Case-control studies

Hans Wolff

Division d’épidémiologie Clinique
Département de médecine communautaire
Hans.Wolff@hcuge.ch
Outline

- Case-control study
- Relation to cohort study
- Selection of controls
- Sampling schemes of controls
Case-control studies (CCS)

- **Genetic factors**
  - Exposed
  - Non exposed

- **Environmental factors**
  - Cases: ill
  - Controls: Not ill

Lifestyle

Presumed cause

Initial situation

February 25, 2004

WHO- Postgraduate course 2004 – CC studies
Case-control studies (CCS)

Lifestyle

Genetic factors

- exposed
- non exposed

Environmental factors

- exposed
- non exposed

Cases: ill

Controls: Not ill

Did they were exposed or not?
### 1. Example: Passive Smoking & Breast Cancer

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Cases</th>
<th>Controls</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexposed</td>
<td>40  22.2%</td>
<td>234  38.7%</td>
<td>1.0</td>
</tr>
<tr>
<td>Passive</td>
<td>140  77.8%</td>
<td>370  61.3%</td>
<td>2.2</td>
</tr>
</tbody>
</table>
Case-Control Design

**SAMPLE**

- **BC Cases**
  - Passive Smokers: 140
  - Non-exposed: 40

- **Controls**
  - Passive Smokers: 370
  - Non-exposed: 234
Presence or absence of disease ... 

... is fixed by design in case-control studies.

- Cases have the disease
- Controls don’t.
- We can NOT compute a risk of disease
- We CAN compute prevalence of exposure in cases and controls
Passive Smoking & Breast Cancer

- **Cases:** all incident breast cancer in Geneva
- **Controls:** random sample of the Geneva female population
- **Exposure:** questionnaire on lifetime history of exposure to passive smoke
Have you ever been exposed?

- ... to passive smoking at least 1 hour per day for at least 1 year? (Yes / No)
- At home? At work? During leisure time?
- If yes, describe each episode of exposure
  - Duration, who, size of the room, etc...
  - *Unexposed* = never active, never passive
What should be always true for a case-control study?

1. Cases and controls are randomized with respect to exposure.
2. Cases are a representative sample of all cases in the general population.
3. Controls are a representative sample of the general population.
4. Cases and controls have the same population of origin.
5. Always start with some cases, then identify their valid controls.
Fundamental conditions for the validity of this case-control design

Cases and controls:
- Originate from Geneva resident, <75 y.
- are sampled independently of their exposure to passive smoke

Solution:
- All incident cases over a given time period
- Controls are a random sample of population
Case Definition

- Incident (= newly diagnosed)
- Between 1/1/92 and 12/31/93
- Resident of Geneva
- Aged < 75 yrs
- Identified: all pathology labs of Geneva
Control Definition

- Never diagnosed with breast cancer
- Between 1/1/92 and 12/31/93
- Resident of Geneva
- Aged < 75 yrs
- Stratified random sample
  - Population controls
    - Why not use hospital controls?
## Prevalence of Passive Smoking

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexposed</td>
<td>40</td>
<td>234</td>
</tr>
<tr>
<td>Passive</td>
<td>140</td>
<td>370</td>
</tr>
</tbody>
</table>
The proportion of passive smoker cases is...

1. \( \frac{40}{234} \)

2. \( \frac{140}{40} \)

3. \( \frac{140}{180} \)

4. \( \frac{370}{234} \)

5. \( \frac{370}{604} \)
## Prevalence of Passive Smoking

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Cases</th>
<th></th>
<th>Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Unexposed</td>
<td>40</td>
<td>22.2</td>
<td>234</td>
<td>38.7</td>
</tr>
<tr>
<td>Passive</td>
<td>140</td>
<td>77.8</td>
<td>370</td>
<td>61.3</td>
</tr>
</tbody>
</table>
## Prevalence of Passive Smoking

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Unexposed</td>
<td>40</td>
<td>22.2</td>
</tr>
<tr>
<td>Passive</td>
<td>140</td>
<td>77.8</td>
</tr>
</tbody>
</table>
The odds of passive smoking in CASES is...

1. \(\frac{140}{40} = 3.5\)

2. \(\frac{77.8}{22.2} = 3.5\)

3. \(\frac{140}{180} = 77.8\)

4. \(\frac{140}{77.8} = 1.8\)

5. Answers 1 or 2
## Odds of Passive Smoking in CASES

<table>
<thead>
<tr>
<th>Smoking history</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexposed</td>
<td>40</td>
<td>22.2</td>
</tr>
<tr>
<td>Passive</td>
<td>140</td>
<td>77.8</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Odds = \( \frac{140}{40} = \frac{77.8}{22.2} = 3.5 \)
### Odds of Passive Smoking in CONTROLS

<table>
<thead>
<tr>
<th>Smoking history</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexposed</td>
<td>234</td>
<td>38.7</td>
</tr>
<tr>
<td>Passive</td>
<td>370</td>
<td>61.3</td>
</tr>
<tr>
<td>Total</td>
<td>604</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Odds = \[\frac{370}{234} = \frac{61.3}{38.7} = 1.6\]
**AR in case-control study?**

Recall

\[ AR_{\text{duration}} = \text{Risk (E+)} - \text{R(E-)} \]

Since risk cannot be computed directly from a case-control study, AR cannot be computed either.
RR in case-control study?

RR = Risk (E+) / R(E-)

Since risk cannot be computed directly from a case-control study, RR cannot be computed either
## Odds Ratio of Passive Smoking

<table>
<thead>
<tr>
<th>Group</th>
<th>Odds</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>3.5</td>
<td>(\frac{3.5}{1.6} = 2.2)</td>
</tr>
<tr>
<td>Controls</td>
<td>1.6</td>
<td>(\frac{1.6}{1.6} = 1.0)</td>
</tr>
</tbody>
</table>

Reference Group
Interpretation of the Odds Ratio (1)

- The odds of being a passive smoker are 2.2 greater in breast cancer cases than in population controls.

Alternatively:

- The odds of breast cancer is 2.2 greater in those exposed to passive smoke than in unexposed.

  WHY?
Case-Control Design

**SAMPLE**

**BC Cases**
- Passive Smokers: 140
- Non-exposed: 40

**Controls**
- Passive Smokers: 370
- Non-exposed: 234

**WHO- Postgraduate course 2004 – CC studies**

February 25, 2004
Imagine ...

you could have done the perfect cohort study instead of the case-control study
Cohort Design (Risk period: 2 yrs)

Female Population of Geneva

Passive Smokers 55,500
- Breast Cancer 140
- No Breast Cancer 55,360

Non-exposed 35,100
- Breast Cancer 40
- No Breast Cancer 35,060
### Odds Ratio of Breast Cancer

<table>
<thead>
<tr>
<th>Breast Cancer</th>
<th>Passive Smokers</th>
<th>Unexposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present (A)</td>
<td>140</td>
<td>40</td>
</tr>
<tr>
<td>Absent (B)</td>
<td>55,360</td>
<td>35,060</td>
</tr>
</tbody>
</table>

- **Odds (A/B):**
  - Passive Smokers: 0.00253
  - Unexposed: 0.00114

- **Odds Ratio:**
  - Passive Smokers: 2.2
  - Unexposed: 1.0 (ref)
Identity of Odds Ratio

- **Case-control study:**
  - Odds ratio of passive smoking = 2.2

- **Cohort study:**
  - Odds ratio of breast cancer = 2.2
    - Same interpretation

- **Identical Odds Ratio in the cohort and in the case-control studies.**
Female Population of Geneva

Passive Smokers

Breast Cancer
140

No Breast Cancer
55,360

Non-exposed

Breast Cancer
40

No Breast Cancer
35,060

Controls

Passive Smokers
180

Non-exposed
40

Passive Smokers
370

Non-exposed
234

$F_1 = 1.0$

$F_2 = 0.005$

$F_3 = 1.0$

$F_4 = 0.005$

$F_n = \text{fraction included into the sample}$
Relation of Case-Control to Cohort Studies

- In a case-control study:
  - **CASES** are sampled among people in the unexposed and passive smokers cohorts who did develop breast cancer
  - **CONTROLS** are sampled among people in the unexposed and passive smokers cohorts who did **not** develop breast cancer
Odds Ratio and Relative Risk

- **Relative Risk** = \[
\frac{140}{55,500} \div \frac{40}{35,100} = 2.2
\]

Note effect of rare disease on denominators

- **Odds Ratio** = \[
\frac{140}{55,360} \div \frac{40}{35,060} = 2.2
\]
Interpretation of the Odds Ratio (2)

- The **ODDS** of breast cancer is 2.2 greater in those exposed to passive smoke than in unexposed.

**Alternatively:**

- The **RISK** of breast cancer is 2.2 greater in those exposed to passive smoke than in unexposed.
Comparison of the OR and RR

Fictive illness with low prevalence

<table>
<thead>
<tr>
<th></th>
<th>Cases (M+)</th>
<th>Controls (M-)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed (E+)</td>
<td>2</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>non-exposed (E-)</td>
<td>1</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>197</td>
<td></td>
</tr>
</tbody>
</table>

\[
RR = \frac{2/100}{1/100} = 2 \quad \text{OR} = \frac{2 / 1}{98 / 99} = 2.02
\]
Comparison of the OR and RR

Fictive illness with high prevalence

<table>
<thead>
<tr>
<th></th>
<th>Cases (M+)</th>
<th>Controls (M-)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed (E+)</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Non-exposed (E-)</td>
<td>25</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
<td>125</td>
<td></td>
</tr>
</tbody>
</table>

\[
RR = \frac{50}{25} = 2 \quad \text{OR} = \frac{50}{50} = 3
\]
Advantages of Case-Control Studies

(1)

- Less expensive …
- Require smaller sample sizes …
- Shorter duration … than prospective study
- Study multiple risk factors for 1 disease
- Easily reproduced in different populations by different investigators
Disadvantages of Case-Control Studies (1)

- Information about exposure is often obtained after the diagnosis is done
  - Example: diet, physical activity
- Dependent on the subject’s memory, which may be affected by the disease
Disadvantages of Case-Control Studies (2)

- Population of origin for cases is difficult to define precisely.
  - Difficult to identify appropriate control group
- Does not provide estimate of risks and attributable risk