GRANT PROPOSALS USING ADMINISTRATIVE/CLAIMS DATA

Margaret Olsen, PhD, MPH
Division of Infectious Diseases
Feb. 15, 2012
DATA SOURCES

• HCUP data
  • Smaller, start-up grants, preliminary data
• Medicare, Medicaid claims
• SEER-Medicare – cancer-related
• MarketScan
  • large employer-based, facility and provider information, outpatient pharmacy claims
• Premier Perspective
  • detailed facility-level information, including daily drug-device utilization (by date, not by time)
• Private Insurer Data
  • UHC, BC/BS
• Others
FIRST STEPS

• Study Section
  • talk to investigators with grants using admin/claims data
  • identify study sections that understand these data

• Preliminary data
  • does not have to involve same data source(s) as proposed in grant
  • can use institutional data
  • must show understanding, ability to use ICD9 and/or CPT codes

• Background
  • discuss pros/cons of using admin/claims data realistically
  • reference papers in literature that reported validation of important ICD9 diagnosis and/or procedure codes relevant to your proposed work and discuss
  • supplement with your previous work, if possible
### VALIDATION OF ICD-9-CM DIAGNOSIS CODES FOR SURGICAL SITE INFECTION

<table>
<thead>
<tr>
<th>Indicators of SSI</th>
<th>SSI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>ICD-9-CM Algorithm</td>
<td>55</td>
</tr>
<tr>
<td>Yes</td>
<td>19</td>
</tr>
<tr>
<td>No</td>
<td>74.3% (55/74)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>82.4% (61/74)</td>
</tr>
<tr>
<td>Wound Culture +</td>
<td>61</td>
</tr>
<tr>
<td>Yes</td>
<td>13</td>
</tr>
<tr>
<td>No</td>
<td>98.4% (61/62)</td>
</tr>
</tbody>
</table>
JUSTIFYING USE OF CLAIMS DATA

• Need to make the case that you can do what you propose given the limitations of claims data
  • Dataset is bigger and more generalizable than institutional or even multi-center study
  • Claims from across spectrum of care (it that is the case with the data you propose to use)
  • Variation in practice

• Also need to make the case that you understand the limitations of claims data and:
  • they are not that important – because of under-coding the estimates you obtain will be minimum estimates or more conservative estimates
    • SSI – the infections missed with claims data algorithms likely minor infections treated with antibiotics only. These infections also more likely to be missed by routine surveillance
  • you have plans to address the limitations
JUSTIFYING USE – VALIDATING CODES

• Validate by chart review
  • only possible with select data sources (BC/BS)
• Validate with additional sources of data
  • Outpatient pharmacy claims – claims for filled prescriptions
    • Comorbidities (e.g., diabetes) with unique medications that would only be used for that indication
    • Infections - antibiotics
    • Other examples?
  • Laboratory data – select data sources with labs (BC/BS)
    • Microbiology cultures to validate diagnosis of infection
    • Rheumatoid factor/CCP to validate new diagnosis of RA
• Validate with additional information from claims
  • Insertion of pacemaker – follow in time and identify subsequent interrogation
  • Other examples?
IDENTIFYING PROXIES FOR UNDER-CODED RISK FACTORS

• Know that specific factor under-coded in claims data (e.g., obesity)
  • Use proxies to help identify
    • Algorithm for cardiometabolic risk factors as proxy for obesity
    • Smoking-related diseases as proxy for previous/current smoking (using ICD-9-CM diagnosis codes and outpatient pharmacy claims)
SENSITIVITY ANALYSES FOR MISCLASSIFICATION OF OUTCOME

• Use estimates for false negative and false positive coding of outcome in sensitivity analyses to determine effect on regression coefficients
  • Estimates for false negative and false positive coding from literature
  • Estimates from validation subset
  • Shows that you have thought about the impact of outcome misclassification

• Misclassification of important confounders

SOURCES OF ADDITIONAL DATA TO AUGMENT ADMIN/CLAIMS DATA

- Social Security Death Index – need identifiers
- Census data
  - www.census.gov
  - Claritas data
  - Merge by geographic area (e.g., zip code, census tract)
  - Create indicators:
    - SES
    - household characteristics
- Area Resource file
  - Merge by geographic area (county)
    - availability of physicians
    - availability of specialty physicians
    - availability and complexity of hospitals and other facilities
PROVIDER DATA TO AUGMENT ADMIN/CLAIMS DATA

- American Hospital Association Annual Survey
  - Detailed information about hospitals - updated yearly
  - Merge by AHA ID
  - Create indicators:
    - Teaching/nonteaching, number of physicians on staff, bed size by unit type (e.g., CTICU), outpatient surgery, etc.
    - Complexity of care (e.g., bone marrow transplants, PET)

- American Medical Association Masterfile
  - Detailed information about MDs and DOs
  - Merge by physician identifier (UPIN for older data, NPI for more recent)
  - Create indicators:
    - years in practice
    - board licensed, specialty, FMG
    - gender, age
OTHER SOURCES OF DATA TO AUGMENT ADMIN/CLAIMS DATA

- Geographic
  - Neighborhood effects on posthospitalization mortality
  - Impact of altitude change on anemia tx response in hemodialysis patients
- Environmental data – EPA
  - Particulate air pollution and survival in COPD pts
  - Ozone and survival in pts with chronic conditions
REVIEW THE LITERATURE – NOT JUST IN YOUR OWN AREA

• Do broader Pubmed search to find examples of novel uses and analyses of claims data

• Collaborate across disciplines
  • Economics
  • Environmental science
  • Geography
  • School of Social Work
  • Other depts/divisions in Medical School

• Get input from investigators using data
How to add strength to claims data analysis

• Objectives

1. Appreciate the importance of quantitative bias analysis in claims data analysis

2. Able to examine and adjust an association for unmeasured confounding

http://sites.google.com/site/biasanalysis/
Definition & types of biases

- A process at any stage of inference tending to produce results that depart systematically from the true values.

- Importance of addressing bias.

- Types:
  - Selection bias
  - Measurement/information bias
  - Confounding

- Presence, magnitude, direction
Typical “bias analysis”

Protective association between maternal multivitamin supplementation during the periconceptual period and acute lymphoblastic leukaemia among children with Down’s syndrome (OR 0.51; CI 0.30 to 0.89).

• Discussion: “Maternal vitamin supplementation was collected by self-report, which is subject to...recall bias. However, a validation study...reported excellent agreement...”
Bias analysis example

Evaluation of Safety Balls and Faceguards for Prevention of Injuries in Youth Baseball

• “Overall, use of safety balls was associated with a reduced risk of ball-related injury (adjusted RR, 0.77; 95% CI, 0.64-0.93).”
• “It is possible that some Little League injuries during the study period did not result in compensated claims and therefore were not included in this study.”
• “For example, if the difference in the underregistration rates between users and nonusers of safety equipment was 10% or less and the proportion of unregistered injuries was no more than 30%, only a modest degree of bias was present in the study findings.”
• “In this scenario, the **corrected RRs** ranged from 0.67 to 0.88 for safety balls ...”
When is quantitative bias analysis valuable?

- When random error is small (i.e., small CI)
- Only limited number of systematic errors

Why are quantitative methods rarely used?

- Lack of focus in epidemiology and biostatistics courses
- No request from the reviewers
- Lack of user-friendly packaged software
Planning for bias analysis

Fig. 2.1 Integration of planning for bias analysis with conventional study design and analysis
Unmeasured & unknown confounders

- 3 components to confounding
- Controlling for confounding
  - Stratification
  - Regression
- 3 types of confounders
  - Measured
  - Unmeasured
  - Unknown

Fig. 5.1 Causal graph showing confounding by religion in the relation between male circumcision and male acquisition of HIV
Bias parameters

**FIGURE 1.** A graphical representation of measured and unmeasured confounding (C) in relation to exposure (E) and disease (D), and of the data sources used to estimate the components of confounding.

- Association between confounder & disease among unexposed
- Strength of the association between confounder & exposure in source population (prevalence)
- Prevalence of confounder in source population
Unknown confounding example

- To determine the association between 5-fluorouracil–based chemotherapy and survival in older patients using a retrospective cohort study.
- 5-fluorouracil therapy: HR: 0.66; 95% CI 0.60, 0.73.

**Survival Associated with 5-Fluorouracil–Based Adjuvant Chemotherapy among Elderly Patients with Node-Positive Colon Cancer**

**Table 4. Sensitivity of the Hazard Ratio to an Unmeasured Binary Confounder***

<table>
<thead>
<tr>
<th>Prevalence of UBC in the Untreated Group, %</th>
<th>Prevalence of UBC in the 5-FU Group, %</th>
<th>UBC Hazard Ratio</th>
<th>5-FU Hazard Ratio, Adjusted for UBC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>10</td>
<td>3.00</td>
<td>1.11 (1.00–1.21)</td>
</tr>
<tr>
<td>60</td>
<td>10</td>
<td>2.00</td>
<td>0.96 (0.88–1.06)</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>2.00</td>
<td>0.90 (0.83–0.99)</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>2.00</td>
<td>0.84 (0.77–0.92)</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>1.75</td>
<td>1.03 (0.94–1.13)</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>1.75</td>
<td>0.81 (0.74–0.88)</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>1.75</td>
<td>0.85 (0.77–0.93)</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>1.75</td>
<td>0.92 (0.84–1.00)</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>1.5</td>
<td>0.77 (0.70–0.84)</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>1.5</td>
<td>0.79 (0.72–0.86)</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>1.5</td>
<td>0.79 (0.72–0.87)</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>1.25</td>
<td>0.72 (0.66–0.79)</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>1.25</td>
<td>0.73 (0.66–0.80)</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>1.25</td>
<td>0.72 (0.65–0.78)</td>
</tr>
<tr>
<td>90</td>
<td>0.5</td>
<td>1.1</td>
<td>0.69 (0.63–0.75)</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>1.1</td>
<td>0.69 (0.63–0.75)</td>
</tr>
</tbody>
</table>

* 5-FU = 5-fluorouracil; UBC = unmeasured binary confounder.

Original Contribution

Neighborhood Conditions and Risk of Incident Lower-Body Functional Limitations among Middle-aged African Americans

Mario Schootman¹, Elena M. Andresen²,³, Fredric D. Wolinsky⁴, Theodore K. Malmstrom⁵, J. Philip Miller⁶, and Douglas K. Miller⁷

• We conducted a series of analyses to challenge the robustness of the findings. First, we conducted a formal sensitivity analysis to assess the extent to which an unmeasured, binary confounder might explain our results. We performed this by varying both the prevalence of an unmeasured, binary confounder in the group with 4–5 fair/poor neighborhood conditions and the incidence of lower-body functional limitations associated with the unmeasured, binary confounder
### TABLE 4. Sensitivity of the odds ratio to an unmeasured binary confounder at 3-year follow-up (2003–2004) in the African-American Health Study

<table>
<thead>
<tr>
<th>Prevalence of unmeasured binary confounder in neighborhood with 4–5 fair/poor conditions (%)</th>
<th>Prevalence of unmeasured binary confounder in neighborhood with 0–1 fair/poor condition (%)</th>
<th>Unmeasured binary confounder risk</th>
<th>Neighborhood risk adjusted for unmeasured binary confounder</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>10</td>
<td>2</td>
<td>2.59</td>
<td>1.34, 5.03</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>2</td>
<td>2.25</td>
<td>1.16, 4.36</td>
</tr>
<tr>
<td>70</td>
<td>10</td>
<td>2</td>
<td>1.98</td>
<td>1.02, 3.85</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>2</td>
<td>1.77</td>
<td>0.92, 3.44</td>
</tr>
<tr>
<td>30</td>
<td>10</td>
<td>3</td>
<td>2.30</td>
<td>1.19, 4.46</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>3</td>
<td>1.84</td>
<td>0.95, 3.57</td>
</tr>
<tr>
<td>70</td>
<td>10</td>
<td>3</td>
<td>1.53</td>
<td>0.79, 2.97</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>3</td>
<td>1.31</td>
<td>0.68, 2.55</td>
</tr>
</tbody>
</table>

OR: 3.07 (95% confidence interval: 1.58, 5.94)
Data sources for unmeasured confounding

- Substudy (internal)
- External to the study
- Guesstimates
Internal data for unmeasured confounding

• Study purpose: Oral contraceptive use and thromboembolism risk

• Internal data
  ▫ Random sample of claims to abstract medical records
  ▫ Measurement of supplemental variables (behaviors, clinical data, lab values)
External data for unmeasured confounding

Adjusting for Unmeasured Confounders in Pharmacoepidemiologic Claims Data Using External Information

*Epidemiology* 2005;16: 17–24

The Example of COX2 Inhibitors and Myocardial Infarction

Sebastian Schneeweiss, Robert J. Glynn, Elizabeth H. Tsai, Jerry Avorn, and Daniel H. Solomon

**FIGURE 1.** A graphical representation of measured and unmeasured confounding (C) in relation to exposure (E) and disease (D), and of the data sources used to estimate the components of confounding.
Prevalence estimates of unmeasured confounders

<table>
<thead>
<tr>
<th>TABLE 1. Prevalence of Potential Confounders Among Medicare Beneficiaries by Anti-inflammatory Drug Use in 1999</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No.</strong></td>
</tr>
</tbody>
</table>
| * |  | * | * |  |  |  |  |  |  | *
| Any COX2 inhibitor | 872 | 70 | 50 | 50 | 69 | 48 | 24 | 9 | 8 | 45 | 46 |
| Celecoxib only | 562 | 68 | 51 | 49 | 69 | 46 | 24 | 8 | 9 | 45 | 46 |
| Rofecoxib only | 244 | 72 | 49 | 51 | 66 | 48 | 19 | 11 | 7 | 49 | 44 |
| Nonselective NSAIDs only | 1302 | 60 | 56 | 44 | 72 | 56 | 24 | 10 | 10 | 51 | 40 |
| Ibuprofen | 281 | 53 | 60 | 40 | 73 | 58 | 23 | 14 | 15 | 50 | 46 |
| Naproxen | 238 | 60 | 57 | 43 | 74 | 55 | 20 | 8 | 8 | 54 | 38 |
| Other NSAIDs only | 677 | 63 | 54 | 46 | 71 | 55 | 25 | 8 | 9 | 50 | 41 |
| Nonusers | 6611 | 57 | 51 | 49 | 69 | 53 | 17 | 9 | 10 | 50 | 40 |

*BMI ≥ 30.
†Includes individuals who used both COX2 inhibitors and nonselective NSAIDs.
# Association of unmeasured confounder and MI

## TABLE 2. Relative Risk Estimates of the Associations Between Selected Potential Confounders and the Incidence of MI from the Medical Literature*

<table>
<thead>
<tr>
<th>Potential Confounder</th>
<th>Relative Risk</th>
<th>Adjustment of Primary Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obesity (BMI ≥ 30 vs. &lt; 30 kg/m²)</td>
<td>1.7⁺</td>
<td>Age-Sex Adjusted: yes</td>
</tr>
<tr>
<td>Aspirin use (use vs. non-use)</td>
<td>0.7⁻</td>
<td></td>
</tr>
<tr>
<td>Smoking (current vs. never)</td>
<td>3.1⁸</td>
<td>Multivariate Adjusted: yes</td>
</tr>
<tr>
<td>Educational attainment (≤ high school vs. &gt; high school)</td>
<td>2.1‖</td>
<td></td>
</tr>
<tr>
<td>Income level (≤ $20,000 vs. &gt; $20,000)</td>
<td>2.1#</td>
<td></td>
</tr>
</tbody>
</table>

*In case of conflicting literature estimates, the more extreme estimate was used. This will potentially lead to an overestimation of the magnitude of bias.

⁺From SHEEP study combining estimates for men (relative risk [RR] = 1.90) and women (RR = 1.50). Confirmed by data from the HPS study (RR about 1.63). The Framingham Study observed slightly lower relative risk estimates.

⁻Combined from the Nurses Health Study (NHS, RR = 0.68) and the Physicians Health Study (PHS = 0.72) on the use of aspirin in primary prevention.

‖SALT trial had an older population than most other trials (mean age = 67 years compared with 46 in NHS and 58 in PHS) but was designed to test the effects of secondary prevention (RR = 0.79).

#From SHEEP study combining estimates for men (RR = 3.29) and women (RR = 2.87). In age ≥65, no significant association was found between current smoking and MI (RR = 1.36). The relative risk of past smoking was imputed to be 1.0, based on conclusions of multiple studies.

⁸From the WHO MONICA study of men and women <65 years (RR = 2.10). In British men 40–59 a RR of 1.24 was observed.

‖From the WHO MONICA study of men and women <65 years (RR = 2.10). In British men 40–59 a RR of 1.24 was observed.

#From the Kuopio Heart Risk Factor Study. Income was split at median, analogous to our cutpoint in the MCBS study (RR = 2.06).
Its effect on the findings

<table>
<thead>
<tr>
<th>Potential Confounder</th>
<th>RR_{CD} (Literature)</th>
<th>Pr(C) (MCBS)</th>
<th>Crude OR_{EC} (MCBS)</th>
<th>Adjusted OR_{EC}(^*) (MCBS)</th>
<th>“True” RR_{ED} (assumed)</th>
<th>Pr(E) (MCBS)</th>
<th>Apparent RR_{ED}(^†)</th>
<th>% Bias(^‡)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obesity (BMI ≥30 vs. &lt;30 kg/m(^2))</td>
<td>1.7</td>
<td>0.24</td>
<td>0.97</td>
<td>0.99</td>
<td>1.00</td>
<td>0.40</td>
<td>0.99</td>
<td>-0.11</td>
</tr>
<tr>
<td>Aspirin use (use vs. non-use)</td>
<td>0.7</td>
<td>0.10</td>
<td>0.89</td>
<td>0.90</td>
<td>1.00</td>
<td>0.40</td>
<td>1.00</td>
<td>0.29</td>
</tr>
<tr>
<td>Smoking (current vs. never)</td>
<td>3.1</td>
<td>0.09</td>
<td>0.71</td>
<td>0.87</td>
<td>1.00</td>
<td>0.40</td>
<td>0.98</td>
<td>-1.97</td>
</tr>
<tr>
<td>Educational Attainment (≤ high school vs. &gt;high school)</td>
<td>2.1</td>
<td>0.71</td>
<td>0.86</td>
<td>0.83</td>
<td>1.00</td>
<td>0.40</td>
<td>0.98</td>
<td>-2.36</td>
</tr>
<tr>
<td>Income level (≤$20,000 vs. &gt;$20,000)</td>
<td>2.1</td>
<td>0.53</td>
<td>0.71</td>
<td>0.92</td>
<td>1.00</td>
<td>0.40</td>
<td>0.99</td>
<td>-1.44</td>
</tr>
</tbody>
</table>

*Age- and sex-adjusted. The adjusted OR_{EC} estimate was used in the subsequent analyses.

†Apparent relative risk between exposure (COX2 use) and disease (MI) if the potential confounders were not controlled, under the assumption that the true relative risk RR_{ED} equals 1.0. The apparent relative risk was also called confounding risk ratio,\(^21\) which equals to apparent RR_{ED} / true RR_{ED}.

‡Bias = \([(\text{apparent } RR_{ED} - \text{true } RR_{ED}) / \text{true } RR_{ED}] \times 100.\)
We also conducted a sensitivity analysis to examine the robustness of our results from influence of unmeasured confounding. We did not observe any material difference in results from these precautionary measures, and therefore we presented results with covariates adjusted in their original categories.
Issues with unmeasured confounding concept

- Many confounders are not categorical
- Often record-level data needed
- Choosing from a range of estimates for bias parameters from literature
- Reporting
- Manuscript & grant submissions
How you can add strength to claims data analysis

http://sites.google.com/site/biasanalysis/
http://www.brixtonhealth.com/pepi4windows.html