

Cohort Studies

Objectives:

At the end of the lesson the students will be able to:

1. Identify and characterize basic Cohort Study design
2. Identify the steps to carry out the prospective cohort study.
3. Calculate measure of association in cohort studies.
4. List the advantages and disadvantages of cohort study.
5. Identify common sources of bias for of cohort study design
6. Identify applications in clinical and in non-clinical settings.
7. Compare prospective with retrospective cohort studies.

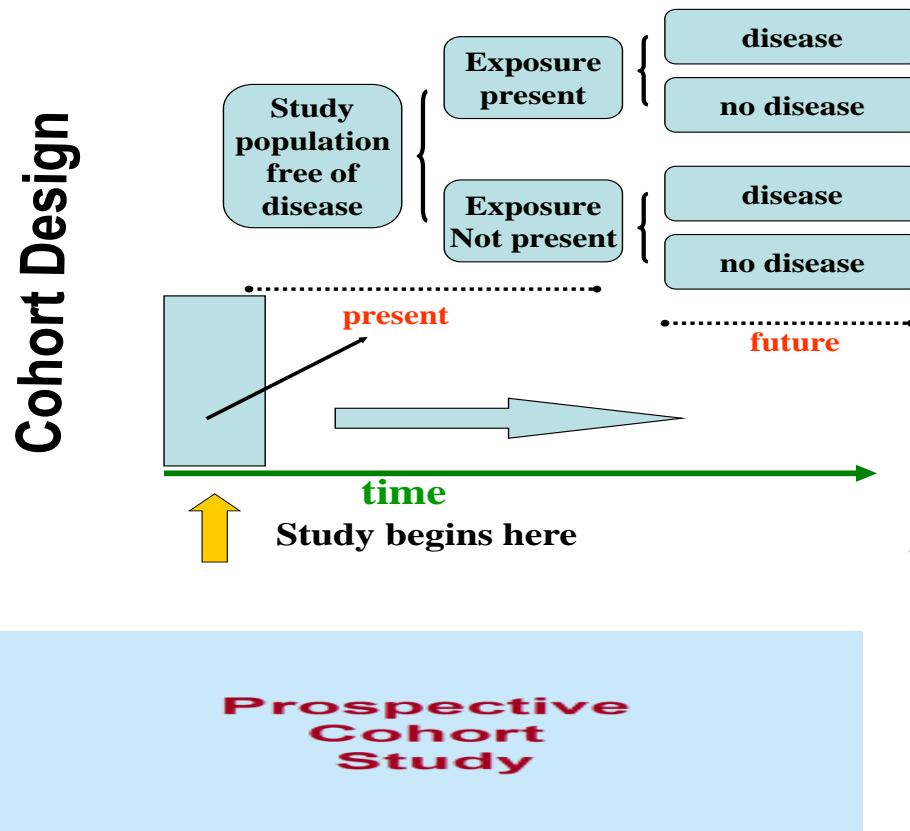
Cohort Studies:

- a. an “observational” design where the investigator categorizes individuals on the basis of **exposure** thus comparing individuals with a known risk factor or exposure with others without the risk factor or exposure. also called follow-up or incidence studies
- b. Looking for a difference in the risk (incidence) of a disease over time.
- c. Study groups are followed up to determine frequency of disease.

What is a cohort?

- Cohort is a group having a common characteristic
- Example: A smoker’s cohort means all are smokers in that group
- A cohort, which is exposed to a suspected factor but not yet developed the disease, is observed and followed over time.
- Then, the incidence of the disease is measured directly.
- The study and the comparison groups should be free from the disease of interest, similar in demographic characteristics.

- Cohort studies provide the best information about the causation of disease and the most direct measurement of the risk of developing disease.
- There are two types of cohort study: *Prospective and Retrospective cohort*.



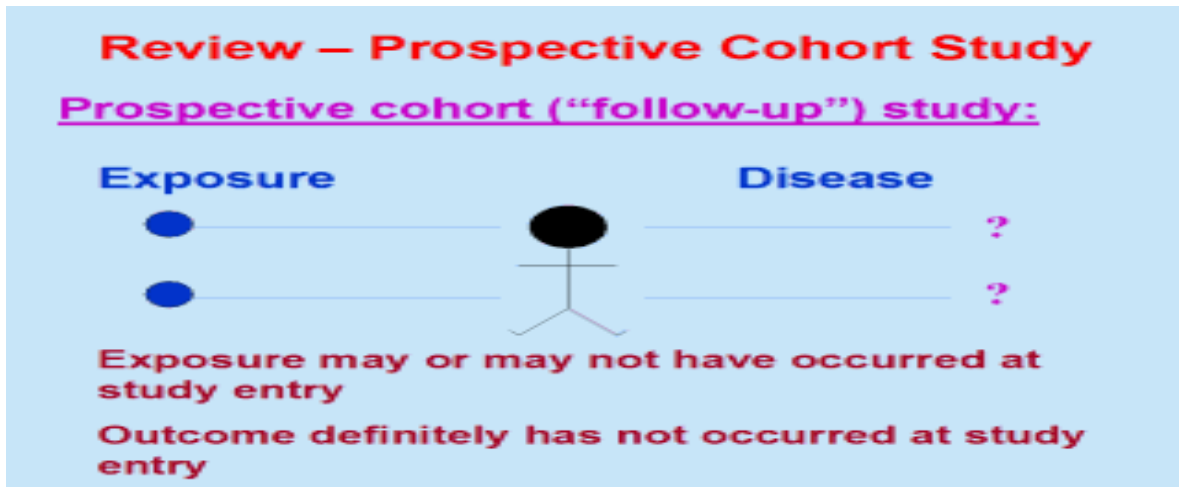
Prospective cohort study (“follow-up”) study:

- Disease free individuals are selected and their exposure status is ascertained
- Subjects are followed for a period of time to record and compare the incidence of disease between exposed and non-exposed individuals.
- Prospective cohort study also called *longitudinal study*.

Prospective cohort study is the best observational design. Why?

- * The investigator proceeds from “E to D” i.e. from cause to effect so he will not face a chicken egg dilemma and the temporal (time) sequence between E and D can be clearly established.

- * It uses a control group to accept or reject the hypothesis between E and D.



Steps to carry out the prospective cohort study:

1. Selection of the cohorts:

- All participants (both exposed and unexposed) in a cohort study must be at risk of developing the outcome.
- Controls should be similar to the exposed in all important aspects, except for the lack of exposure. This will reveal the background rate of the outcome in the community.
- Choice of cohort depends on the nature of disease under investigation.
- Selection of cohort depends on exposure:
 - If the exposure is common (e.g. smoking), the cohort will be selected from the general population and this is a heterogenous group. Either all the population or a representative sample is selected.
 - If the exposure is rare (exposure to radiation) the study subjects will be selected from a homogenous group as they are all experienced a similar exposure (e.g. radiologists exposed to X ray). Then the study subjects should be compared with external group who are not exposed this is called external comparison.

-**Prospective cohort studies** reduce the possibility that the results will be biased by selecting subjects for the comparison group who may be more or less likely to have the outcome of interest, because in a cohort study the outcome is not known at baseline when exposure status is established. In contrast, case-control studies, and to a lesser extent retrospective cohort studies may be subject to selection bias

2. Obtaining data on exposure:

- **Sources of data on exposure are:**

- a. Interviews or questionnaires from cohort members
- b. Review of medical records: e.g. dose of radiation, kinds of surgery, details of vaccination or medical treatment can be obtained from medical records.
- c. Medical examination or special test: e.g. blood pressure, serum cholesterol or ECG.
- d. Environmental survey: e.g. the level of air pollutants.

3. Follow up:

- This is a challenge; Drop outs affect the study's validity.
- Drop outs are not random events. If the likelihood of dropping out is related to the exposure and outcome, then bias can result.
- To optimize follow up, try to get a stable population, motivate them and do regular contacting and tracing.
- Procedures required may be:
 - A- Periodic medical examination.**
 - B- Routine surveillance of records.**
 - C- Periodic mailed questionnaire, phone calls, home visits.**

4. Analysis and interpretation:

Issues in analysis:

The basic analysis involves:

- Calculation of incidence rates among the exposed

- Calculation of incidence rates among the non-exposed

Tabulation of data

	Outcome		Total	Incidence (a/a+b) (c/c+d)
	Diseased	Not diseased		
Exposed	a	b	a + b	
Non-exposed	c	d	c + d	

1.Estimation of risks:

$$\text{Relative risk} = \frac{\text{Incidence among exposed}(I_e)}{\text{Incidence among non exposed}(I_o)}$$

$$RR = \frac{\left(\frac{a}{a+b} \right)}{\left(\frac{c}{c+d} \right)}$$

- The relative risk (also called the risk ratio) is measure of association between incidence of disease and certain exposure The risk ratio is used in assessing the likelihood that an association represents a causal relationship. For example, the risk ratio of lung cancer in long-term heavy smokers compared with non-smokers is approximately 20.
- **Relative risk answers the question:**
“How many times a person who is exposed to risk factor is at risk of developing disease compared to non-exposed?”

Calculation of Relative Risk

	Outcome		
Exposure	Present	Absent	TOTAL
Present	a	b	a+b
Absent	c	d	c+d
TOTAL	a+c	b+d	a+b+c+d

Interpretation of Relative Risk (RR)

RR=1: No association between exposure and disease incidence rates is identical between groups

RR> 1: Positive association (increased risk) exposed group has higher incidence than non-exposed group

RR< 1: Negative association (protective effect) non-exposed group has higher incidence

2-Attributable risk percent: (ARP):

- The AR indicates to what extent the disease under study is attributed to the exposure.
- AR is the difference between the incidence rate of the disease among the exposed (I_e) and that among the non exposed (I_0). It is expressed as a percent.
- $$ARP = \frac{(I_e - I_0)}{I_e} \times 100$$

- **Retrospective Cohort study**

- In retrospective cohort study, the investigator goes back 10- 20 years ago to select the study subjects (exposed to suspected factor, E_e) and the comparison group (not exposed to suspected factor E₀) from the existing records.
- Retrospective design is effective for diseases with a long development time. Sometimes referred to as **historical cohort studies**
- Historical cohort studies track people forward in time from exposure to outcome.

Comparison between Prospective versus Retrospective

Prospective cohort	Retrospective study
1.Exposure may or may not have occurred at the time of the study but the disease definitely didn't occur.	1-Both exposure and disease have occurred and the investigator has to make sure that Exposure occurred before the disease status.
2- It begins in the present and continue in the future	2- It begins in the past and continue to the present
3- Time consuming.	3- Save time.
4- Expensive	4- Cheap.

- **Advantages of cohort studies**

1. Valuable in rare exposures.
2. Can study multiple effects of a single exposure.
3. Exposure happened before outcome (time relation is established) (Temporality)
4. Can calculate incidence rates.
5. Can quantify Risk, Relative risk, & Attributable Risk
6. Dose response ratio can be calculated.
7. Low potential for bias(confounding , selection) than case-control study

8. The study itself may alter participants' behavior.

- **Disadvantages of cohort studies**

1. Attrition (loss to follow up), loss of the participants (migration, deaths, lack of interest) may affect validity of results.
2. Measurement errors, multiple interviews, tests
3. Involve a large sample
4. Inefficient for evaluation of rare diseases.
5. Takes a long time.
6. Expensive
7. Ethical problems
8. Loss of experience staff and loss of funds.
9. They are not good for diseases with a long latency.

- **Disadvantages of Retrospective Cohort Studies**

1. As with prospective cohort studies, they are not good for very rare diseases.
2. If one uses records that were not designed for the study, the available data may be of poor quality.
3. There is frequently an absence of data on potential confounding factors if the data was recorded in the past.
4. It may be difficult to identify an appropriate exposed cohort and an appropriate comparison group.
5. Differential losses to follow up can also bias retrospective cohort studies.

Calculation of Relative Risk

Example 1: Calculate the relative risk of the Smoking and low birth weight?

Smoking status	Birth Weight		
	<2500 g	≥2500 g	TOTAL
Smoker	120	240	360
Non-smoker	60	580	640
TOTAL	180	820	1000

Answers

- 1. Incidence of LBW among smokers

$$= \frac{120}{360} \times 1,000 = 333.3$$

- 2. Incidence of LBW among non-smokers

$$= \frac{60}{640} \times 1,000 = 93.8$$

- 3. **Relative risk** for having a LBW baby among smokers versus non-smokers

$$= \frac{333.3}{93.8} \approx 3.6$$

Example 2:

A group of male individuals are classified according to their work stress level into high stress (n=200), low stress (n=200). The two groups are similar in all other aspects as age, education, social class.... They are followed up for 10 year period. MI are detected among 20 of high stress and among 2 of low stress

Analysis and interpretation

A. Tabulation

B. Calculate relative risk

c. Interpretation

d. Attributable risk percent: (ARP):

Answer:

	MI		
Work Stress level	Yes	No	TOTAL
High stress level	20	180	200
Low stress level	2	198	200
TOTAL	22	278	400

B.

Incidence rate among high stress worker (Ie) =

$$a \backslash a+b = 20 \backslash 20+180 \times 100 = 10\%$$

Incidence rate among low stress worker (Io) =

$$c \backslash c+d \times 100$$

$$= 2 \backslash 2+198 \times 100 = 1\%$$

$$\underline{\underline{RR = I_e \setminus I_o = 10 \setminus 1 = 10}}$$

C. MI occurs 10 times in high stress level than low stress level.

D. ARP =

$$I_e - I_o$$

$$\text{-----} \times 100 =$$

$$I_e$$

$$10 - 1 \setminus 10 \times 100 = 90 \%$$

Stress is response for 90% of MI.

Example 3:

The table below shows the incidence of flue among individuals who have been vaccinated or not and subsequently either develop influenza or not. According to this calculate the following

A. RR

B. ARP

C. Interpretation the result

Table 3c. 2 × 2 Table Illustrating Association

	Influenza	No Influenza	Total	
Not Vaccinated	35	5	40	
Vaccinated	10	70	80	
Total	45	75	120	Incidence rate

among non-vaccinated: =

Number of new cases of a disease occurring in the population during a specified period of time
----- x100

Number of persons exposed to risk of developing the disease during that period of time

$$A \setminus a+b \times 100 = 35 \setminus 40 \times 100 = 87.5 \%$$

Incidence rate among vaccinated: =

Number of new cases of a disease occurring in the population during a specified period of time
----- x100
Number of persons exposed to risk of developing the disease during that period of time

$$C \div c + d \times 100 = 10 \div 80 \times 100 = 12.5\%$$

$$A. RR = I_e \div I_o = 87.5 \div 12.5 = 7$$

Non- vaccinated individual 7 times more liable for Influenza.

B. ARP

$$I_e - I_o$$

$$----- \times 100 =$$

$$I_e$$

$$87.5 - 12.5$$

$$----- \times 100 = 85.7$$

$$87.5$$

Good Luck