Data Linkages to Improve Health Outcomes

An End User Perspective

Deborah Schrag MD
Department of Epidemiology and Biostatistics
Department of Medicine
Memorial Sloan-Kettering Cancer Center
Overview

• Types of research questions
• Examples of linkage attempts
• Challenges encountered
• Wish list
Health Services Research

- Examines relationships between need, demand, supply, delivery and outcomes of health care:
  - Disparities in care
  - Access/barriers to care
  - Technology dissemination
  - Quality measurement
  - Efficiency of care delivery
Layers of Data

- Facilities
- Providers
- Diseased Populations
- Source Populations
Evaluating the Quality of Health Care

- Community Attributes
- Health Risks and Behaviors
- Structure of Delivery Systems
- Process of Care
- Person Attributes
- Health Outcomes

Community Attributes

Health Risks and Behaviors

Structure of Delivery Systems

Process of Care

Person Attributes

Health Outcomes
The Spectrum of Sources for Clinical Research Data

- **Population-Based data**
  - All patients in NY State with lung cancer
  - Data Source: Cancer Registry and Census Data

- **Quasi-population-based data**
  - All patients in NY State with lung cancer covered by Oxford
  - Data Source: Medical records and claims for care from Oxford

- **Non-population-based data**
  - All NY State patients evaluated at MSKCC with lung cancer

- **Health Services Research Strategy**:
  - Simultaneous use of various data sources and juxtaposition of analyses can reveal opportunities for improving health care delivery
Implementation Gap

- Efficacy – Effectiveness = Implementation Gap
- Need to understand reasons for gaps
- Identify important and remediable sources of variation

  • Endogenous to patients
  • Endogenous to MDs
  • Endogenous to health care system
Adjuvant Chemotherapy Use By Age
Stage III Colon Cancer in SEER-Medicare

<table>
<thead>
<tr>
<th>Age</th>
<th>% Treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>65-69</td>
<td>78</td>
</tr>
<tr>
<td>70-74</td>
<td>73</td>
</tr>
<tr>
<td>75-79</td>
<td>58</td>
</tr>
<tr>
<td>80-84</td>
<td>34</td>
</tr>
<tr>
<td>85-89</td>
<td>11</td>
</tr>
</tbody>
</table>

P (adj) < .001
Stage III Adjuvant Chemotherapy
By Marital Status

(p adj <0.001 HR for single vs. married 0.62, 95% CI 0.54-0.72)
Linking to Providers

• Why don’t all patients get chemotherapy?
• Do they refuse?
• Do they see a medical oncologist post-operatively?
• UPINs on CMS claims can be linked to CMS specialty codes but the data are incomplete
Most patients who forego adjuvant therapy make treatment decisions without oncology consultations. Provider specialty cannot be precisely specified—some oncologists are internists.
Wish List: Provider Characteristics

• Linkage of UPINs on claims data to files describing physician characteristics
• AMA data is better than CMS data
• ABIM/ACS is better than AMA
• State-level data is most complete and most difficult to obtain
Pharmacy Claims

- Oral chemotherapy?
- Anti nausea medications?
- Adherence to therapy
- Pain control

Wish list:
- Part D data
- Medicaid data
- Private claims data sets
Taxonomy of Data Custody Types

- Custom Data
- Federal Data
- "Private" Data with public relevance
- State Level Data

"Private" Data with public relevance

- Federal Data
- Custom Data

State Level Data

- Federal Data
- Custom Data
Capacity for Mammography in the US

- US women age 40-80 need mammograms
- Many women unscreened
- Large racial disparities
- Lack of facilities and radiologists are potential reasons for sub-optimal use

- Does lack of capacity explain geographic variation in use? Racial disparities?

- Does capacity for mammography predict breast cancer incidence and mortality?

- Geocoding
Data Sources

• Where are the facilities? FDA accreditation data
• Where are the radiologists? AMA/States
• Where are women unscreened? BRFSS, Medicare
• Where are the high rates of breast cancer? SEER

• Data desired at census tract level
• Where to start to obtain permissions?
• Approval from one agency or many?

• Central clearinghouse, clearly delineated procedures would help
Area versus Person-Level Data

• Access to granular area level data helps most health services researchers

• Privacy/security concerns involve less risk

• Enables researchers to extract greater information from their own person-level data sets
Layers of Data Access

- Individual Patient data
- Anonymized patient data linked to unit area
- Area level data: State, County, Zip code, Census tract
Repetitive Common Tasks

• Geographic variation:
• Where are?

• Patients
• Providers/services
• Disparities
• Mortality rates
Wish List

• Access to chloropleth maps
  • By county, zip code, census tract
  • Useful for common data elements census/survey data results
• Shared resource for investigators
• ArcGIS software
Fragmentation of Care

- Do patients with chronic conditions in NY State where there are many hospitals consolidate their care or is it fragmented across multiple institutions?

- Is fragmentation higher in the Medicaid program?

- SPARCS: statewide discharge database
  - Available from states
  - Some states have data available from NCHS

- State discharge data not linked to patient residence/census data

- Medicaid enrollment data
Medicaid

• Largest component of state budgets

• Health care for poorest, often sickest members of society

• Untapped resource because of complexity of data structure, organization and access, completeness

• Enrollment versus process

The perfect as the enemy of the good
### 1998 Incident Primary Cancers Reported to CCR and % in Medi-Cal

<table>
<thead>
<tr>
<th>Tumor Site</th>
<th>Incident Cancer Cases from the California Cancer Registry</th>
<th>Cases in CCR and Medi-Cal</th>
<th>As % of All Cases, Age 18-64</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All 1998 Cases</td>
<td>% age 18-64 at Diagnosis</td>
<td>1998 Cases, Age 18-64</td>
</tr>
<tr>
<td>Breast</td>
<td>20,864</td>
<td>54.6%</td>
<td>11392</td>
</tr>
<tr>
<td>Lung</td>
<td>17,004</td>
<td>30.6%</td>
<td>5203</td>
</tr>
<tr>
<td>Colorectal</td>
<td>10,254</td>
<td>27.3%</td>
<td>2796</td>
</tr>
<tr>
<td>Cervical</td>
<td>1,690</td>
<td>79.9%</td>
<td>1350</td>
</tr>
<tr>
<td>Prostate</td>
<td>19,001</td>
<td>29.5%</td>
<td>5605</td>
</tr>
<tr>
<td>Testis</td>
<td>937</td>
<td>93.6%</td>
<td>877</td>
</tr>
<tr>
<td>Uterine</td>
<td>3,587</td>
<td>48%</td>
<td>1722</td>
</tr>
<tr>
<td>Bladder</td>
<td>5,452</td>
<td>26.5%</td>
<td>1445</td>
</tr>
<tr>
<td>Hepatoma</td>
<td>1,570</td>
<td>53.3%</td>
<td>836</td>
</tr>
<tr>
<td>Stomach</td>
<td>2,552</td>
<td>32.2%</td>
<td>822</td>
</tr>
</tbody>
</table>
## Duration and Timing of Medi-Cal Enrollment in Relation to Cancer Diagnosis

<table>
<thead>
<tr>
<th>Duration of Medicaid Enrollment over 24-month interval</th>
<th># of patients</th>
<th>% of cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire 24 months</td>
<td>3443</td>
<td>47%</td>
</tr>
<tr>
<td>13-23 months</td>
<td>1671</td>
<td>23%</td>
</tr>
<tr>
<td>7-12 months</td>
<td>1540</td>
<td>22%</td>
</tr>
<tr>
<td>1-6 months</td>
<td>610</td>
<td>8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medicaid Enrollment Status at month of diagnosis</th>
<th># of patients</th>
<th>% of cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES, Enrolled during month of diagnosis</td>
<td>5364</td>
<td>74%</td>
</tr>
<tr>
<td>NO, First enrolled after month of diagnosis*</td>
<td>1698</td>
<td>23%</td>
</tr>
<tr>
<td>NO, Enrolled prior to diagnosis but not during month of diagnosis</td>
<td>202</td>
<td>3%</td>
</tr>
</tbody>
</table>

*Of these, 905/1698=53% enrolled within 3 months of diagnosis.
## How often do California Medicaid claims corroborate cancer site specific diagnoses reported to the CCR?

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th># of CCR-Medicaid enrollees with 1998 cancer diagnoses</th>
<th># (%) of Patients diagnosed at anytime during 1998 with a corroborating diagnosis code recorded in 1998 Medicaid claims files</th>
<th># of Patients Enrolled in Medicaid in 1998 with CCR cancer diagnosis in the first 6 months of 1998</th>
<th># (%) of Patients from preceding column who have a cancer diagnosis code recorded in 1998 (entire year) Medicaid claims files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>1014</td>
<td>770 (76%)</td>
<td>549</td>
<td>436 (80%)</td>
</tr>
<tr>
<td>Lung</td>
<td>997</td>
<td>751 (75%)</td>
<td>523</td>
<td>416 (80%)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>557</td>
<td>394 (71%)</td>
<td>274</td>
<td>207 (76%)</td>
</tr>
<tr>
<td>Cervical</td>
<td>287</td>
<td>207 (72%)</td>
<td>146</td>
<td>117 (80%)</td>
</tr>
<tr>
<td>Prostate</td>
<td>268</td>
<td>176 (66%)</td>
<td>127</td>
<td>93 (73%)</td>
</tr>
<tr>
<td>Testis</td>
<td>78</td>
<td>60 (77%)</td>
<td>38</td>
<td>28 (74%)</td>
</tr>
<tr>
<td>Uterine</td>
<td>176</td>
<td>124 (70%)</td>
<td>86</td>
<td>64 (74%)</td>
</tr>
<tr>
<td>Bladder</td>
<td>113</td>
<td>86 (76%)</td>
<td>53</td>
<td>43 (81%)</td>
</tr>
<tr>
<td>Liver</td>
<td>288</td>
<td>180 (63%)</td>
<td>151</td>
<td>99 (66%)</td>
</tr>
</tbody>
</table>
Medicaid Data

• SEER-Medicaid Data
  • Attempted link in California
  • 2 years to obtain data sets
  • Denominator file structure limits ability to identify cohorts of the chronically poor

• Challenges:
  • Retroactive enrollment
  • Chronic vs. episodic poverty
  • Spend downs—illness precipitates enrollment
  • Variation in states thresholds/generosity
  • Definition of an HMO
Wish List: Medicaid Data

• Consistent definitions in Medicaid enrollment files
  • What does managed care mean?
  • When are claims itemized?
• Linkages of Medicaid data files to state discharge abstracts
• Geocoding of where Medicaid beneficiaries reside
• Linkage to pharmacy data
• Linkage to census tract socioeconomic variables
Priorities

- Coordination of procedures for obtaining access to data and the review process
- Standardization of reporting rules (e.g. N must not be less than 10)
- Develop categorization schema for types of linkages
- Central clearinghouse/index describing linkages that exist as well as those that are possible
- Facilitate federation of state data
- Chloropleth maps for use in commons based systems
- Work with states to facilitate analyses of Medicaid enrollment and claims files