What is a case control study?

Tarani Chandola
Social Statistics
University of Manchester
Imagine....

- It is 1950
- You suspect an association b/w smoking and lung cancer
- How would you show this?
Cases

- Find cases of lung cancer (where?)
- What is the exposure?

*Hypothetical scenario:*

- You find 100 cases of lung cancer
- 60 of them were heavy smokers (>25 cigs/day)
- Does this show an association between smoking and lung cancer?
Controls

• Where can you find similar groups of patients to lung cancer patients?

*Hypothetical scenario:*
- You find 100 control patients who do not have lung cancer
- 20 of them were heavy smokers
- So 20% of controls and 60% of cases were heavy smokers
- Does this show an association between smoking and lung cancer?
Comparing cases and controls

• Populations of cases and controls differ
• So cannot compare % who are exposed (different denominators/populations)
• But you can compare the odds of having lung cancer for two groups: the exposed and the unexposed
Doll and Hill (1952)

• Data was collected from hospitalized patients from more than 20 hospitals in London over a four year period (April 1948 and February 1952).
• Investigators asked hospital personnel to contact them whenever a patient was admitted to the hospital with a new diagnosis of lung cancer (cases).
• Investigators also selected a random sample of patients from the same hospitals, but with different illnesses (controls).
• Cases and controls were interviewed about their smoking habits.
Doll and Hill (1952)

• 1465 case-patients were interviewed for the study all under the age of 75, 1357 men and 108 women.
• Because of the disproportionate ratio of men to women (25 to 2), only men were included in the final study.
• An equal number of controls (1357) were interviewed about their smoking history.
• Investigators divided smoking history into groups based on the average number of cigarettes smoked per day.
### Doll and Hill (1952)- results

<table>
<thead>
<tr>
<th>Cigarettes smoked daily</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>10-14</td>
<td>565</td>
<td>706</td>
</tr>
<tr>
<td>15-24</td>
<td>445</td>
<td>408</td>
</tr>
<tr>
<td>25+</td>
<td>340</td>
<td>182</td>
</tr>
<tr>
<td>All Smokers</td>
<td>1,350</td>
<td>1,296</td>
</tr>
<tr>
<td>Total</td>
<td>1,357</td>
<td>1,357</td>
</tr>
</tbody>
</table>

What numbers will you compare?
Doll and Hill (1952) - analysis

<table>
<thead>
<tr>
<th>Cigarettes smoked daily</th>
<th>Odds if case</th>
<th>Odds if control</th>
<th>Odds of smoking</th>
<th>Odds of smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7/1350</td>
<td>61/1296</td>
<td>=7/1350</td>
<td>=61/1296</td>
</tr>
<tr>
<td>10-14</td>
<td>565/1350</td>
<td>706/1296</td>
<td>=565/1350</td>
<td>=706/1296</td>
</tr>
<tr>
<td>15-24</td>
<td>445/1350</td>
<td>408/1296</td>
<td>=445/1350</td>
<td>=408/1296</td>
</tr>
<tr>
<td>25+</td>
<td>340/1350</td>
<td>182/1296</td>
<td>=340/1350</td>
<td>=182/1296</td>
</tr>
<tr>
<td>All Smokers</td>
<td>1350</td>
<td>1296</td>
<td>1357</td>
<td>1357</td>
</tr>
</tbody>
</table>

Total 1357
Doll and Hill (1952) - Odds Ratios

<table>
<thead>
<tr>
<th>Cigarettes smoked daily</th>
<th>Cases</th>
<th>Odds of smoking if case</th>
<th>Controls</th>
<th>Odds of smoking if control</th>
<th>Odds of being a case given exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>0.01</td>
<td>61</td>
<td>0.05</td>
<td>0.02/0.05</td>
</tr>
<tr>
<td>10-14</td>
<td>565</td>
<td>0.42</td>
<td>706</td>
<td>0.54</td>
<td>0.42/0.54</td>
</tr>
<tr>
<td>15-24</td>
<td>445</td>
<td>0.33</td>
<td>408</td>
<td>0.31</td>
<td>0.33/0.31</td>
</tr>
<tr>
<td>25+</td>
<td>340</td>
<td>0.14</td>
<td>182</td>
<td>0.14</td>
<td>0.25/0.14</td>
</tr>
<tr>
<td>All Smokers</td>
<td>1350</td>
<td>1.00</td>
<td>1296</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### Doll and Hill (1952)- Odds Ratios

<table>
<thead>
<tr>
<th>Cigarettes smoked daily</th>
<th>Cases</th>
<th>Controls</th>
<th>Odds of smoking if case</th>
<th>Odds of smoking if control</th>
<th>Odds ratio</th>
<th>Odds ratio (rescaled)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>61</td>
<td>0.01</td>
<td>0.05</td>
<td>0.11</td>
<td>1.00</td>
</tr>
<tr>
<td>10-14</td>
<td>565</td>
<td>706</td>
<td>0.42</td>
<td>0.54</td>
<td>0.77</td>
<td>6.97</td>
</tr>
<tr>
<td>15-24</td>
<td>445</td>
<td>408</td>
<td>0.33</td>
<td>0.31</td>
<td>1.05</td>
<td>9.50</td>
</tr>
<tr>
<td>25+</td>
<td>340</td>
<td>182</td>
<td>0.25</td>
<td>0.14</td>
<td>1.79</td>
<td>16.28</td>
</tr>
</tbody>
</table>

As you move up the groups, from no cigarettes per day to 25+ cigarettes per day, the values of the odds ratio rise steadily, consistent with a dose-response relationship between the exposure (daily number of cigarettes smoked) and the strength of the association (odds ratio).
Odds ratio

• An approximation of relative risk
• RR $\rightarrow$ “how much more likely are exposed people to get the disease than the unexposed”
• OR $\rightarrow$ “how much more likely were people with the disease to have been exposed than those without the disease”
Calculation of odds ratio

\[
\begin{array}{cccc}
\text{Disease} & + & - & \text{(cases)} \\
\text{Exposure} & + & a & c \\
\text{Exposure} & - & b & d
\end{array}
\]
Calculation of odds ratio in 2 x2 table

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease + (cases)</th>
<th>Disease - (controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure +</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Exposure -</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Odds ratio = \( \frac{a}{c} : \frac{b}{d} = \frac{ad}{bc} \)

= cross-product ratio
• What types of bias in a case-control study?
Selection Bias

• Selection of appropriate controls
  – hospital based studies
  – community based studies
  – Selection bias with cases
• Selection of cases
• Matched case-control studies
  – Cases and controls often differ in important aspects (age, sex, ethnicity, behaviours...)
  – These can confound the study
  – One way to eliminate such differences is matching controls to cases on these factors
  – More than 1 control per case can be used
Information Bias

• retrospective measurement of exposure may be problematic:
  – inaccurate - will underestimate any relation
  – reporting bias - overestimation
  – reverse causation - spurious association
Why carry out a case control study?
Strengths of case-control studies

- Quick (cases already exist, no need to wait)
- Cheap (not necessary to examine large number of people)
- Can examine many exposures
- Suitable to study rare diseases
- Suitable to study stable exposures (e.g., genetic markers)
Weaknesses of case-control studies

• Not suitable for rare exposure
• Cannot calculate incidence risk or death rates
• Prone to selection bias
• Prone to misclassification of exposure
• Prone to reverse causation (people with disease may change their behaviour)
References:

