Hard to separate the effects of the individual constituents because of the high correlation between predictors and measurement error.
- We should account the compositional nature of the data.
- We should take advantage of the hierarchy of the geographical locations and the hierarchy of the chemical constituents.
- Interactions can also be important.
Are Results Reproducible?
Are current epidemiological studies reproducible?

The research community

<table>
<thead>
<tr>
<th>Statistical analysis implementation</th>
<th>NMMAPS (1987-2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not reported</td>
<td>21</td>
</tr>
<tr>
<td>By hand</td>
<td>0</td>
</tr>
<tr>
<td>Use of software package</td>
<td>48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method of processing measured data</th>
<th>NMMAPS (1987-2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not reported</td>
<td>43</td>
</tr>
<tr>
<td>By hand</td>
<td>1</td>
</tr>
<tr>
<td>Use of software package</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome data reported to be available</th>
<th>NMMAPS (1987-2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure data reported to be available</td>
<td>0</td>
</tr>
<tr>
<td>Code for statistical analysis available</td>
<td>0</td>
</tr>
<tr>
<td>Code for processing measured data available</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research component</th>
<th>What we have done</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>The entire NMMAPS database is available to the public via the iHAPSS website and the NMMAPS data package for R; the data are available under a “full access” class of license.</td>
</tr>
<tr>
<td>Methods</td>
<td>A full compendium written in ( \text{LaTeX} ) and R is available for download.</td>
</tr>
<tr>
<td>Documentation</td>
<td>We have outlined our data-processing pipelines on the iHAPSS website, and papers/technical reports are available for download.</td>
</tr>
<tr>
<td>Distribution</td>
<td>We use the World Wide Web to disseminate our data and software.</td>
</tr>
</tbody>
</table>

*Details at [http://www.biostat.jhsph.edu/~rpeng/reproducible/](http://www.biostat.jhsph.edu/~rpeng/reproducible/).  
† NMMAPS, National Morbidity, Mortality, and Air Pollution Study, iHAPSS, Internet Health and Air Pollution Surveillance System.

Peng, Dominici, Zeger AJE 2006
NMMAPSdata package for R

• R is a free software environment for statistical analysis and graphics
• NMMAPSdata package contains the entire updated (1987—2000) NMMAPS database as an add-on module for R
• Supplemental code available online for reproducing canonical NMMAPS analysis and other analyses
• iHAPSS: Internet-based Health and Air Pollution Surveillance System
  - http://www.ihapss.jhsph.edu/

Peng, Welty, 2006, R news
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.


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
Reproducible Research
(www.biostat.jhsph.edu/MCAPS)

Medicare Air Pollution Study (MCAPS), 1999--2002

Benefits:
- Verifying published findings
- Conducting alternative analyses of the same data
- Eliminating uninformed criticisms which do not match data
- Expediting interchange of ideas among investigators

Supplementary Materials

FREQUENTLY ASKED QUESTIONS about the MCAPS study

Press release from The Johns Hopkins Bloomberg School of Public Health

Materials for Reproducing Study Results

County-specific estimates
Discussion

• Identify the harmful sources and/or chemical constituents of PM$_{2.5}$ is among the MOST important questions in air pollution epidemiology
• Sources are not measured
• PM$_{2.5}$ chemical components are measured poorly, but at a national scale
• This is a real opportunity for statisticians for methods development
• Reproducibility must be a necessary component of these investigations