Epidemiology

Epidemiology: The word Epidemiology is derived from two Greek words **Epi** means upon and **demos** means people. It is the study of the distribution and determinants of disease in the human population. It is also called the "**Natural history of the disease**".

Epidemiology has to deal with the mass phenomena of diseases as they affect the population. It deals with all mass and infectious diseases as regards to morbidity, mortality, time and place, and phases of biological gradients of the disease.

Incidence rate: The incidence is the number of new cases of a disease added during a given period of time. The incidence rate is this number, expressed in terms of a unit of population under study.

$$\text{Annual Incidence Rate} = \frac{\text{Number of cases occurring during a period of } T \text{ years}}{\text{Average number of person living in the community during period } T}$$

The average incident rate is defined as

$$\text{Average Incident Rate} = \frac{f_A}{\left(N_A - \frac{1}{2} f_A\right) \times T}$$

where, $T = \text{Year}$

$f_A = \text{Number of persons constructed diseases A during a period of length time unit}$

$N_A = \text{Initial population in the community exposed to disease } A$

$$\therefore \text{Average Incident Rate} = \frac{\text{Number of cases occuring the period}}{\text{Mid year population}}.$$

Prevalence: The prevalence of a disease expresses the total number of cases (old and new) known to have existed during a given time. The prevalence rate is this number expressed per unit of population under study. Prevalence rate can be subdivided into

- **Point Prevalence:**

  The point prevalence is the frequency of the cases (old and new) at a given instant of time. It is a kind of cross-sectional study at any point in time.

$$\text{Point Prevalence Rate} = \frac{\text{Number of existing cases of disease } A \text{ at a specific time}}{\text{Number of persons in the community at that time}} \,..$$

- **Period Prevalence:**

The period prevalence is a measure that expresses the total number of cases (old and new) of disease known to have existed, having a prescribed period of time.

$$\text{Period prevalence Rate} = \frac{\text{Number of existing cases of disease } A \text{ in a given period of time}}{\text{Average number of persons living in the community during this period}} .$$

Descriptive study or cross-sectional study: Descriptive studies are concerned with observing and distribution and progression of disease in the population. It is carried out in order to determine the frequency of disease in the population in terms of person, place, and time. A descriptive study usually comprises observation made at a point of time called a cross-sectional study or prevalence study.

Retrospective study: It is basically a comparative study as we compare a group of individuals with disease (case) to that with the comparable group having no disease (control) in a situation where all the relevant events (cause & effect) have already occurred.

It is called a retrospective study because it compares both the cases and controls with regard to the presence of some causal element in their past experience.

Design of case-control retrospective study:

	Cases (with disease)	Controls (without disease)
Were exposed	a	b
Were not exposed	c	d
Total	$a+c$	$b+d$
Proportion exposed	$\frac{a}{a+c}$	$\frac{b}{b+d}$

Prospective Study: A prospective study is a longitudinal study that starts with a group of healthy study population of exposed and non-exposed individuals; all considered to be free of a given disease and are followed up to find out whether persons with the characteristic (exposed) develop the disease more frequently than those who do not have the characteristic (non-exposed).

A prospective study needs a clear hypothesis to test and is a large-scale; time-consuming, costly undertaking, and the investigator must have to wait for the disease to appear in the study population. In a prospective study, several approaches to the selection of a group are possible. The group may be heterogeneous with respect to exposure consisting of members with high, low, or non-exposure.

The study was carried out on a selected special exposure to follow as to see

Whether	Disease develops	The disease does not develop	Total	Incidence rate of disease
Exposed	a	b	$a+b$	$\frac{a}{a+b} = \text{incidence in exposed}$
Non-exposed	c	d	$c+d$	$\frac{c}{c+d} = \text{incidence in non - exposed}$

Advantages of Prospective Study:

- The group can be defined as a set which are alike a comparable in all respects for the characteristic whose relationship to the disease is to be studied.
- It can obtain the current information on the frequency of a disease or condition in the particular population studied.
- It can be demonstrative of the strength of association between a factor and disease giving a more direct estimation of the risk from exposure to each factor.

Advantages of Retrospective Study:

- It is relatively quick, easy and cheap.
- A significant number of cases can be assumed and a variety of hypotheses can be rapidly.
- It provides a practical approach to studies of the cause of certain rare diseases.
- It is relatively easy to start from a group of patients with an established disease and to make inquiries about their family. Their background personal chart eristic and habits to investigate several possible factors that may have an association with the disease.

The disadvantage of Retrospective Study:

The disadvantage to retrospective studies is that they need information about past events which may not be available from patient disease records or may be inaccurately recorded.

Distinguish between retrospective and prospective studies.

Retrospective study	Prospective study
In a retrospective study, comparisons are made to compare between groups of individuals with the disease (case) to those with the comparable group having no disease (control).	In prospective studies, comparisons are made with their difference in exposure (exposed, non-exposed) to the rate at which the disease has developed.
Retrospective study can be carried out in two separate groups essentially homogeneous in exposure.	In a prospective study, the group may be heterogeneous.
It has the advantages of being relatively quick, easy, and cheap.	Compared with a retrospective study, it is complex and time-consuming.
In retrospective study assumed that the incidence rate is low.	In a prospective study, no assumption is needed.
In a retrospective study, finding the relative risk is very complicated, so we use the odds ratio.	In a prospective study, we can find relative risk.

Risk Factor: The factors whose presence or absence is associated with an increased likelihood that the disease may or possibly will develop at a later stage are called risk factors.

Relative Risk: The ratio of incidence between exposure and incidence among non-exposure is called relative risk. It is expressed as

$$\text{Relative Risk} = \frac{\text{The incidence among exposure}}{\text{The incidence among non - exposure}}.$$

If $RR = 1$, it indicates that the incidence of exposure is equal to the incidence of non-exposure.

If $RR > 1$, it implies that the incidence of non-exposure is greater than the incidence of exposure.

Attributable Risk: The difference between the incidence rate between exposure and the incidence rate among non-exposure is called attributable risk. Mathematically it can be written as

$$AR = (\text{The incidence rate among exposure}) - (\text{The incidence rate among non - exposure}).$$

Odds Ratio: In a retrospective study, we do not know the incidence in the exposed population or the incidence in the non-exposed. Hence we cannot calculate the relative risk directly from a retrospective study. In a retrospective study, the estimate of relative risk is

$$RR = \frac{\text{Incidence in exposed}}{\text{Incidence in non-exposed}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}.$$

This can be obtained using the odds ratio of the incidence of the disease in low, as the situation with most chronic diseases.

$$RR = \frac{a/a+b}{c/c+d} = \frac{a/b}{c/d} = \frac{ad}{bc} = \text{odds ratio},$$

It can be used in a retrospective study as an estimate of the relative risk if we apply the following three assumptions:

- The incidence of the disease is low.
- The cases are representative of all cases with regard to exposure.
- Controls are representative of the reference population with respect to exposure.

The odd ratio is also known as the cross-product ratio since it is obtained by multiplying the diagonal from the (2×2) table.

Find the relative risk and odds ratio for the following table:

	A	$\bar{A}$	Total
F	n_{11}	n_{12}	n_{1+}
$\bar{F}$	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n_{++}

Where $F = \text{Factor present}$
$\bar{F} = F\text{actor absence}$
$A = \text{Disease present}$
$\bar{A} = \text{Disease absent}$

Solution: Let us consider

$$P(A|F) = P_1 = \frac{n_{11}}{n_{1+}}, \quad P(A|\bar{F}) = P_2 = \frac{n_{21}}{n_{2+}}$$

$$P(\bar{A}|F) = Q_1 = \frac{n_{12}}{n_{1+}}, \quad P(\bar{A}|\bar{F}) = Q_2 = \frac{n_{22}}{n_{2+}}.$$

So, the relative risk of disease A is

$$RR_A = \frac{\text{probabiliy of getting the disease } A \text{ when factor } F \text{ is present}}{\text{probabiliy of getting the disease } A \text{ when factor } F \text{ is absence}} = \frac{P_1}{P_2} = \frac{n_{11}/n_{1+}}{n_{21}/n_{2+}}.$$

We know that,

$$\text{Odds} = \frac{\text{Probabiliy of success}}{\text{Probabiliy of failure}}$$

$$\therefore \text{Odds}\left(\text{when F present}\right) = \frac{P_1}{Q_1}$$

$$\therefore \text{Odds}\left(\text{when F absent}\right) = \frac{P_2}{Q_2}$$

$$\therefore \text{Odds ratio, w} = \frac{\text{odds in the presence of the factor}}{\text{odds in the absence of the factor}} = \frac{P_1/Q_1}{P_2/Q_2} = \frac{P_1Q_2}{P_2Q_1} = \frac{\frac{n_{11}}{n_{1+}}\cdot\frac{n_{22}}{n_{2+}}}{\frac{n_{21}}{n_{2+}}\cdot\frac{n_{12}}{n_{1+}}}$$

$$= \frac{n_{11}n_{22}}{n_{12}n_{21}} = \text{Cross product ratio}$$

Why Odd ratio is called a cross-product ratio?

The odds ratio is also known as the cross-product ratio since it is obtained by multiplying diagonally in a (2×2) table.

Estimating risk from studies of matched pairs: Four types of case-control pairs are

Concordant pairs:

- Pairs in which both the case and the controls were exposed.
- Pairs in which neither the case nor the control was exposed.

Disconcordant pairs:

- Pairs in which the case was exposed but the control was not.
- Pairs in which the control was exposed and the case was not.

These probabilities are shown in the following (2×2) table (the figure in each cell represents not individual subjects but case-control pairs)

	Controls		
case	Exposed	*a* (concordant)	*b* (disconcordant)
	Non-exposed	*c* (disconcordant)	*d* (concordant)

Calculation of the ratio in such a matched pair study is based on the disconcordant pairs only (b & c). The concordant pairs (a & d in which cases and controls were both exposed or both non-exposed) are ignored.

$$\therefore \text{Odds ratio} = \frac{\text{The number of pairs in which the case was exposed and controls was not exposed}}{\text{The number of pairs in which the control was exposed and the case was not}}$$

$$= \frac{b}{c}$$

Attributable risk for the exposed group: The extent to which the incidence of disease in a group of exposed persons can be attributed to the exposure expressed as the arithmetic or absolute difference in incidence rates between the exposed and non-exposed group known as the attributable risk for exposed.

$$\text{This incidence in exposed group} = \text{incidence not due to the exposure} + \text{incidence due to the exposure}$$

and, $\text{incidence in non-exposed group} = \text{incidence not due to the exposure}$.

Therefore, the incidence in the exposed group which is attributable to the exposure can be calculated by

$$(\text{incidence in exposed group}) - (\text{incidence in non-exposed group}).$$

And the proportion of the total incidence in the exposed group which is attributable to the exposure, can be calculated as

$$\text{Attributable risk} = \frac{(\text{incidence in exposed group}) - (\text{incidence in non-exposed group})}{\text{incidence in exposed group}}.$$

Attributable risk for the total population: The extent to which the incidence of disease in a total population (exposed & non-exposed) can be attributed to the specific exposure is the attributable risk for the total population. The incidence in the population which is due to the exposure can be calculated by subtracting

$$(\text{incidence of total population}) - (\text{incidence in non-exposed group}).$$

And the proportion of the incidence in the total population which is attributable to the exposure can be calculated by

$$\frac{(\text{incidence of total population}) - (\text{incidence in non-exposed group})}{\text{incidence of total population}}.$$

Sensitivity: Sensitivity is the ability of the test to detect the condition when it is present. Sensitivity is defined as the percent of those who have the disease and so indicated by the test i.e.

$$\text{i.e. Sensitivity} = \frac{\text{True positives}}{\text{True positive} + \text{False negative}} = \frac{\text{True positives}}{\text{All these with the disease}}.$$

Specificity: Specificity is the ability of the test of difference cases in which the condition is present from those in which it is absent. Specificity is defined as the percentage of those who do not have the disease and are so indicated by the test

$$\text{i.e. Specificity} = \frac{\text{True negative}}{\text{True negative} + \text{False positive}} = \frac{\text{True negative}}{\text{All those without the disease}}.$$

Mortality: Mortality is the demographical measurement that deals with the total process of death and it brings change to the population.

Infant mortality rate (IMR): The infant mortality rate is the ratio of the total number of children dead under one year and the total number of live births in that year, i.e.

$$IMR = \frac{D_{(0-1)\text{year}}}{B} \times 1000.$$

Child mortality rate (CMR): The child mortality rate is the rate of the total number of deaths under 5 years during a specific period and the total number of births occurring in that period, i.e.

$$CMR = \frac{\text{Total number of deaths under 5 years during a specified period}}{\text{Total number of births occuring in that period}}.$$

Standardized mortality rate (SMR): Standardized mortality rate is defined as the rate of observed deaths to the expected number of deaths in occupation or industry had the age-specific death rate for the total population prevailed in that. It is calculated by the formula

$$SMR = \frac{\text{Observed deaths}}{\text{Expected deaths}} \times 100 = \frac{d^0}{\sum P_a^0 ASDR_a^+} \times 100$$

where, $d^0 = \text{Total deaths in occupation } 0$

$P_a^0 = \text{Total population at age } a \text{ in occupation } 0$

$ASDR_a^+ = \text{Age specific death rate at age } a \text{ in total population}$

Standardized death rate (SDR): This is the number of deaths that have occurred per 1000 in some standard population as per age-specific death rates of the given communities, i.e.

$$SDR = \frac{\sum(\text{standard population in each age group}) \times (\text{specific death rates in each corresponding age group})}{\text{Total standard population}} \times 1000$$

.