Administrative Data: A Valuable Tool for Modeling Health Service Utilization and Outcomes for Chronic Disease

Lisa M. Lix, PhD P.Stat.
University of Saskatchewan

Workshop on Dynamic Modeling for Health Policy

July 24, 2009
Outline

- Background
- Administrative data and data repositories
- Strengths and limitations
- Identifying obesity cases
- Obesity-related chronic disease
  - Identifying disease cases
  - Challenges
- Data linkage
- Further research opportunities
- Conclusions
Acknowledgements

- This research is funded by the Canadian Institutes of Health Research, the Lupina Foundation, and the Manitoba Health Research Council.
Background

✧ Population-based administrative data have multiple uses in the study of obesity and obesity-related chronic diseases

• Monitor demographic, socioeconomic, and temporal variations in prevalence and incidence
• Detect geographic clusters
• Compare health service use and costs for cases and controls
• Investigate quality and/or processes of care for cases and controls
A Definition of Administrative Data

✧ Data collected for purposes of health system monitoring, financial management and provider remuneration

✧ Not originally intended for research
Administrative Data Repositories

Core components of a typical repository

- Registry – population identification, demographic characteristics
- Hospital database – abstracts (summaries) of hospital events, including diagnoses and procedures
- Physician database – billing claims used for fee-for-service remuneration – include service information and diagnosis(es)
- Vital statistics database – deaths, births
Based on the figure by Roos, Menec, & Currie (2004)
Strengths of Administrative Data

✧ Population-based
✧ Longitudinal histories can be constructed
✧ Relatively inexpensive compared to primary data
✧ Address multiple policy-relevant questions
  • Health disparities
  • Equity of access
  • Quality of care
  • Processes and outcomes of care
Limitations of Administrative Data

- Do not contain information about body composition (e.g., height, weight) or risk factors (physical activity; food consumption)

- Variations in data quality over time, across geographic areas, and across databases

- Changes in the way data are captured/recorded
  - WHO’s International Classification of Diseases: Change from ICD-9 to ICD-10
Identifying Obesity Cases: Diagnoses

✧ ICD-9-CM

- 278.00: Obesity, unspecified
- 278.01: Morbid obesity
- V85: Body Mass Index (BMI), Kilograms per meters squared
  - V85.4x: Body Mass Index 40 and over, adult

✧ ICD-10-CA

- E66.0: Obesity due to excess calories
- E66.1: Drug-induced obesity
- E66.2: Extreme obesity with alveolar hypoventilation
- E66.8: Other obesity
- E66.9: Obesity, unspecified
Validity of Obesity Diagnoses


- US study comparing BMI calculated from medical record with ICD-9-CM obesity diagnoses in hospital data
- Children/youth aged 2-20 years
- Under-reporting of obesity diagnoses in hospital data:
  - sensitivity: 8.0%
  - specificity: 99.8%
Identifying Obesity Cases: Procedures

✧ **ICD-9-CM**
  - 44.38: Laparoscopic gastroenterostomy
  - 44.39: Other gastroenterostomy
  - 44.95: Laparoscopic gastric restrictive procedure

✧ **ICD-10-CA: Canadian Classification of Interventions (CCI)**
  - 1.NF.78^^: gastric banding
## Surgical Procedures

<table>
<thead>
<tr>
<th></th>
<th>Number of Procedures</th>
<th>Mean Age (Years)</th>
<th>% Female</th>
<th>Mean LOS (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>British Columbia</td>
<td>109</td>
<td>41</td>
<td>88</td>
<td>3.0</td>
</tr>
<tr>
<td>Alberta</td>
<td>224</td>
<td>37</td>
<td>91</td>
<td>5.5</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>41</td>
<td>39</td>
<td>83</td>
<td>10.3</td>
</tr>
<tr>
<td>Ontario</td>
<td>303</td>
<td>40</td>
<td>84</td>
<td>4.9</td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>47</td>
<td>39</td>
<td>85</td>
<td>6.4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>724</strong></td>
<td><strong>39</strong></td>
<td><strong>87</strong></td>
<td><strong>5.2</strong></td>
</tr>
</tbody>
</table>

*Excludes Manitoba and Quebec because of differences in data collection methodology and provinces/territories where fewer than five procedures were performed annually.

Source: Discharge Abstract Database, CIHI.

Obesity-Related Chronic Diseases

- Hypertension
- Diabetes
- Coronary artery disease
- Other cardiovascular disease
- Hypercholesterolemia
- Osteoarthritis
- Stroke
- Cancers: colon, breast, endometrial, kidney, esophageal
Identifying Disease Cases

- Constructing a case definition
- Validating a case definition
- Comparing case definitions over time and across jurisdictions
Constructing a Case Definition

✦ Elements

- Type of data source
- Number of years of data
- Diagnosis and/or prescription codes
- Number of contacts with health system
- Accruing cases over time
Osteoarthritis, Crude Prevalence Estimates, Manitoba, 2004/05 – 2005/06
Assumption Underlying Case Accrual

Disease Absent → Disease Present

Dead
Diabetes, Illustration of Case Accrual, Yukon Territory
Validating a Case Definition

✧ Gold Standard Exists
  • Estimate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV)

✧ Gold Standard Does Not Exist
  • Biased gold standard
  • Capture-recapture methods
Estimating Sensitivity, Specificity, PPV, NPV

<table>
<thead>
<tr>
<th>Admin Data</th>
<th>Has Disease</th>
<th>Does Not Have Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has Disease</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Does Not Have Disease</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

Sensitivity = \( \frac{A}{A+C} \times 100 \)

Specificity = \( \frac{D}{B+D} \times 100 \)

PPV = \( \frac{A}{A+B} \times 100 \)

NPV = \( \frac{C}{C+D} \times 100 \)
Validating Osteoporosis Case Definitions, Manitoba Bone Mineral Density Tests as the Gold Standard

<table>
<thead>
<tr>
<th># years</th>
<th>Algorithm</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>YI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1+ P</td>
<td>69.4</td>
<td>92.7</td>
<td>91.7</td>
<td>72.2</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>2+ P</td>
<td>34.1</td>
<td>99.1</td>
<td>97.8</td>
<td>56.3</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>1+ Rx</td>
<td>78.5</td>
<td>90.1</td>
<td>90.3</td>
<td>78.2</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>1+ H or 1+ P</td>
<td>69.4</td>
<td>92.7</td>
<td>91.7</td>
<td>72.2</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>1+ H or 2+ P</td>
<td>34.1</td>
<td>99.1</td>
<td>97.8</td>
<td>56.3</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>1+ H or 1+ P or 1+ Rx</td>
<td>89.4</td>
<td>86.1</td>
<td>88.3</td>
<td>87.5</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>1+ H or 2+ P or 2+ Rx</td>
<td>77.7</td>
<td>91.8</td>
<td>91.7</td>
<td>77.9</td>
<td>0.70</td>
</tr>
<tr>
<td>2</td>
<td>1+ P</td>
<td>74.0</td>
<td>90.0</td>
<td>89.6</td>
<td>74.8</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>2+ P</td>
<td>43.9</td>
<td>96.9</td>
<td>94.3</td>
<td>59.7</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1+ Rx</td>
<td>82.2</td>
<td>88.5</td>
<td>89.3</td>
<td>81.0</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>1+ H or 1+ P</td>
<td>74.0</td>
<td>90.0</td>
<td>89.6</td>
<td>74.8</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>1+ H or 2+ P</td>
<td>43.9</td>
<td>96.9</td>
<td>94.3</td>
<td>59.7</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>1+ H or 1+ P or 1+ Rx</td>
<td>92.0</td>
<td>83.1</td>
<td>86.4</td>
<td>89.9</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>1+ H or 2+ P or 2+ Rx</td>
<td>83.4</td>
<td>89.5</td>
<td>90.3</td>
<td>82.2</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Note: P = physician billing claims; H = hospital separation abstracts; Rx = prescription drug data; YI = Youden’s index
Gold Standard Does Not Exist

- Biased gold standard
  - Adjust the estimates of sensitivity, specificity, PPV, NPV based on the estimated amount of bias in the gold standard.
  - Average the estimates of sensitivity, specificity, PPV, NPV across several biased gold standards to arrive at a ‘true” estimate.
Validating Osteoarthritis Case Definitions Using the Canadian Community Health Survey, Cycle 3.1

<table>
<thead>
<tr>
<th># Years</th>
<th>Algorithm</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1+ P</td>
<td>27.5</td>
<td>96.8</td>
<td>48.9</td>
<td>92.4</td>
</tr>
<tr>
<td>2</td>
<td>2+ P</td>
<td>16.4</td>
<td>98.9</td>
<td>61.6</td>
<td>91.5</td>
</tr>
<tr>
<td>3</td>
<td>1+ H or 2+ P</td>
<td>16.8</td>
<td>98.8</td>
<td>60.6</td>
<td>91.5</td>
</tr>
<tr>
<td>4</td>
<td>1+ H or 2+ P or (1 P &amp; 2+ Rx)</td>
<td>22.9</td>
<td>98.0</td>
<td>56.3</td>
<td>92.1</td>
</tr>
<tr>
<td>2</td>
<td>1+ P</td>
<td>35.4</td>
<td>94.8</td>
<td>42.9</td>
<td>93.0</td>
</tr>
<tr>
<td>6</td>
<td>2+ P</td>
<td>23.8</td>
<td>97.9</td>
<td>55.4</td>
<td>92.1</td>
</tr>
<tr>
<td>7</td>
<td>1+ H or 2+ P</td>
<td>23.9</td>
<td>97.8</td>
<td>54.3</td>
<td>92.1</td>
</tr>
<tr>
<td>8</td>
<td>1+ H or 2+ P or (1 P &amp; 2+ Rx)</td>
<td>31.2</td>
<td>96.6</td>
<td>50.0</td>
<td>92.7</td>
</tr>
</tbody>
</table>

Note: P = physician billing claims; H = hospital separation abstracts; Rx = prescription drug data
### Validating Hypertension Case Definitions Using the Canadian Community Health Survey, Cycle 3.1

<table>
<thead>
<tr>
<th># Years</th>
<th>Algorithm</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1+P</td>
<td>56.7</td>
<td>96.2</td>
<td>81.8</td>
<td>88.2</td>
</tr>
<tr>
<td></td>
<td>2+P</td>
<td>40.9</td>
<td>98.6</td>
<td>89.6</td>
<td>84.9</td>
</tr>
<tr>
<td></td>
<td>1+H or 1+ P</td>
<td>57.8</td>
<td>96.2</td>
<td>81.7</td>
<td>88.4</td>
</tr>
<tr>
<td></td>
<td>1+ H or 2+ P</td>
<td>42.5</td>
<td>98.5</td>
<td>89.3</td>
<td>85.2</td>
</tr>
<tr>
<td></td>
<td>1+ H or 1+ P or 1+ Rx</td>
<td>78.5</td>
<td>92.4</td>
<td>75.5</td>
<td>93.5</td>
</tr>
<tr>
<td>2</td>
<td>1+P</td>
<td>68.6</td>
<td>93.0</td>
<td>74.5</td>
<td>90.9</td>
</tr>
<tr>
<td></td>
<td>2+P</td>
<td>55.1</td>
<td>97.0</td>
<td>84.4</td>
<td>87.9</td>
</tr>
<tr>
<td></td>
<td>1+H or 1+ P</td>
<td>69.9</td>
<td>92.8</td>
<td>74.1</td>
<td>91.2</td>
</tr>
<tr>
<td></td>
<td>1+ H or 2+ P</td>
<td>56.7</td>
<td>96.7</td>
<td>83.5</td>
<td>88.2</td>
</tr>
<tr>
<td></td>
<td>1+ H or 1+ P or 1+ Rx</td>
<td>81.4</td>
<td>89.1</td>
<td>68.9</td>
<td>94.1</td>
</tr>
</tbody>
</table>

Note: P = physician billing claims; H = hospital separation abstracts; Rx = prescription drug data
Gold Standard Does Not Exist

✧ Capture-recapture methods
  • Regression models or non-parametric estimation methods used to estimate the size of the population by estimating the number of cases captured in individual data sources and in the overlap of two or more data sources
  • Assumptions:
    • High-quality data linkage
    • Homogeneity of capture across data sources
    • Independence of data sources
### Osteoporosis Prevalence Estimates, Capture-Recapture Models for Physician and Prescription Data, Manitoba

<table>
<thead>
<tr>
<th></th>
<th>1999/00</th>
<th>2000/01</th>
<th>2001/02</th>
<th>2002/03</th>
<th>2003/04</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14765</td>
<td>17599</td>
<td>20018</td>
<td>22799</td>
<td>24727</td>
</tr>
<tr>
<td>N (95% CI)</td>
<td>(14608, 14923)</td>
<td>(17408, 17785)</td>
<td>(19821, 20224)</td>
<td>(22580, 23007)</td>
<td>(24506, 24962)</td>
</tr>
<tr>
<td>Capture Rate (%)</td>
<td>71</td>
<td>72</td>
<td>74</td>
<td>75</td>
<td>76</td>
</tr>
</tbody>
</table>

### Non-Parametric Method

<table>
<thead>
<tr>
<th></th>
<th>1999/00</th>
<th>2000/01</th>
<th>2001/02</th>
<th>2002/03</th>
<th>2003/04</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14957</td>
<td>17576</td>
<td>19908</td>
<td>22485</td>
<td>24516</td>
</tr>
<tr>
<td>N (95% CI)</td>
<td>(14665, 15250)</td>
<td>(17273, 17878)</td>
<td>(19598, 20218)</td>
<td>(22177, 22793)</td>
<td>(24199, 24833)</td>
</tr>
<tr>
<td>Capture Rate (%)</td>
<td>71</td>
<td>72</td>
<td>74</td>
<td>76</td>
<td>77</td>
</tr>
</tbody>
</table>
Data Linkage: Administrative and Survey Data

✧ Behavioral risk factors
  • Physical activity
  • Food consumption
✧ Socio-demographic and environmental determinants
✧ Quality of life
Data Linkage: Survey and Administrative Data

✧ Potential barriers
  • Lack of a unique personal identifier
  • Data privacy legislation
  • Permission for linkage

✧ Survey availability: one time versus ongoing collection
Further Research Opportunities

❖ New data fields in administrative data
  • Height, weight
❖ Data quality and completeness
  • Develop methods
  • Feedback to data custodians
  • Training in data collection techniques
❖ Expand data repositories
  • Public health services
    • Contacts with dieticians, psychologists
    • Physical environment: Availability of green spaces, crime
❖ New classification systems
  • Symptoms, not just diagnoses
Conclusions

✧ Administrative data have limited value for estimating prevalence/incidence of obesity

✧ Obesity-related surgical procedures are few in number and may not be consistently captured in all jurisdictions

✧ Administrative data have significant value for population-based research about obesity-related chronic diseases
**Conclusions**

✧ The use of administrative data in obesity-related chronic disease research requires careful consideration of a valid methodology to construct case definitions.

✧ Data quality issues can create significant inefficiencies in the use of administrative data for research.

✧ Existing data repositories require expansion to include new data fields and data sets that will help to address questions of current interest in the policy environment.
Conclusions

✧ State-of-the art developments in the use of administrative data and data repositories for obesity research require a team approach: clinicians, epidemiologists, statisticians, computer scientists, social scientists, health services researchers, public health professionals, policy analysts.