Estimation of Drug Exposure: Ascertainment and definition

Sonia Hernández-Díaz, MD, DrPH
Department of Epidemiology, Harvard School of Public Health

Agenda
- Exposure Ascertainment
  - Sources of data
  - Misclassification & Differential Misclassification
  - Quantification of errors:
    - Validation
    - Sensitivity analysis
- Exposure Definition
  - Randomized clinical trials (RTC)
  - Observational studies
    - Exposure start date (New user / prevalent user, immortal time bias)
    - Exposure discontinuation (adherence / selection bias)

Sources of data on drug use

Rx drugs
- Medical Record
- Dispensing Records
- Pharmacy
- Claims
- Interview

OTC drugs
- Prescriber
- Pharmacy
- Super-market Medicines cabinet
- Insurance
- Drug used by subject
- Interview

Misclassification
- False positives (adherence / compliance)
- False negatives (OTC, sharing, missing specialist or hospital prescriptions, or incomplete recall)

Los Angeles Times
A quarter of new prescriptions go unfilled, especially when the drugs are for symptomless condition.

Reference:

Misclassification – timing

Exposure
- Probability of exposure after a prescription
- Uncertain start date
- Intended duration
- Uncertain stop date
- Use start
- Use stops
- Filling

Quantification of misclassification
- Reliability: comparison of same instrument at different times or collected by different individuals
- Agreement: comparison of different sources, none of them superior (e.g. medical records and interview)
- Validity / accuracy: Comparison of a method with a gold standard (e.g. questionnaire with blood levels):
  - Sensitivity: correct classification of users
  - Specificity: correct classification of non-users
Quantification of misclassification

**Validation studies**

- Interview responses vs. prescription dispensation?

<table>
<thead>
<tr>
<th>Prescriptions</th>
<th>Exposed</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>Agreement</td>
<td>Imperfect recall</td>
</tr>
<tr>
<td>Unexposed</td>
<td>Other drug sources</td>
<td>Incompleteness</td>
</tr>
</tbody>
</table>

4/11/2012 Hernández-Díaz

**Quantification of misclassification**

**PPV**
- $a/(a+b)$

**NPV**
- $d/(c+d)$

**Gold Standard**

<table>
<thead>
<tr>
<th>Gold Standard</th>
<th>Exposed</th>
<th>Unexposed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>$a$</td>
<td>$b$</td>
<td>$a+b$</td>
</tr>
<tr>
<td>Unexposed</td>
<td>$c$</td>
<td>$d$</td>
<td>$c+d$</td>
</tr>
<tr>
<td>Total</td>
<td>$a+c$</td>
<td>$b+d$</td>
<td>$a+b$</td>
</tr>
</tbody>
</table>

- There may be a tradeoff between the specificity and the sensitivity of an exposure definition, e.g. assume full compliance with days supply vs. using prescription dates only, or consider exposed only if agreement by two sources vs. exposed if one or more sources suggest exposure.

4/11/2012 Hernández-Díaz

**Example: Naproxen prescription vs. use**

Sensitivity = 80%; Specificity = 90%; Prevalence = 5%

<table>
<thead>
<tr>
<th>Gold Standard (use)</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>40</td>
<td>95</td>
<td>135</td>
</tr>
<tr>
<td>Unexposed</td>
<td>10</td>
<td>855</td>
<td>865</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>950</td>
<td>1000</td>
</tr>
</tbody>
</table>

Use based on claims=135/1000=13.5%

4/11/2012 Hernández-Díaz

**Example: Naproxen prescription vs. use**

Sensitivity = 90%; Specificity = 90%; Prevalence = 5%

<table>
<thead>
<tr>
<th>Gold Standard</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>45</td>
<td>95</td>
<td>140</td>
</tr>
<tr>
<td>Unexposed</td>
<td>5</td>
<td>855</td>
<td>860</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>950</td>
<td>1000</td>
</tr>
</tbody>
</table>

Use based on claims=140/1000=14%

4/11/2012 Hernández-Díaz

**Example: Naproxen prescription vs. use**

Sensitivity = 80%; Specificity = 95%; Prevalence = 5%

<table>
<thead>
<tr>
<th>Gold Standard</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>40</td>
<td>47</td>
<td>87</td>
</tr>
<tr>
<td>Unexposed</td>
<td>10</td>
<td>903</td>
<td>913</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>950</td>
<td>1000</td>
</tr>
</tbody>
</table>

Use based on claims=87/1000=8.7%

4/11/2012 Hernández-Díaz

**Example: Naproxen prescription vs. use**

Sensitivity = 80%; Specificity = 95%; Prevalence = 5%

<table>
<thead>
<tr>
<th>Gold Standard</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>40</td>
<td>47</td>
<td>87</td>
</tr>
<tr>
<td>Unexposed</td>
<td>10</td>
<td>903</td>
<td>913</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>950</td>
<td>1000</td>
</tr>
</tbody>
</table>

Use based on claims=87/1000=8.7%

4/11/2012 Hernández-Díaz
Example: Naproxen prescription vs. use
Sensitivity = 80%; Specificity = 99%; Prevalence = 5%

<table>
<thead>
<tr>
<th>Claims</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>40</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Unexposed</td>
<td>10</td>
<td>940</td>
<td>950</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>50</td>
<td>950</td>
<td>1000</td>
</tr>
</tbody>
</table>

Use based on claims=50/1000=5%

Effects of Specificity and Exposure Prevalence and on the PPV
- With rare exposures, the positive predictive value is driven by the specificity of the method
- When use is low, there are few true positives in the population, so false positives can be large compared to the number of true positives

Differential misclassification
- Validity affected by the outcome:
  - Recall bias (retrospective interviews, e.g. case-control)
  - Differential ascertainment (more info for cases)
- Prospective collection usually non-differential except when factors associated with recording of information are also associated with outcome (e.g. non-independent errors for exposure and outcome)

Impact on effect estimates
- Epidemiologists generally:
  - Quantify random errors (chance): 95% CI
  - Recognize other sources of errors
    - Confounding (and use fancy methods for reducing it)
    - Selection bias
    - Measurement error (misclassification)

Sensitivity analyses
- How sensitive are my results to errors?
- Towards a “full disclosure of uncertainty”
  - Account for uncertainty from:
    - Chance
    - Biases
Sensitivity analyses

- Quantification of effects
  - Direction (Non-differential misclassification usually bias effect estimate towards the null, except when variable has >2 categories)
  - Magnitude

Exposure prevalence among cases and controls:

\[ X = 1 \text{ if exposed, } 0 \text{ if not} \]

\[ X^* = 1 \text{ if classified as exposed, } 0 \text{ if not.} \]


Cell counts for controls and for cases:

- \( B_1 \) = Subjects truly exposed
- \( B_0 \) = Subjects truly unexposed
- \( M_0 \) = Control total = \( B_1 + B_0 \)
- \( M_1 \) = Case total = \( A_1 + A_0 \)

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E=1 )</td>
<td>( A_1 )</td>
<td>( B_1 )</td>
</tr>
<tr>
<td>( E=0 )</td>
<td>( A_0 )</td>
<td>( B_0 )</td>
</tr>
<tr>
<td><strong>True</strong></td>
<td>( M_1 )</td>
<td>( M_0 )</td>
</tr>
</tbody>
</table>
Case Control

\[ B_1^* = \text{Subjects classified as exposed} = \text{SeB}_1 + \text{FpB}_0 \]
\[ B_0^* = \text{Subjects classified as unexposed} = \text{FnB}_1 + \text{SpB}_0 \]

\[ M_0 = \text{Control total} = B_1 + B_0 = B_1^* + B_0^* \]
\[ M_1 = \text{Case total} = A_1 + A_0 = A_1^* + A_0^* \]

\[ E=1 \quad \text{Case Control} \]
\[ \begin{array}{cc}
A_1 & B_1^* \\
A_0 & B_0^* \\
\end{array} \]

True

\[ E=0 \quad \text{Case Control} \]
\[ \begin{array}{cc}
A_1^* & B_1^* \\
A_0^* & B_0^* \\
\end{array} \]

Observed

4/11/2012

Sensitivity analyses

Cell counts:
\[ B_i = (\text{SpB}_i^* - \text{FpB}_i) / (\text{SeSp} - \text{FnFp}) \]
\[ B_0 = M_0 - B_1 \]
\[ A_i = (\text{SpA}_i^* - \text{FpA}_i) / (\text{SeSp} - \text{FnFp}) \]
\[ A_0 = M_1 - A_1 \]

Need:
- Exact sensitivity and specificity
- Estimates from validation studies
- Reasonable estimates from published data (e.g. 63% to 100%)
- Relative reporting accuracy for controls AND cases (e.g. ratio = 0.7 and 0.9)

4/11/2012

Example: Folic Acid Antagonists and Risk of Neural Tube Defects (NTDs)

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>27</td>
<td>67</td>
</tr>
<tr>
<td>No</td>
<td>1215</td>
<td>6593</td>
</tr>
</tbody>
</table>

OR: 2.2 (1.4, 3.4)
Sensitivity analyses

**Conclusion:**
- Non-differential misclassification would have biased results towards the null
- Under-reporting limited impact
- Over-reporting (false positives) enormous impact

False positives are unfilled or unused prescriptions when using records/claims

**Exposure Definition**
- **Randomized clinical trial (RTC)**
  - Intention to treat (ITT), as randomized: Assigned to treatment A vs. assigned placebo or treatment B
    - Valid statistical test (interpretation of p-value)
    - Underestimates drug effects when participants do not fully adhere to their assigned treatment. (With active reference groups, ITT can overestimate a treatment’s effect in the presence of differential adherence.)
    - Non-conservative for safety


- **As treated:**
  - Classification is based on the treatment that participants actually received. Participants may not adhere to assigned treatment, some of those assigned to placebo may decide to take treatment, and some of those assigned to treatment may decide not to take it.
Example of Exposure Definition

HR: 1.24 (95% CI: 1.00, 1.54)


Exposure Definition

- Observational studies
  - Exposure discontinuation (adherence) LIKE RCT
  - Define exposure start date (New user, time bias)

Hernán et al. Observational Studies Analyzed Like Randomized Experiments. An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease. Epidemiology 2008;19:766-779

ITT

Adherence adjusted

Hernán et al. Observational Studies Analyzed Like Randomized Experiments. An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease. Epidemiology 2008;19:766-779
**Exposure Definition**

- **Observational studies - Define exposure start date**
  - Prevalent users:
    - Underascertainment of adverse effects occurring early in therapy: “survivors” of the early period of treatment, which can introduce substantial bias if risk varies with time (e.g., aspirin and GI bleeding, HRT and CV events)
    - Time-varying confounding affected by exposure. Prevalent use affected by initiation and continuation (adherence factors)
    - Prevalent users will dominate the effect of the few new users, except for new drugs.

- **Example – Statins and CV events**
  - Pooled, multivariate adjusted mortality hazard ratio for statin use was:
    - 0.84 (95% CI: 0.77, 0.91) in RCT
    - 0.77 (95% CI: 0.65, 0.91) in studies that compared incident users with nonusers
    - 0.70 (95% CI: 0.64, 0.78) in studies that compared a combination of prevalent and incident users with nonusers
  - The greater the proportion of prevalent statin users in observational studies, the larger the discrepancy between observational and randomized estimates.

---

**Exposure Definition - New-user design**

- Restricts the analysis to persons under observation at the start of treatment. Excludes prevalent users, i.e., loss of sample size and power.
- Follow-up begins at initiation of therapy, or t0, after a minimum period of nonuse (washout), like RCT.
- Baseline confounders: patient characteristics obtained at before t0.
- Allows description of time-varying hazards.

**Exposure Definition - Immortal time bias**

- **Immortal time**: period of follow-up during which death (event) cannot occur (e.g., the period between cohort entry and the first prescription)
Exposure Definition- Immortal time bias

Definition 1: Subjects exposed or unexposed to a drug according to whether or not they receive a prescription for this drug within a certain period after cohort entry (e.g. Aspirin and CV event 1-yr after 1st MI)
- Misclassify the immortal time as exposed, when in fact it is unexposed.
- Rate ratio biased downward, rate ‘exposed’ include unexposed immortal person-time.

Exposure Definition - Immortal time bias

Definition 2: Exposed subjects enter the cohort at the time of their first prescription. All other subjects are considered unexposed and their cohort entry is defined arbitrarily by some ‘index date’.
- Exclude the immortal time, when in fact it is unexposed.
- Rate ratio biased downward, rate ‘unexposed’ exclude immortal person-time.

Exposure Definition- Immortal time bias

Definition 3: Count person-time exposed as exposed and person-time unexposed as unexposed.
- Are exposed and unexposed comparable? That’s another issue

Exposure Definition- Immortal time bias

Can result in underestimated risks or apparently beneficial effects
- Extent of the bias will depend directly on the amount of total person-time misclassified or excluded
- A proper exposure definition and person-time approach to data analysis eliminates this bias. Don’t define exposure status after looking in the future
- Immortal time bias can affect observational cohort and case-control studies

Conclusions

- Know the validity of your data (questionnaire, medical records or claims)
- For rare exposures, favor specificity (<false positives)
- Estimate the robustness of your results (sensitivity analysis and true uncertainty around estimates)
- Define your exposure and research question
- Be aware of prevalent users and immortal person time
- Deal with adherence appropriately