Case-Control Studies

Case-Control Study: Definition

- A group of individuals with a disease (cases) are compared with a group of individuals without the disease (controls).
- Also called a retrospective study because it starts with people with disease and looks backward for previous exposures which might be relevant to the development of the disease.

Case-control studies

- Case-control study is a method of sampling a population in which cases are identified and a sample of the source population that gave rise to the cases is also identified and enrolled (controls).

What is the source population

- The population from which the cases are derived.
- The source population can be difficult to identify.

Design of Case-Control Studies

- Cases should be representative of all those with disease in the source population.
- Controls should be representative of all those without disease in source population.

<table>
<thead>
<tr>
<th></th>
<th>cases</th>
<th>controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>exposed</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>unexposed</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Total cases = a + c  Total controls = b + d
Case Ascertainment

- Where can you find cases?

Selection of Study Cases

- Hospitals
- Physicians practices
- Disease Registries
- HMOs
- Others

Selection of Study Cases

- Incident or Prevalent Cases?????

Incident vs. prevalent cases

- Incident cases-newly diagnosed - takes time for them to occur - study takes a long time to complete

- Prevalent cases-existing cases - available now but may include many long survivors - not representative of all cases

Cases

Population-based cases:
- Random sample (ideally)
- All individuals from a defined geographic/political area

Hospital-based/medical care system cases:
- In-patient (admitted) patients to hospitals/clinics
- Regional disease registries (e.g., cancer reg.)
- Records of large ambulatory-care practices (e.g., MD offices, HMO's)

Case Ascertainment

- Advantages/disadvantages of finding cases -
- In the community?
- In the hospital?
- From medical/physician records?
- From death certificates?
How can we ascertain disease status?

**How do we know a case is truly a case?**

- Case registries
- Physician's medical records
- Death certificates
- Hospital admission/discharge records
- Pathology department log books
- Health Insurance Forms

*Any source that includes information about disease status (accurate)*

---

**Case control studies**

- Finding a comparison population - controls...........

---

**Case-control studies**

- Why do we need controls???

**Why do we need controls?**

Purpose: to compare the history of exposure among the cases with that of individuals from the same source population who are free of the disease (at the time of the study).

- Controls are a sample of the population that produced the cases- they provide information on the exposure distribution of the source population.
- Selection of an appropriate comparison (control) group is the most difficult and critical issue in the design of case-control studies.

---

**Case-control studies**

- Where can you find controls?

**Selecting Controls**

- Should be a sample of the population that produced the cases (source population or study base)
- Controls, if they had gotten the disease, would have been a case (in the study)
- Controls must be sampled independent of exposure status
Sources of Controls

- **General population controls**
  *Controls come from the population*

- **Hospital-based controls**
  *Controls come from the hospitals in which cases arise*

- **Special control series**
  *Controls come from family/friends/relatives*

General Population Controls

- Used when cases are selected from the population in which the study is based, or
- As an alternate control group in hospital-based case-control studies

**Ways to collect information**

- Canvassing households in targeted neighborhoods
- Random-digit telephone dialing
- Identification of individuals from population registers or voting lists

General Population Controls: Advantages

- Can be selected randomly
- Minimizes potential for selecting certain individuals with specific characteristics that may alter results
- Representative of the source population

General Population Controls: Disadvantages

- Can be costly and time-consuming
- Population lists (for sampling) are not always available
- Quality of information may differ between cases and controls
- Healthy individuals may be less motivated to participate
- A subset of the population actually are diseased

Hospitalized Based Controls: Advantages

- Easily available
- Readily identifiable
- More likely to be aware of prior exposures
- More likely to be affected by same intangible factors that influenced cases to come to particular hospital
- More willing to participate

Hospitalized Based Controls: Disadvantages

- People who are in hospital are there because they are SICK
- Disease may be associated with the exposure of interest
- **may not be representative of the source population in terms of exposure (may smoke more etc)**
Hospital Controls

- Use all other patients or selected diagnoses?

- Who should be excluded?

- Who is different from the general population - cases or controls?????

Special control groups

Controls come from spouse/family/friends/relatives

Advantages:
- Healthy
- More likely to participate
- May also offer a degree of control regarding potential confounding factors

Disadvantage:
- Potential for inadvertently selecting a group too similar to cases

Which control group to choose?

- No perfect control group

- Select several control groups if feasible (given cost and logistic constraints)

- If study results are consistent across control groups, this suggests that results are real/valid

- If different effects are observed, it still provides useful information as to nature of the association or possible biases that might be present

Selecting Controls

- Should they be living or dead?

- If the case is deceased, should you select a control who is also deceased?

- Advantage - Proxy respondents for both

- Disadvantage - not representative of the overall (source) population

Multiple Controls

- May be of same type or different - one set of hospital controls and one set of community controls

- Using more controls of the same type: there are often more controls available than cases - increases power of study - Most increase is gained up to a ratio of 4 controls/case

Number of controls per case

In studies in which a sufficiently large number of cases and controls are available and participate, the optimal ratio of cases:controls is 1:1.

When relatively few cases can be found (n cases if "fixed") can increase power by increasing number of controls:case. However, note law of diminishing returns:

![Graph showing diminishing returns](image_url)
Matching
- Selecting controls so that they are similar to cases in certain characteristics which may be associated with the disease being studied (age, sex, race etc)
- Group matching (frequency matching)
- Individual matching (pair matching)

Group Matching
- Selecting controls so the proportion of controls with a characteristic such as age or sex is the same proportion as in the case group
- Must select cases first then select controls in the same proportions for selected variables

Individual (Pair) Matching
- A control(s) is selected for each case
- Similar in specified characteristics (age, race, sex etc)
- Often used with hospital controls (example: next admission of same sex and race, year of birth +/- 2 yrs)

Problems with Pair Matching
- Finding the matching control - more difficult the more characteristics matched
- Cannot study a characteristic which the cases and controls are matched on
- Unplanned matching may occur - example: community controls may be similar in SES

Matching
Conclusion:
- Match only on characteristics/variables that are risk factors for the disease which are not to be investigated in the study