Matched Case-Control Study

Duanping Liao, MD, Ph.D

Email: Dliao@psu.edu

Phone: 531-4149
Learning Objectives

• Strategies to select cases and controls and to analyze data in pair-matched case-control studies.

• Identify strengths and limitations of case-control studies.
Case-Control Study

An epidemiological study in which a group of persons with the disease of interest (case group) and a group of persons similar to the case group but not having the disease (control group) are selected to compare the proportion of persons exposed to a risk factor of interest in order to elucidate the causal relationship of the risk factor of interest and the disease.

Interpretation of Odds Ratio - same as relative risk:

- OR = 1: exposure is not related to disease.
- OR > 1: a positive association (E is associated with increased risk of D).
- OR < 1: a negative association (E is associated with a lower risk of D).
What we need to know about a case-control study

From Table 2 of the AJE Paper (Costello et al)

1974 – 1989 Exposed to both agent vs. not exposed

Odds ratio =\( \frac{74 \times 113}{39 \times 93} = 2.31 \)

<table>
<thead>
<tr>
<th>2nd step: Past Exp to both Pesticides</th>
<th>Exposure</th>
<th>Diseased (Cases)</th>
<th>Non-diseased (controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
<td>74</td>
<td>39</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>93</td>
<td>113</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>167</td>
<td>152</td>
</tr>
</tbody>
</table>

The authors reported OR=2.14 (1.24, 3.68). Why different from 2.31 above?

They adjusted for age, sex, nonwhite race, education, and smoking status! Why?
Matching

• Select controls who are identical to cases on potential confounders

  – Pair match, frequency match

  – Better control for confounding, especially when the distributions of a confounder do not have much overlap between the cases and the source population (e.g., if cases of myocardial infarction tend to be older)

  – Matched controls represent the source population within each level of the matched variable
• Pair-matched data (Koepsell and Weiss, P 389)

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposed</td>
<td>Non-exposed</td>
</tr>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

  – Odds ratio = b/c

• 1:n matched data, see a biostatistics or advanced epidemiology textbook
• Pair-matched data: estrogens and endometrial carcinoma; matched on age at diagnosis (within 4 years) and year of diagnosis (within 2 years) (Hennekens and Buring, P 300)

<table>
<thead>
<tr>
<th>Controls</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>Exposed</td>
<td>Non-exposed</td>
<td>Total</td>
</tr>
<tr>
<td>Exposed</td>
<td>39</td>
<td>113</td>
<td>152</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>15</td>
<td>150</td>
<td>165</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>263</td>
<td>317</td>
</tr>
</tbody>
</table>

– Odds ratio = 113/15 = 7.5

• Unmatched analysis of pair-matched data: estrogens and endometrial carcinoma (Hennekens and Buring, P 302)

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>152</td>
<td>54</td>
<td>206</td>
</tr>
<tr>
<td>Non-exposed</td>
<td>165</td>
<td>263</td>
<td>428</td>
</tr>
<tr>
<td>Total</td>
<td>317</td>
<td>317</td>
<td>634</td>
</tr>
</tbody>
</table>

– Odds ratio = (152*263)/(54*165)= 4.5, biased towards null
– The stronger the confounding, the stronger the bias.
• Pair-matched data: Tobacco use and acoustic neuroma (Bergenheim M, et al, AJE, 2012;175:1243-1251)

• Matched on age (within 5 years), sex, and place of residence

• 2 controls were attempted to be enrolled for 1 case

• Analysis: conditional logistic regression stratified by matched set with adjustment for potential confounders (Table 2 on P 1247)
• **Strengths (compared to cohort studies)**
  - Efficient
    Small sample size (rare disease)
    Less time (disease with long induction and latent period)
    Less expensive (small N, efficient for exposure that is expensive to measure)

• **Limitations**
  - Inefficient for rare exposure
  - Selection bias (Are cases representative? Do controls represent the source population?)
  - Challenge in measurement of exposure (measure exposure after the occurrence of disease)
  - Difficulty in determine temporality
  - Only one disease is studied
  - Incidence of disease cannot be studied, though OR is intended to estimate the relative risk (incidence ratio).
Summary of Observational Studies

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

Things to remember:

- Low response rate in cross-sectional study
- Low follow-up rate in cohort study
- Selection bias in case-control study
- Information (misclassification) bias in all studies
- Confounding bias in all studies
- Use the most appropriate analytical methods
- Work with your epidemiologists and biostatisticians **early.**