Lec 6 **Epidemiology**

Analytic studies

Observational studies :

The investigator simply observed the natural course of the events , noting who is exposed and non-exposed and who has and has not developed the outcome of interest (passive investigator) . We have

A-Case-control study .

B-Cohort study .

C-Interventional studies (clinical trial) . The investigator himself will allocate the exposure (active investigator) .

CASE-CONTROL STUDY . (always retrospective)

Subjects are selected on the basis of whether they do (cases) or do not (control) have a particular disease (outcome) under study , the groups are then compared with respect to the proportion having a history of an exposure or characteristic of interest .

Starting point

outcome

	+ve outcome	-ve outcome	
exposed	a	b	a+b
Not-exposed	c	d	c+d
	present	free	N

Start from here

Risk factor present ← ← ← case (do have the disease)

Risk factor absent ←←←

Risk factor present ← ← ← control (free from disease)

Risk factor absent ←←←←

Past ← (pass our questionnaire retrospectively) ← present

SO IN CASE-CONTROL STUDY STARTING POINT FROM OUTCOME (DISEAS—OR NOT DISEASE) .

Strength (advantages) .

1-Quick and inexpensive .

2-Depend on already available data .

3-suitable for rare disease . (rare disease as in ALZHEIMER or MULTIPLE SCLEROSIS or any other rare disease (very low

prevalence) , we can not wait until a case appears so we use available cases .
4-Suitable for disease of long latency period . The latency period is the period between the exposure to a certain risk factor and the appearance of the disease (like the incubation period in infectious diseases) , so if we take a disease as MI duo to smoking we are starting from a group that already has the disease (the cases) , and ask them about the exposure (retrospectively) (exposure mean smoking) . while in a cohort study we have no cases (no disease) because we start with exposure .
5-Can examine multiple etiological factors for a single disease . EX, CHD , With smoking , diet , physical exercise ----etc –
6-relatively quick results .
7-smaller number of subjects .
Limitations :
1-Not establish the temporal relation- ship between exposure and outcome . (Both exposure and disease have already occurred) .
2-Can not compute the (INCIDINCE) .
3-More prone for bias compared with other analytic studies .
4-In efficient for the evaluation of rare exposure .
5-incomplete information .
6-problms of selecting control group and matching variable s.
7-not representative to general population .
8- confounder .

Definition of cases :
1-Homogenous disease entity (clear definition of the disease) .
EX, When we say (CA UTERUS) , what was mean , CA body of uterus or CA cervix . Because CA body of uterus in women with high socioeconomic class . BUT CA cervix in women with low socioeconomic status .
2- Depend on strict diagnostic criteria .
EX,MI diagnosed by chest pain more than 30 min, ECG changes and enzyme changes .

Sources of selection of cases .
1-Hospital – based case- control study .
Advantages : common , easy , and inexpensive .
Limitations : only sever cases of the disease enter the hospital (selection bias) .

2-Population based case-control .

Advantages : avoid selection bias and describe the picture of the disease in the population .

Limitations : difficult , costly , and not routinely done .

Selection of control .

It is important to select appropriate control group (free of disease under study) . The control group should be selected to be comparable to the cases and there is no control fit for all situations . We have the following types of control :

1-Hospital control :

Advantages : common , easy , more willing to cooperate and inexpensive.

Limitations : They are disease .

2- General population control :

Advantages : represent healthy population .

Limitations : difficult , costly and not routinely done as it is difficult to contact with healthy individuals .

Estimation of risk (is there an association) .

In case- control study we can not calculate the incidence (absolute risk), because we start with already disease population (cases) and non disease (control) people . Hence , we use the (odds ratio) OR which is measure using the following formula , OR = ad/bc .

EG:

To determine the relationship between smoking and CHD , patients with CHD had been compared to the patients from orthopedic department , the two groups were matched for age and sex and were asked about smoking history ,the following had been found .

Coronary heart disease

	+ve	- ve	
smoking	112	176	288
No smoking	88	224	312
total	200	400	600

1- What is the design of the study ?

Case- control

2-Is there any relation between smoking and CHD ?

We measure the association by estimating the OR

OR=ad/bc = 112x 224 / 176x 88 x 100= 1.62 those who have CHD were 1.62 time more exposed to smoking than those free of disease .
3- Calculate a measure of the excess risk of CHD in smoking that is attributable to their smoking .

AR%= OR-1/ OR X 100 = 1.62-1/ 1.62 X100 = 38% .

The analysis and interpretation: To illustrate the study design, we identify a number of children who are suffering from acute respiratory infection (say 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). Suppose the number of controls is 380. Now, for children in both groups, the smoking habits of their parents are ascertained through careful interviewing of these parents. We try to know whether parent(s) do smoke or not and if they do, what is the number of cigarettes smoked per day. Suppose we found that the parents of 170 cases and 200 controls were smokers.

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

History of smoking	Cases of pneumonia	Children without pneumonia	Total
Positive			370
Negative			250
Total	240	380	620

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

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

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

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.

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

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

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

AR%= OR-1/ OR X 100

It is possible to estimate the attributable risk (population attributable risk proportion – which is different from our definition above in that the PARP includes the prevalence of the specific exposure in the population) in case control studies by using the following formula:

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

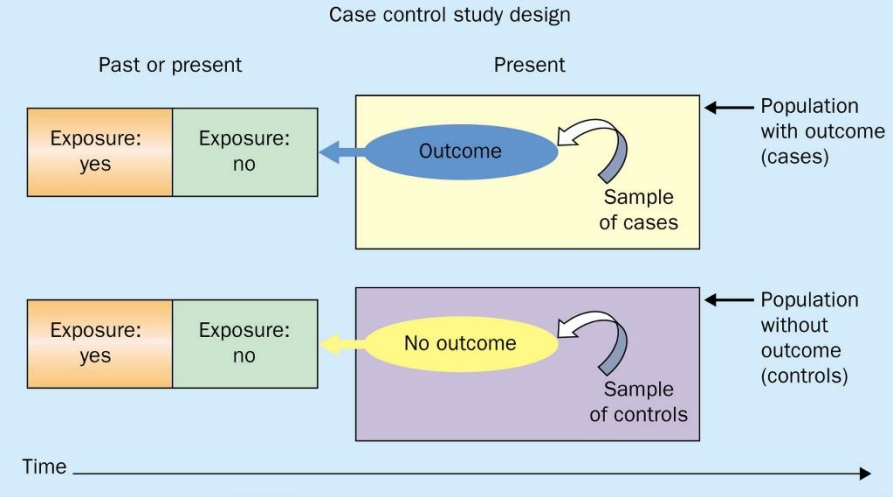
Where r = Odds ratio

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%)

The attributable risk = $\frac{0.4\ (2.2 - 1)}{\text{-----------------}} = \frac{0.48}{\text{-----}} = 0.33$ or 33%

$0.4\ (2.2 - 1\ ) + 1 \quad 1.44$



Cohort study . (always prospective).
Types of cohort study .
1-prospective .
2-retrospective cohort .it is more important in study of long latency period diseases .
3-historical cohort .
The group(s) is defined on the basis of presence or absence of exposure to a suspected risk factor for a disease (outcome) . All subjects then followed over a period of time to assess the occurrence of that outcome .
All subjects considered to be free of a given disease .

Starting Point

(OUTCOME)

	+ve	-ve	
exposed	a	B	a+b
no-exposed	c	D	c+d
	present	Free	

In cohort study start from exposure and follow up .

Present →→→ future

Exposed to risk factor →→ Develop outcome(a)
→→ Not develop outcome(b)

Non-exposed to risk factor →→ Develop outcome(c)
→→ Not develop outcome(d)

Strength : (Advantage)

1- Establish the temporal relationship between exposure and outcome .
2-Allow direct measurement of incidence .
3-Examine multiple effects of single exposure .
4-Suitable for rare exposure .
5-Provide information on confounders .
6-very accurate .

Limitations :

1-Expensive and time consuming .
2-Validity of the results can be seriously affected by losses to follow up(bias) .
3-Insufficient for the evaluation of rare disease .
4-large numbers of subjects required .

Selection of exposed population

1-For rare exposure (e.g radiation) , we need specially exposed people (as individual in certain occupation , or exposed to special event or treatment) .
2- For common exposure (e.g smoking) we need specially chosen people (can obtain complete information e.g --- members in certain profession as doctors , teachers , -- etc or student in particular collage or school) .

The source of exposure data can obtained from .

1-Pre-existing data .
2-From the study subject .
3-Direct physical examination .
4-Direct investigation .
5-Direct measurement of environment .

Selection of comparison population

1-Internal comparison group , used when a single cohort enter the study.

2-External comparison group , when specially exposed cohort enters the study .
3-Multiple comparison groups , very valid .

Estimation of risk (is there any association)

In cohort study , we can measure the incidence (absolute risk) directly .
The incidence among exposed (Ie) and incidence among non-exposed (Ie-) . Relative Riske (RR) , is the measurement of association in cohort study and can be calculated as follow .

RR=Ie/ Ie- , if RR =1 no association . if RR>1 positive association . if RR < 1 negative association (inverse association , possible protective) .
Attributable risk (AR) provides information about absolute effect of the exposure . AR=(Ie)-(Ie-) . AR indicate the number of cases among the exposed that can be attributed to the exposure itself .
Attributable risk percent (AR%)= AR/ Iex100% estimate the proportion of the disease among the exposed that can be attributed to the exposure itself.

Ex:- To determine the relation ship between cigarette smoking and CHD , 8000 healthy individual age > 45 years were enrolled in a study . 3000 of them were smoker , within 10 years , 84 of the smoker and 87 0f non-smokers develop CHD .
1-What is the design of the study ?
2-Draw 2x2 table .
3-Is there any relation between smoking and CHD?

A/1- Cohort study .
2-

CHD(OUTCOME)

	+VE	-VE	
Exposed +ve	**84**	**2916**	**3000**
Non exp ve-	**87**	**4913**	**5000**
	171	**7829**	**8000**

3- RR= 2.8/ 1.74 = 106

Example

A study was carried out to ascertain the relationship of parental

smoking to the risk of acute respiratory infection among children aged less than five years. A total of 800 children of smoking parents and 1200 of nonsmoking parents were followed up for six months. During the follow up period, 592 of the first group and 636 of the second group developed at least one attack of acute respiratory infection. Do these results suggest that parental smoking predisposes children to acute respiratory infection?

The analysis:

	pneumonia	No pneumonia	
exposed			800
Non exposed			1200
total			

The first step is to calculate the incidence rate of infection in the two cohorts.

Incidence rate among children exposed to parental smoking

$$= \frac{592}{800} \times 1000 = 740/1000$$

Incidence rate among children not exposed to parental smoking

$$= \frac{636}{1200} \times 1000 = 530/1000$$

It is clear that the incidence rate of infection is greater among the exposed group than the non exposed group

The second step is to measure the strength of association between parental smoking and infection by calculating the relative and attributable risk.

$$\text{The relative risk (RR)} = \frac{\text{Incidence rate among exposed}}{\text{Incidence rate among non exposed}}$$

$$= \frac{740}{530} = 1.4$$

This means that the risk of infection among children exposed to parental smoking is 1.4 times greater than the risk of infection among children not exposed to parental smoking.

In addition to the relative risk, we also measure the attributable risk which represents the fraction of risk that could be attributed to the exposure under study.

Attributable risk = Incidence rate among exposed – incidence rate among non exposed.

= 740 – 530 = 210 / 1000

This fraction is equivalent to the magnitude of reduction in the incidence rate if the exposure is eliminated, and can be expressed as percentage reduction (preventive fraction) as follows:

$$\text{Percentage reduction in the risk} = \frac{210}{740} \times 100 = 28.4\%$$

Notice that the relative risk is not very high which means a relatively moderate association, and therefore the expected reduction in the risk (incidence rate) among children who parents smoke in response to the elimination of the exposure, i.e., stopping parental smoking is relatively moderate (only 28.4%).
As in case control studies, a statistical test is performed such as the chi-squared test to show whether the association is statistically significant or not.
Note: In case of multiple exposures (the disease is related to multiple risk factors) a more sophisticated analysis is carried out to determine the relative effect or contribution of each risk factor. Logistic regression analysis and stepwise multiple regression analyses are commonly used.

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**
10. Risk to subjects	**Usually none**	**Risk of not removing exposure**

In case control studies few causal criteria may be fulfilled such as the size of the Odds ratio, the time sequence (but this is a problem in many instances) and dose response relationship. In cohort studies, the relative risk, the attributable risk, the time sequence and the dose response relationship are all possible to ascertain.