มหาวิทยาลัยราชภัฏนครปฐม

# CHAPTER 5

## Epidemiology Study Designs

Episode 5.2: 1) Cohort study

Email: wanpenw@webmail.npru.ac.th

# CHAPTER CONTENTS



## Epidemiological Study Designs

2

Cohort study or Prospective study

3

Case-control study or Retrospective study

# Prospective or cohort study

มหาวิทยาลัยราชภัฏนครปฐม
Nakhon Pathom Rajabhat University

## Definition

- A prospective or cohort study is an **analytical epidemiological study**, a longitudinal study that follows over time a group of similar individuals (cohorts) who differ with respect to certain exposures (factors) under study, to determine how these exposures (factors) affect rates of a certain outcome (disease).

## Objectives

- To study and test the relationship between the factors that are expected to cause disease and disease progression.

## Keywords

- An analytical epidemiological study
- Observation
- Follow-up of healthy samples
- Starting from free of disease
- Follow forward until the expected disease (outcome) arising

# The nature of the cohort study

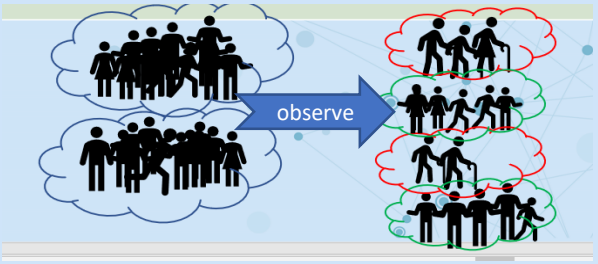


**Question of a study**

- Exposure/Factor

?

- Outcome/Disease




**Initially, the study population was divided into individuals who did not have the disease.**

- Exposed group

follow-up

- 

no disease

- Non-exposed group

follow-up

**The follow-up was used for a long enough time to measure whether the suspected factor was causing the disease.**

The pathogenesis of that factor:

- : Induction period
- : Incubation period

**Disease / Outcome**

follow-up

- Incidence of the exposure group
- (**disease**)

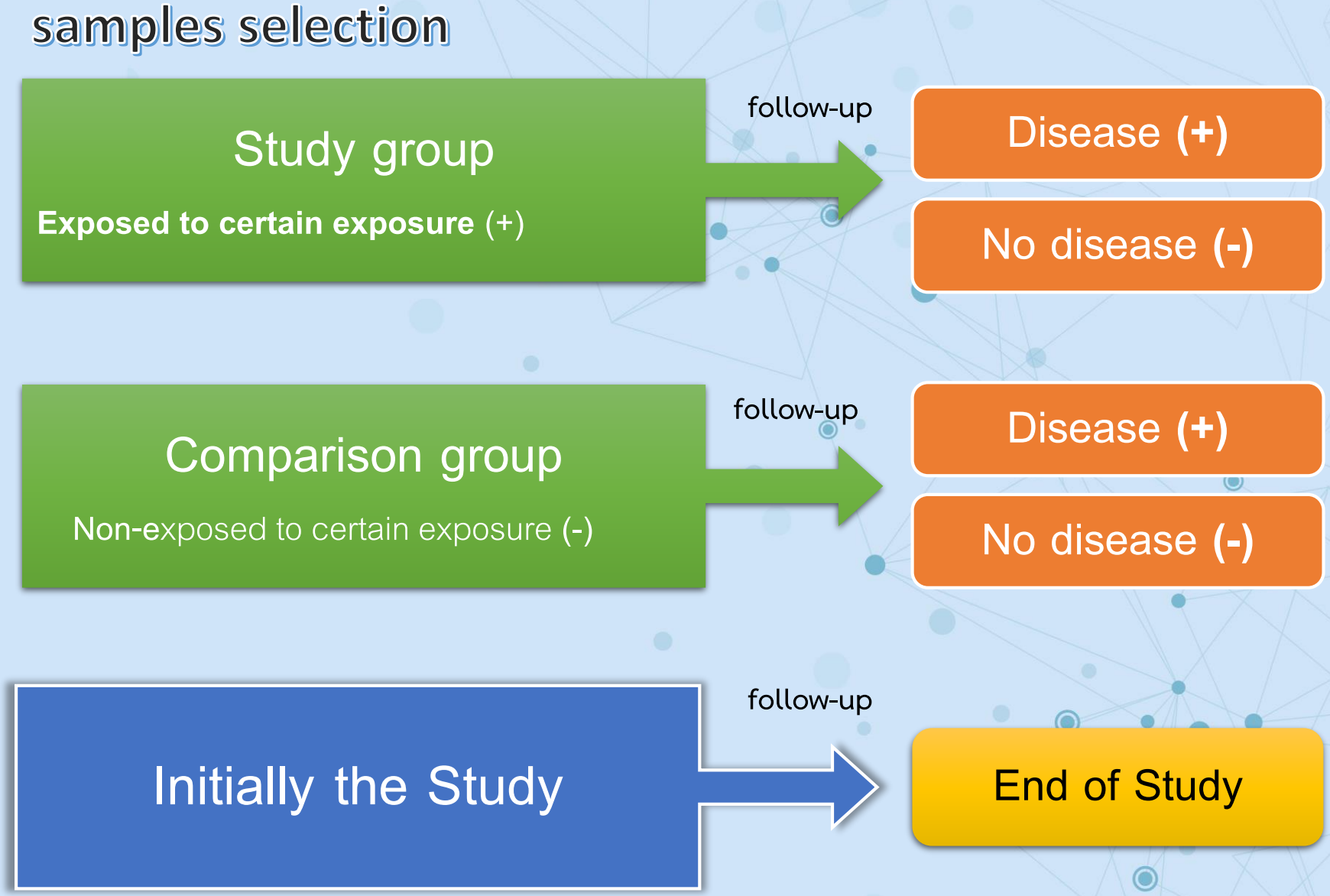
follow-up

- Incidence of the non-exposure group
- (**disease**)

# The nature of the cohort study

observation without manipulation

test the relationship between the exposure (factor) and disease



# Data collection and measurement of exposures and outcomes



## Measurement of Exposures

- Questionnaire
- Interview form
- Biochemical or laboratory tests
- Physical examination
- Special examination
- - x-ray
- - EKG
- Surveying the environment

## Measurement of outcomes

- Questionnaire
- Biochemical or laboratory tests
- Physical examination
- Special examination
- - x-ray
- - EKG
- Death information
- - death certificate
- - Department of Provincial Administration
- Illness information
- - medical records
- - disease registration record
- - patient report  (directly)

# Data collection and calculation

| Study design | Time | | |
|---|---|---|---|
| | Past | Present | Future |
| Cohort study or Prospective study | - | Collecting personal data and **natural exposure** | Collecting personal data and **outcome** |
| | | Initially the Study | End of Study |

2x2 table

| Exposure | Disease (+) | No disease (-) | Total | Calculation outcome |
|---|---|---|---|---|
| Exposed (+) | a | b | a + b | **I ex** = a / (a+b) |
| Non-exposed (+) | c | d | c + d | **I non** = c / (c+d) |
| Total | **a + c** | **b + d** | **a + b + c + d** | |

## Calculation

Incidence rate of exposure group (I ex) = a / (a + b)

Incidence rate of non-exposure group (I non) = c / (c + d)

Relative Risk or Risk Ratio = (I ex) / (I non) or $\frac{a / (a+b)}{c / (c+d)}$

# Outcomes: Incidence Rate of Both Groups



1. Cumulative Incidence (CI):

$$= \frac{\text{Number of new cases of disease or injury during specified period}}{\text{Time each person was observed, totaled for all persons}} \times K$$

The value of $k$ (constant) can be 100, 1000, 10,000

2. Incidence Density (ID) or Person-time Incidence

$$\text{Person-time rate} = \frac{\text{Number of new cases of disease or injury during specified period}}{\text{Sum of the person-time of the at-risk population (Total person-time at risk)}} \times K$$

# Person-time at risk calculation...1

*A study group: exposed group*

## Example

Investigators enrolled 500 men who were smokers in a study group and followed them annually for four years to determine the incidence rate of hypertension. After one year, one had a new diagnosis of hypertension, but 50 had been lost to follow-up. After two years, two had a new diagnosis of hypertension, and another 10 had been lost to follow-up. After three years, another five had new diagnoses of hypertension, and 20 had been lost to follow-up. After four years, another 6 had new diagnoses with hypertension, and 30 more had been lost to follow-up.

# Person-time at risk calculation...2

Person-time at risk = Number Persons × Period of Time

| Number | Duration (Year) | Lost FU. | Incidence | Person-time at risk |
|---|---|---|---|---|
| Y 0 = 500 | 0 | 50 | - | 50x0 = 0 |
| Y1 = 450 | 1 | 10 | 1 | 10x1 = 20 |
| Y2 = 440 | 2 | 20 | 2 | 20x2 = 40 |
| Y3 = 410 | 3 | 30 | 5 | 30x3 = 90 |
| Y4 = 390 | 4 | - | 6 | 390x4 = 1,560 |

Numerator = number of new cases of hypertension = 1 + 2 + 5 + 6 = 14

Denominator = Person-time at risk = 0+20 + 40 + 90 + 1,560 = 1,710

Incidence Density (ID) or Person-time Incidence of Exposed Group
= 14/1,710 = .008 / person-year (Or 8/1,000 person-years)

# Person-time at risk calculation...3

Comparison group: non-exposed

## Example

Investigators enrolled 500 men who were non-smokers in a comparison group and followed them annually for four years to determine the incidence rate of hypertension. After one year, no one had a new diagnosis of hypertension, but 10 had been lost to follow-up. After two years, one had a new diagnosis of hypertension, and another 20 had been lost to follow-up. After three years, another one had new diagnoses of hypertension, and 5 had been lost to follow-up. After four years, another 2 had new diagnoses with hypertension, and 50 more had been lost to follow-up.

# Person-time at risk calculation...4

Person-time at risk = Number Persons × Period of Time

| Number | Duration (Year) | Lost FU. | Incidence | Person-time at risk |
|---|---|---|---|---|
| Y 0 = 500 | 0 | 10 | - | 10x0 = 0 |
| Y1 = 450 | 1 | 20 | 0 | 20x1 = 20 |
| Y2 = 440 | 2 | 5 | 1 | 5x2 = 10 |
| Y3 = 410 | 3 | 50 | 1 | 50x3 = 150 |
| Y4 = 390 | 4 | - | 2 | 425x4 = 1,700 |

Numerator = number of new cases of hypertension = 0 + 1 + 1 + 2 = 4

Denominator = Person-time at risk = 0+20 + 10 + 150 + 1,700 = 1,880

Incidence Density (ID) or Person-time Incidence of non-exposed Group
= 4/1,880 = .002 / person-year (Or 2/1,000 person-years)

# Relative Risk or Risk Ratio (RR) (ค่าความเสี่ยงสัมพัทธ์)



$$\text{Relative Risk (RR)} = \frac{Risk\ of\ disease\ (incidence\ proportion)\ in\ exposed\ group}{Risk\ of\ disease\ (incidence\ proportion)\ in\ comparison\ group\ (non-exposed)}$$

$Risk\ of\ disease\ in\ exposed\ group$ (I ex) = 14/1,710 = .008 (8/1,000 person-years)

$Risk\ of\ disease\ in\ non-exposed$ (I non) = 4/1,880 = .002 (2/1,000 person-years)

Relative Risk or Risk Ratio = (I ex) / (I non) , = .008/.002 x 1,000 = 4

The RR of 4.0 means that people who smoke are 4 times more likely to develop hypertension than those who do not smoke.

**Relative Risk or Risk Ratio (RR) Interpretation**

- A risk ratio of 1.0 indicates identical risk among the two groups.
- A risk ratio greater than 1.0 indicates an increased risk for the group in the numerator, usually the exposed group.
- A risk ratio less than 1.0 indicates a decreased risk for the exposed group, indicating that perhaps exposure actually protects against disease occurrence.

# The biases of the cohort study design



**1. Selection bias**

1.1 Choosing an inappropriate study group can be solved by sampling

1.2 Inappropriate selection of comparison groups can be solved by using an internal comparison group.

**2. Information bias**

2.1 Misclassification of exposures or risk factors

2.2 Inaccurate classification of outcome or disease

2.3 Lost of follow-up

3. Confounding bias such as age, smoking

3.1 Bias arising during follow-up, can be solved by clarification and understanding of the selected samples and by emphasizing the importance of cooperation during participation in the study.

3.2 Study design and data collection biases can be solved by assigning both groups to have the same interference factor characteristics.

These biases can be solved by good planning

# Advantages and limitations of the cohort study



**Advantages**

- Incidence can be directly calculated
- The relative risk (RR) can be directly estimated
- Temporal association of the exposure with the outcome can be found
- Disseminated certain biases like recall bias and interviewer's bias
- More than one outcome of the risk factor can be studied

**Limitations**

- Require resources, time, money, and personnel
- Rare diseases limitation
- Long periods of follow up needed, lost of follow-up issue
- Ethical issue, wait and watch may be unethical
- Only one or a few risk factors can be studied.

# Summary



Cohort or Prospective study

- This is a prospective follow-up analytical study without any action on the study interests.
- Typically examines multiple health effects of exposure after observation of a group of people with factors and a group of people without factors and followed for a period of time to see if the “incidence rate” of those people with factor factors differs from the comparison group that does not have the factors studied.
- Follow-up must be long enough to measure whether the hypothesized factor is pathogenic, and must be at least equal to the induction period of the factor or the incubation period.

มหาวิทยาลัยราชภัฏนครปฐม
Nakhon Pathom Rajabhat University

# Thank You for Your Attention

Email: wanpenw@webmail.npru.ac.th

to be continued