Introduction of case-control study

Hirohide Yokokawa, M.D., Ph.D.
Department of General Medicine,
Juntendo University School of Medicine
What is case-control Study?

- Overview of study design
- Purpose of case-control study

Major types of case-control studies

Procedure of case-control study

Risk assessment in case-control study

Example of case-control study
Overview of study designs

Meta-Analysis of RCT

RCT

Cohort study

Case-Control study
A case-control study is designed to help determine if an exposure is associated with an outcome.

It is always retrospective because it starts with an outcome then traces back to investigate exposures.
Basic design of cohort study

Define Study Participants

EXPOSED

DISEASE DEVELOPS

DISEASE DOES NOT DEVELOP

NOT EXPOSED

DISEASE DEVELOPS

DISEASE DOES NOT DEVELOP

Present

Future
Basic design of case-control study

- **EXPOSED**
  - **HAVE THE DISEASE**
    - "CASES"
  - **DO NOT HAVE THE DISEASE**
    - "CONTROLS"
- **NOT EXPOSED**
  - **HAVE THE DISEASE**
  - **DO NOT HAVE THE DISEASE**
Compared to prospective cohort studies, case-control study tends to be less costly and shorter in duration. In several situations they have greater statistical power than cohort studies, which must often wait for a 'sufficient' number of disease events (target disease) to accrue.
## Advantage and Disadvantage of Case-Control Study compared to Cohort study

<table>
<thead>
<tr>
<th></th>
<th>Cohort Study</th>
<th>Case-Control Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measures</strong></td>
<td>Incidence Rate, Relative Risk (RR)</td>
<td>Odds Ratio (OR) only</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>Expensive</td>
<td>Inexpensive</td>
</tr>
<tr>
<td><strong>Study term</strong></td>
<td>Long term</td>
<td>Short term</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>Need large sample</td>
<td>Powerful with small sample cases</td>
</tr>
<tr>
<td><strong>Exposure</strong></td>
<td>Good for rare exposure</td>
<td>Limited to rare exposure</td>
</tr>
<tr>
<td><strong>Disease</strong></td>
<td>Poor potential for rare Possible for several disease</td>
<td>Good for rare disease Only one disease</td>
</tr>
<tr>
<td><strong>Causal</strong></td>
<td>Potentially strong</td>
<td>Potentially less strong</td>
</tr>
<tr>
<td><strong>Generalization</strong></td>
<td>Possibly generalizable</td>
<td>Probably not generalizable</td>
</tr>
</tbody>
</table>
Which study design do you choose?

A Vietnamese neurologist has a hypothesis that high coffee consumption in youth may be associated with Pick disease in middle or elder age.
Which study design do you choose?

A Vietnamese clinician has a research question whether the ratio of LDL and HDL cholesterol (L/H ratio) is associated with cardiovascular events among Vietnamese diabetic patients.
Review feasibility of the planned study considering study question, target disease, exposure, cost, study period, etc.!
Major types of case-control study

Population-based case-control study
⇒ Cases and controls are recruited from population.

Case-control study nested in cohort study
⇒ Cases and controls are registered in a cohort study.

Hospital based case-control study
⇒ Cases and controls are patients who are hospitalized or outpatients.
Procedure of case-control study

1. "CASES" - HAVE THE DISEASE
2. "CONTROLS" - DO NOT HAVE THE DISEASE

EXPOSED

NOT EXPOSED

EXPOSED

NOT EXPOSED
① Identification of CASEs and CONTROLS (Present).
② Measurement of exposure and determination of EXPOSED or NON-EXPOSED (Past).
③ Expected findings if the exposure is associated with DISEASE
Cases; Cases should be identified from patients with your interests systematically.

Controls; Controls should be recruited from appropriate sources.

You are highly recommended to define “Eligible criteria” and “Exclusion criteria” prior to recruiting of cases and controls.
Exposure history can be collected by the past information using interview, biomarkers or medical records.

NOTICE; Bias and Confounding
“Any systemic error in the design, conduct or analysis of a study that results in a mistaken estimate of an exposure’s effect on the risk of disease.”
Types of bias

- Selection bias
  - Results from an unrepresentative sample in sampling.

- Detection bias
  - Occurs when a phenomenon is more likely to be observed for a particular set of study subjects.

- Information bias (Recall bias, Family information bias)
  - Occurs when we place too much attention on information, even when it is not strictly relevant.
<table>
<thead>
<tr>
<th>Some types and sources of information bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias in abstracting</td>
</tr>
<tr>
<td>Bias in interviewing</td>
</tr>
<tr>
<td>Bias from surrogate interviews</td>
</tr>
<tr>
<td>Recall bias</td>
</tr>
<tr>
<td>Reporting bias</td>
</tr>
</tbody>
</table>
“A third variable or a mediator variable, can adversely affect the relation between exposure and outcome.”
Causal

Due to Confounding

https://explorable.com/confounding-variables
The observed association between “eating meat” and “colon cancer” may result from confounding by smoking.
### Before

**Study design**
- Randomization (Intervention)
- Restriction (Cohort, Case-control)
- Matching (Case-control)

### After

**Data analysis**
- Adjustment
- Stratification
- Multivariate analysis
Restriction

A method that limits participation in the study to individuals who are similar in relation to the confounder

Problems

- Reduces eligible population
- Limits generalizability
Matching

Controls and cases are similar in variables, which may be related to the topic we are studying BUT are not of interest in themselves.

Problems
- Reduces eligible population
- Limits generalizability
Example of matching

Cases

No.1
No.2
No.3
No.4
No.5

Controls

No.1
No.2
No.3
No.4
No.5

Matched by SEX and AGE

EXPOSED

NOT EXPOSED

NOT EXPOSED
Problems with Matching

- Practical limit on how many you can match on
- Cannot analyze the association for the matched variables
- Sometimes difficult, expensive (Requires large number of cases and controls)
- Over matching (Matched by factors strongly related to the exposure which is your main interest.)
Odds = prevalence / (1 – prevalence)
Odds ratio = (Odds in cases) / (Odds in controls)
   = (a/c)/(b/d) = ad/bc
Interpretation of Odds ratio

<table>
<thead>
<tr>
<th>OR</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR = 1</td>
<td>No association</td>
</tr>
<tr>
<td>OR ≥ 1</td>
<td>Risk factor</td>
</tr>
<tr>
<td>OR ≤ 1</td>
<td>Protective factor</td>
</tr>
</tbody>
</table>

95% confidence interval (CI); A 95% probability which the interval includes the true odds ratio (OR)

If 95% CI range includes “1”, it is not statistically significant since it could be either a risk factor (OR ≥ 1) or a protective factor (OR ≤ 1). If 95% CI range is greater than 1, the exposure is a significant risk factor (OR ≥ 1) with a probability of higher than 95%.