CHAPTER 1

PRINCIPLES & METHODS OF EPIDEMIOLOGY

1.1 INTRODUCTION

Biostatistics: is defined as, the application of statistical methods on the biological and medical data.

Epidemiology: is defined as **“**the studies of the **distribution** and **determinants** of **health** and **health related** events in a specified population, and the **application** of this study for the promotion of health, prevention and control of diseases.”

Key terms in this definition of epidemiology are:

☞ ***Study***

Epidemiology is a scientific discipline, sometimes called “the basic science of public health.” It has, at its foundation, sound methods of scientific inquiry. Epidemiology is data-driven and relies on a systematic and unbiased approach to the collection, analysis, and interpretation of data.

☞ ***Distribution***

Epidemiology is concerned with the **frequency** and **pattern** of health events in a population.

- **Frequency:** refers not only to the number of health events such as the number of cases of meningitis or diabetes in a population, but also to the relationship of that number to the size of the population. The resulting rate allows epidemiologists to compare disease occurrence across different populations.
- **Pattern:** refers to the occurrence of health-related events by **time**, **place**, and **person**. ***Time*** patterns may be annual, seasonal, weekly, daily, hourly, weekday versus weekend, or any other breakdown of time that may influence disease or injury occurrence.

  Place: patterns include geographic variation, urban/rural differences, and location of work sites or schools.

  Personal characteristics include demographic factors, which may be related to risk of illness, injury, or disability such as age, sex, marital status, and socioeconomic status, as well as behaviors and environmental exposures.

☞ ***Determinants***

Are any factors, whether event, characteristic, or other definable entity, that brings about a change in health and health related conditions (biological, chemical, physical, social, cultural, economic, genetic and behavioral). It refers to, why diseases occur in certain places? In a certain period? Or in a certain population groups?

☞ ***Health Related States and Events***

The most ambitious definition of health is that proposed by “WHO” in 1948: “health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.”

Epidemiology is concerned not only with disease but events like birth, death, migration injuries**,** causes of death, behaviors such as use of tobacco, positive health states, reactions to preventive regimes and provision and use of health services.

☞ ***Specified Populations***

The focus of epidemiology is mainly on the population rather than individuals include those with identifiable characteristics, such as occupational groups.

☞ ***Application***

Epidemiology is an applied science. The ultimate purpose of all epidemiological studies is the prevention and control of health problems.

Scope of Epidemiology

Originally, epidemiology was concerned with epidemics of communicable diseases and epidemic investigations.

Later - extended to endemic communicable diseases and non-communicable diseases

At present epidemiologic methods are being applied to:

- ✓ Infectious and noninfectious diseases
- ✓ Injuries and accidents
- ✓ Nutritional deficiencies
- ✓ Maternal and child health
- ✓ Environmental health
- ✓ Cancer, Health behaviors, etc

1.2 USES OF EPIDEMIOLOGY

Some important uses of epidemiology are listed as following

1. *Description of Health Status of Populations*

Descriptive epidemiology provides information of health status of populations by:

- ✓ Measure of disease frequency (quantify disease)
- ✓ Assess distribution of disease by person, place and time (answer questions of who, where and when)
- ✓ Formulation of hypothesis concerning causal and preventive factors tested bay analytic epidemiology

- ✓ Provide information health and health care policy and planning and resource allocation

2. ***Identify Determinants of Disease***

- ⦿ The purpose of epidemiology is identification of determinants or causes of disease. Determinants may genetic factor, behavioral or cultural factor, health care facility, environmental factors (physical, biological, social...) by hypothesis testing and establish causation of disease. Among the determinants, the most amenable for modification is the physical environment.
- ⦿ Providing physical environment implies **safe water**, **sanitary conveniences**, **clean environment**, **Good food etc.**

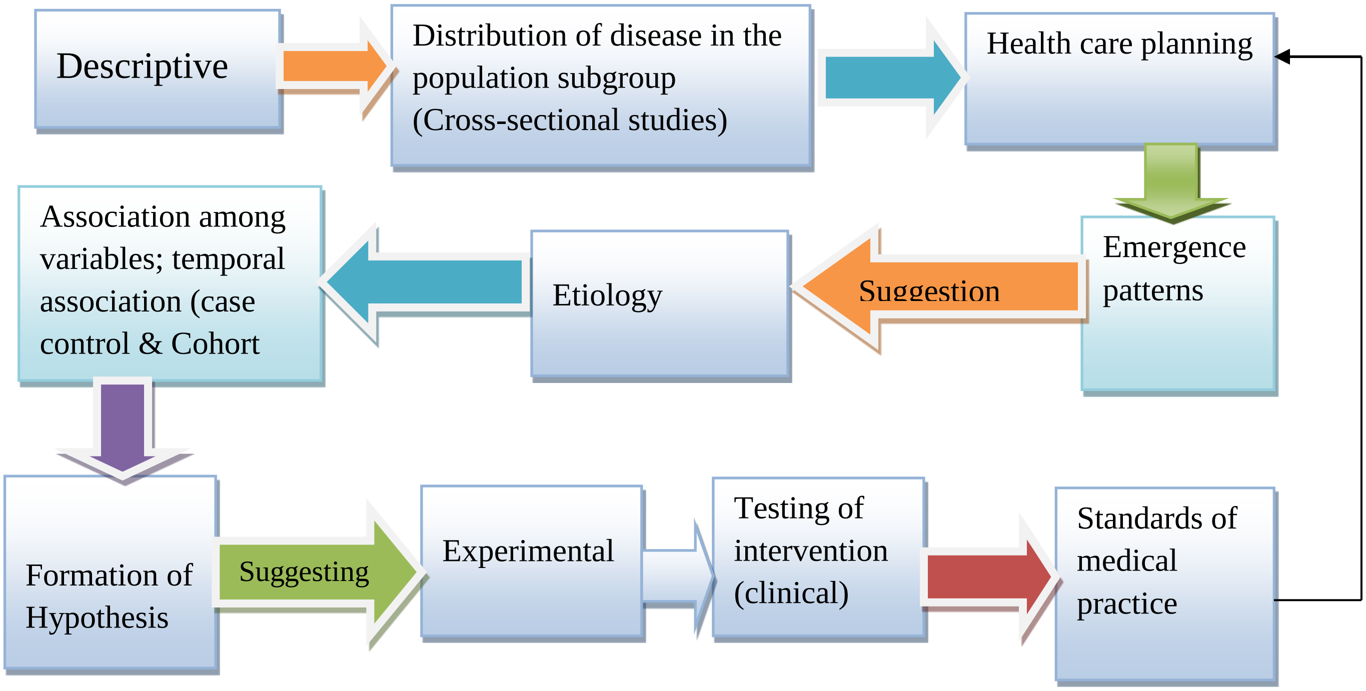
This involves a lot of spending and therefore it is often said that public health is "**purchasable**".

3. ***Evaluation of the Intervention Program***

Any program designed to prevent and control disease must be evaluated its efficiency, effectiveness and feasibility impact.

4. ***Classification of Disease*****:** Diseases are classified based on the route of transmission and severity (chronic or infectious).

The following flow chart shows the uses of epidemiology


Descriptive
Distribution of disease in the population subgroup (Cross-sectional studies)
Health care planning
Emergence patterns
Suggestion
Etiology
Association among variables; temporal association (case control & Cohort
Formation of Hypothesis
Suggesting
Experimental
Testing of intervention (clinical)
Standards of medical practice

1.3 BASIC ASSUMPTIONS IN EPIDEMIOLOGY

1. Nonrandom distribution of diseases i.e. the distribution of disease in human population is not random or by chance &
2. Epidemiology is also based on the assumption that diseases have causal and preventive factors and these can be identified by studying human populations at different places and times. Since distribution of diseases is not random or by chance, we need to identify what factors lead to the higher level of occurrence of a disease in one area as compared to others. So human diseases have causal and preventive factors

1.4 SOME BASIC CONCEPTS

Clinical Versus Community Medicine

- Clinical medicine is concerned with diagnosing and treating diseases in individual patients, while
- Community medicine is concerned with diagnosing the health problems of a community, and with planning and managing community health services.
- Public health - a science and an art of preventing disease, prolonging life, and promoting health and efficiency through organized community effort for sanitation, control of communicable disease, health education, etc.

Community Diagnosis

- Is the process of identification and detailed description of the most important health problems of a given community
- Information on the health and disease of a defined community is gathered through community diagnosis

Methods of Community Diagnostics

1) Discussion with community leaders and health workers
2) Survey of available health records
3) Field survey
4) Compilation and analysis of the data

It is impossible to address all the identified problems at the same time because of resource scarcity. Therefore, the problems should be put in the order of priority using a set criterion.

Criteria for Priority Setting

- ✓ **Magnitude** (amount or frequency) of the problem
- ✓ **Severity** (to what extent is the problem disabling, fatal,....)
- ✓ **Feasibility** (availability of financial and material resource, effective control method)
- ✓ **Community concern** (whether it is a felt problem of the community)
- ✓ **Government concern** (policy support, political commitment)

1.5 SOME EPIDEMIOLOGIC CONCEPTS: MORTALITY RATES

"Mortality is the fundamental factor in the dynamics of population growth and causes of death."

- ➢ Back to the 19th century, the major causes of mortality were influenza, pneumonia, tuberculosis and gastroenteritis.
- ➢ Now days in the 21st century, the main causes of death are Ischemic heart disease, stroke, lower respiratory infections and chronic obstructive lung cancer.

Mortality rates

The occurrence of death in the different subgroup of individuals may yield clues to the existence of a health problem.

In general, there are three types of measure of occurrence i.e. **Ratio**, **Proportion** and **Rates**

The most important epidemiological tool used for measuring occurrences is the rate. However, ratios and proportions are also used

Ratio: quantifies the magnitude of one occurrence or condition in relation to another.

Example: The ratio of male to female patients treated per day.

Proportion: is a type of ratio, which quantifies occurrences in relation to the population in which these occurrences take place. Ex. Male births/ (Male births + Female births)

Rate: is a proportion with time element (crude birth rate, crude death rates)

Mortality in a population can be monitored through a variety of measures

To calculate rate, we must have an estimate of the population at risk during a specific time period for the denominator but ratio and proportion does not require this.

The three types of measures of mortality rate are:

- ⦿ Crude death rates
- ⦿ Category/case specific death rate
- ⦿ Standardized death rate

Reading assignment: discuss on the advantage and disadvantage of the three types of mortality measures and their calculation.

Remark: Ratios, proportion, and rates are used in infectious disease epidemiology to describe morbidity (disease) and mortality (death).

1.6 MEASURE OF DISEASE FREQUENCY (INCIDENCE AND PREVALENCE RATES)

The prevalence rate and the incidence rate are two measures of morbidity (illness)

A. PREVALENCE

Prevalence rate measures the number of people in a population who have a disease at a given time. It includes both new and old cases. There are two types of prevalence rates; **Point** Prevalence rate & **Period** Prevalence rate.

Point Prevalence: is the amount of disease present in a population at a single point in time.

$$\textbf{Point PR=} \frac{\text{Number of people with the disease or condition at a specified time}}{\text{Population during the same time period}} * \mathbf{10^n}$$

Example 1: A review of patients reported to the tuberculosis registry in city A revealed that as of July 1, 2005 there were 35 cases that had not yet completed therapy. The most recent population estimate for that city was 57,763.

The prevalence of TB in that city on July 1, 2005 was: (35/57,763) X 10^4 = **6 per 10,000 people**

Period prevalence: refers to prevalence measured over an interval of time. It is the proportion of persons with a particular disease or attribute at any time during the interval (week, year, decade, or any other specified period). Period prevalence is calculated as:

$$\textbf{Period PR=} \frac{\text{All new and pre - existing cases (old + new) during a given time period}}{\text{Population during the same time period}} *10^n$$

Example 2: Two surveys were done in the same community 12 months apart. Of 5000 people surveyed the first time, 25 had antibody to histoplasmosis. Twelve months later, 35 had anti bodies, including the original 25. Calculate the prevalence at the first survey and at the 12 month interval.

Prevalence at the first survey is point prevalence and calculated as:

Point PR= (25/5000)*10^3=5.

This implies that there are 5 antibodies people per 1000 population.

Prevalence at the 12 month interval is period prevalence and find as:

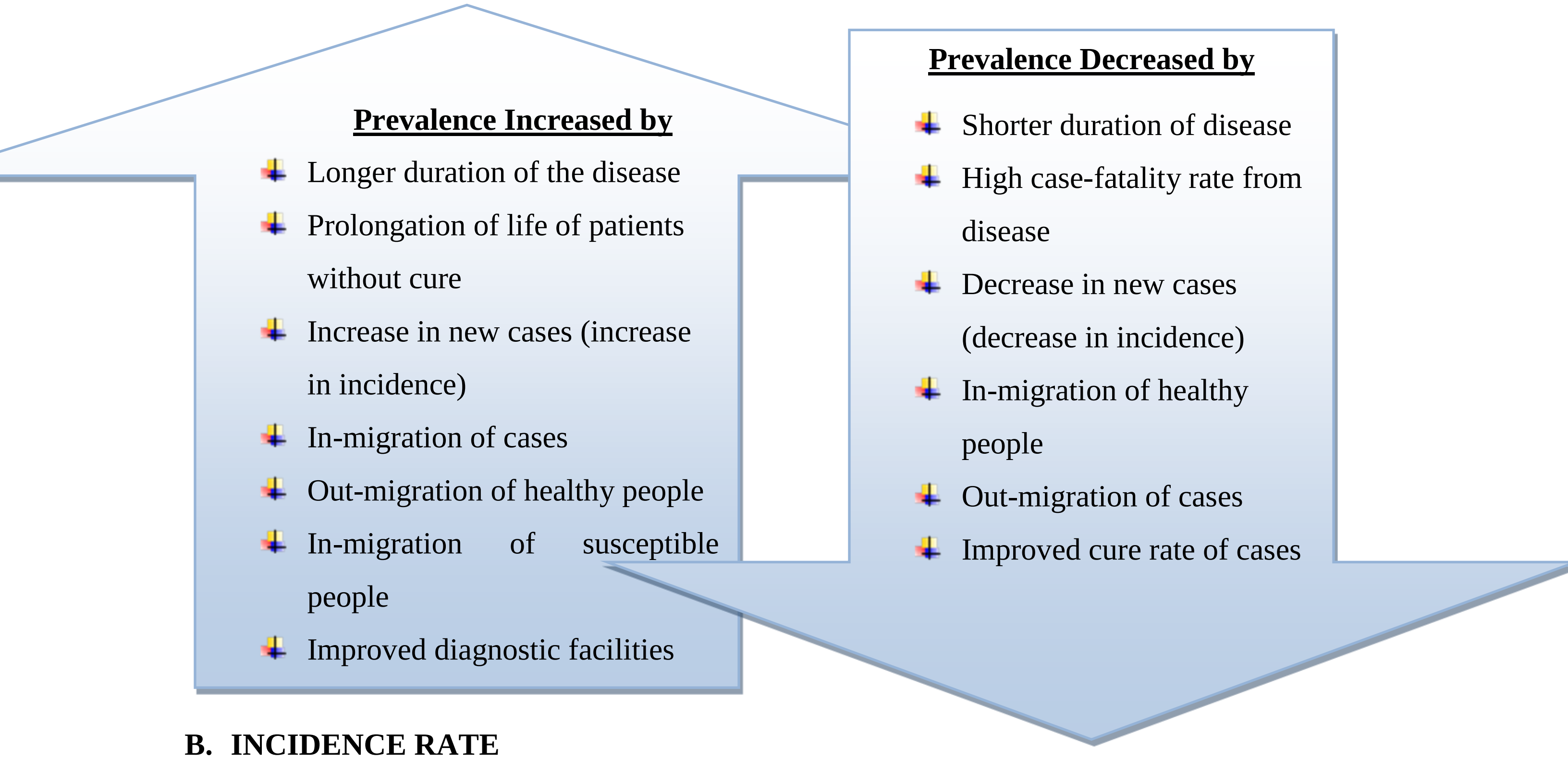
Antibodies positive at the 12 month interval is 35 (25 old cases + 10 new cases) the total population was 5,000

Period PR=35/5,000*1000=7 antibody to histoplasmosis peoples per 1,000 population.

Characteristics of prevalence

- Cause and effect measured simultaneously implies impossible to infer causation
- Useful for planning (beds, clinics, workforce needs)
- High prevalence doesn't mean necessarily high risk, could reflect increased survival (improved care, behavioral change, long duration)
- Low prevalence could reflect a rapidly fatal process, rapid cure of disease or low incidence
- Easy to obtain

Factors Affecting Prevalence



Prevalence Increased by

- Longer duration of the disease
- Prolongation of life of patients without cure
- Increase in new cases (increase in incidence)
- In-migration of cases
- Out-migration of healthy people
- In-migration of susceptible people
- Improved diagnostic facilities

Prevalence Decreased by

- Shorter duration of disease
- High case-fatality rate from disease
- Decrease in new cases (decrease in incidence)
- In-migration of healthy people
- Out-migration of cases
- Improved cure rate of cases

B. INCIDENCE RATE

Incidence refers to the rate at which new events occur in a population. Incidence takes into account the variable time periods during which individuals are disease-free and thus "at risk" of developing the disease. In the calculation of incidence, the numerator is the number of new events that occur in a defined time period, and the denominator is the population at risk of experiencing the event during this period. **Incidence is a measure of risk.**

- ✓ Two commonly used types of incidence are **cumulative incidence rate (Probability of developing disease, risk or incidence proportion)** and **incidence Density**

Cumulative Incidence (CI)

Cumulative incidence is the proportion of an initially disease free population that develops disease, becomes injured, or dies during a specified (usually limited) period of time.

Cumulative Incidence (CI) is calculated as follows

$$CI = \frac{\text{Number of new cases of disease or injury during specified period}}{\text{Number population at risk during the beginning of the period}} * 10^n$$

Remark: The people who are **susceptible** to a given disease (conditions) are called the population **at risk**.

Example: In the study of diabetics, 100 of the 189 diabetic men died during the 13-year follow-up period. Calculate the risk of death for these men or cumulative incidence or risk.

Solution:

Numerator = 100 deaths among the diabetic men

Denominator = 189 diabetic men

Risk (CI) = (100 / 189) x $10^2 \approx$ 53 deaths per 100 diabetic men

Example: In an outbreak of gastroenteritis among attendees of a corporate picnic, 99 persons ate potato salad, 30 of whom developed gastroenteritis. Calculate the risk of illness among persons who ate potato salad.

Solution:

Numerator = 30 persons who ate potato salad and developed gastroenteritis

Denominator = 99 persons who ate potato salad $10^n = 10^2 = 100$

Risk (CI) = (30 / 99) x 100 = 0.303 x 100 = 303 developed gastroenteritis per 1000 persons who ate potato salad.

Incidence Density (ID) or Person Time Rate

Is an incidence rate whose denominator is calculated using **person time unit**. Similar to other measure of incidence, the numerator of the incidence density is the number of new cases in the population. The denominator is the sum of each individual's time at risk or the sum of the time that each person remained under observation. Incidence density (ID) is calculated as follows

$$ID = \frac{\textit{Number of new cases of disease or injury during specified period}}{\textit{Time each person was observed, totaled for all persons}} * 10^n$$

Each subject contributes a specific person-time of observation (days, months, years) to the denominator. Many researchers assume that persons lost to follow-up were, on average, disease-free for half the year, and thus contribute ½ year to the denominator. Similarly, persons diagnosed with the disease contribute ½ year of follow-up during the year of diagnosis.

- ✓ Sometimes detailed information about time at risk is unavailable for each member of the population.
- ✓ It may nevertheless be possible to estimate total person-time at risk by multiplying the average size of the population at risk by duration of the observation period.
- ✓ Total person-time ~ (Avg. size of population at risk) X (Length of observation period).

Often the mid-period size of the population at risk is used as an estimate of the average population at risk

Example: Investigators enrolled 2,100 women in a study and followed them annually for four years to determine the incidence rate of heart disease.

The study results could be described as follows: No heart disease was diagnosed at the first year. Heart disease was diagnosed 1 woman at the second year, 7 women at the third year, and 8 women at the fourth year of follow-up.

100 women were lost to follow-up by the first year, another 99 were lost to follow-up at the second years, another 793 were lost to follow-up at third years, and another 392 women were lost to follow at the 4th years, leaving 700 women who were followed for four years and remained disease free. Calculate the incidence rate of heart disease among this cohort. Assume that persons diagnosed with disease and those lost to-follow-up were disease free for half of the year and thus contribute 1/2 year to the denominator.

Then number of new cases diagnosed=0+1+7+8=16

The person-years of observation:

(2,000+1/2*100)+(1,900+1/2*1+1/2*99)+(1,100+1/2*7+1/2*793)+(700+1/2*8+1/2*392)=6,400 person-years of observation.

A second way to calculate the person-years of observation is to turn the data around to reflect how many people were followed for how many years, as follows

- ✓ At the end of 4th year, 700 disease free women and 8 disease women + 392 lost follow up (400 women) contribute 700*4.0 years + 400*3.5 years = **4,200** person-years.
- ✓ At the end of 3rd year (7+793) =800 women contribute 800*2.5 years=**2,000** person-years
- ✓ At the end of 2nd year (1+99) =100 women contribute 100*1.5 years=**150** person-years
- ✓ At the end of 1st year (0+100) =100 women contribute *0.5 years=**50** person-years
  - ➢ Then Total = 4200+200+150+50 = **6,400** person-years of observation

Person-time rate = $ID = \frac{Number\ of\ cases\ during\ 4-years\ study}{Time\ each\ person\ was\ observed,\ totaled\ for\ all\ persons} *10^{n}$

=16/6,400*10^{n}=0.0025*10^{n}, if n=4, there were 25 cases per 10,000 persons per years

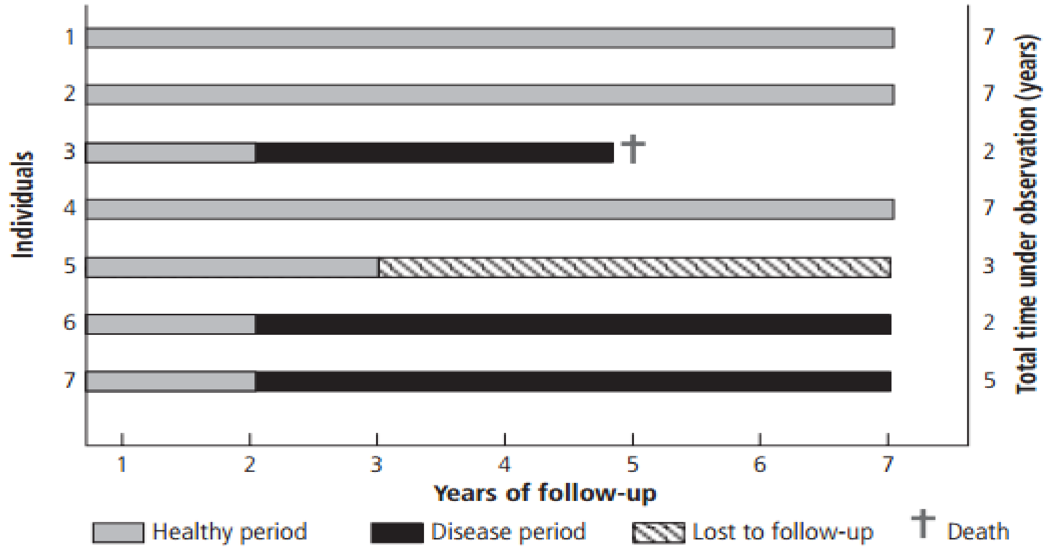
In contrast, the cumulative incidence or incidence proportion can be calculated as 16 / 2,100 = 76 cases per 10,000 populations during the four-year period, or an average of 19 cases per 10,000 per year (76 divided by 4 years). The incidence proportion underestimates the true rate because it ignores persons lost to follow-up, and assumes that they remained disease free for all four years.

Example: The figure bellow illustrates the various measures of disease. This hypothetical example is based on a study of seven people over seven years.

Calculate cumulative incidence, incidence density and prevalence at the start of year four.

- **Cumulative Incidence:** is the number of new events in the population at risk (3) divided by the number of people in the same population free of the disease at the beginning of the study period (7), i.e. 43 cases per 100 persons during the seven years;
- **Incidence density** of the disease during the seven-year period is the number of new events (3) divided by the sum of the lengths of time at risk of getting the disease for the population (30 person years), i.e. 1 cases per 10 person years;
- **Prevalence** depends on the point in time at which the study takes place; at the start of year 4 is the ratio of the number of people with the disease (2) to the number of people in the population observed at that time (6), i.e. 33 cases per 100 persons.

Figure 2.3. Calculation of disease occurrence



1.7 MEASURE OF ASSOCIATION

In epidemiologic studies, we are often interested in knowing how much more likely an individual is to develop a disease (condition) if he or she is exposed to a particular factor than the individual who is not so exposed. The strength of association between exposure and outcome is mostly assessed by calculating relative risk (risk ratio) and odds ratio.

****Risk*** is the probability or likelihood that an event will happen.

A characteristic (factor) that influence whether or not an event occurs is called risk factor or exposure.

A. RELATIVE RISK (RR)

- ✓ The relative risk (also called the risk ratio) is the ratio of the risk of occurrence of a disease among exposed people to that among the unexposed. RR shows the magnitude of association between exposure & disease
- ✓ Indicates the likelihood of developing the disease in exposed group relative to non-exposed
- ✓ RR can also be used to compare risks of death, injury, and other possible outcomes of the exposure.
- ✓ It is a direct measurement of a risk to develop the outcome of interest and usually used in cohort and experimental studies

RR is computed as: $RR = \frac{\text{Incidence Rate among Exposed group } (I_1)}{\text{Incidence Rate among non Exposed group } (I_0)}$

Epidemiologic data are often presented in the form of two-by-two (contingency) table as follow.

Exposure status	Disease status		
	Yes (+)	No (-)	Total
Exposed	a	b	a+b
Non exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

Therefore, relative risk is computed as RR = $\frac{\mathbf{a/(a+b)}}{\mathbf{c/(c+d)}} = \frac{\mathbf{a(c+d)}}{\mathbf{c(a+b)}}$

Remark: *RR* is a number between 0 and ∞.

Example ; Among 2390 women aged 16 to 49 years who were free from bacteriuria, 482 were OC users at the initial survey in 1973, while 1908 were not. At a second survey in 1976, 27 of the OC users had developed bacteriuria, as had 77 of the nonusers. Calculate the measure of association and interpret it.

We can construct the 2x2 table as follow:

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

Then, RR $= \frac{a(c+d)}{c(a+b)} = \frac{27*1908}{77*482} = 1.4$

Interpretation: women who used oral contraceptive had 1.4 times higher risk or more likely to developing bacteriuria than when compared to non-users.

Example: Data from a cohort study of postmenopausal hormone use and coronary heart disease among female nurses

Postmenopausal hormone use	**Coronary heart disease**		
	Yes	No	Person years
Yes	30		54,308.7
No	60		51,477.5
Total	90		105,786.2

RR $= \frac{IDe}{IDo} = \frac{30/54308.7}{60/51477.5} = 0.5$

Interpretation: Women who used postmenopausal hormones had 0.5 times, or only half, the risk of developing coronary heart disease compared with nonusers. This means that postmenopausal hormone is a protective factor.

B. ODDS RATIO (OR)

An odds ratio (OR) is another measure of association that quantifies the relationship between an exposure by calculating the ratio of the odds of exposure among the cases to that among the controls.

- ✓ It is an indirect measure of a risk in a disease of rare occurrence
- ✓ It is usually used in a case-control and cross-sectional studies
- ✓ Cases and controls are predetermined and we are calculating to determine whether cases or controls are more exposed to postulated risk factors.

Exposure status	Disease status		
	Yes (+)	No (-)	Total
Exposed	A	b	a+b
Non exposed	C	d	c+d
Total	a+c	b+d	a+b+c+d

Referring to the four cells (2x2) table, the odds ratio is calculated as

$$\textit{Odds ratio (OR)} = \frac{\text{Odds of exposed among cases (a/c)}}{\text{Odds of exposed among controls (b/d)}} = \frac{\boldsymbol{ad}}{\boldsymbol{bc}}$$

Note: The interpretation is nearly the same as seen in relative risk except some technical aspects in the calculation of odds. For rare disease; a+c ≈ c and b+d ≈ d, the RR ≈ ${}^{a/c}\!/_{b/d}$ = OR

Example: Of 156 women with Myocardial Infarction (MI), 23 were current OC users at the time of their hospital admission. Of the 3120 control women without MI, 304 were current OC users. Calculate the measure of association and interpret it.

	Myocardial Infarction (MI)		
Current OC users	Diseased (Yes)	No disease (No)	Total
Users	23	304	327
Non users	133	2816	2949
Total	156	3120	3276

Then, $OR = \frac{\boldsymbol{ad}}{\boldsymbol{bc}} = \frac{\mathbf{23*2816}}{\mathbf{304*133}} = \mathbf{1.6}$

Interpretation: The odds of Myocardial Infarction among women using OC, is 1.6 times higher than women without oral contraceptive use.

Interpretation of measure of association (RR/ OR)

RR/ OR > 1, the exposure is risk

RR/ OR = 1, there is no association

RR/ OR < 1, the exposure is preventive

1.8 MEASURE OF IMPACT (ABSOLUTE MEASURES)

A. Attributable Risk (AR) or Risk Difference

Absolute risk or the attributable risk or risk difference is a measure of association indicating absolute difference of diseases in exposed group than unexposed group, assuming the association between the exposure and disease is causal.

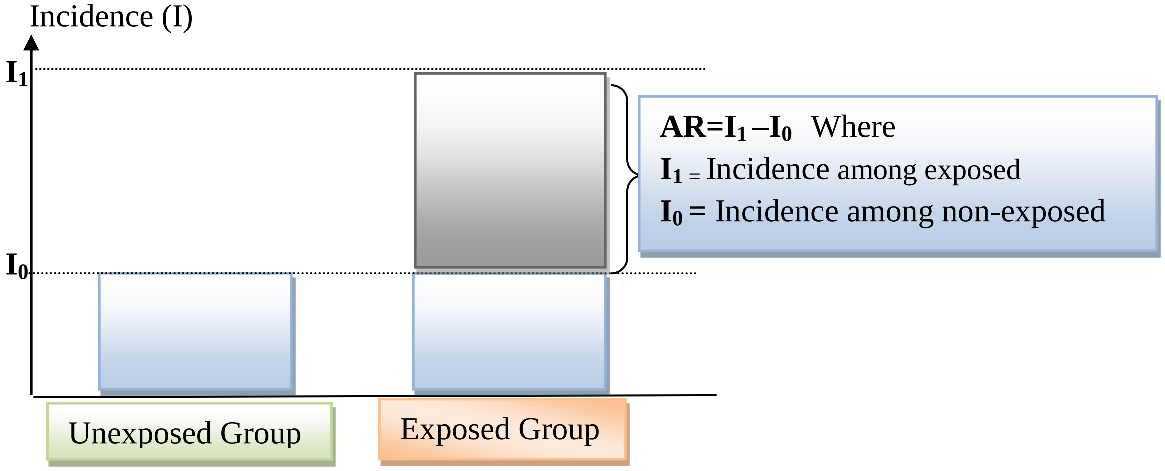
☞ Is also called as excess risk of developing diseases among exposed groups

☞ Risk difference (RD) indicates how much of the risk is due to (attributable to) the exposure alone

☞ The attributable risk is the difference between the disease rate in exposed groups and the disease rate in non-exposed groups

☞ Quantify the excess risk in the exposed that can be attributable to the exposure

Graphically we can show the excess risk or AR as follow.



AR=$I_1 - I_0$ Where
I_1 = Incidence among exposed
I_0 = Incidence among non-exposed

***Example*[**]:** Among 2390 women aged 16 to 49 years who were free from bacteriuria, 482 were OC users at the initial survey, while 1908 were not. At a second survey, 27 of the OC users had developed bacteriuria, as had 77 of the nonusers. Calculate the excess risk or attributable risk for bacteriuria.

Solution: We can construct a table as follow.

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

Incidence among exposed group is given as (I_1): $I_1 = {}^{27}/_{482} = 0.05602$

Incidence among exposed group is given as (I_0): $I_0 = {}^{77}/_{1908} = 0.04036$

Then the risk difference (AR) is given by: AR= I_1- I_0= 0.056-0.04= 0.01566 or ≈16 per 1000 OC users.

Interpretation: The excess occurrence of bacteriuria among OC users attributable to their OC use is 16 per 1000 OC users.

B. Attributable Risk Percent (AR %)

- Attributable proportion, also known as attributable risk percent, is a measure of the public health impact of causative factors.
- It is the proportion of additional cases observed because of the exposure
- It represents the expected reduction in disease if the exposure could be eliminated
- AR % is an attributable risk expressed as a percentage of risk in exposed groups

It is calculated as: $AR\% = \frac{\mathrm{I_1-I_0}}{\mathrm{I_1}} * 100 = \frac{\mathrm{AR}}{\mathrm{I_1}} * 100$

Example: refer example** previously done about bacteriuria and OC user and find AR%.

$I_1 = 0.056$, I_1- I_0 = 0.016. Then, $AR\% = \frac{\mathrm{I_1-I_0}}{\mathrm{I_1}} * 100 = \frac{0.016}{0.056} * 100 = 28.57\%$ per year

Interpretation: If OC use causes for bacteriuria, about 29 % of bacteriuria among women who use OC can be attributed to their OC use and can be eliminated if they did not use oral contraceptives.

C. Population Attributable Risk (PAR)

- PAR shows the effect of eliminating the exposure on the population as a whole.
- PAR takes into account not only the actual incidence rate of the outcome but also the prevalence rate of the exposure.
- **PAR=$\mathbf{I}_{\text{population}}$ – $\mathbf{I}_{\text{unexposed}}$** or PAR = AR X prevalence rate of the exposure

Example: From example** about bacteriuria, calculate PAR.

The PAR of bacteriuria associated with OC use (Table 1) is:

PAR = I_T - I_o = 104/2390 – 77/1908 = 316 per 100,000 per year

Alternatively: PAR = (AR) (PR) = $1566/10^5$ X (482/2390) = 316 per 100,000 per year PR is exposure prevalence rate.

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 women.

D. Population Attributable Risk Percent (PAR %)

PAR% estimates the proportion of disease in the study population that is attributable to the exposure, and thus could be eliminated if the exposure were eliminated. Population attributable risk percent is the proportion of the risk in the population that is related to the exposure to the postulated risk factor

$$PAR\% = \frac{\text{PAR}}{\text{Incidence rate in total population}} * 100$$

Example: PAR = 17.8 per 105 per year

Mortality rate in non-smokers = 7 per 105

Mortality rate in the total population = 24.8 per 105 per year

Calculate PAR %. Then PAR % = $\frac{\text{17.8 per 105 per year}}{\text{24.8 per 105 per year}} * 100 = 71.8\%$

Interpretation: 72% of deaths from lung cancer occurring in the general population could be prevented by eliminating cigarette smoking.

Interpretation of measure of Impact (AR/PAR)

AR/ PAR > 0, the exposure is attributing

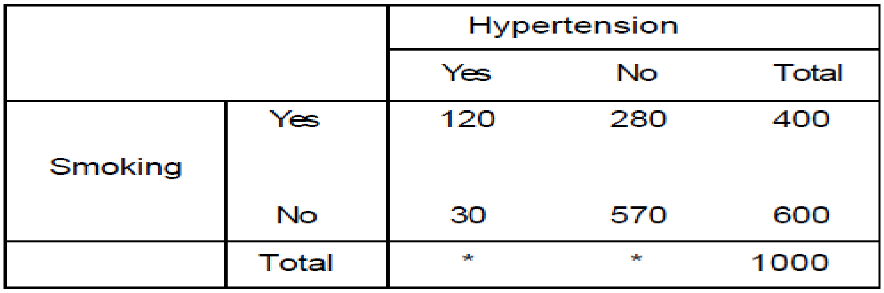
AR/ PAR = 0, there is no attribution

AR/ PAR < 0, the exposure is not attributing (it is preventive)

- Both AR and PAR are used to estimate the effect on disease incidence of eliminating a given risk factor,
- AR estimates reduction in disease incidence only in those exposed,
- PAR estimates reduction in disease incidence in the population as a whole.

Self-Exercise

Suppose that a cohort study of 400 smokers and 600 non-smokers documented the incidence of hypertension over a period of 10 years. The following table summarizes the data at the end of the study period:



		Hypertension		
		Yes	No	Total
Smoking	Yes	120	280	400
	No	30	570	600
	Total	*	*	1000

Based on the above information, calculate and interpret the following measures of association:

1. Prevalence of hypertension
2. Relative risk (RR)
3. Attributable risk (AR) and/or preventive fraction (PF)
4. Attributable risk percent (AR%)
5. Population attributable risk (PAR)
6. Population attributable risk percent (PAR%)

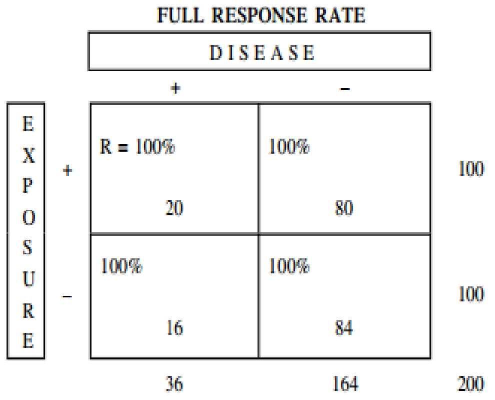
1.9 RESPONSE BIAS

There are many different types of bias that might lead you to either underestimate or overestimate the size of a relative risk or odds ratio, and it is important to try to anticipate potential sources of bias and avoid them. The illustration on the next page shows the impact of one kind of possible bias: ***ascertainment or response bias***.

Assume that you have the following situation. Of 100 people exposed to a risk factor, 20% develop the disease and of a 100 people unexposed, 16% develop the disease, yielding a relative risk of 1.25, as shown in the illustration. Now imagine that only 60% of the exposed respond to follow-up, or are ascertained as having or not having the disease, ***a 60% response rate among the exposed***.

Assume further that all of the ones who don't respond happen to be among the ones who ***don't*** develop disease. The relative risk would be calculated as 2.06. Now imagine that only 60% of the non-exposed reply, ***a 60% response rate among the non-exposed,*** and all of the non-exposed who don't respond happen to be among the ones who don't have the disease. Now the relative risk estimate is 0.75.

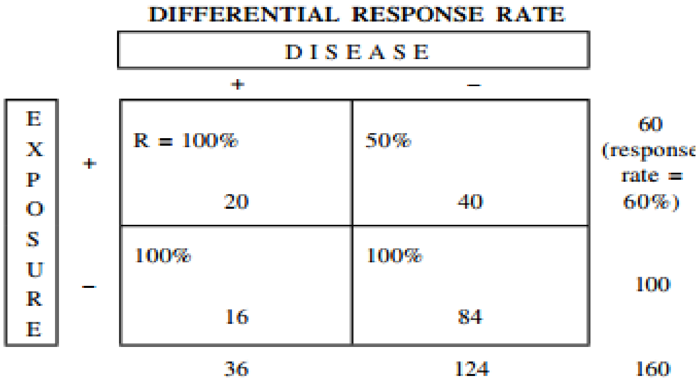
To summarize, you can get conflicting estimates of the relative risk if you have differential response rates. Therefore, it is very important to get as complete a response or ascertainment as possible. Please refer the following charts.



FULL RESPONSE RATE

EXPOSURE		DISEASE +	DISEASE −	
	+	R = 100% 20	100% 80	100
	−	100% 16	100% 84	100
		36	164	200

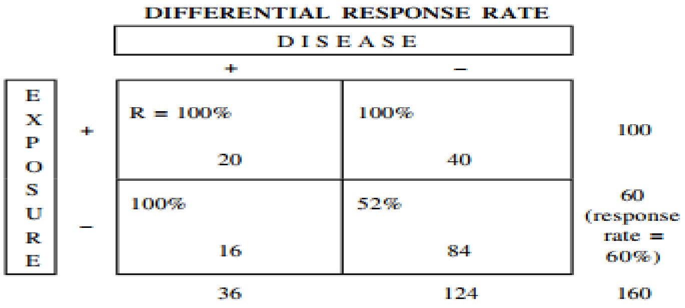
$$RR=\frac{20/100}{16/100}=\frac{.20}{.16}=1.25$$



DIFFERENTIAL RESPONSE RATE

EXPOSURE		DISEASE +	DISEASE −	
	+	R = 100% 20	50% 40	60 (response rate = 60%)
	−	100% 16	100% 84	100
		36	124	160

$$RR=\frac{20/60}{16/100}=\frac{.33}{.16}=2.06$$



DIFFERENTIAL RESPONSE RATE

EXPOSURE		DISEASE +	DISEASE −	
	+	R = 100% 20	100% 40	100
	−	100% 16	52% 84	60 (response rate = 60%)
		36	124	160

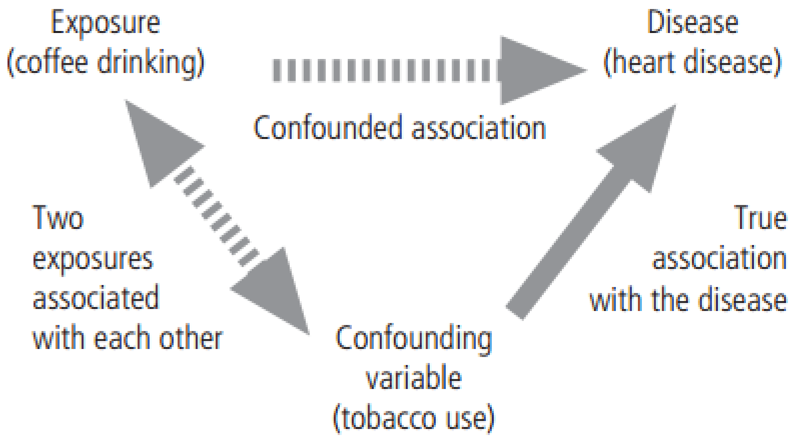
$$RR=\frac{20/100}{16/60}=\frac{.20}{.266}=.75$$

1.10 CONFOUNDING VARIABLES

A ***confounding variable*** is one that is closely associated with both the independent variable and the outcome of interest in those unexposed. For example, a confounding variable in studies of coffee and heart disease may be smoking. Since some coffee drinkers are also smokers, if a study found a relationship between coffee drinking (the independent variable) and development of heart disease (the dependent variable), it could really mean that it is the smoking that causes heart disease, rather than the coffee. In this example, smoking is the confounding variable.

If *both* the confounding variable and the independent variable of interest are closely associated with the dependent variable, then the observed relationship between the independent and dependent variables may really be a reflection of the ***true*** relationship between the confounding variable and the dependent variable.

The following figure show about Confounding: relationship between coffee drinking (exposure), heart disease (outcome), and a third variable (tobacco use)



Control of Confounding

Several methods are available to control confounding, either through study design or during the analysis of results.

- ✓ The methods commonly used to control confounding in the design of an epidemiological study are:
  - ➢ ***Randomization***
  - ➢ ***Restriction***
  - ➢ ***Matching***
- ✓ At the analysis stage, confounding can be controlled by:
  - ➢ Stratification
  - ➢ Statistical Modeling.

Randomization

In experimental studies, randomization is the ideal method for ensuring that potential confounding variables are equally distributed among the groups being compared.

The sample sizes have to be sufficiently large to avoid random maldistribution of such variables. Randomization avoids the association between potentially confounding variables and the exposure that is being considered.

Restriction

One way to control confounding is to limit the study to people who have particular characteristics.

For example, in a study on the effects of coffee on coronary heart disease, participation in the study could be restricted to nonsmokers, thus removing any potential effect of confounding by cigarette smoking.

Matching

Matching is used to control confounding by selecting study participants so as to ensure that potential confounding variables are evenly distributed in the two groups being compared.

For example, in a case-control study of exercise and coronary heart disease, each patient with heart disease can be matched with a control of the same age group and sex to ensure that confounding by age and sex does not occur.

Matching can be expensive and time-consuming, but is particularly useful if the danger exists of there being no overlap between cases and controls, such as in a situation where the cases are likely to be older than the controls.

Stratification and Statistical Modeling

In large studies, it is usually preferable to control for confounding in the analytical phase rather than in the design phase. Confounding can then be controlled by stratification, which involves the measurement of the strength of associations in well-defined and homogeneous categories (strata) of the confounding variable.

If age is a confounder, the association may be measured in, say, 10-year age groups; if sex or ethnicity is a confounder, the association is measured separately in men and women or in the different ethnic groups. Methods are available for summarizing the overall association by producing a weighted average of the estimates calculated in each separate stratum.

Although stratification is conceptually simple and relatively easy to carry out, it is often limited by the size of the study and it cannot help to control many factors simultaneously, as is often necessary. In this situation, multivariate statistical modeling is required to estimate the strength of the associations while controlling for several confounding variables simultaneously; a range of statistical techniques is available for these analyses.