This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike License. Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.
Section D

Another Non-Randomized Study Design: The Case-Control Design
Case-Control Study

- Researchers were interested in studying the association between alcohol consumption and esophageal cancer.

- Esophageal cancer is a rare condition—a prospective study would require a huge number of subjects.

- Another approach—choose subjects whose cancer status is known at the time of recruitment into the study.
  - In this scenario, researchers chose 200 cases and 775 controls and asked about alcohol consumption.
Case-Control Study

- Study results: case/control status by alcohol consumption per day

<table>
<thead>
<tr>
<th></th>
<th>&gt; 80 gm/day</th>
<th>≤ 80 gm/day</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>96</td>
<td>104</td>
<td>200</td>
</tr>
<tr>
<td>Control</td>
<td>109</td>
<td>666</td>
<td>775</td>
</tr>
<tr>
<td>Totals</td>
<td>205</td>
<td>770</td>
<td>975</td>
</tr>
</tbody>
</table>
Alcohol/Eosophageal Cancer

- Important questions
  - Can we estimate the prevalence of esophageal cancer based on the results for this study?
  - Can we calculate the probability of cancer if you drink more than 80 grams of alcohol per day using this case-control study?
  - Can we compute the relative risk of cancer for those who drink > 80 grams of alcohol per day as compared to those who drink ≤ 80 grams per day?
Important Caveat in Case-Control Studies

- In case-control studies, the individuals with the disease (the cases) have been over-sampled.

- The percentage of subjects in your study who have disease are greater than in the population: hence the prevalence/risk in the sample is an overestimate of actual prevalence/risk, usually by a large factor.

- “Prevalence” in the sample is a function of the design of the study: in this example researchers set prevalence (risk) at . . .

\[
\frac{200}{200 + 775} = \frac{200}{975} \approx .21
\]
Important Caveat in Case-Control Studies

- The percentage of the population who have disease from a case-control study (i.e., the risk/prevalence of disease) cannot be correctly estimated from a case-control study.

- Hence, you cannot estimate relative risk (RR) relating disease to exposure of interest.

- CANNOT compute relative risk from case-control study.

- CAN compute odds ratio from case-control study.
Recall, the estimated odds ratio of an outcome compares the observed odds of the outcome for two groups of individuals and is a function of the risk for each group.

\[ OR = \frac{\hat{p}_1 / (1 - \hat{p}_1)}{\hat{p}_2 / (1 - \hat{p}_2)} \]
Case-Control Study

- Quick approach to computing odds ratio from a 2x2 table:
  diagonal cross products!

<table>
<thead>
<tr>
<th></th>
<th>&gt; 80 gm/day</th>
<th>≤ 80 gm/day</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>96</td>
<td>104</td>
<td>200</td>
</tr>
<tr>
<td>Control</td>
<td>109</td>
<td>666</td>
<td>775</td>
</tr>
<tr>
<td>Totals</td>
<td>205</td>
<td>770</td>
<td>975</td>
</tr>
</tbody>
</table>

\[
\text{Odds Ratio Estimate (OR)} = \frac{96 \times 666}{109 \times 104} \approx 5.60
\]
Alcohol and Esophageal Cancer

- Interpretation
  - Individuals with high alcohol consumption (> 80 grams/day) are over five times the odds of esophageal cancer compared to individuals with low alcohol consumption
Important Caveat in Case-Control Studies

- The odds ratio is very close to what the relative risk would be if you had performed a cohort study (provided the disease was rare, say < 1/100)

- If the disease is not rare, OR still follows same direction as RR, but may not be a very accurate estimate of RR
In the alcohol-esophageal cancer example, 5.64 is an estimate of the odds ratio based on a limited sample of data.

It is not the population parameter odds ratio.

Confidence intervals can be calculated that give the range of plausible values for the population odds ratio.

If the 95% confidence interval for the odds ratio does not include one, it suggests that there is a significant association (p < .05).
How can you test if the population odds ratio is one or not?
- Fisher’s exact test
- Chi-square test (approximation)
“cci” command syntax—same setup as “csi” command that we saw in SR1

cci a b c d

Where a, b, c, d from appropriate 2x2 table:
Using Stata

- **Alcohol/esophageal cancer example**
  - `cci 96 105 109 666`

- Recall the 2x2 table:

```
  Yes  No
Yes  96  104  200
No  109  666  775
```

```
205  770
```
Using Stata

- Results from Stata

```
. cci 96 104 109 666

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Unexposed</th>
<th>Total</th>
<th>Proportion Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>96</td>
<td>104</td>
<td>200</td>
<td>0.4800</td>
</tr>
<tr>
<td>Controls</td>
<td>109</td>
<td>666</td>
<td>775</td>
<td>0.1406</td>
</tr>
<tr>
<td>Total</td>
<td>205</td>
<td>770</td>
<td>975</td>
<td>0.2103</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Point estimate</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratio</td>
<td>5.640085</td>
<td>3.937435 8.061794</td>
</tr>
<tr>
<td>Attr. frac. ex.</td>
<td>.8226977</td>
<td>.7460276 .8759581</td>
</tr>
<tr>
<td>Attr. frac. pop</td>
<td>.3948949</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{chi2(1)} = 110.26 \quad \text{Pr}>\text{chi2} = 0.0000 \]
```
## Results from Stata

```
.cci 96 104 109 666
```

<table>
<thead>
<tr>
<th>Cases</th>
<th>Exposed</th>
<th>Unexposed</th>
<th>Total</th>
<th>Proportion Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>96</td>
<td>104</td>
<td>200</td>
<td>0.4800</td>
</tr>
<tr>
<td>Controls</td>
<td>109</td>
<td>666</td>
<td>775</td>
<td>0.1406</td>
</tr>
</tbody>
</table>

| Total       | 205     | 770       | 975   | 0.2103             |

<table>
<thead>
<tr>
<th>Point estimate</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>

| Odds ratio | 5.640085 | 3.937435 | 8.061794 (exact) |
| Attr. frac. ex. | .8226977 | .7460276 | .8759581 (exact) |
| Attr. frac. pop | .3948949 |           |                 |

chi2(1) = 110.26  Pr>chi2 = 0.0000
Using Stata

**Results from Stata**

```
. cci 96 104 109 666

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Unexposed</th>
<th>Total</th>
<th>Prop. Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>96</td>
<td>104</td>
<td>200</td>
<td>0.4800</td>
</tr>
<tr>
<td>Controls</td>
<td>109</td>
<td>666</td>
<td>775</td>
<td>0.1406</td>
</tr>
<tr>
<td>Total</td>
<td>205</td>
<td>770</td>
<td>975</td>
<td>0.2103</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Point estimate</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratio</td>
<td>5.640085</td>
<td>3.937435 8.061794 (exact)</td>
</tr>
<tr>
<td>Attr. frac. ex.</td>
<td>0.826977</td>
<td>0.7460276 0.8759581 (exact)</td>
</tr>
<tr>
<td>Attr. frac. pop</td>
<td>0.3948949</td>
<td></td>
</tr>
</tbody>
</table>

chi2(1) = 110.26 Pr>chi2 = 0.0000
```
Using Stata

- The 95% CI for the OR of esophageal cancer for those consuming > 80 grams of alcohol per day compared to those consuming 80 grams or less is 4.0 to 8.0
Odds Ratio and Case-Control Studies

- Why would we even bother calculating the odds ratio when we can calculate relative risk?
  - The odds ratio turns out to be important because you can calculate it either in cohort studies or case-control studies
  - The relative risk can only be calculated from cohort studies

- Luckily, as we saw in SR1, the odds ratio informs us about risk

- If the outcome of interest is rare overall then the odds ratio is a good estimate for the relative risk
Odds Ratio and Case-Control Studies

- Recall:
  - $H_o: p_1 = p_2$  
  - $H_a: p_1 \neq p_2$

  - $H_o: RR = 1$
  - $H_a: RR \neq 1$

  - $H_o: OR = 1$
  - $H_a: OR \neq 1$

- All three hypotheses testing for disease exposure relationship