A Bayesian Analysis of the Relationship Between Exposure to Fine Particulate Matter and Cardiovascular Mortality

Catherine A. Calder
Department of Statistics
The Ohio State University

Christopher H. Holloman, Steven M. Bortnick, Warren Strauss, Michele Morara
Statistics and Data Analysis Systems
Battelle Memorial Institute

Division of Epidemiology and Biostatistics Seminar
November 23, 2004

This research was funded by a 2003 Battelle Internal Research and Development Fund.
Outline

• Background
• Our Hierarchical Model
  – Level 1: Relating Monitor Readings to Ambient PM Concentrations
  – Level 2: Relating Ambient Concentrations to Personal Exposure
  – Level 3: Relating Personal Exposure to Cardiovascular Mortality
• Results
• Model Comparisons
• Discussion
Background - PM

**Particulate Matter (PM)** - generic term for airborne particles

\[ \text{PM}_{2.5} = \text{PM}_{\text{fine}} = \text{ambient particles less than } 2.5 \text{ microns (} \mu \text{m} \text{) in aerodynamic diameter (includes organic compounds, metals, water droplets, sulfate, nitrate, ammonium, hydrogen ions, and elemental carbon)} \]

**Historical Monitoring Focus**

Total Suspended Particulate (TSP) → **PM\(_{10}\)** → **PM\(_{2.5}\)**

- until the mid-1980s
- late 1980s - late 1990s
- last 4-5 years
Background - Health Effects

Effects of particulate matter (PM) on health are difficult to study directly.
- composition of PM is complex
- hard to measure PM exposure directly

Previous studies have found (after adjusting for confounding factors) a correlation between PM and acute health effects:
- Non-accidental Mortality (Goldberg et al. 2001)
- Cardiovascular Deaths (Ostro et al. 2000, Hoek et al. 2001)
- Elderly Deaths (Katsouyanni et al. 2001)
- Morbidity (Schwartz 1999, Zanobetti et al. 2000)

Traditional Modeling Approach: relate ambient levels measured by monitors to health outcomes such as mortality, morbidity, asthma, etc.
Background - Modeling

Classical Model for Relating PM to Health Outcomes

- Usually a single centrally located monitor. Sometimes combine data from several monitors.

PM$_{2.5}$ Data from Monitors → Poisson GAM → Health Effects

- Relate monitor readings of ambient PM$_{2.5}$ level to mortality or morbidity in a single city or county.

Confounders (e.g., weather, long-term change in mortality rate)
Background - Modeling

These models are useful for relating ambient PM$_{2.5}$ to mortality, but they are difficult to adapt for relating exposure to mortality since exposure can’t be measured over large populations.

Exposure Link: HEI Study (Dominici et al. 2001)

– Relates ambient PM$_{10}$ measurements to county-level mortality counts in Baltimore, MD from 1987 to 1994.
– Incorporates information on PM exposure, when exposure data is not available.
Our Hierarchical Model

PM$_{2.5}$ Data from Monitors

True Ambient PM$_{2.5}$ Surface

Group Exposure Levels

Health Effects

- Examine all ambient PM$_{2.5}$ monitors over the region of interest.
- Estimate the true ambient surface over the entire region (more than just one site or county).
- Infer exposure levels from true ambient level using a simulator.
- Relate group exposure levels to group health outcomes.
Our Hierarchical Model

- Examine all ambient PM$_{2.5}$ monitors over the region of interest.
- Estimate the true ambient surface over the entire region (more than just one site or county).
- Infer exposure levels from true ambient level using a simulator.
- Relate group exposure levels to group health outcomes.

We can make inference on the direct connection between exposure and mortality.

PM$_{2.5}$ Data from Monitors

True Ambient PM$_{2.5}$ Surface

Group Exposure Levels

Health Effects
Our Hierarchical Model

PM$_{2.5}$ Data from Monitors

True Ambient PM$_{2.5}$ Surface

Group Exposure Levels

Health Effects

confounders

A single, coherent statistical model

- Examine all ambient PM$_{2.5}$ monitors over the region of interest.
- Estimate the true ambient surface over the entire region (more than just one site or area).
- Infer exposure levels from true ambient level using a simulator.
- Relate group exposure levels to group health outcomes.
Our Hierarchical Model

Data:
• Mortality (8 NC counties over 1096 days)
• PM$_{2.5}$ monitors (23 monitors)
• Meteorology (17 stations)
• Demographics
• Activity Patterns

Statistical Techniques:
• Spatial Modeling
• Incorporation of an Exposure Simulator (exposure data unavailable)
• Poisson GAM relating exposure to mortality

66,945 parameters
Level 1: Relating Monitor Readings to Ambient Levels

PM\textsubscript{2.5} Data from Monitors

True Ambient PM\textsubscript{2.5} Surface

Group Exposure Levels

Health Effects

- PM\textsubscript{2.5} observations are made at a network of ambient monitors in and around the region of interest.
- Ambient level exists at all locations.
- Interpolation is performed by assuming the true ambient surface is smooth and that the monitors aren’t too inaccurate.
PM$_{2.5}$ concentration readings ($\mu g/m^3$ local conditions) taken from the EPA’s AQS database.
Level 1: Relating Monitor Readings to Ambient Levels
Level 1:  
Relating Monitor Readings to Ambient Levels

Notation:  
- $X_t$ - monitor readings at time $t$  
- $\psi_t$ - ambient level at time $t$

- Monitor readings are normally distributed around the ambient surface.  
  \[
  X_t(s) \mid \psi_t(s), \sigma_x^2 \sim N(\psi_t(s), \sigma_x^2)
  \]
- The ambient surface is spatially correlated.  
  \[
  \psi_t(s_1, \ldots, s_{n_p}) \mid \theta, \Sigma \sim N((m_t(\theta, s_1), \ldots, m_t(\theta, s_{n_p})), \Sigma)
  \]
- The ambient level depends on some covariates.  
  \[
  m_t(\theta, s) = \theta_0 + \theta_1 \text{(maximum temperature}_t(s)) + \theta_2 \text{(average wind speed}_t(s)) + \theta_3 \sin\left(\frac{2\pi t}{365} + \theta_4\right).
  \]
  (Weather data obtained from the National Climatic Data Center (NCDC)).
Level 1: Relating Monitor Readings to Ambient Levels

Semi-variogram for PM2.5

Exponential Covariance Function:

\[ V(d) = 2.87 + 52.06 \left( 1 - \exp\left( -\frac{d}{6.78} \right) \right) \]

Assumptions:

1. No temporal dependence
2. Covariance is constant over time
Level 1: Relating Monitor Readings to Ambient Levels

Advantages:

- Estimation of ambient surface in counties with no monitors.
- Accounts for uncertainty in monitor readings (incorporate an informative prior).
- Straightforward to include the effects of weather and seasonal cycles for modeling the ambient surface.
- Accounts for varying levels of uncertainty depending on the number of monitor readings on any given day.

Limitations:

- Requires modeling assumptions to specify properties of surface (form of the mean level model, smoothness).
Level 2:
Relating Ambient Levels to Group Exposure

- Use an exposure simulation approach similar to the SHEDS-PM model (Burke, Zufall, Özkaynak, 2001).
- Generate a random sample of individuals in each county of interest (using demographic information).
- Match individuals with activity records from CHAD.
- For each day and in each county, calculate the true average exposure given the average ambient level and the activity data.
Level 2: Relating Ambient Levels to Group Exposure

True Average Ambient PM$_{2.5}$ Level (County and Day Specific)

Activity Information (County Specific)

Demographics

CHAD (NHAPS)

Exposure Simulator

True Average Exposure (County and Day Specific)
Level 2: Relating Ambient Levels to Group Exposure

Notation: 
- $\overline{\psi}_{ct}$ - Average ambient PM$_{2.5}$ level in county $c$ at time $t$
- $Z_{ct}$ - Average exposure level in county $c$ at time $t$

- True average exposure in a county is normally distributed around the value predicted by the simulator.

$$Z_{ct} \mid \overline{\psi}_{ct}, \sigma_z^2 \sim N(\xi(\overline{\psi}_{ct}, A_c), \sigma_z^2)$$
Level 2: Relating Ambient Levels to Group Exposure

Our Exposure Simulator:
(based on SHEDS-PM, Burke et al. 2001)

\[ \phi_{ct,i} = \frac{1}{1440} \left[ m_{oi} \bar{\psi}_{ct} + \sum_{j=1}^{n} m_{ji} L_{ct,j} \right] + \frac{m_{\text{smoke},i}}{1440} \times L_{\text{smoke}} + \frac{m_{\text{cook},i}}{1440} \times L_{\text{cook}} \]

Levels in Indoor Microenvironments: \[ L_{ct,j} = w \bar{\psi}_{ct} + (1 - w)L^*_j \]

(w - “ambient influence parameter” (between 0 and 1)

- \( w = 1 \) implies that indoor and ambient levels are equivalent.
- \( w = 0 \) implies that the ambient level has no influence on indoor levels.)
# Level 2: Relating Ambient Levels to Group Exposure

<table>
<thead>
<tr>
<th>Indoor Environment</th>
<th>$L_e^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential</td>
<td>2.89μg/m$^3$</td>
</tr>
<tr>
<td>Office</td>
<td>4.50μg/m$^3$</td>
</tr>
<tr>
<td>School</td>
<td>9.80μg/m$^3$</td>
</tr>
<tr>
<td>Store</td>
<td>12.70μg/m$^3$</td>
</tr>
<tr>
<td>Vehicle</td>
<td>34.30μg/m$^3$</td>
</tr>
<tr>
<td>Restaurant</td>
<td>14.80μg/m$^3$</td>
</tr>
<tr>
<td>Bar</td>
<td>14.80μg/m$^3$</td>
</tr>
</tbody>
</table>
Level 2: Relating Ambient Levels to Group Exposure

Generating Activity Pattern:

- Sample 100 random individuals from each county that are representative of that county based on demographic information (age, sex, and employments) obtained from Census 2000.

- Match each individual with an activity diary in NHAPS-A or NHAPS-B (CHAD) studies. Records are selected in a way such that all days of the week are represented equally.

Predict the exposure level for each individual in each county.

Average Exposure Levels: \[ \xi(\overline{\psi}_{ct}, A_c) = \frac{1}{100} \sum_{i=1}^{100} \phi_{ct,i} \]
Level 3: Relating Exposure to Mortality

- Use the standard Poisson GAM form replacing monitor readings with exposure.

1. PM$_{2.5}$ Data from Monitors
2. True Ambient PM$_{2.5}$ Surface
3. Group Exposure Levels
4. Health Effects

**confounders**
Level 3: Relating Exposure to Mortality

Mortality Data: daily mortality counts (death from cardiovascular causes) from The Odum Institute at the University of North Carolina

Modeling Approach: Standard Poisson Generalized Additive Model (GAM)

\[
\text{Mortality}_{ct} \sim Poi(\lambda_{ct} E_c)
\]

\[
\log(\lambda_{ct}) = \mu + \sum_{k=0}^{3} Z_{c,t-k} \beta_k + \sum_{k=1}^{p} (\text{Confounders}_{ck}) \gamma_k
\]
Our Hierarchical Model

- PM$_{2.5}$ Data from Monitors
- True Ambient PM$_{2.5}$ Surface
- Group Exposure Levels
- Health Effects

- Model is fit using a Markov chain Monte Carlo (MCMC) algorithm run for 750,000 iterations (plus 250,000 ‘burnin’ iterations)
Results

- Effects of PM$_{2.5}$ exposure on the relative risk of mortality (from left to right) the same day, the following day, two days later, and three days later.

- Mean and 95% CI for $w$ (ambient influence parameter):
  0.8106 (0.7120, 0.9061)
Results

- Empirical estimate of the correlation between ambient levels and exposure as a function of the ambient influence parameter ($w$).

$w = 1$ implies that indoor and ambient levels are equivalent.

$w = 0$ implies that the ambient level has no influence on indoor levels.
Results

- Boxplots of the effect of PM$_{2.5}$ exposure on the relative risk of mortality, as a function of the ambient influence parameter ($w$):

$w = 1$ implies that indoor and ambient levels are equivalent.

$w = 0$ implies that the ambient level has no influence on indoor levels.
Model Comparisons

1. Our Exposure Simulator

\[ Z_{ct} | \bar{\psi}_{ct}, \sigma^2_z \sim N(\xi(\bar{\psi}_{ct}, A_c), \sigma^2_z) \]

2. Simple Exposure Simulator

\[ Z_{ct} | \bar{\psi}_{ct}, \sigma^2_z \sim N(\bar{\psi}_{ct}, \sigma^2_z) \]

3. No Exposure Simulator

\[ Z_{ct} = \bar{\psi}_{ct} \]
Model Comparisons

### Beta_0

![Boxplot for Beta_0](image)

### Beta_1

![Boxplot for Beta_1](image)

### Beta_2

![Boxplot for Beta_2](image)

### Beta_3

![Boxplot for Beta_3](image)
Discussion

Contributions of Hierarchical Modeling:

- We can specify a single sophisticated model in simple stages.
- Data can be inserted at any stage of the model.
- We can include populations often underrepresented in epidemiological studies via spatial modeling.
- It is possible to incorporate deterministic (and stochastic) simulators within a statistical model.
- Alterations to the model are simple – we can substitute different parts without respecifying the entire model or worrying about being able to fit it.
- Bayesian modeling allows probabilistic statements about parameters – directly applicable to cost/benefit analysis.
Discussion

Extensions

• More sophisticated exposure simulator
  – random baseline levels
  – temporally varying activity patterns
  – different values of $w$ for the various indoor microenvironments

• Space-time modeling of the ambient PM concentrations
  – temporal dependence
  – temporally varying spatial dependence structure (non-separable)
  – co-pollutant information