ANALYTICAL STUDIES
Case-Control study
by
Dr. Asaad Q. Al-Yassen
Assist. Professor of Dermato-
Epidemiology & Occupational Health
Head of Department of Family
and Community Medicine
College of Medicine
University of Basrah

Objectives
The aim of this lecture is to understand:
1. The basic design of case control study

2. The principles of selecting cases and controls

3. The concept of matching

4. Analysis and interpretation

5. The advantages and disadvantages of case control studies

What is case-control study?
The case-control study is a retrospective, analytic study. It is an observational rather than interventional study.
i.e. It is an observational epidemiological study of persons with the disease (or other outcome variable) of interest and a suitable control (comparison / reference) group of persons without the disease.
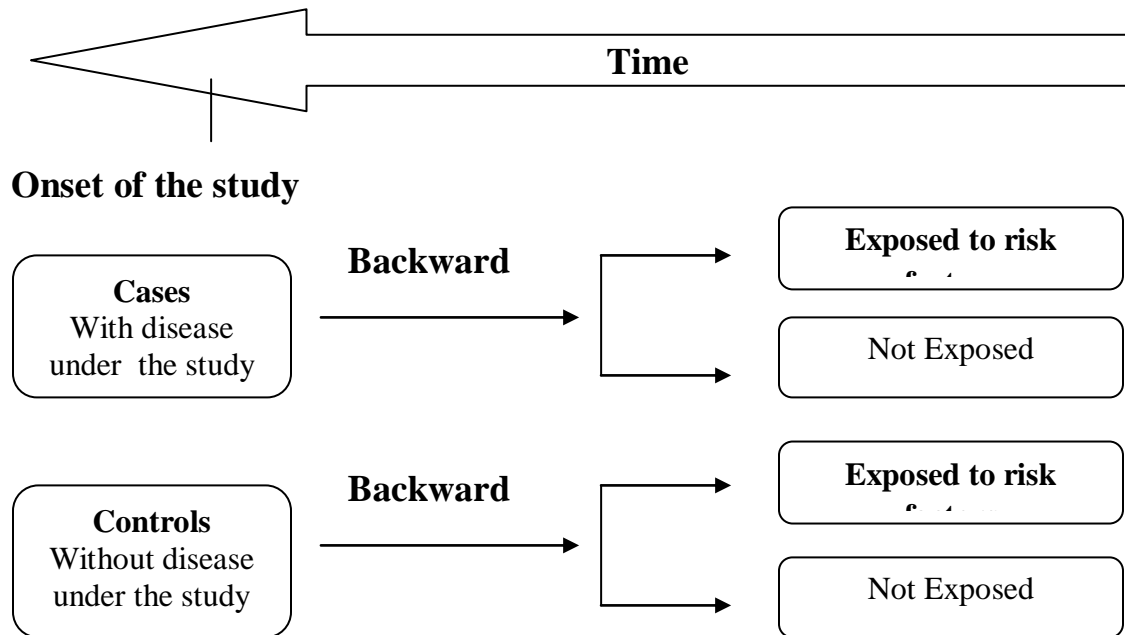
Cases and controls has three distinct features:
1. Both exposure and outcome have occurred before the start of the study.
2. The study proceeds backwards from effect to cause.
3. It uses a control or comparison group to support or refute an inference.

Information regarding exposure is collected and can be evaluated in terms of presence or absence of exposure or in terms of level of exposure.
Exposure factors can be external (an environmental exposure) or internal (a subject's genetic makeup).

Two groups of persons are studied. The first consists of subjects who have the disease under study at the time of the beginning of the study (cases). The second group consists of subjects who are free from the disease under study (controls). Both cases and controls are preferably matched for age and sex. Sometimes they are matched for other variables but overmatching is to be avoided.

```
<------------------------  Time  ---------------------->
```

Onset of the study

Cases With disease under the study	**Backward** →	**Exposed to risk** Not Exposed
Controls Without disease under the study	**Backward** →	**Exposed to risk** Not Exposed

The basic steps in conducting a case control study are:

1. Selection of cases and controls.
2. Matching for known confounding variables (at least age and sex).
3. Measurement of past exposure in both groups.
4. Analysis and interpretation.

Identification of Cases
To conduct a case–control study, we must start by identifying a group of people who have the disease in question, typically called cases**.**

Sources of cases
- Hospital

- Physicians clinic

- Registries or clinic records listing all patients having a certain disease.

- Community [in a predefined group such as schools, and some industrial plants].

Identification of Controls
- Individuals without the disease.

- Controls should be representative of the source population from which cases were derived.

- The control should be at risk of the disease.

- The controls should resemble the cases in all respects except for the presence of disease

Sources of Controls
- **Random sample of controls** from the source population from which the cases came. (*ideal, represents exposure distribution in the general population*)
- **Special groups** such as friends, neighbors or relatives of the cases.

- **Hospital- or clinic-based controls** are frequently used but often are not representative of the source population. *[more susceptible to bias (they have a disease)]*

Matching: The process by which we select controls in such a way that they are similar to cases with regard to certain pertinent selected variable which are known to influence the outcome of disease and if not adequate for comparability could distort or confound the result.

Analysis of data
1. Arrangement of contingency table (2×2 table)

2. Calculation of prevalence of exposure in cases and in controls

3. Calculation of Odds ratio
4. Statistical test – Chi squared test

Odds ratio:
   -   A measure of association typically used to quantify the strength of association between a potential risk factor (exposure) and an outcome.
   - It is frequently used in case–control studies, in which incidence rates cannot be calculated.

Interpretation of Odds ratio

* OR = 1 **-** Odds of exposure among cases and control are same.
- Exposure is not associated with disease

* OR > 1 - Odds of exposure among cases are higher than controls
- Exposure is positively associated with disease.

* OR < 1 - Odds of exposure among cases are lower than controls
- Exposure is negatively associated with disease.

To illustrate the study design, we identify a number of children who are suffering from acute respiratory infection (pneumonia) . Suppose the number is 240. An equal or more number of children matched for age and sex but are free from acute respiratory infection at the time of the study is also selected (controls).

The analysis and interpretation:

The first step is to present the data in a 2x2 table

History of smoking	Cases with Pneumonia	Children without Pneumonia	Total
Positive	170	200	370
Negative	70	180	250
Total	240	380	620

The second step is to calculate the percentage of smokers (exposed) among parents of cases and controls.

$$\textbf{Percentage of smokers among parents of cases} = \frac{\textbf{170}}{\textbf{240}} \times \textbf{100} = \textbf{70.8\%}$$

$$\textbf{Percentage of smokers among parents of controls} = \frac{\textbf{200}}{\textbf{380}} \times \textbf{100} = \textbf{52.6\%}$$

It is clear that the habit of smoking was more frequent among parents of cases as compared to parents of controls. Cases were more likely to be children of smoking parents.

The third step is to measure the strength of association between parental smoking and acute respiratory infection. This is achieved by calculating a proxy measure to the relative risk. This measure is called the Odds ratio.

$$\text{The Odds Ratio} = \frac{\textbf{Cases exposed X Controls not exposed}}{\textbf{Cases not exposed X Controls exposed}}$$

$$= \frac{\textbf{170 X 180}}{\textbf{70 X 200}} = \textbf{2.2}$$

This means that the risk of acute respiratory infection among children of smoking parents is nearly double the risk among children of nonsmoking parents.

The fourth step is to perform a suitable statistical test to ascertain any significant association. Chi-squared test is the usual test performed on such data.

Note:
It is possible to estimate the attributable risk in case control studies by using the following formula:

$$\text{Attributable risk} = \frac{\textbf{b (r - 1)}}{\textbf{b (r − 1) + 1}}$$

Where **r** = Odds ratio

     b = The proportion of people in the general population with the risk factor.

If we assume that the proportion of smokers in the general population is 0.4 (40%)

$$\textbf{The attributable risk} = \frac{0.4\,(2.2 - 1)}{0.4\,(2.2 - 1\,) + 1} = \frac{0.48}{1.48} = 0.32 \text{ or } 32\%$$

Advantages

1. Efficient in time & cost ·
2. Etiology of long latency period can be studied
   e.g. chronic diseases ·
3. Multiple exposures could be studied·
4. Useful for studying rare disease

Disadvantages:

1. Cannot be used to compute incidence rates. ·

2. Not easy to elucidate the chronologic order of the exposure and disease. ·

3. They are not good for studying rare exposures, and can study only one outcome ·

4. Disease and exposure have already occurred and potential for bias remains

 - selection
 - recall
 - interviewer

Comparison of case – control and cohort studies

Item of comparison	Case-control	Cohort
1. No. of subjects	Small	Large
2. Time	Short	Long
3. Cost	Lower	Higher
4. Organization	Easier	More difficult
5. Interpretation of results	More difficult	Easier
6. Usefulness for rare disease	Useful	Not useful
7.Bias is likely in ascertainment	of exposure	of diagnosis
8. Usefulness for risk measurement	Less useful	More useful
9. Usefulness for causal criteria	Less useful	Very useful

A fourfold table

Prospective (Cohort)

Retrospective (Case-Control)

EXPOSURE

DISEASE

cases present | controls absent

present exposed | a | b

absent Not exposed | c | d

Total | Total

Mausner, 1985