Doing Environmental Epidemiologic Research with Electronic Health Records

Brian S. Schwartz, MD, MS
January 10, 2013
Overview

1. The future of environmental epidemiology – some thoughts
2. Types of data available from health systems
3. The Geisinger Health System, the Environmental Health Institute (EHI), the HMO Research Network
4. Examples of our ongoing research
   - Use of patient data
   - Use of other secondary data sources for environmental characterization
5. A bit more detail
   - MRSA and animal feeding operations
   - Type 2 diabetes and coal abandoned mine lands
6. Questions
The Future of Environmental Epidemiology

1. A need to evaluate complex causal pathways – physical environment, social, behavioral, genetic, epigenetic, lifespan, critical exposure periods …
2. … and thus need large sample sizes …
3. … will create complex data streams – longitudinal, multilevel, highly multidimensional …
4. … multiple, varied environmental exposures …
5. … of symptoms, signs, biometrics, diseases …
6. … with methodological constraints (e.g., no RDD) …
7. … in a setting of funding constraints …
8. … and thus a need to leverage existing, secondary and new data sources
EHR Data in Epidemiology

- EHR: longitudinal digital record of patient health information generated by clinical encounters in a care delivery setting
- ARRA 2009 committed substantial funds to increase use of EHR and claims data to improve practice & research
- Relatively low cost for large sample size, longitudinal data
- Prospective studies: broader range of clinical outcomes
- Genomic studies: for patient phenotyping
- Surveillance projects: near real-time data
- Comparative effectiveness research: compare clinical interventions in shorter time and with fewer costs than in prospective clinical trials
  - So useful that some conclude that continued CER success is now largely dependent on health information technology
Claims Data in Epidemiology

- Claims data are created by payers from bills generated by providers seeking payment for services rendered
- Sources: private insurers, Medicare, Medicaid, DoD, Dept. of Veterans Affairs
- Access is becoming easier
- Data on all inpatient & outpatient services while enrolled in a health plan, but not the outcome of these services
- Unlike EHR data (data limited to care received at one health system), claims include data on all covered services received regardless of provider
- Common use in pharmacoepidemiology and cost studies
The Environmental Health Institute

• The Geisinger Center for Health Research and the JHBSPH Dept. of Environmental Health Sciences
  – MOU signed Jan 2007, work began Feb 2007
  – Required Business Associate Agreement and Data Use Agreement

• Environmental epidemiologic research
  – Geisinger region is loaded with environmental challenges that offer opportunities for study
  – Proof of concept, then results
The Geisinger Clinic

- 40+ community practice clinics and 4+ hospitals
- 400,000+ primary care patients representative of the general population in the region
  - 2M+ specialty care patients
- EHR since 2001, > 11 years of data
- Across a large, varied geography (40+ counties)
- Patients can use the health system with any health insurance
- Recent partnerships with the Guthrie and Susquehanna Health Systems
- 30% of primary care patients have Geisinger Health Plan insurance – can get claims data
**HMO RN:** a consortium of health care delivery organizations with both defined patient populations and formal, recognized research capabilities

19 members, ~20 million patients
Geisinger EHR – Epic Software

- Primary care and specialty patients
- Inpatient, outpatient, emergency, and telephone encounters
- Socio-demographics, health insurance (surrogate for SES)
- Vital signs, doctor orders, problem list
- Laboratory tests, medications
- Procedures, imaging
  - Results may be in secondary databases
- ICD-9 codes accompany encounters, labs, procedures, medications, and orders
# Overview of Environmental Epidemiologic Studies to Date

<table>
<thead>
<tr>
<th>Health Outcome</th>
<th>MACRO Environmental Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Marcellus shale development</td>
</tr>
<tr>
<td>Asthma</td>
<td>✓**</td>
</tr>
<tr>
<td>Chronic rhinosinusitis (CRS)</td>
<td>✓</td>
</tr>
<tr>
<td>Diabetes</td>
<td>✓</td>
</tr>
<tr>
<td>Methicillin resistant <em>S. aureus</em> (MRSA)</td>
<td></td>
</tr>
<tr>
<td>Obesity</td>
<td>✓</td>
</tr>
</tbody>
</table>

* In relation to AEUs at home farm and crop field application of manure
** Also early efforts on cardiovascular disease, injuries, adverse pregnancy outcomes
Methods Common to All Studies

- Obtain patient data from EHR
- Geocode patients – automated and manual
- Consider how environment contributes to disease burden
  - Define exposure of interest – INDIVIDUAL or CONTEXTUAL measure (if latter, define relevant context)
- Use geographic information systems (GIS) to create exposure metrics
  - Get maps of points, lines, polygons, & metadata, or geocode data as needed
  - Create commonly used metrics: density, diversity, design, accessibility (distance), clustering
- Link exposure and patient measures
- Perform biostatistical analysis – person, place, time
Data Sources

Patient data – EHR
Environmental data – next
New Development
ROAD SEGMENTS: average block size, road mile density, intersection density, connectivity, walkability
Food Establishments, COMMERCIAL DATA: density, diversity, accessibility to different types of food RETAIL and food SERVICE (e.g., fast food density, distance to closest grocery store)

**Food Environment: Food Service & Retail in GHS’s 31 Counties**

Dun and Bradstreet, InfoUSA (Business Analyst): purchase, geocode, create metrics; known problems with the data
Local Physical Activity Opportunity Environment (LPAOE).

Municipal boundaries, GHS’s 31 counties, 4287 LPAOE points.
Productive Natural Gas Wells in the Marcellus Shale

8,233 well permits with 2,257 wells in production
Source: PA DEP, Well Production Report, December 2011

Natural gas yields in productive wells (1000's of cubic feet (Mcf))
- Permitted but not in production
- 1 - 50,000 (644 wells)
- 50,001 - 300,000 (826 wells)
- 300,001 - 3,000,000 (787 wells)

- Geisinger Health System service area
- Guthrie Healthcare System service area
- Susquehanna Health service area

Center for Health Research, Geisinger Health System
Animal feeding operations (antibiotic use in animal feeds) and risk of methicillin-resistant Staphylococcus aureus

(PhD thesis research of Joan Casey)
A New MRSA: Community Associated (CA-)

• Since mid-1990s, large increase in MRSA infections in persons lacking prior contact with the healthcare system
• Shortly after, were recognized to be new MRSA strains
• Were rapidly disseminated among US general population, now affect patients with and without contact with healthcare system
• These new strains cause different clinical syndromes, particularly skin and soft tissue infections (SSTIs)
  – Incidence of SSTIs in US has been increasing
• These new strains now account for the majority of MRSA infections
• Large reservoirs of MRSA isolates now exist outside healthcare facilities
Identification of MRSA Cases Using Electronic Health Records

446,480 GHS\(^a\) primary care patients, 2001-2010

G1\(^b\)

Positive culture for MRSA or positive culture for \(S.\)\(\text{aureus}\) with oxacillin resistance

N = 3145

Confirmed not nasal colonization\(^c\)

12 PL only

Specific MRSA Code

N = 754

<table>
<thead>
<tr>
<th>ICD-9 codes</th>
<th>Description</th>
<th>n(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V12.04</td>
<td>History of MRSA</td>
<td>46</td>
</tr>
<tr>
<td>041.12</td>
<td>MRSA infection</td>
<td>125</td>
</tr>
<tr>
<td>482.42</td>
<td>MRSA pneumonia</td>
<td>11</td>
</tr>
<tr>
<td>038.12</td>
<td>MRSA septicemia</td>
<td>4</td>
</tr>
</tbody>
</table>

EMR code

EP884: MRSA

N = 571

G2

From ICD-9 or EP code

G3

ICD-9 code for \(S.\)\(\text{aureus}\) with V09.0 resistance code\(^e\)

N = 195

<table>
<thead>
<tr>
<th>ICD-9 codes</th>
<th>Description</th>
<th>n(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>041.10</td>
<td>Staph infection(^f)</td>
<td>93</td>
</tr>
<tr>
<td>041.11</td>
<td>(S.)(\text{aureus}) infection</td>
<td>153</td>
</tr>
<tr>
<td>482.41</td>
<td>(S.)(\text{aureus}) pneumonia</td>
<td>18</td>
</tr>
<tr>
<td>038.10</td>
<td>Staph septicemia(^f)</td>
<td>2</td>
</tr>
<tr>
<td>038.11</td>
<td>(S.)(\text{aureus}) septicemia</td>
<td>21</td>
</tr>
</tbody>
</table>

18 PL only

73 history codes; V12.04\(^#\)

757 Diagnosed prior to 2005\(^h\)

CA-MRSA criteria applied\(^i\)

HA

N = 1521

CA

N = 1713
Criteria for Identification of HA- and CA-MRSA Cases

Figure 1

446,480 GHS primary care patients, 2001-2010

From laboratory table

G1

Positive culture for MRSA or positive culture for S. aureus with oxacillin resistance N = 3145

G2

Specific MRSA Code N = 754

ICD-9 codes Description

041.12 MRSA infection
038.12 MRSA septicemia
482.42 MRSA pneumonia

G3

ICD-9 code for S. aureus with VU9.0 resistance code
N = 195

ICD-9 codes Description

041.11 S. aureus infection
038.11 S. aureus septicemia
482.41 S. aureus pneumonia
V09.0 Infection with microorganisms resistant to penicillin

CA-MRSA disqualification pathway

MRSA culture or diagnosis for an inpatient ≥ 3 days after admission

G1 = 169
G2 = 9
G3 = 8

Hospitalization in year prior to MRSA culture or diagnosis

G4 = 1008
G5 = 218
G6 = 173

Dialysis in year prior to MRSA culture or diagnosis

G7 = 50
G8 = 7
G9 = 11

Surgery in year prior to MRSA culture or diagnosis

G10 = 700
G11 = 120
G12 = 91

Residence in a nursing home in year prior to MRSA culture or diagnosis

G13 = 135
G14 = 3
G15 = 3

Indwelling catheter or subcutaneous device at time of MRSA culture or diagnosis

G16 = 151
G17 = 16
G18 = 17

Lab Culture HA n = 1099
Lab Culture CA n = 1276

Specific ICD-9 HA n = 277
Specific ICD-9 CA n = 419
ICD-9 with V09.0 HA n = 150
ICD-9 with V09.0 CA n = 18

HA, N = 1521; CA, N = 1713
Annual Incidence of HA and CA-MRSA and SSTIs, Geisinger Health System, 2001-2009

<table>
<thead>
<tr>
<th>Year</th>
<th>Combined MRSA</th>
<th>CA-MRSA</th>
<th>HA-MRSA</th>
<th>SSTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>3404</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2002</td>
<td>3576</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2003</td>
<td>3677</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2004</td>
<td>3934</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2005</td>
<td>4020</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2006</td>
<td>4091</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2007</td>
<td>4416</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2008</td>
<td>4434</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
<tr>
<td>2009</td>
<td>4523</td>
<td>0</td>
<td>0</td>
<td>4523</td>
</tr>
</tbody>
</table>
Obtained Nutrient Management Plans for CAFOs, CAOs, VAOs, for swine & dairy/veal operations
\[ n = 147 \text{ home} + 271 \text{ importing fields with address; geocoded} \rightarrow \text{circular buffer} \]

\[ n = 180 \text{ home operations with aerial photos; Google Earth} \rightarrow \text{Shapefile} \rightarrow \text{ArcMap} \]

\[ n = 131 \text{ importing fields, only township; randomly select point on appropriate land use type} \rightarrow \text{circular buffer} \]

Three methods for crop fields
Associations of **Seasonal Crop Field Manure Exposure** with HA-MRSA, CA-MRSA and SSTI (full multilevel model\(^a\))

<table>
<thead>
<tr>
<th></th>
<th>HA-MRSA Adjusted OR (95% CI)</th>
<th>CA-MRSA Adjusted OR (95% CI)</th>
<th>SSTI Adjusted OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Swine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Q2</td>
<td>1.19 (0.97-1.46)</td>
<td>1.08 (0.89-1.31)</td>
<td>1.03 (0.88-1.20)</td>
</tr>
<tr>
<td>Q3</td>
<td>1.26 (1.03-1.55)</td>
<td>1.25 (1.04-1.52)</td>
<td>1.22 (1.05-1.41)</td>
</tr>
<tr>
<td>Q4</td>
<td>1.29 (1.04-1.60)</td>
<td>1.38 (1.13-1.68)</td>
<td>1.37 (1.18-1.60)</td>
</tr>
<tr>
<td>(p)</td>
<td>0.01</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><strong>Dairy/veal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Q2</td>
<td>0.83 (0.68-1.03)</td>
<td>0.97 (0.80-1.18)</td>
<td>0.91 (0.78-1.06)</td>
</tr>
<tr>
<td>Q3</td>
<td>0.93 (0.76-1.13)</td>
<td>0.91 (0.75-1.10)</td>
<td>0.85 (0.73-0.99)</td>
</tr>
<tr>
<td>Q4</td>
<td>0.77 (0.62-0.97)</td>
<td>1.25 (1.02-1.53)</td>
<td>1.02 (0.87-1.19)</td>
</tr>
<tr>
<td>(p)</td>
<td>0.06</td>
<td>0.06</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Abbreviations: CA-MRSA, community-associated methicillin-resistant *S. aureus*; HA-MRSA, healthcare-associated MRSA; OR, odds ratio; SSTI, skin and soft tissue infection

\(^a\) For multinomial models \(n = 5788\) and for binomial model \(n = 5809\), all models control for sex, age, race/ethnicity, ever-smoking status, antibiotic prescription in the prior two years, residential minor civil division, and community socioeconomic deprivation

\(^b\) P value for linear trend
Coal abandoned mine lands (chronic environmental contamination) and risk and severity of type 2 diabetes mellitus

(PhD thesis research of Ann Liu)
Community Stress, Psychosocial Hazards, and EPA Decision-Making in Communities Impacted by Chronic Technological Disasters

Psychosocial stress has emerged as an important
Stephen R. Couch, PhD, and Charlton J. Coles, PhD

Critical Biological Pathways for Chronic Psychosocial Stress and Research Opportunities to Advance the Consideration of Stress in Chemical Risk Assessment

Emerging evidence suggests that psychosocial stress
Bruce S. McEwen, PhD, and Pamela Tucker, MD

Cumulative Risk Assessment for Combined Health Effects From Chemical and Nonchemical Stressors

Cumulative risk assessment
Ken Sexton, ScD, and Stephen H. Linder, PhD
What are the contextual effects of living in this community?
Abandoned Coal Mine Lands (AML) in Pennsylvania

Acid mine drainage

Waste pile

Subsidence

Abandoned structure

Photos from PA DEP
Selecting the HbA1c Outcomes of Interest

Healthy → Pre-diabetes → Diabetes → Complications

- 100 mg/dL ≤ FBS ≤ 125 mg/dL
- HbA1c (screening)
- ICD-9 code
- HbA1c (monitoring)
- Rx

- Mean duration = 159 days
- Mean duration = 117 days
- Mean duration = 1534 days

HbA1c (pre-therapeutic)
- n = 7337
- Mean = 7.51%

1st ICD-9 diabetes code
- n = 17,959
- Mean = 7.64%

HbA1c (post-ICD-9)

HbA1c (last-ever)

Mean duration = 1732 days
Do community conditions constrain the health care system?
• Patients are on their own > 99% of the time
• Community factors influence diet, activity, & stress
• The best health care may have little impact on patient outcomes
  – The wrong behaviors are enabled …
  – The right behaviors are constrained …
  – … by community conditions

Joe may only be able to get so far in managing his diabetes even with the help of four specialists, a dietician and trainer.
DOCTOR: “Joe, I want you to eat healthier foods.”
JOE: “Ok Doctor Smith.”

Photos courtesy of J. Feng
EHR Challenges

• Patient must seek care
• Cannot exactly determine if patient is under observation and is well or has left care
• Persons enter and exit cohort at any time
• As a secondary data source, data are not perfect
  – ICD-9 coding has known problems; some data in text
• Many variables desired for analysis are not available (e.g., certain test results, income)
• Cannot link patients (i.e., mother-child, siblings)
• No measures of environmental exposures
• Large learning curve for use; much processing
Thank you for listening