Building sound methodology for analysis of water and health data

Elena Naumova, Jyotsna Jagai

Tufts University School of Medicine, Boston MA USA
March 14th, 2007
Discussion Questions

- Do watersheds represent drinking water source for the surrounding community?

- Do these studies suggest that an increase in rainfall will lead to an increase in the number of cases of waterborne disease? Why or why not? If not, why should we pursue these types of studies? If yes, what can be done based on these results?

- What are the advantages and disadvantages of using survey design methodology overall? For studying environmental epidemiology?

- Discuss the pros and cons of using large datasets such as the EPA outbreak data?
Goals and Objectives

Main goal: To examine temporal and spatial patterns of gastrointestinal diseases (GID) in the US elderly with respect to climate and water indicators at the national and regional levels.

Objectives: To assess whether spatial and temporal variability in GID in the elderly is associated with climate and water indicators.

- Static assessment
- Dynamic assessment
- Methodology illustration using records of hospitalization for Salmonella Infection
CMS Databases

Centers for Medicare & Medicaid Services (CMS) databases

Total U.S. population 2000  279,583,437
U.S. elderly (≥65yo) 2000  34,795,453 (12.5%)
Medicare beneficiaries ~ 95%

Number of CMS records  62,301,882
Number of CMS records/year  12,460,376

GID: ICD-9M codes: 001-009, 558.9, 787

CMS GID records  3,470,526
CMS GID records/year  694,105
## GID Composition

<table>
<thead>
<tr>
<th>Disease Category</th>
<th>ICD 9M</th>
<th>Total Counts</th>
<th>Females %</th>
<th>Annual Rate *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cryptosporidiosis</td>
<td>007.(2, 4, 8)</td>
<td>174</td>
<td>66%</td>
<td>0.10</td>
</tr>
<tr>
<td>Shigellosis</td>
<td>004</td>
<td>871</td>
<td>66%</td>
<td>0.49</td>
</tr>
<tr>
<td>Giardiasis</td>
<td>007.1</td>
<td>1,707</td>
<td>63%</td>
<td>0.96</td>
</tr>
<tr>
<td>Campylobacteriosis</td>
<td>008.43</td>
<td>3,369</td>
<td>59%</td>
<td>1.90</td>
</tr>
<tr>
<td>Salmonellosis</td>
<td>003</td>
<td>8,905</td>
<td>60%</td>
<td>5.03</td>
</tr>
<tr>
<td><em>Clostridium difficile</em></td>
<td>008.45</td>
<td>448,914</td>
<td>62%</td>
<td>253.60</td>
</tr>
<tr>
<td><strong>All GID</strong></td>
<td><strong>001 – 009</strong></td>
<td><strong>633,786</strong></td>
<td><strong>65%</strong></td>
<td><strong>358.04</strong></td>
</tr>
<tr>
<td>Nonspecific gastroenteritis</td>
<td><strong>558.9</strong></td>
<td><strong>549,483</strong></td>
<td><strong>71%</strong></td>
<td><strong>310.41</strong></td>
</tr>
<tr>
<td>Symptoms involving digestive system</td>
<td><strong>787</strong></td>
<td><strong>2,363,893</strong></td>
<td><strong>64%</strong></td>
<td><strong>1335.41</strong></td>
</tr>
</tbody>
</table>

* rate per 100,000 elderly
# GID Composition

<table>
<thead>
<tr>
<th>Disease Category</th>
<th>ICD 9M</th>
<th>Total Counts</th>
<th>Females %</th>
<th>Annual Rate *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cryptosporidiosis</td>
<td>007.(2, 4, 8)</td>
<td>174</td>
<td>66%</td>
<td>0.10</td>
</tr>
<tr>
<td>Shigellosis</td>
<td>004</td>
<td>871</td>
<td>66%</td>
<td>0.49</td>
</tr>
<tr>
<td>Giardiasis</td>
<td>007.1</td>
<td>1,707</td>
<td>63%</td>
<td>0.96</td>
</tr>
<tr>
<td>Campylobacteriosis</td>
<td>008.43</td>
<td>3,369</td>
<td>59%</td>
<td>1.90</td>
</tr>
<tr>
<td>* rate per 100,000 elderly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>** Clostridium difficile **</td>
<td>008.45</td>
<td>448,914</td>
<td>62%</td>
<td>253.60</td>
</tr>
<tr>
<td>All GID</td>
<td>001 – 009</td>
<td>633,786</td>
<td>65%</td>
<td>358.04</td>
</tr>
<tr>
<td>Nonspecific gastroenteritis</td>
<td>558.9</td>
<td>549,483</td>
<td>71%</td>
<td>310.41</td>
</tr>
<tr>
<td>Symptoms involving digestive system</td>
<td>787</td>
<td>2,363,893</td>
<td>64%</td>
<td>1335.41</td>
</tr>
</tbody>
</table>

Naumova EN / WSSS Seminar/ March 2008
Spatial Distributions

- Successful CMS records’ geocoding at the county level ~93%
- Aggregation of counties with < 1000 elderly persons (from 3141 to 2794 counties)
- Supplementary data from Census, ESRI, FEMA, PRISM, USGS at the county level:
  - Demographic indicators (Total population; Population density; Percent of elderly)
  - Socio-economic indicators (Number of nursing homes and hospitals; Number of nursing home residents and hospital bed)
  - Environmental indicators (Watershed hydrologic regions; Köppen Climate Classification; Monthly temperature and precipitation; Domestic water use)
Aging Communities and Distribution of GID

Geographic distribution of aging communities

- Total Population
- Population Density
- Percent of Elderly

Percentage of Total U.S. Population Ages 65 and Over

<table>
<thead>
<tr>
<th>Total U.S. Population Ages 65 and Over</th>
<th># Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1% - 16.7%</td>
<td>537</td>
</tr>
<tr>
<td>16.8% - 21.0%</td>
<td>346</td>
</tr>
<tr>
<td>21.1% - 35.7%</td>
<td>173</td>
</tr>
</tbody>
</table>

Average Annual Rate of Gastrointestinal Infectious Diseases (ICD-9M 558.9) for the U.S. Elderly, Ages 65 and Over

<table>
<thead>
<tr>
<th>Hospitalization Rate per 1,000 people (Counts)</th>
<th># Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4 - 1.9</td>
<td>346</td>
</tr>
<tr>
<td>2.0 - 3.1</td>
<td>537</td>
</tr>
<tr>
<td>3.2 - 4.8</td>
<td>346</td>
</tr>
<tr>
<td>4.8 - 7.8</td>
<td>910</td>
</tr>
<tr>
<td>7.9 - 30.6</td>
<td>324</td>
</tr>
</tbody>
</table>

Naumova EN / WSSS Seminar/ March 2008
GID Seasonality

- *Salmonella* Infection: 72% variability is explained by seasonal and trend components*

\( Y(t) = \exp \{ \beta_0 + \beta_1 \sin(2\pi \omega t) + \beta_2 \cos(2\pi \omega t) + \beta_3 \sin(4\pi \omega t) + \beta_4 \cos(4\pi \omega t) + \beta_5 f(t) \} \)
Seasonal Maps of *Salmonella* Infection

**Winter: October 15 – April 14**

**Average Annual Winter Rate of Salmonellosis Hospitalizations for the U.S. Elderly, Ages 65 and Over**

**Summer: April 15 – October 14**

**Average Annual Summer Rate of Salmonellosis Hospitalizations for the U.S. Elderly, Ages 65 and Over**

- Dot-density maps for *Salmonella* Infection for the Winter and Summer periods of 1998 – 2002 superimposed on thematic maps of percent of elderly at the county level (Census 2000)
Watershed Regions

Hydrological Unit Codes 2 represent the largest aggregation of watersheds, hydrologic regions

01: NEW ENGLAND
02: MID ATLANTIC
03: SOUTH ATLANTIC-GULF
04: GREAT LAKES
05: OHIO
06: TENNESSEE
07: UPPER MISSISSIPPI
08: LOWER MISSISSIPPI
09: SOURIS-RED-RAINY
10: MISSOURI
11: ARKANSAS-WHITE-RED
12: TEXAS-GULF
13: RIO GRANDE
14: UPPER COLORADO
15: LOWER COLORADO
16: GREAT BASIN
17: PACIFIC NORTHWEST
18: CALIFORNIA
Watershed Regions

Hydrological Unit Codes 2 represent the largest aggregation of watersheds, hydrologic regions

01: NEW ENGLAND
02: MID ATLANTIC
03: SOUTH ATLANTIC-GULF
04: GREAT LAKES
05: OHIO
06: TENNESSEE
07: UPPER MISSISSIPPI
08: LOWER MISSISSIPPI
09: SOURIS-RED-RAINY
10: MISSOURI
11: ARKANSAS-WHITE-RED
12: TEXAS-GULF
13: RIO GRANDE
14: UPPER COLORADO
15: LOWER COLORADO
16: GREAT BASIN
17: PACIFIC NORTHWEST
18: CALIFORNIA
Use of Water Supply

Water Use Ratio = \( \frac{\text{Total population with self served water supply}}{\text{Total population with public water supply}} \)

**Self-supplied water use** — water withdrawn from a ground-water or surface-water source by a user rather than being obtained from a public supply.

Data Source: USGS, 2000
Köppen Climate Classification

- Based on monthly temperature and total precipitation
- Major Classifications
  - A – Tropical Rain Climates
  - B – Arid Climates
  - C – Temperate Rain Climates
  - D – Boreal Forest and Snow Climates
  - E – Cold Snow Climates
- Original classification is based on data averaged over 30 years
- 23 categories in U.S.
- 16 categories when aggregated by county
**Salmonella Infection in US Elderly**

The annual rate of hospitalization due to *Salmonella* infection has doubled in counties with predominantly self-served water usage and with warm and wet climate compared to cold and dry areas.

Relative Risk, associated with 10°C:

\[ \text{RR} = 1.35 \quad \text{CI}_{95\%} = [1.19, 1.53] \]

(this model includes elderly population, % of elderly, % of nursing home residents; climate and watersheds characteristics)
Summary

- Examined over 3 mln records of GID hospitalizations (ICD-9M: 001-009, 558.9, 787) recorded by the CMS in 1998-2002

- Developed a model to assess effects of temperature and precipitation on GID hospitalizations rate with an adjustment for climate and watershed characteristics

- Developed a system of dynamic maps to demonstrate spatio-temporal patterns in GID
Summary

- **Relevant publications:**
  


  Naumova EN, Egorov AI, Morris RD, Griffiths JK. The elderly are a sensitive subpopulation for waterborne Cryptosporidium infection: hospitalizations for gastroenteritis in the elderly before and during the 1993 Milwaukee cryptosporidiosis outbreak. Emerging Infectious Diseases. 2003. 9(4):418-25.

- **Website:**
  
  http://www.tufts.edu/med/informid/