Statistical methods for misclassified outcomes and exposures in data from electronic medical records

Rebecca Hubbard
Associate Investigator
Group Health Research Institute

September 29, 2014
Acknowledgments

National Cancer Institute grant R21CA158510 and U01CA152959
National Cancer Institute contract HHSN261201100031C
Breast Cancer Surveillance Consortium
The promise of using electronic medical records for research

• Information on care and outcomes as practiced in the community
• Large, generalizable patient population
• Structured data can potentially be readily combined across sites
• Administrative data have the potential to cover an even broader population, easily pooled, standardized
• NIH, FDA, CMS, and PCORI have sponsored several initiatives to assemble research networks using EMR data
The challenge of using electronic medical records for research

- Data not collected for research purposes may not be research quality
- Many of the differences between administrative/clinical data and research data give rise to misclassification
  - Research and clinical definitions may differ
  - Not all data elements of interest may be available in structured form (or at all)
  - Administrative data contain even coarser information (e.g., receipt of medical tests but not results)
Using EMR data to study cancer screening

• Several NIH and PCORI initiatives study cancer screening using EMR data
• Large, multi-site studies allow pooling of data to facilitate study of rare outcomes
• Indicative of screening performance in community practice
• But careful consideration must be given to data source
Objectives

• Discuss challenges encountered in research using EMR data with a focus on misclassification in exposure and outcome assessment
• For each example we will,
  • Describe the problem and why it arises
  • Briefly discuss alternative statistical methods to address the challenge
  • Present illustration using cancer screening as the motivating example
Estimating utilization rates under misclassification
Estimating utilization rates

- Utilization rates are commonly used for quality assessment and public reporting.
- From a research perspective, also important to understand screening rates in various populations.
- However, if utilization is measured imperfectly than true rates may be under- or over-estimated.
- **Example**: screening colonoscopy cannot be definitively identified from administrative data.
  - A variety of algorithms have been developed to identify screening colonoscopies using codes.
  - These have operating characteristics varying from sensitivity of 70-90% and specificity of 60-90%.
Bias in estimates of utilization rates

- Let \( Y^* \) represent an imperfect, EMR-based utilization measure and \( Y \) represent the true measure.
- Bias in utilization estimates will vary as a function true screening rate \( P(Y = 1) \), sensitivity \( Se(Y^*) \), and specificity \( Sp(Y^*) \):
  \[
P(Y^*=1) - P(Y=1) = Se(Y^*) \times P(Y=1) + (1-Sp(Y^*)) \times (1-P(Y=1)) - P(Y=1)
  \]

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Estimated screening rate (%)</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fisher</td>
<td>90</td>
<td>58</td>
<td>70.1</td>
<td>11.6</td>
</tr>
<tr>
<td>El-Serag</td>
<td>70.1</td>
<td>71.6</td>
<td>52.8</td>
<td>-5.7</td>
</tr>
<tr>
<td>Adams</td>
<td>88.5</td>
<td>90.5</td>
<td>55.7</td>
<td>-2.8</td>
</tr>
</tbody>
</table>
Bias corrections for utilization rate estimates

- Bias can be corrected if
  - Sensitivity and specificity are known (through direct algebraic manipulation of the formula for bias)
  - Algorithm returns predicted probabilities known to be well-calibrated (by averaging the probabilities)
- We demonstrate the relative performance of an uncorrected and two bias-corrected approaches using a simulation study
- We simulated data for 10,000 populations of size 10,000
- Assumed 59% of individuals were truly screened
- Algorithm for identifying screening had sensitivity and specificity of 0.9
Simulation study results comparing approaches to estimating utilization rates

Provider profiling with misclassified outcomes
Estimating provider-specific outcome measures

- Related to case of utilization rates but estimates are needed for individual providers, potentially with much smaller sample sizes
- Provider profiling, quantitative evaluation of medical providers, has been gaining momentum
- Specifically called out in the Affordable Care Act as a key element of healthcare reform
- However, error in outcome measures may induce bias in estimates of provider performance or diminish our ability to distinguish good from poor providers
- **Example**: Estimating breast cancer detection rates (CDR) for mammography facilities
  - Existing algorithm for CDR with Se = 94.0%, Sp = 99.9%
  - Using an EMR-based algorithm for screen-detected breast cancers, can we identify providers failing to meet performance targets?
Bias and precision in provider profiling

• As discussed for estimating utilization rates, error in outcome ascertainment will induce bias
• However, additional complications arise due to potentially small sample sizes available per provider
• Can poor providers be identified in a setting of:
  • Misclassified outcomes?
  • Small provider volumes?
• Bias can be corrected as previously described
• Imprecision/instability due to small sample sizes can be addressed using shrinkage estimators (Bayesian or random effects)
• We illustrate this problem using a simulated population of 1000 providers with volume and performance distribution based on data from the Breast Cancer Surveillance Consortium
Classification using unadjusted EMR-based measure
Simulation study results

Sensitivity and specificity of provider classification relative to a performance benchmark

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted ML</td>
<td>65.27</td>
<td>84.74</td>
</tr>
<tr>
<td>Unadjusted Bayes</td>
<td>15.75</td>
<td>99.61</td>
</tr>
<tr>
<td>Bias-adjusted ML</td>
<td>75.99</td>
<td>80.3</td>
</tr>
</tbody>
</table>

The primary factor in this setting is the small sample size, error in the outcome plays relatively less of a role.

Estimating exposure-outcome relationships under outcome misclassification
Estimating exposure-outcome relationships

• In some contexts absolute utilization rates may not be needed and measures of relative utilization between populations may be of greater interest.

• **Example:** Association between patient characteristics and screen-detected breast cancer
  • Use administrative data algorithm to identify screen-detected breast cancers
  • How do algorithm operating characteristics affect estimates of relative risk of screen-detected breast cancers in different patient characteristics?
Estimating relative risks

- We have seen that absolute measures of outcome risk must be adjusted to avoid bias

\[
RR_{True} = \frac{P(Y = 1|X = 1)}{P(Y = 1|X = 0)}
\]

\[
RR_{EMR} = \frac{P(Y = 1|X = 1) \times Se(Y^*) + (1 - P(Y = 1|X = 1)) \times (1 - Sp(Y^*))}{P(Y = 1|X = 0) \times Se(Y^*) + (1 - P(Y = 1|X = 0)) \times (1 - Sp(Y^*))}
\]

- When the outcome is assessed via EMR data, relative risk estimates will be unbiased if algorithm specificity is perfect
- Note: This assumes non-differential operating characteristics across patient sub-groups
Estimating other measures of association

- It is also possible to eliminate bias in odds ratios.
- For a dichotomous exposure an explicit expression is available:

$$OR = \frac{P(Y^* = 1|X = 1) - (1 - Sp(Y^*))}{Se(Y^*) - (1 - P(Y^* = 1|X = 1))} \div \frac{P(Y^* = 1|X = 0) - (1 - Sp(Y^*))}{Se(Y^*) - (1 - P(Y^* = 1|X = 0))}$$

- Using EM algorithm, can also obtain adjusted estimates for:
  - Continuous exposures
  - RRs or ORs adjusted for confounders
  - Operating characteristics vary by sub-group
  - Unknown operating characteristics

Relative risk of screen-detected breast cancer by patient race/ethnicity

- Screen-detected breast cancers identified using EMR-based algorithm
  - Se = 92.9%, Sp = 99.9%
- BCSC data provided administrative data as well as gold-standard clinical outcomes for comparison

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>$\text{RR}_\text{EMR} (95% \text{ CI})$</th>
<th>$\text{RR}_\text{True} (95% \text{ CI})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>White, non-Hispanic</td>
<td>1.0 (ref)</td>
<td>1.0 (ref)</td>
</tr>
<tr>
<td>Black</td>
<td>1.11 (0.85, 1.45)</td>
<td>1.14 (0.87, 1.49)</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>0.89 (0.56, 1.40)</td>
<td>0.94 (0.59, 1.48)</td>
</tr>
<tr>
<td>Other/mixed/unknown</td>
<td>0.85 (0.65, 1.10)</td>
<td>0.83 (0.63, 1.10)</td>
</tr>
</tbody>
</table>

Discussion and conclusions
Discussion

- EMR data have enormous promise as a source of information
- However, careful consideration must be given to the source and quality of data
- Many sources of missing data and measurement error due to conflicting clinical/administrative and research definitions and data needs
- Misclassification may lead to bias or loss of precision with the severity of the problem depending on the objective of the study, prevalence of the outcome or exposure, and operating characteristics of the EMR-based measure
- EMR data combining clinical and administrative data are uniquely valuable because
  - Detailed clinical data can be used to create measures and obtain operating characteristic estimates
  - These can then be applied in a larger population with administrative data only
Discussion

- Informatics and machine learning community have led the way in mining EMR databases
- Important sound statistical thinking is employed when using data derived from these approaches
- Problems similar to those encountered across studies of observational data
  - In many cases, existing methods for observational data apply
  - Potential to harness methods for missing data, measurement error, joint modeling, etc.
  - Opportunities also exist for novel methodology development
Thank you

hubbard.r@ghc.org
References


