Lecture 9: Case-Control Studies

Mathew Reeves, BVSc, PhD
Credit to Michael Collins, MD, MS
and Michael Brown, MD, MS

Observational Studies
- Investigator has no control over exposure
  - Descriptive
    - Case reports & case series (Clinical)
    - Cross-sectional (Epidemiological)
  - Analytical
    - Cohort
    - Case-control
    - Ecological

Objectives - Concepts
- Define and identify case reports and case series
- Define, understand and identify (CCS)
  - Distinguish CCS from other designs (e.g., retrospective cohort)
- Understand the principles of selecting cases and controls
- Understand the analysis of CCS
  - Calculation and interpretation of the OR
- Understand the concept of matching
- Understand the origin and consequence of recall bias
  - Example of measurement bias
- Advantages and disadvantages of CCS
Case Report and Case Series

- Profile of a clinical case or case series which should:
  - Illustrate a new finding,
  - Emphasize a clinical principle, or
  - Generate new hypotheses

- Not a measure of disease occurrence!

- Usually cannot identify risk factors or the cause (no control or comparison group)
  - Exception: 12 cases with salmonella infection, 10 had eaten cantaloupe

Occasionally case reports or case series become very important...

- Famous Examples:
  - A report of 8 cases of GRID, LA County (MMWR 1981)
  - A novel progressive spongiform encephalopathy in Cattle (Vet Record, October 1987)
    - Clinical and pathologic findings of 6 cases reported
  - Twenty five cases of ARDS due to Hantia-virus, Four Corners, US (NEJM, 1993)
Case-Control Studies (CCS)

- An alternative observational design to identify risk factors for a disease/outcome.
- Question:
  - How do diseased cases differ from non-diseased (controls) with respect to prior exposure history?
  - Compare frequency of exposure among cases and controls
  - Effect —— cause.
- Cannot calculate disease incidence rates because the CCS does not follow a disease-free population over time

Case-control Study – Design

Select subjects on the basis of disease status

```
Disease

+  -
Exp +  a   b
Exp -  c   d
```

Schematic diagram of case-control study design

Example CCS - Smoking and Myocardial Infarction

Study: Desert island, population = 2,000 people, prevalence of smoking = 50% [but this is unknown], identify all MI cases that occurred over last year (N=40), obtain a random sample of N=40 controls (no MI). What is the association between smoking and MI?

<table>
<thead>
<tr>
<th>MI</th>
<th>+</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smk+</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Smk-</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

OR = a/d + b/c = 30/20 = 3.0 (same as the RR)

Examples of CCS

- Outbreak investigations
  - What dish caused people at the church picnic to get sick?
  - What is causing young women to die of toxic shock?
- Birth defects
- Drug exposure and heart tetralogy
- New (unrecognized) disease
  - DES and vaginal cancer in adolescents
  - Is smoking the reason for the increase in lung CA (1945's)
- Four CCS implicating smoking and lung cancer appeared in 1950s, establishing the CCS method in epidemiology

Essential features of CCS design

- Directionality
  - Outcome is exposure
- Timing
  - Retrospective for exposure, but case ascertainment can be either retrospective or prospective
- Rare or new disease
  - Design of choice if disease is rare or if a quick "answer" is needed (cohort design not useful)
- Challenging
  - The most difficult type of study to design and execute
- Design options
  - Population-based vs. hospital-based
Selection of Cases

- Requires case-definition:
  - Need for standard diagnostic criteria e.g., AMI
  - Consider severity of disease? e.g., asthma
  - Consider duration of disease
    - prevalent or incident case?

- Requires eligibility criteria
  - Area of residence, age, gender, etc

Sources of Cases

- Population-based
  - Identify and enroll all incident cases from a defined population
    - e.g., disease registry, defined geographical area, vital records

- Hospital-based
  - Identify cases where you can find them
    - e.g., hospitals, clinics.
  - But......
  - Issue of representativeness?
  - prevalent vs incident cases?

Selection of Controls

- Controls reveal the ‘normal’ or ‘expected’ level of exposure in the population that gave rise to the cases.

- Issue of comparability to cases – concept of the “study base”
  - Controls should be from the same underlying population or study base that gave rise to the cases?
  - Need to determine if the control had developed disease would he or she be included as a case in the study?
    - if no then do not include

- Controls should have the same eligibility criteria as the cases
Sources of Controls

- **Population-based Controls**
  - ideal, represents exposure distribution in the general population, e.g.,
    - driver's license lists (16+)
    - Medicare recipients (65+)
    - Tax lists
    - Voting lists
    - Telephone RDD survey
  - But if low participation rate = response bias
    (selection bias)

Sources of Controls

- **Hospital-based Controls**
  - Hospital-based case control studies used when population-based studies not feasible
  - More susceptible to bias

  - **Advantages**
    - similar to cases? (hospital use means similar SES, location)
    - more likely to participate (they are sick)
    - efficient (interview in hospital)
  - **Disadvantages**
    - do they have disease?
    - Don't select if risk factor for their disease is similar to the disease under study e.g., COPD and Lung CA
    - are they representative of the study base?

Other Sources of Controls

- **Relatives, Neighbors, Friends of Cases**

  - **Advantages**
    - similar to cases wrt SES/education/neighborhood
    - more willing to co-operate

  - **Disadvantages**
    - more time consuming
    - cases may not be willing to give information?
    - may have similar risk factors (e.g., smoke, alcohol, golf)
• Odds of exposure among cases = \( a / c \)
• Odds of exposure among controls = \( b / d \)

### Disease

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp +</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Exp -</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>( d_1 )</td>
<td>( d_2 )</td>
</tr>
</tbody>
</table>

### Analysis of CCS

The OR as a measure of association

- The only valid measure of association for the CCS is the Odds Ratio (OR).
- Under reasonable assumptions (the rare disease assumption) the OR approximates the RR.
- \( \text{OR} = \frac{\text{Odds of exposure among cases (disease)}}{\text{Odds of exposure among controls (non-disease)}} \)
  - Odds of exposure among cases = \( a / c \)
  - Odds of exposure among controls = \( b / d \)
  - \( \text{Odds ratio} = \frac{a/c}{b/d} = \frac{a \cdot d}{b \cdot c} \) (cross-product ratio)

### Example CCS - Smoking and Myocardial Infarction

Study: Desert island, population = 2,000 people, prevalence of smoking = 50% (but this is unknown), identify all MI cases that occurred over last year (N=40), obtain a random sample of N=40 controls (no MI). What is the association between smoking and MI?

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smk +</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Smk -</td>
<td>10</td>
<td>40</td>
</tr>
</tbody>
</table>

\( \text{OR} = \frac{a \cdot d}{b \cdot c} = \frac{30 \cdot 40}{20 \cdot 10} = 3.0 \) (same as the RRI)
Odds Ratio (OR)

- Similar interpretation as the Relative Risk
- OR = 1.0 (implies equal odds of exposure - no effect)
- ORs provide the exact same information as the RR if:
  - cases represent the target population
  - cases represent all cases
  - non-random ascertainment bias (e.g., case-control study is undertaken with population-based sampling)

- Remember:
  - OR can be calculated for any design but RR can only be calculated in RCT and cohort studies
  - The OR is the only valid measure for CCS
  - Publications will occasionally mislabel OR as RR (or vice versa)

Controlling extraneous variables (confounding)

- Exposure of interest may be confounded by a factor that is associated with the exposure and the disease i.e., is an independent risk factor for the disease

How to control for confounding

- At the design phase
  - Randomization
  - Restriction
  - Matching

- At the analysis phase
  - Age-adjustment
  - Stratification
  - Multivariable adjustment (logistic regression modeling, Cox regression modeling)
Matching is commonly used in CCS

- Control an extraneous variable by matching controls to cases on a factor you know is an important risk factor or marker for disease
  - Example:
    - Age (within 5 years)
    - Sex
    - Neighbourhood
  - If factor is fixed to be the same in the cases and controls then it can't confound

Matching

- Analysis of matched CCS needs to account for the matched case-control pairs
  - Only pairs that are discordant with respect to exposure provide useful information
  - McNemar's OR = b/c
  - Conditional logistic regression

- Can increase power by matching more than 1 control per case e.g., 4:1
  - Useful if few cases are available

Matched CCS - Discordant pairs

Match 40 controls to 40 cases of AMI so they have the same age and sex. Then classify according to smoking status.

```
   Controls
     +  -
   +  32  20
   -  10  18

80
```

McNemar's OR = \( \frac{b \cdot c}{a \cdot d} \) = \( \frac{20 \cdot 10}{32 \cdot 18} \) = 2.9
Over-matching

- Matching can result in controls being so similar to cases that all of the exposures are the same

- Example:
  - 8 cases of GRID, LA County, 1981
  - All cases are gay men so match with other gay men who did not have signs of GRID
  - Use 4:1 matching ration i.e. 32 controls
  - No differences found in sexual or other lifestyle habits

Recall Bias

- Form of measurement bias
  - Presence of disease may affect ability to recall or report the exposure.
  - Example – exposure to OTC drugs during pregnancy use by moms of normal and congenitally abnormal babies.
  - To lessen potential:
    - Blind participants to study hypothesis
    - Blind study personnel to hypothesis
    - Use explicit definitions for exposure
    - Use controls with an unrelated but similar disease
      - e.g., heart attack (cases), hypovolans (controls)

Other issues in interpretation of CCS

- Beware of reverse causation
  - The disease or sub-clinical manifestations of it results in a change in behaviour (exposure)

- Example:
  - Obese children found to be less physical active than non-obese children
  - Multiple sclerosis patients found to use more multi-vitamins and supplements
**CCS - Advantages**

- Quick and cheap (relatively)
  - so ideal for outbreaks
    (http://www.cdc.gov/hs/casestudies/casestudies.htm)
- Can study rare diseases (or new)
- Can evaluate multiple exposures (fishing trips)

**Case-control Studies - Disadvantages**

- uncertain of E ————— D relationship (esp. timing)
- cannot estimate disease rates
- worry about representativeness of controls
- inefficient if exposures are rare
- Bias:
  - Selection
  - Confounding
  - Measurement (especially recall bias)