CHAPTER 11—CASE CONTROL STUDIES
CASE CONTROL STUDIES

- Epidemiological studies
- Experimental studies
- Observational studies
- Case control studies
  - Case comparison
  - Case referent
- Retrospective study
WHEN TO USE CASE CONTROL STUDIES

- Exposure data are difficult/expensive to obtain
- Disease is rare
- Disease has long induction/latent period
- Little is known about disease
- Underlying population is dynamic
CLASSIFICATION OF STUDIES

- When disease is identified
- When exposure/treatment is recognized
- When analysis is conducted
CLASSIFICATION OF CASE CONTROL STUDY

- When disease is identified—PRESENT
- When exposure/treatment is recognized—PAST
- When analysis is conducted—PRESENT
FIGURE 11.1
Schematic for Case Control Studies
CASE CONTROL ASSUMPTIONS

- Frequency disease population is small
- Cases/controls are representative
- Can’t calculate relative risk (RR) directly
CHANGING VIEW OF CASE CONTROL STUDIES

- Traditional (pre-1980s)
  - Case control as inferior alternative to cohort
  - Cohort: cause $\rightarrow$ effect
  - Case control: cause $\leftarrow$ effect, therefore backward
  - Called TROHOC studies (cohort spelled backward)

CHANGING VIEW OF CASE CONTROL STUDIES

- Modern view
  - Case control study as a method of sampling population that gave rise to cases
  - Control group provides info of exposure distribution in population that gave rise to cases
  - Can calculate relative risk directly
    - Exposed/unexposed numerators come from cases
    - Exposed/unexposed denominators come from sample of population (controls) + cases

SELECTING CASES

- Cases usually have the disease
- Define cases specifically
  - Signs/symptoms
  - Clinical exams
  - Diagnostic tests
- Err on being restrictive rather than inclusive
SOURCES OF CASE IDENTIFICATION

- Clinic patient rosters
- Death certificates
- Surveys
- Cancer/birth defect registries
- Use incident, not prevalent, cases when possible—why?
SOURCES OF CONTROLS

- Population controls
  - Tax lists, voter registration, driver’s license rosters, telephone directories, random digit dialing
  - Come from same population base but time consuming and expensive

- Hospital/clinic controls
  - Illnesses in control must be unrelated to exposure
  - Illnesses in control must have same referral pattern to healthcare facility

- Dead controls—when some cases deceased

- Friend/spouse/relative controls
  - Likely to share same socioeconomic status, race, etc., but cases reluctant to nominate and may share same habits

THE BALANCE IN SELECTING CONTROLS

- Matching versus finding controls
EXPOSURE

- Characteristics/events that increase or decrease probability of disease/disability/death

- Protective and malicious exposures
SOURCES OF EXPOSURE INFORMATION

- In-person/telephone interview
- Questionnaires
- Medical/pharmacy/registry/employment/insurance/birth/death records
- Biological specimens—biomarkers

CASE CROSSOVER STUDY

- “When brief exposure causes transient change in risk of a rare acute-onset disease”
  - Risk of acute myocardial infarction (AMI) immediately following exertion
  - Risk of motor vehicle (MV) collision while using cell phones
  - Risk of unsafe sex following consumption of alcohol
- Cases serve as their own controls
- Exposure frequency during hazard period compares to nonhazard period (control)
  - Hazard period defined above as one hour before AMI
    - Get data on physical exertion inside and outside the one-hour window
  - Hazard period defined above as ten minutes before MV accident
    - Compare phone activity during ten-minute window to a “control” period of time
  - Get risk of exposure during hazard period compared to control period, e.g., risk of sex without and with alcohol consumption

Relative risk = number of times more likely cases are to get disease than controls given exposure
MEASURING RELATIVE RISK

The $2 \times 2$ table

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Disease (cases)</th>
<th>No Disease (controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>With factor</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Without factor</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

**Odds ratio**—estimate of relative risk $= \frac{ad}{bc}$
### FIGURE 11.2
Case Control Study Design

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cases</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>
FIGURE 11.3
Family History and Breast Cancer Risk

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cases</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>123</td>
<td>51</td>
</tr>
<tr>
<td>No</td>
<td>3,960</td>
<td>4,032</td>
</tr>
</tbody>
</table>

\[ \text{OR} = \frac{123 \times 4,032}{51 \times 3,960} = 2.43 \]

### FIGURE 11.4
Induced Abortion and Breast Cancer Risk

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cases</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>95</td>
<td>63</td>
</tr>
<tr>
<td>No</td>
<td>208</td>
<td>181</td>
</tr>
</tbody>
</table>

\[
OR = \frac{95 \times 181}{63 \times 208} = 1.3
\]

**SOURCE:** Daling et al. (1996). Used with permission of Oxford University Press.
**FIGURE 11.5**
Dose-Response Relationship Between Duration of Lactation and Premenopausal Breast Cancer

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Cases</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 3 months</td>
<td>Yes: 203</td>
<td>375</td>
</tr>
<tr>
<td></td>
<td>No: 602</td>
<td>1,009</td>
</tr>
<tr>
<td>OR = 0.91</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure 4-12 months</th>
<th>Cases</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes: 195</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>No: 602</td>
<td>1,009</td>
<td></td>
</tr>
<tr>
<td>OR = 0.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure 13-24 months</th>
<th>Cases</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes: 106</td>
<td>251</td>
<td></td>
</tr>
<tr>
<td>No: 602</td>
<td>1,009</td>
<td></td>
</tr>
<tr>
<td>OR = 0.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SOURCE:** Data from Newcomb et al. (1994).
The Odds Ratio (OR)

- An estimate of relative risk
  - When cases/controls are representative
  - If disease prevalence is small
- If > 1.0, then increased risk
- If < 1.0, then decreased risk (protective)
**Attributable Fraction**

- How much is risk factor responsible for disease?
- Proportion of disease attributed to risk factor
- Attributable fraction (AF) = \( \frac{p(\text{OR} - 1)}{p(\text{OR} - 1) + 1} \times 100\% \)
  - \( p \) = proportion with risk factor
  - \( \text{OR} \) = odds ratio
ATTRIBUTABLE FRACTION EXAMPLE

- A study of the risk factors for lung cancer (Kreuzer et al. 1998)
  - ORs of 15.9 and 29.9 for males and females, respectively
  - Assume smoking prevalence: men (36.8 percent), women (21.5 percent)

- Female smokers: 29.9 times as likely to get lung cancer; smoking responsible for 86.1 percent of disease
- Male smokers: 15.9 times as likely to get lung cancer; smoking responsible for 84.5 percent of disease

\[
\text{AF (males)} = \frac{0.368(14.9)}{0.368(14.9)+1} \times 100 = 84.5\%
\]

\[
\text{AF (females)} = \frac{0.215(28.9)}{0.215(28.9)+1} \times 100 = 86.1\%
\]
ATTRIBUTABLE FRACTION—OTHER EXAMPLES

- Human papillomavirus and cervical cancer in Morocco: OR = 61.6 and AF = 92 percent

- Why is AF low?
  - Low odds ratio
  - Low prevalence of the risk factor
  - Induced abortion and breast cancer (Daling et al. 1996)
    - OR = 1.4, 30 percent get abortion, AF = 10.7 percent
CROSS-SECTIONAL STUDY (PREVALENCE STUDY)

- Usually, survey done at a particular point in time—snapshot of the population
- Both exposure and disease outcome determined simultaneously
- Cases of disease are prevalent (not incident) cases
<table>
<thead>
<tr>
<th>Exposure</th>
<th>Divorce</th>
<th>Ulcers</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>100</td>
<td>Yes</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>400</td>
<td>No</td>
<td>450</td>
<td></td>
</tr>
</tbody>
</table>
HOW TO KNOW STUDY IS VALID

- Eliminate alternative explanations
  - Bias
  - Confounding
  - Random error
BIAS

- Definition: systematic error that leads to incorrect/invalid estimate of association
  - Easier to avoid bias than remove or fix bias
- Types of bias
  - Selection bias
  - Observation (or misclassification) bias
- How to evaluate for bias?
  - Identify source
  - Estimate magnitude
  - Assess direction

Too many names

- More than 100 different kinds of bias with all sorts of names
- I like to think of two broad categories
  - **Selection bias**—how groups (disease or exposure) are selected
  - **Misclassification bias**—whether a subject is incorrectly classified on the basis of disease or exposure
SELECTION BIAS

- Definition: error due to systemic differences in characteristics between those selected for study and those not (Last)
  - How subjects are chosen
  - Case control—if different criteria related to exposure
  - Retrospective cohort—if selection of exposed or unexposed group related to outcome

SELECTION BIAS WITH CASE CONTROL

- If using hospitalized patients, bias is called Berkson bias
- Suppose you are trying to associate exposure (X) with disease A using disease B as controls
- Bias occurs if A, B, and X have different probabilities of admission to hospital
- No bias if:
  - Can’t be hospitalized simply because have X; or
  - Rate of admission to hospital is same for A and B
EXAMPLE OF BERKSON BIAS

Suppose you want to test if depression (X) is a risk factor for breast cancer (A) using pneumonia patients (B) as controls.
HOW TO PREVENT SELECTION BIAS?

- Define study population independent of disease, not after cases appear (if possible, define study population prior to follow-up)

- Get same information from cases and controls (selection criteria the same)

- Don’t let disease influence the availability of information, e.g., in occupational cohort study if those who develop disease have records sent to their physicians and they are misplaced

- Don’t let disease influence loss of subjects to follow-up, e.g., in cohort study if those who don’t develop disease drop out of study more than those who do
INFORMATION BIAS

- Definition: the means of obtaining information about subjects is inadequate → incorrect
- Misclassification bias
- Bias in abstracting records
- Bias in interviewing
- Bias from surrogate interviews
- Surveillance bias
- Recall bias
- Reporting bias
Misclassification by disease or exposure

Four types:

- Nondifferential misclassification bias of disease
- Differential misclassification bias of disease
- Nondifferential misclassification bias of exposure
- Differential misclassification bias of exposure

- Differential bias: opposite directions
- Nondifferential bias: same direction
NONDIFERENTIAL MISCLASSIFICATION BIAS OF DISEASE

- Definition of disease gap—too narrow or too wide
- Bias underestimates effect
- Shifts relative risk toward 1.0
DIFFERENTIAL MISCLASSIFICATION BIAS OF DISEASE

- Classification as cases/controls depends on exposure
- Exposed more likely to be cases
- Tendency without blinding
- Fen-Phen example
- Bias unpredictable in terms of effect on odds ratio
NONDIFFERENTIAL MISCLASSIFICATION BIAS OF EXPOSURE

- Misrepresent exposure status
- Cases/controls both overestimate and underestimate
- Bias underestimates effect
- Shifts relative risk (or odds ratio) toward 1.0
DIFFERENTIAL MISCLASSIFICATION BIAS OF EXPOSURE

- Misrepresentation of exposure different for cases/controls
- Either cases/controls misrepresent or bias is in different directions
- Also called “recall bias”
EXAMPLE OF DIFFERENTIAL MISCLASSIFICATION BIAS OF EXPOSURE

- Breast cancer and abortion in the Netherlands
- Two regions where Roman Catholics are 63 percent of population
- Two regions where Roman Catholics are 28 percent of population
- Willingness to report abortion

Overall adjusted OR for both areas = 1.9
HOW TO INTERPRET THIS STUDY

- Odds ratio in Roman Catholic areas exaggerated
- Cases report exposure
- Controls misrepresent exposures
- Effect exaggerated: relative risk shifted
CONFOUNDING

- Definition: mixing of effects between exposure, outcome, and third variable (confounder)
- Factors other than exposure that ↑↓ risk of disease
- Failure of comparison group to reflect exposed group (as if it were unexposed)
- Can lead to biased results

CRITERIA FOR CONFOUNDER

- Associated with exposure?
- Associated with disease?
- Not intermediate step in causal pathway

CONFOUNDING

Exposure or risk factor

Confounder

Disease
CONFOUNDING: A MATTER OF STRENGTH (MAGNITUDE) AND DIRECTION

- **Magnitude of confounding** = \( \frac{(RR \text{ crude} - RR \text{ adjusted})}{RR \text{ adjusted}} \)
  - Size bias depends on degree of association
  - Effect of multiple confounders may be large

- **Direction of confounding**
  - Exaggerate (positive confounding)
  - Hide (negative confounding)

Confounding Example 1

Positive Confounding

Oral contraceptive use + Myocardial infarction

Smoking

Confounding Example 2

Negitive Confounding

Oral contraceptive use  +  Myocardial infarction

Obesity