بسم الله الرحمن الرحيم

Epidemiological and Research Studies

X

20-11-2024

Prof DR. Waqar Al – Kubaisy

Epidemiological and Research Studies

Cohort Study

Cohort Study

begins with group of people free of disease & classified into subgroups
a group of individuals **exposed to a risk factor**
a group who are **unexposed to the risk** factor
are followed over time (often years)

Issues in the design of cohort studies understand the differences from a CCS,
*Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (RR,AR
*appreciate its strengths and weaknesses.

Cohort Study

Also called: follow up study, or incidence studies,

Definition: Study in which persons,

- **based on their exposure to a determinant and**
- **free of the disease outcome at the start** of the study
- **are followed in time** to **assess the occurrence of the disease outcome**

- It begins with a group of people who **are free of disease**
- and who are **classified into subgroups according to exposure** to a **potential cause of disease or outcome**
- **Cases** are excluded at the beginning
- **Variables of interest are specified and measured** and the
- **whole cohort is followed up to see how the subsequent**
- **development of new cases of the disease** (or other outcome) **differs** between the groups
- **with** and **without exposure**

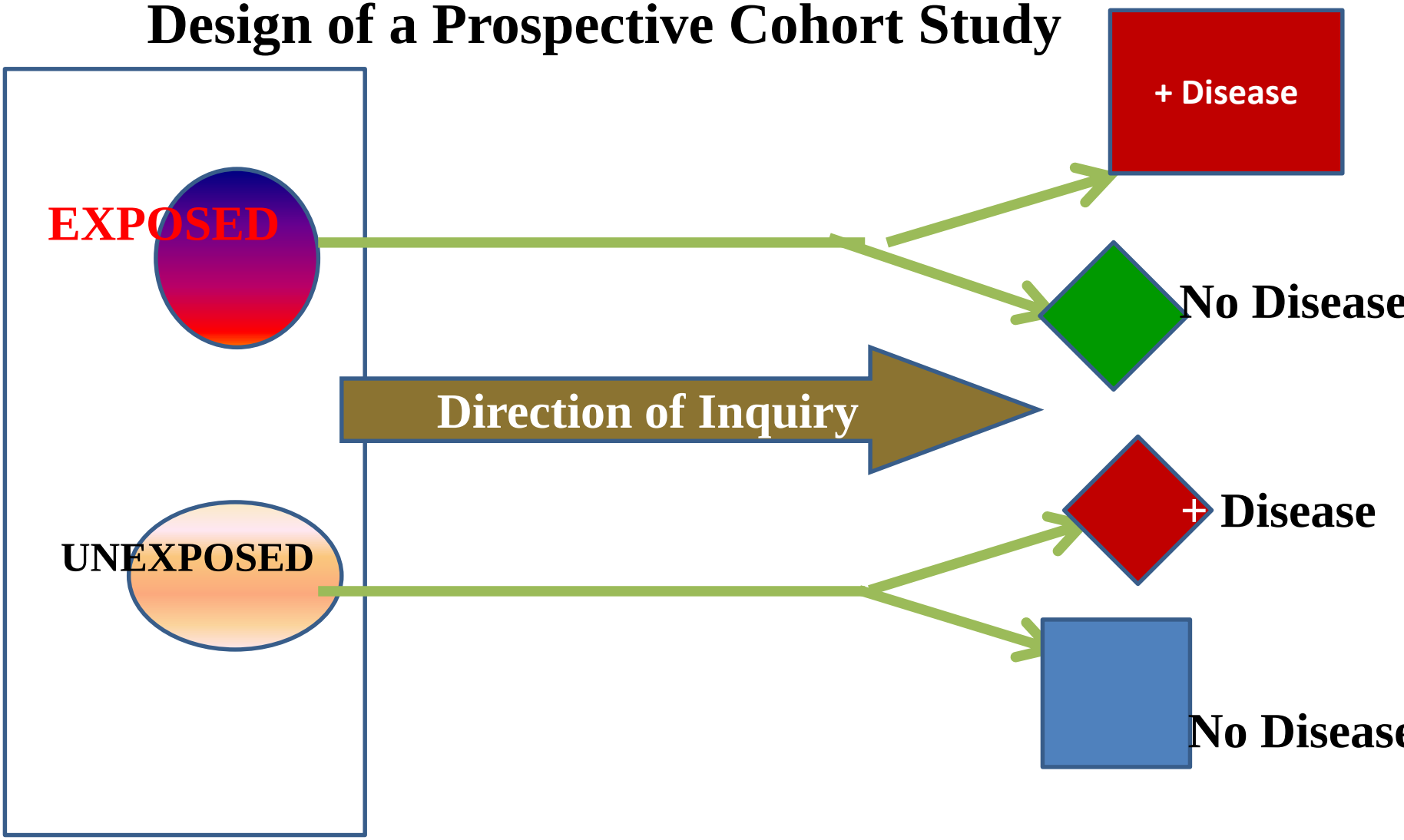
Cohort studies are a **form of longitudinal study** design that **flows from** the **exposure to outcome**.

In a cohort study,

- **Group** of individuals **exposed to a putative risk factor**
- **Group** who are **unexposed to the risk** factor **are**
- **followed over time** (often years)
- **to determine the occurrence of disease**.
- The **incidence of disease**
- in the **exposed group** is **compared with** the
- **incidence** of disease in the **unexposed group**.

- Therefore **Relative risk (incidence risk or incidence rate)** is used **to assess whether the exposure and disease are causally linked**.
- **Cohort studies** be **prospective**

Design of a Prospective Cohort Study



It begins with group of people **free of disease** and **classified into subgroups**

a group of individuals **exposed to a risk factor**

a group who are **unexposed to the risk** factor

are followed over time (often years)

- **Cohort studies** be **prospective**
- **A prospective cohort** study is also called a
- **concurrent cohort study**,

where the **subjects have been followed up for a period and the outcomes of interest are recorded**

1. Issues in the design of cohort studies:

- **Selection of study groups**
- **Measuring exposure**
- **Measuring outcome**
- **Methods of follow-up**

*Issues in the design of cohort studies understand the differences from a CCS,
*Potential bias in cohort studies
*Analysis of cohort studies
*calculate the basic measures (relative risk, attributable risk etc
*appreciate its strengths &weaknesses.

- **Selection of study groups**

❖ The aim of a cohort study is to **select study participants who are identical with the exception**

❖ **of their exposure status**.

❖ **All study participants must be**

- **All study participants must be**
- **Free of the outcome under investigation and**
- **have the potential to develop the outcome** under investigation.

Measuring exposure

- **Levels of exposure** (e.g. packs of cigarettes smoked per year)are
  - **measured for each individual at baseline**
  - **at the beginning of** the study **and**
  - **Assessed at intervals during the period of follow-up**.
- **When several** exposures are being considered simultaneously, the **non-exposed group** should comprise all those with none of the risk factors under investigation.
- **A particular problem** occurring in cohort studies is **whether individuals in the control group are truly unexposed.**

For example, study participants may start smoking or they may fail to correctly recall past exposure.

❖ **Similarly**, those in the **exposed group** **may change** their **behavior in relation to** the exposure such as diet, smoking or alcohol consumption.

❑ **Exposure data may be obtained** from a number of sources including:

❑ **Medical** or **Employment records**, **standardized** questionnaires, interviews and by **physical examination**

❑Measuring outcome

❑ **Outcome measures may be obtained from various sources, including**

- **Directly from the participant**
- **Routine surveillance of cancer** registry data,
- **Death Certificates**, **Medical records**

❖ **Method used to ascertain outcome**

❖ **must be identical for both exposed and unexposed groups**

Issues in the design of cohort Selection of study groups
Measuring exposure
Measuring outcome
Methods of follow-up

❑ Methods of follow-up

❖ **The follow-up of study participants in a cohort study is a major challenge.**

- **A great deal of cost and time is required to ensure follow-up** of cohort members
- and to **update measures of exposures and confounders**,
- in addition to **monitoring participants' health outcomes**

The failure to collect outcome data for all members of the cohort will affect the validity of study results

- . .

1 Issues in the design of cohort studies understand the differences from a CCS,
2 Potential bias in cohort studies
3 Analysis of cohort studies
4 calculate the basic measures (relative risk, attributable risk etc
5 appreciate its strengths and weaknesses.

2- Potential sources of bias

A major source of potential bias in cohort studies is due to:

- **losses to follow-up.**
- **Cohort members may; die, Migrate, Change jobs or**
- **Refuse to continue to participate** in the study.

In addition, losses to follow-up **may be related to the**

- **exposure, outcome or both**. For example, individuals who develop the outcome may be less likely to continue to participate in the study.
- **The degree to which losses** to follow-up are correlated with exposure and outcome will **lead to serious bias** in the **measures of effect of exposure and outcome**

- **A major source of potential bias in cohort studies arises from**

- **The degree of accuracy with**

- **The degree of accuracy with which subjects have been classified with respect to their exposure or disease status.**
- **Differential misclassification can lead to an over or underestimate of the effect between exposure and outcome**

Analysis of cohort studies

- **Analysis of a cohort study uses either**
- **the risk or the rate ratio of disease in the exposed cohort**
- **compared with the rate or risk in the unexposed cohort.**

- **Risk estimates:**

To estimate risk of event to occur when exposed to a risk factor.

Relative risk (RR)

$$RR = \frac{\text{proportion of disease in exposed group}}{\text{proportion of disease in unexposed group}}$$

Attributable risk can be calculated ???

Population attributable Risk PAR

Relative risk (RR)

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

Exposure	+Ve Out com	no outcome	Total
positive	a	b	
Negative	c	d	
Total			

RR Used in cohort study

- **The risk is the relative incidence in the exposed and non exposed group**

Risk estimates

$$RR = \frac{\text{proportion of disease in exposed group}}{\text{proportion of disease in unexposed group}}$$

Example: A cohort study of smoking and cancer of the pancreas was conducted for 90,049 individuals and followed for 1 year. Consist of 27,042 smokers of them 42 were developed CA pancreas, while only 7 developed the CA pancreas among non smokers group . Is smoking is a risk factor for CA pancreas

Exposure	CA pancreas	no CA pancreas	Total
positive	a 42	b	27,042
Negative	C 7	d	
Total			90,049

	Cancer of the pancreas	No disease	Total	Incidence rate
Smokers	42	27,000	27,042	1.5/1000/yr
Non-smokers	7	63,000	63,007	0.1/1000/yr
Total	49	90,000	90,049	

the data, taken from a **cohort** study **to investigate the association between smoking and cancer of the pancreas,** the **relative and attributable risk** can be calculated as follows:

	Cancer of the pancreas	No disease	Total	Incidence rate
Smokers	42	27,000	27,042	1.5/1000/yr
Non-smokers	7	63,000	63,007	0.1/1000/yr
Total	49	90,000	90,049	

Rate Ratio = $\frac{\text{Incidence rate in exposed group (r1)}}{\text{Incidence rate in unexposed group (r0)}}$

RR = 1.5/0.1 = 15

The RR of 15 indicates that
the risk of cancer of the pancreas is 15 times higher among smokers than non-smokers.

Relative risk (RR)

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

Exposure	+Ve Out com	no outcome	Total
positive	a	b	
Negative	c	d	
Total			

RR Used in cohort study

- **The risk is the relative incidence in the exposed and non exposed group**

Risk estimates

$$RR = \frac{\text{proportion of disease in exposed group}}{\text{proportion of disease in unexposed group}}$$

Example: A cohort study of smoking and cancer of the pancreas was conducted for 90,049 individuals and followed for 1 year. Consist of 27,042 smokers of them 42 were developed CA pancreas, while only 7 developed the CA pancreas among non smokers group . Is smoking is a risk factor for CA pancreas

Exposure	CA pancreas	no CA pancreas	Total
positive	a 42	b	27,042
Negative	C 7	d	
Total			90,049

	Cancer of the pancreas	No disease	Total	Incidence rate
Smokers	42	27,000	27,042	1.5/1000/yr
Non-smokers	7	63,000	63,007	0.1/1000/yr
Total	49	90,000	90,049	

the data, taken from a **cohort** study **to investigate the association between smoking and cancer of the pancreas,** the **relative and attributable risk** can be calculated as follows:

	Cancer of the pancreas	No disease	Total	Incidence rate
Smokers	42	27,000	27,042	1.5/1000/yr
Non-smokers	7	63,000	63,007	0.1/1000/yr
Total	49	90,000	90,049	

Rate Ratio = $\frac{\text{Incidence rate in exposed group (r1)}}{\text{Incidence rate in unexposed group (r0)}}$

RR = 1.5/0.1 = 15

The RR of 15 indicates that
the risk of cancer of the pancreas is 15 times higher among smokers than non-smokers.

interpretation

Relative risk = incidence a / incidence b

If both are equal then it is 1 (no risk)

If a > b then it is more than one, it is risky

If a < b then it is less than one, protective

Difference between **two incidence** rates is called

attributable risk=

Incidence of disease rate among exposed- incidence of disease rate among non-exposed **X100**

Incidence rate of disease among exposed

In the previous example: **Attributable risk** = $\frac{1.5-0.1}{1.5}$ X100

= 93% (this is the risk CA pancreas attributing to smoking

interpretation

Attributable risk = $\frac{\text{incidence a} - \text{incidence b}}{\text{incidence a}}$

If both are equal then it is 0 (no risk)

If a > b then it is **more than zero, it is risky**

If a < b then it is **less than zero, protective**

Attributable risk can be useful as a measure of the public health impact of a particular exposure

Risk difference (attributable risk)

the risk difference tells you the amount of disease that potentially could be prevented if the risk factor could be eliminated

Attributable risk can be useful as a measure of the public health impact of a particular exposure

Population Attributable Risk (PARs)

PAR **tells us about the amount of extra disease occurring in the exposed group because of exposure.**

How **much of disease in the whole community** can be attributed to the exposure

$$PAR = I_T - I_0$$

I_T is the rate in the population

I_0 is the rate in the unexposed group

Population Attributable Risk

PAR estimate the excess rate of disease in the total study population of exposed and non-exposed **individuals that is attributable to the exposure.**

PAR, helps determine which exposures have the most relevance to the health of a community

Population AR Versus AR

AR tell us how much disease in exposed group can be attributed to exposure

PAR: how much disease in the whole population can be attributed to exposure

The population attributable-risk percent (PAR%)

PAR% expresses the **proportion of disease in the study population that is attributable to** the exposure **and thus could be eliminated (removed) if the exposure were eliminated**

$$PAR\% = \frac{PAR}{I_T} \times 100$$

Example

Data from a cohort study of oral contraceptive (OC) use and bacteriuria among women aged 16-49 years

	Bacteriuria		
	Yes	No	Total
OC use			
Yes	27	455	482
No	77	1831	1908
Total	104	2286	2390

Data from D. A. Evans et al., Oral contraceptives and bacteriuria in a community-based study. *N. Engl. J. Med.* 299:536, 1978.

The population attributable risk of bacteriuria associated with OC use can therefore be calculated as:

$PAR = I_T - I_0 = 104/2390 - 77/1908 == 316/10^5$/year

Thus, if OC use were stopped, the-excess annual incidence rate of bacteriuria that could be eliminated among women in this study is 316 per 100,000.

$$\text{Relative Risk}\,(RR) = \frac{27/482}{77/1908} = 1.4$$

4. Strengths and weaknesses of cohort studies

1 Issues in the design of cohort studies
understand the differences from a CCS,
2 Potential bias in cohort studies
3 Analysis of cohort studies
4 calculate the basic measures (relative risk, attributable risk etc
5 appreciate its strengths &weaknesses.

Strengths

- ✓ **Multiple outcomes** can be measured for any one exposure.
- ✓ **Can look at multiple exposures**.
- ✓ **Exposure is measured before the onset of disease**
- ✓ **Good for measuring rare exposures,** for example among different occupations.
- ✓ **Demonstrate direction of causality.**
- ✓ **Can measure incidence**

4. Strengths and weaknesses of cohort studies

Weaknesses

- Costly and time consuming.
- Prone to bias due to loss to follow-up.
- Prone to confounding.
- Participants may move between one exposure category
- Knowledge of exposure status may bias classification of the outcome.
- Being in the study may alter participant's behaviour.
- Poor choice for the study of a rare disease.
- **Classification of individuals (exposure or outcome status) can be affected by changes in diagnostic procedures.**

Thank you for attention

example

In a study in the United States of America, the incidence rate of stroke was measured in a population of women who were 30–55 years of age and free from coronary heart disease, stroke and cancer in 1976. A total of 274 stroke cases were identified in eight years of follow-up .

Never smoked : 70 cases among 395 594

Ex-smoker : 65 cases among 232 712

Smoker: 139 cases among 280 141

Calculate

--Relative for smoking

-attributable risk for smoking (ignore ex-smoker

In a study in the United States of America, the incidence rate of stroke was measured in a population of women who were 30–55 years of age and free from coronary heart disease, stroke and cancer in 1976. A total of 274 stroke cases were identified in eight years of follow-up .

Never smoked : 70 cases among 395 594
Ex-smoker : 65 cases among 232 712
Smoker: 139 cases among 280 141
Calculate
-Incidence for each group
-Relative for smoking
-attributable risk for smoking (ignore ex-smoker

Smoking category	Number of cases of stroke	Person-years of observation (over 8 years)
Never smoked	70	395 594
Ex-smoker	65	232 712
Smoker	139	280 141
Total	274	908 447