Guidelines for Appropriate OAI Data Use

Yuqing Zhang, Boston University

December 6, 2007
OAI Design and Analysis Issues

- Many of the variables are clustered by knee.
- Many of the variables are collected longitudinally over time on a person and/or knee.
- Expensive MRI measurements may not be available for all participants – may make case-control design (nested within cohort) more attractive.
Consequences

- Need to weigh the advantages and disadvantages of various design strategies.
- Failure to properly analyze clustered or longitudinal data can cause p-values and confidence intervals to be seriously incorrect.
Agenda

- Yuqing Zhang to discuss design considerations with a 15 minute question and answer period
- Chuck McCulloch to discuss analysis issues with a 15 minute question and answer period
Study Design
Types of Epidemiologic Study

- Cross-sectional Study
- Cohort Study
  - Longitudinal cohort study
- Case-control Study
  - Matched case-control study
Cross-sectional Study

- Assess presence or absence of both risk factor and disease at the same point in time among study population and examine their association.

- Measure of association

\[
\text{Prevalence odds ratio} = \frac{\text{odds of disease among exposed subjects}}{\text{odds of disease among non-exposed subjects}}
\]
Cross-sectional Study

In a survey of 1000 subjects, 200 subjects had ROA in at least one knee. Of those with knee OA, 80 reported knee pain; and of those without knee ROA, 100 reported knee pain.

<table>
<thead>
<tr>
<th>ROA</th>
<th>Subjects with Knee Pain</th>
<th>Total No. Subjects</th>
<th>Prevalence of Knee Pain (%)</th>
<th>Prevalence odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>700</td>
<td>100</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>100</td>
<td>800</td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td>120</td>
<td>80</td>
<td>200</td>
<td>40.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cohort Study

Subjects are classified on the basis of presence or absence of exposure to a particular risk factor and followed for a specified period of time to determine development of disease in each group.

Measure of association

$$RR = \frac{CI_e}{CI_u}$$
Cohort Study

3000 subjects without knee ROA were followed up for 2 years. Among 2000 subjects with BMI $\geq 25$ kg/m$^2$, 200 developed knee ROA. Among subjects with BMI $< 25$ kg/m$^2$, 50 developed knee ROA.

<table>
<thead>
<tr>
<th>BMI (kg/m$^2$)</th>
<th>No. subjects Developing Knee ROA</th>
<th>Total No. Subjects</th>
<th>Risk of ROA Over 2 years (%)</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 25$</td>
<td>50</td>
<td>1000</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$\geq 25$</td>
<td>200</td>
<td>2000</td>
<td>10.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Longitudinal Cohort Study

- Studies in which outcome variable is repeatedly measured over time (i.e., outcome variable is measured in same individual on several different occasions)
  - Observations of one individual over time are not independent
  - Special statistical techniques are required to take into account correlation between repeated observations
  - It can assess effect of change of exposure over time on change in outcome variable
Longitudinal Cohort Study

- Both knee pain (WOMAC) and BMI were assessed repeatedly (5 times) among participants
  - Is greater BMI associated with higher WOMAC score (cross-sectional association)
  - Is change in BMI associated with change in WOMAC score (longitudinal association)
    - Is increasing BMI associated with increasing WOMAC score
    - Is decreasing BMI associated with decreasing WOMAC score
Case-control Study

- Subjects with a disease of interest are selected as the cases, and a sample of subjects from the source population which gives rise to cases are selected as the controls.
- ‘Nested case-control’ study: cases and controls are drawn from the defined sample of a cohort study.
- Measure of association
  \[ \text{OR (odds ratio)} = \frac{a \times d}{b \times c} \]

K. Rothman, Epidemiology: an Introduction, 2002
Case-control Study

- Of 100 knees with incident knee pain at follow-up visit (cases), 40 knee had BMLs present at baseline examination; among 100 knees without pain (controls), 10 had BMLs present at baseline examination.

<table>
<thead>
<tr>
<th>Frequent Knee Pain</th>
<th>Presence of BMLs at Baseline</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>90</td>
<td>1.0</td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>6.0</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Matched Case-control Study

- Subjects are selected in such a way that some potential confounders (i.e., matching variables) are distributed in an identical manner among cases and controls
  - **Explicit matching**
    - e.g., each case and control pair was matched by study center, sex age (± years), etc.
  - **Implicit matching**
    - e.g., knee with pain (case) vs. contra-lateral knee without pain (control)
Should We Analyze Knees or Persons?
Background

- In a knee OA study, data are often collected at both the person-level and the knee-level.
- Analysis for knee-specific outcome variable presents special issue in statistical inference owing to correlation between two knees within a person.
Observation Unit (Knees vs. Person)

• **Research Questions**
  • Do subjects with SxOA take longer time to complete 15-m walk than those without SxOA?
  • Is greater BMI associated with an increased risk of knee ROA?
  • Is presence of synovitis associated with prevalent knee pain?
  • Are changes of severity of bone marrow lesions (BMLs) associated with pain fluctuation?

Cost of Additional Data Collection
• Outcome
• Exposure
• Confounding

• Residual Confounding
Research Questions

- Do subjects with SxOA take longer time to complete 15-m walk than those without SxOA?
  - Outcome is a person-based

- Is greater BMI associated with an increased risk of knee ROA?
  - Outcome can be either person-based or knee-based
Person-Based Approach

- Observation unit for the outcome variable is the person
  - Worst knee
  - None, unilateral, or bilateral

- Model outcome as a function of a set of risk factors
Pros and Cons

- Statistical software is readily available (logistic regression, linear regression)
- Effect estimate for site-specific risk factor might be biased
- Statistical power might be sacrificed
Research Question

- Is greater BMI associated with an increased risk of knee ROA?
- Is presence of synovitis associated with prevalent knee pain?
Knee-Based Approach

• Observation unit for the outcome variable is the knee

• Model outcome in each knee as a function of a set of risk factors, while accounting for correlation between two knees within a person
**Pros and Cons**

- Offer improved statistical power
- Provide new insights into etiology of risk factor
- Statistical software is readily available (Generalized Estimating Equations (GEE))
- Increase cost of data collection in both knees
Research Questions

• Is presence of synovitis associated with prevalent knee pain?
Cost of Data Collection

• All subjects in OAI had knee MRI taken repeatedly
• Cost of reading all MRIs is extreme
  • To reduce cost, read MRIs of cases and a sample of controls
Traditional Approach

Study design: case-control study

- Knees with pain *(cases)*
- Knees without pain *(controls)*
- Assess presence of synovitis (or its severity)
- Evaluate association between synovitis and prevalent knee pain using logistic regression model adjusting for confounding factors
Pros and Cons

- Reduce cost of data collection
- Statistical model is readily available (logistic regression model)
- Residual confounding is a concern
Which is better control knee?
Person-Matched Case-Control Design

- Eligible subjects
  - Knee with pain (*case*)
  - Contra-lateral knee without pain (*control*)

- Case and control are implicitly matched by person-level risk factors
Pros and Cons

• Reduce cost of data collection
• Eliminate effect of person-level confounding factors
• Statistical model is readily available (conditional logistic regression)
• Unable to examine effect of change in risk factor on pain fluctuation
• Findings may not apply to persons with bilateral knee pain
Research Questions

- Are changes of severity of synovitis associated with pain fluctuation?
Potential Confounding

• Pain is subjective measurement
• There is a natural variability in pain sensitivity, perception and tolerance to pain stimuli
• Many factors affecting pain variability are often not collected in the study
  • genetic predisposition
  • socio-cultural environment
  • prior experience
  • expectations
  • current mood status
• Use each knee as its own control to minimize potential confounding between persons
Traditional Approach

Study design: longitudinal study

- Collect data on presence of pain for each knee repeatedly
- Collect data on presence (or severity) of synovitis for each knee repeatedly
- Analyze severity of synovitis and its change in relation to pain fluctuation adjusting for confounding factors
Pros and Cons

• Allow investigators to assess relation of severity of synovitis and its change to occurrence of knee pain

• Statistical model is readily available (GEE or MIXED model)

• Cost of assessing synovitis from MRI for all subjects is too high

• Residual confounding is still a concern
Self-Matched Case-Control Design

• Each knee serves as its own control (akin to case-crossover study design)
• Included are only knees that have pain at one or more, but *not all*, examinations
• Synovitis, either severity or presence, is assessed for each knee repeatedly
Eligible Knees

• Knees which had pain in **at least one, but not all**, visits (i.e., exams)
  • *Case-visit*: a knee had pain at a scheduled visit
  • *Control visit*: the **same** knee that did not have pain at a scheduled visit
<table>
<thead>
<tr>
<th>Knees</th>
<th>Baseline</th>
<th>12-month</th>
<th>24-month</th>
<th>Eligible</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td><img src="image1" alt="Baseline Image" /></td>
<td><img src="image2" alt="12-month Image" /></td>
<td><img src="image3" alt="24-month Image" /></td>
<td>pain -&gt; no pain</td>
</tr>
<tr>
<td>6</td>
<td><img src="image4" alt="Baseline Image" /></td>
<td><img src="image5" alt="12-month Image" /></td>
<td><img src="image6" alt="24-month Image" /></td>
<td>pain -&gt; no pain</td>
</tr>
<tr>
<td>7</td>
<td><img src="image7" alt="Baseline Image" /></td>
<td><img src="image8" alt="12-month Image" /></td>
<td><img src="image9" alt="24-month Image" /></td>
<td>no pain -&gt; pain</td>
</tr>
<tr>
<td>8</td>
<td><img src="image10" alt="Baseline Image" /></td>
<td><img src="image11" alt="12-month Image" /></td>
<td><img src="image12" alt="24-month Image" /></td>
<td>no pain -&gt; pain</td>
</tr>
</tbody>
</table>

The diagram shows the status of knees at baseline, 12-month, and 24-month assessments, indicating transitions between pain and no pain.
Pros and Cons

• Allow investigators to examine effect of change in severity of synovitis on pain fluctuation

• Statistical model is readily available (conditional logistic regression model)

• Reduce cost of MRI reading

• Minimizing residual confounding

• Unable to identify risk factors for persistently painful or never painful knees
Summary

- No single study design can fit every research question in OAI
- When choosing study design, one must consider:
  - Cost
  - Internal validity
    - Confounding
    - Bias
  - Generalizability or external validity (???)
- Collaboration and group effort are required
  - Clinician, basic scientist, statistician, epidemiologist

It takes a village!!!!
Thank you for your attention.