PCORI Methodology Standards: Academic Curriculum
Module 6a: Imperfect Exposure
Variables

Category 3: Data Integrity and Rigorous Analyses

Prepared by Jodi Segal, MD, MPH
Scott Zeger, PhD

Presented by Jodi Segal, MD, MPH
When using secondary data, you don’t have control over the data collection

This raises issues of missing data
  - See Modules for Standards for Preventing and Handling Missing Data

Quality of data is uncertain
In This Module

- Review some definitions
- Discuss examples of measurement error and its impact on study results
- Identify possible corrections
- Review quantitative methods for correction
- Review examples of validation studies using medical records
Definitions

- **Reliability**—same answer each time when assessed with same instrument (might not be true or right)

- **Agreement**—same answer when assessed with a different instrument (might not be true or right)
Reliability refers to the consistency of results when the same data collection method is used on the same individual.

Example:
- You wish to learn by survey about a patient’s alcohol exposure history
  - If the same question is posed to the same participant, would the answer be the same?
  - If the answer is yes, the question yields reliable data.

Good surveys are developed to enhance reliability by encouraging different recall processes.

Surveys often ask about an exact date, temporal sequencing, correlation with other events, or adequacy of the response to treatment.
On what date did you first use estrogen?
- Unlikely to get a valid response because this is not a landmark event

When did you go through menopause?
Was estrogen started immediately upon menopause?
- Temporally tying the exposure to a condition

Were you menopausal during your last hospitalization?
- Tying the diagnosis to a landmark event

How confident are you that you have been on estrogen for more than three years but fewer than eight years?
Definitions Again

- **Reliability**—same answer each time when assessed with same instrument (might not be true or right)

- **Agreement**—same answer when assessed with a different instrument (might not be true or right)
Cohen’s Kappa Statistic: A Measure of Agreement

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Medical record</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>On drug</td>
</tr>
<tr>
<td>On drug</td>
<td>( A )</td>
</tr>
<tr>
<td>Not on drug</td>
<td>( C )</td>
</tr>
<tr>
<td></td>
<td>( n_1 )</td>
</tr>
</tbody>
</table>

**Accuracy** = \( \frac{A+D}{N} \)

**Chance agreement (expected)** = \( \frac{(n_1 \times m_1) + (n_2 \times m_2)}{N^2} \)

**Kappa statistic** = \( \frac{\text{accuracy} - \text{chance agreement}}{1 - \text{chance agreement}} \)
Why a Kappa Statistic?

- You could envision using Kappa statistic when there is no “gold standard” test
- You really don’t know which the better method of assessment is
- You still won’t know how valid the exposure (or outcome) data are
- Nevertheless ... you can have some confidence in your data if two ways of assessing exposure (or outcome) give you a similar answer (high kappa)
- Kappa statistic is used often when you are thinking about accuracy of diagnostic tests that have no gold standard
“Malnutrition is common as liver disease progresses. However, an accepted method to screen these patients for malnutrition is lacking. The 6-question undernutrition screening tool was developed for professionals without nutrition training to identify a decline in the nutrition status of patients with liver cirrhosis.”

“... a cross-sectional investigation was completed to compare the agreement between the undernutrition screening tool and nutrition assessment by a registered dietitian (RD).”

“The RD assessment identified undernutrition in 82% of patients (95% CI, 60%-95%). The κ statistic indicated a fair agreement (0.34) between the screening tool and RD assessment.”

It is typically not smart to dichotomize naturally continuous or polytomous measurements, such as...

- Systolic blood pressure (continuous)
- PQ-9 depression score (ranges discretely from 1 to 27 points)

There is a substantial loss of information when we do

Therefore, correlation coefficients are better than kappa statistics as measures of agreement when we have continuous measures.
Questionable validity about an exposure or outcome will lead to problems of misclassification. You will call people exposed (or unexposed) and be wrong.

Quantitatively, this is described as sensitivity and specificity.

Determining validity requires comparison to an assessment that is considered to be a gold standard—the “truth.”

Rarely is there a gold standard. Often called a tarnished standard, alloyed gold standard, or reference standard.
Exposure Validity

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Insurance claims</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Claim for drug</td>
</tr>
<tr>
<td>On drug</td>
<td>A</td>
</tr>
<tr>
<td>Not on drug</td>
<td>C</td>
</tr>
</tbody>
</table>

- **Sensitivity** = $\frac{A}{(A + C)}$
  - In words: *Sensitivity* measures the degree to which the questionnaire correctly identifies individuals who have used the drug (according to the reference standard)

- **Specificity** = $\frac{D}{(B + D)}$
  - In words: *Specificity* measures the degree to which the questionnaire correctly identifies individuals who are not using the drug (according to the reference standard)
But ...

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Insurance claims</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Claim for drug</td>
</tr>
<tr>
<td>On drug</td>
<td>A</td>
</tr>
<tr>
<td>Not on drug</td>
<td>C</td>
</tr>
</tbody>
</table>

Sensitivity and specificity are not really what you want

You really want to know the **predictive values**:
- Probability that someone classified as exposed is truly exposed and that someone classified as unexposed is truly unexposed
To examine validity of the screening tool for assessing undernutrition relative to the registered dietitian’s nutrition assessment, “… sensitivity, specificity, positive predictive value, and negative predictive value were calculated …”

“Sensitivity and specificity of the screening tool were 72% and 75%, respectively”

Positive predictive value was 93% (“a positive screen strongly indicates undernutrition is present”)

Negative predictive value was 37.5% (“a negative screen does not reliably indicate lack of undernutrition”)
You need to know how imperfect exposure classification impacts your estimate of risk of the outcome of interest that is attributable to the exposure.
Equations (see epidemiological textbooks) allow you to recalculate the odds ratio or relative risk attributable to the exposure with reclassification of people on the basis of what you know about the sensitivity and specificity of the exposure measure.

You need to make assumptions in the calculations about whether you think the misclassification is differential or nondifferential.

- **Differential misclassification**—the rates of misclassification of exposure are different in the cases and the controls.
- **Nondifferential misclassification**—the rates of misclassification of exposure are the same in the cases and the controls.

There are also Bayesian methods for doing this that rely on information from previous studies to make distributions about the sensitivity and specificity.
When planning research, you get to choose how you will measure the exposure of interest

No measure is perfect
  - Everything only approximates the actual exposure that you wish to measure

Researchers must describe the chosen measure in applications and what they know about its reliability and validity and any measures of agreement with other measures

If these are not known, the researcher might need to do preparatory work to establish them

Researcher might choose to adjust in analyses for imperfect exposure assessment