



**FACULTY OF MEDICINE  
MANSOURA NATIONAL UNIVERSITY**



# SEMESTER 4



**STAFF OF MANSOURA NATIONAL UNIVERSITY**

**2024**



**FACULTY OF MEDICINE  
MANSOURA NATIONAL UNIVERSITY**



# **RESEARCH, EVIDENCE-BASED AND BIostatISTICS**



## Introduction to research

### **Research Design** **The 8 steps model**

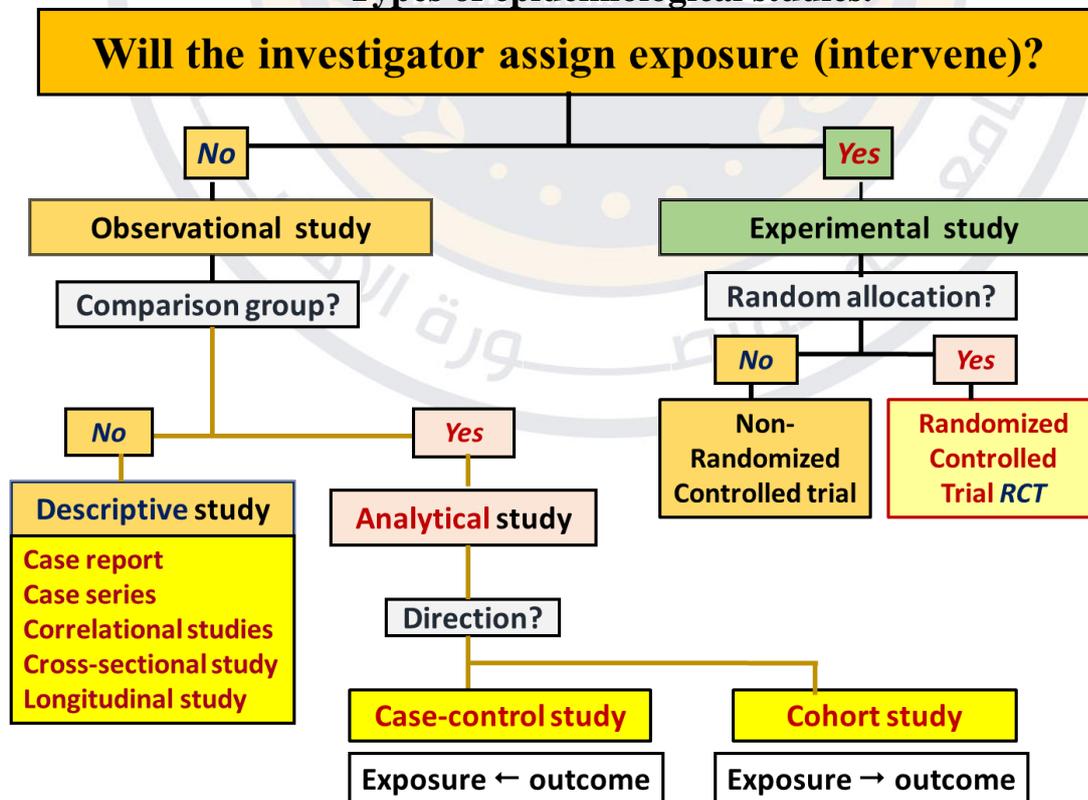
#### A. Steps in planning research study:

- Step 1:** Formulate a **research problem**.
- Step 2:** Research **design**.
- Step 3:** Construct **tools** for data collection.
- Step 4:** Select a **sample (size & method)**.
- Step 5:** Write research protocol (**proposal**).

#### B. Steps in conducting study:

- Step 6:** **Collecting** data.
- Step 7:** **Processing** data.
- Step 8:** **Writing** research report.

### Types of epidemiological studies.



## Descriptive studies

### Uses

- ☒ First phase in the epidemiological investigation.
- ☒ Describes the **pattern**, **characteristics** and **distribution** of a disease or health problem in the population.
- ☒ Give data about:
  - When the disease occurs (Time).
  - Where the disease occurs (Place).
  - Who is getting the disease (Person).
- ☒ **Formulating (not testing) research hypotheses** (It is the 1<sup>st</sup> step in the search for determinants or risk factors).

### Types of descriptive studies:

#### 1- Case Reports:

Presentation of a single case that is newly reported or has unique finding e.g.

- ☒ Newly described disease.
- ☒ Unexpected or new therapeutic effect.
- ☒ Link between diseases.

#### 2- Case Series:

- ☒ Describe a number of similar cases with a given disease in one report.
- ☒ May describe unusual variations of a disease and May indicate the start of an epidemic.
- ☒ A major trigger for further research.

#### 3- Ecological studies:

- ☒ **Looking for associations (correlation) between exposures & outcomes in population** rather than in individuals.
- ☒ **Use already collected population data** (e.g. vital statistics, censuses and national health surveys).
- ☒ **Comparing populations in different places at the same time or in a time series by comparing the same population in one place at different times.**

☒ **Examples:**

- Mortality from CHD & per capita sales of cigarettes.
- High incidence of MI & consumption of dietary fat & fast food.
- Negative correlation between access to efficient ANC & maternal mortality rate.

**4- Cross-Sectional Studies (Prevalence studies):**

- ☒ It is **observational** study that **carried out once** (snapshot of a population) at a **single point in time**.
- ☒ Both **exposure** (risk factors) and **outcome** (diseases) **are present** (we cannot determine if exposure preceded disease or not).
- ☒ It measures **prevalence**, not incidence of disease.
- ☒ Used to study conditions that are relatively *frequent with long duration* of expression (non-fatal, chronic conditions).

$$\text{Prevalence rate} = \frac{\text{Number of new and preexisting cases of disease during specified period}}{\text{Population examined during the specified period}} \times 10^{11}$$

Advantages of Cross-sectional Studies	Disadvantages of Cross-sectional Studies
<ul style="list-style-type: none"> <li>☒ Used to study conditions that are relatively frequent with long duration (<u>chronic conditions</u>).</li> <li>☒ Good for generating <u>hypotheses</u> about the <u>cause of disease</u>.</li> <li>☒ Can estimate <u>prevalence rates</u> and <u>exposure proportions</u> in the population.</li> <li>☒ Relatively <u>easy, quick and inexpensive</u>.</li> <li>☒ <u>No follow up</u>, relatively <u>easy, quick and inexpensive</u>.</li> <li>☒ It is the first step to develop <u>evidence for causal association</u>.</li> </ul>	<ul style="list-style-type: none"> <li>☒ It is not useful for studying: <ul style="list-style-type: none"> <li>• Acute diseases.</li> <li>• Diseases with seasonal variations.</li> <li>• Highly fatal diseases.</li> <li>• Rare diseases.</li> </ul> </li> <li>☒ Can't estimate <u>incidence rate</u>.</li> <li>☒ It gives very little information about the <u>natural history of diseases</u>.</li> <li>☒ Cannot determine if exposure preceded disease or not.</li> <li>☒ Not differentiate between causes of disease &amp; factors associated with disease.</li> <li>☒ Not provide solid evidence for <u>causal association</u> as it does not determine if really exposure preceded disease or not.</li> </ul>



## 5- Longitudinal (incidence) studies:

- ☒ Repeated observations (follow-up) in same community over prolonged period to identify new cases of disease.
- ☒ Follow up and re-examination have the following problems:
  1. Loss to follow-up.
  2. Difficulty in maintaining standards and stability of clinical and laboratory examination over a long period of time.
- ☒ It is used to measure:
  1. Incidence rate.
  2. Natural history of dis. & its final outcome (case fatality, survival).
  3. Risk factors of disease.





## Analytical studies

✚ Basic Question in Analytic Epidemiology: Are exposure & outcome (disease) linked?

- ☒ These studies are used to test an etiologic hypothesis such as smoking and Lung cancer; excess carbohydrates and obesity
- ☒ Analytical studies always require the use of control group.

✚ Types of analytical studies:

### A. Case-control studies (Retrospective Studies)

☒ **Definition:**

It is an “observational” in which we assess the frequency of exposure to specific risk factor (suspected etiological factors) in patients who have developed a disease

- ☒ It is compared with that of controls who do not have the disease
- ☒ Case-control studies provide a relatively simple way to investigate causes of diseases, especially rare diseases
- ☒ The investigator is looking backward from the disease to a possible cause (retrospectively) by direct questioning and or extracted from clinical records

☒ **Steps to conduct case control study:**

1. **Identify the study group (cases):** Define case & criteria for inclusion & exclusion of cases
2. **Identify controls:** (needed for comparison)
  - Controls must be free from the studied disease.
  - Controls must be matched with cases for certain characteristics known to influence the outcome of the disease (confounding factors) e.g. age, sex, social class ...
  - Sources of controls:
    - General population
    - Hospital controls
    - Special control series (Family members- friends- neighbors)



3. Summarize your data in 2x2 (association) tables:

- Summarize frequencies of disease & exposure
- Calculate association

Exposure	Disease status		Total
	Cases	Controls	
Yes (+)	(a) Diseased with exposure	(b) Not diseased with exposure	(a+b) Total exposed
No (-)	(c) Diseased without exposure	(d) Not diseased without exposure	(c+d) Total non-exposed
Total	(a+c) Total cases	(b+d) Total control	(a+b+c+d) Grand total

- Proportion of the exposed among cases (P1) =  $a/a + c$
- Proportion of the exposed among control (P2) =  $b/b + d$
- **Relative contribution = P1 – P2**

It represents relative contribution of the suspected cause to the total frequency of the disease

- Odds Ratio = odd of exposure among cases / odd of exposure among controls

$$\text{Odds Ratio} = a/c / b/d = ad/bc$$

**It is the indirect estimation of the risk**

- **Interpretation of the odds Ratio (OR):**
  - **OR = 1** Exposure is not associated with outcome or disease.
  - **OR > 1** Increased exposure accompanies increased outcome
  - **OR < 1** Increased exposure accompanies decreased outcome.



☒ **Advantages of case control studies:**

1. Easy to carry out.
2. Quick & cheap.
3. Can be used in rare diseases.
4. Allows the study of several risk factors.
5. Useful in the study of disease with a long latency.
6. Does not require large samples.
7. Can prove hypothesis (Exposure & Disease are related).
8. Can estimate risk (odds ratio).

☒ **Disadvantages of Case Control Study:**

1. Cannot calculate prevalence or incidence rates.
2. Not useful in rare exposure.
3. Liable to bias or mistakes

**N.B: Bias:** Any systematic error in the design, conduct or analysis of a study that can result in wrong results.





## Analytic studies II

### Cohort study

#### Definition:

- ☒ Cohort is a group having a common characteristic, e.g. a smoker's cohort means all are smokers in that group.
- ☒ An observational prospective (longitudinal or follow up) study in which we compare
- ☒ exposed group (individuals with a risk factor) with
- ☒ non exposed group (others without the risk factor)
- ☒ as regards the incidence of a disease over time.

#### Steps of cohort:

1. First we exclude cases of disease under investigation.
2. The free cohort, divided into 2 groups:
  - **Exposed group:** individuals exposed to risk factor.
  - **Control group:** individuals not exposed to this factor.
3. Both groups are followed up over a sufficient period of time.  
Therefore the cohort should be stable, cooperative & accessible to the investigator.
4. If the incidence of disease among exposed group is higher than its incidence among non exposed group, this supports the etiological hypothesis



## Basic analysis of cohort study:

Exposure	Disease status		Total
	Present	Absent	
Yes (+)	(a) with exposure with disease	(b) with exposure without disease	(a+b) Total exposed
No (-)	(c) Without exposure with disease	(d) without exposure Without disease	(c+d) Total non- exposed
Total	(a+c) Total with disease	(b+d) Total without disease	(a+b+c+d) Grand total

### ☒ Basic analysis involves: calculation of:

1. Overall incidence =  $a + c / a+b+c+d$
2. Incidence rate among the exposed ( $I_e$ ) =  $(a/a+b)$
3. Incidence rate among the non-exposed ( $I_n$ ) =  $(c/c+d)$
4. Relative risk (RR) =  $\frac{\text{Incidence among exposed } (I_e)}{\text{Incidence among non exposed } (I_n)}$

RR answers the question: "How many times exposed person is at risk of developing disease compared to non-exposed?"

5. Attributable risk (AR) =  $I_e - I_n$

AR answers the question: "How much of the studied disease can be attributed to exposure". **“Proportion of disease in a population that would be eliminated if risk factor is eliminated”**.

### ☒ Interpretation of Relative Risk (RR):

- **RR = 1:** No association between exposure & disease.
- **RR > 1:** Positive association (increased risk) i.e. exposed group has higher incidence than non-exposed group.
- **RR < 1:** Negative association (protective effect) i.e. non-exposed group has higher incidence.



**Example**

	Developed CHD	Do not Develop CHD	Total
Smokers	60	40	100
Non-smokers	20	80	100

**Answer**

- Incidence in smokers =  $60/100$
- Incidence in non-smokers =  $20/100$
- Relative risk =  $60/20 = 3$  (smokers are at a higher risk of developing CHD 3 times than non-smokers).
- Attributable risk =  $60-20/100 = 40/100$  (40 out 100 of CHD cases among smokers is attributed to their smoking).



### Differences between case-control & cohort studies

Case-control studies	Cohort studies
Proceed from outcome to cause ( from disease to risk factor)	Proceed from cause to outcome (from risk factor to disease)
Compares people with disease & those without disease	Compares exposed with non exposed
Retrospective	Prospective
<b>☒ Aims</b> to prove or disprove that suspected cause occurs more frequently in diseased than non diseased	to prove or disprove that suspected disease occurs more frequently in exposed than non exposed.
<b>☒ Advantages</b> <ol style="list-style-type: none"> <li>Cheap &amp; quickly done.</li> <li>Does not require large sample.</li> <li>Useful in studying rare diseases.</li> <li>Can study several risk factors.</li> <li>Can estimate risk (odds ratio)</li> </ol>	<ol style="list-style-type: none"> <li>Less bias in selection of control.</li> <li>Methods can be standardized.</li> <li>Study several outcomes.</li> <li>Valuable in rare exposure.</li> <li>incidence rate and relative risk can be calculated</li> </ol>
<b>☒ Drawbacks</b> <ol style="list-style-type: none"> <li>Liable to bias.</li> <li>Not useful in rare exposures.</li> <li>Uncertain data due to incomplete records of past events &amp; unstandardized observation.</li> <li>Difficulty to be sure that the association is causal or not.</li> </ol>	<ol style="list-style-type: none"> <li>Expensive and time consuming.</li> <li>Needs a very large sample even with common diseases.</li> <li>Delayed results if latent period is long.</li> <li>Prolonged follow up can cause drop out of cases and loss of standardization.</li> </ol>



## Summary and wrap-up

- ☒ Cohort study is a type of **observational analytic study** designs.
- ☒ The participants **do not have** the outcome of interest to begin with. They are selected based on the exposure status of the individual.
- ☒ If the exposure is rare, then a cohort design is an efficient method .
- ☒ Losses during Follow-up of the study participants is very important source of bias.
- ☒ These studies are used to estimate the incidence rate and relative risk.





## Experimental studies

### + Types of the studies

#### ☒ Non-intervention studies:

- Descriptive studies.
- Comparative (analytical) studies.

#### ☒ Intervention studies:

- Experimental studies.

### + What is intervention study?

A prospective study comparing the effect and value of intervention (s) against a control in human being. It confirms etiological hypothesis & assess effectiveness of preventive measures & new therapies

### + Two approaches

1. **Addition of possible causal agent (therapeutic)** e.g. testing new drug, implantation of organ. May be dangerous or fatal for human, for practical and ethical reasons.
2. **Protection from causative agent:** by removing agent from environment (smoking) or administering a protective measure (preventive) e.g. fluoridation of water supplies or vaccination. These are safe and done on human.

### + Characteristics of experimental study

- ☒ **Manipulation:** the researcher does something to one group of subjects in the study.





- ☒ **Control:** the researcher introduces one or more control group(s) to compare with the experimental group.
- ☒ **Randomization:** the researcher takes care to randomly assign subjects to the control and experimental groups. (Each subject is given an equal chance of being assign in either group)

#### + Intervention studies

Investigator determines which individuals are exposed to factor of interest (intervention arm) and which are unexposed (control arm).

#### + Stages and phases of clinical trials:

**A. Stage 1 (preclinical studies = pre-phase I): involves lab. animals**

**B. Stage 2: involves human participants**

1. Phase I trials
2. Phase II trials
3. Phase III trials
4. Phase IV trials (post-marketing surveillance)

#### + Stage 1 (preclinical studies = phase 0 = pre-phase I) :

☒ **In vitro and lab. animals. We look at five things:**

- Pharmacokinetics.
- Pharmacodynamics.
- Drug metabolism.
- Lethal dose (LD50).
- Teratogenic effects.





## **Stage 2: involves human participants**

### **☒ Phase I trials (20-100 subjects):**

- Not randomized, volunteers.
- Drugs with serious SE can be tested on seriously ill patients who have failed to respond to current established therapy.
- To assess safety, pharmacokinetics, and safe dose range.

### **☒ Phase II trials (100-200 subjects):**

- Therapy is still promising after phase I.
- Objectives are to set & test dose necessary for pharmaco-dynamic effects, to evaluate potential effectiveness (preliminary efficacy) and to determine optimal method of administration.

### **☒ Phase III trials (the classical phase) (500-1500 subjects):**

- Randomized double blind controlled trial with adequate sample size & power.
- Aim to assess efficacy and additional safety
- Used to evaluate whether a new product should be licensed for public use.
- Provide decision makers with scientific evidence about relative effectiveness and safety of competing treatments.

### **☒ Phase IV trials :**

Conducted after treatment is approved for general use. Aim to assess:

- Effectiveness
- Drug safety: long term effects, surveillance for rare SE.
- Drug interactions with other drugs or diets.
- Pharmaco-epidemiology: distribution & determinant of drug use.
- Pharmaco-economics: cost-effectiveness of drug.
- Benefits & harms in presence of comorbidity.





### **+ Conducting a randomized Clinical trial**

- A. Formulate a hypothesis.
- B. Select participants and get informed consent.
- C. Allocate subjects to comparison groups.
- D. Administer treatment and measure outcome.
- E. Analyze data

### **+ Allocate subjects to comparison groups**

- By randomization into: study & control groups. (means each subject in reference group has an equal chance to be present in either groups)
- Study group exposed to intervention.
- Control group: No treatment or Placebo
- Both groups must be matched

### **+ Blindness**

- Means ensuring that a person “investigator, data collector, or analyst” remains unaware of which arm a subject has been allocated to.
  - **Single-blind:** the subject participating in the trial.
  - **Double-blind:** the subject & investigators (clinician, interviewers, laboratory personnel).
  - **Triple blind:** the subject, investigators & the data analysts.





## Medical statistics 1

### + Definition

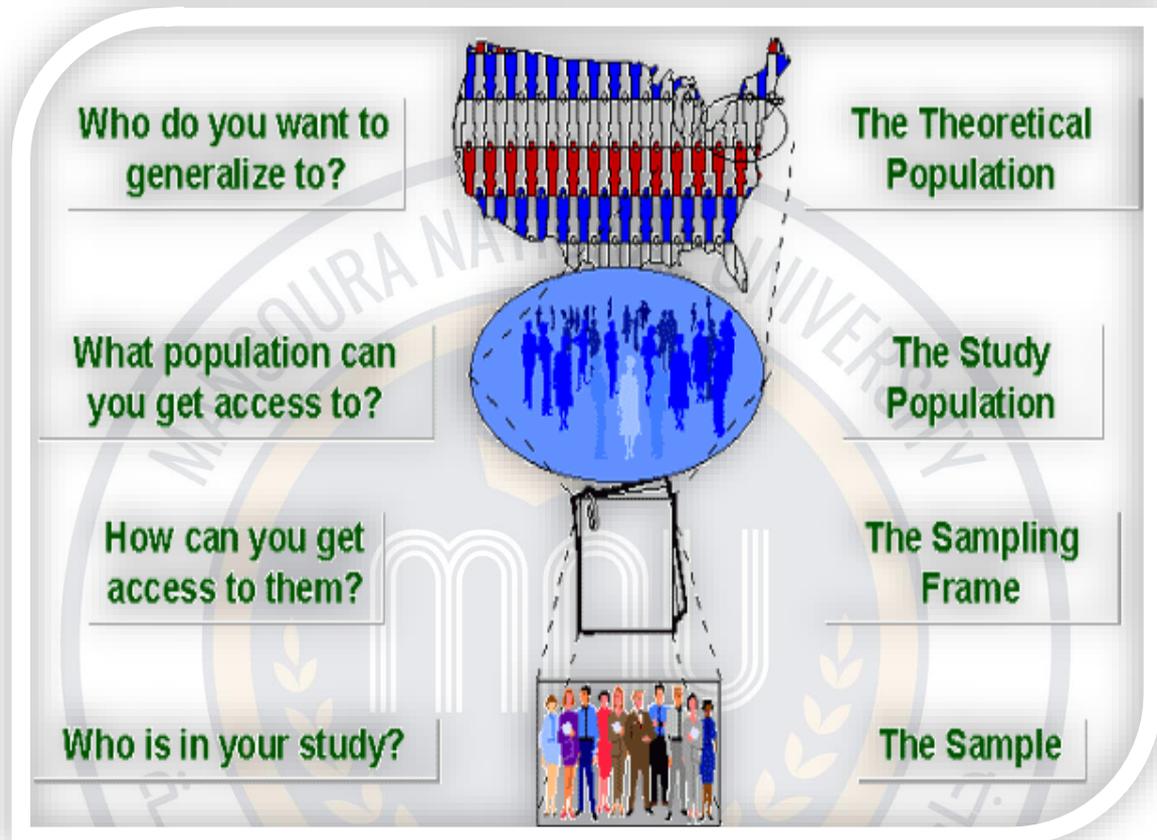
It is the study of methods of collecting, presenting (descriptive statistics), analysing and evaluating conclusions from data (inferential statistics).

### + Importance:

- ☒ It presents facts
- ☒ It simplifies mass of figures
- ☒ It reduces the volume of data
- ☒ It facilitates comparison
- ☒ It helps in:
  - formulating and testing hypothesis
  - formulation of suitable policies.
  - measuring the standard of health



## Population & samples & sampling techniques



### + Sample:

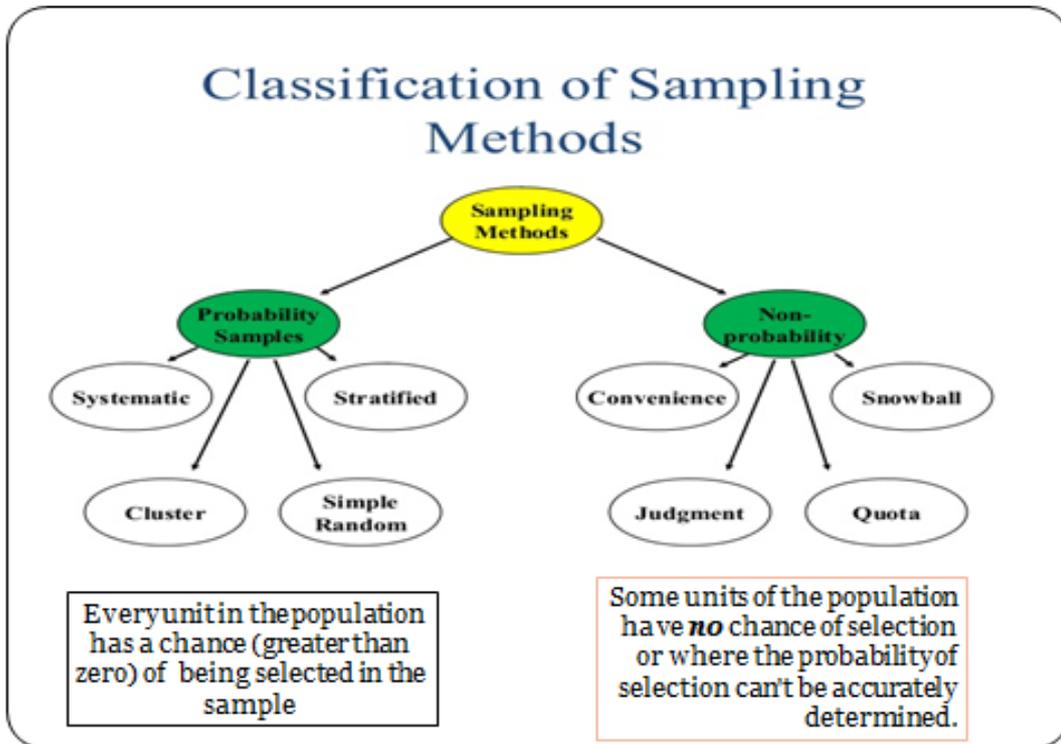
is a subset of population that is used to gain information about the entire population. A good sample : representative-adequate-unbiased

### + Why Sampling?

- ☒ Lower cost
- ☒ Saves time
- ☒ Provides more intensive and accurate investigations and information.

## What happens when there is no sampling?

Selection Bias (non-representative sample): systematic difference between the characteristics of the people selected for a study and those who are not.



## Sampling techniques :

- Non-Probability
- Probability

Population characteristics	Appropriate sampling technique
I. Population is a homogeneous mass of individuals	Simple Random Sample
II. Population is heterogeneous, consists of definite strata each of which is different, characteristics	Stratified Random Sample
III. Sample unit is a group not an individual <ul style="list-style-type: none"> <li>• They are selected randomly from all groups of same type</li> <li>• All members of selected group will be included in the study</li> </ul>	Cluster Sample
IV. Population is a confined community Select sample units at regular intervals from this list every 3 <sup>rd</sup> or 5 <sup>th</sup> or 10 <sup>th</sup> <<<<The start is randomly >>>>	Systematic Random Sample
V. Population is distributed over a large geographical area as in national surveys	Multi-stage Random Sample

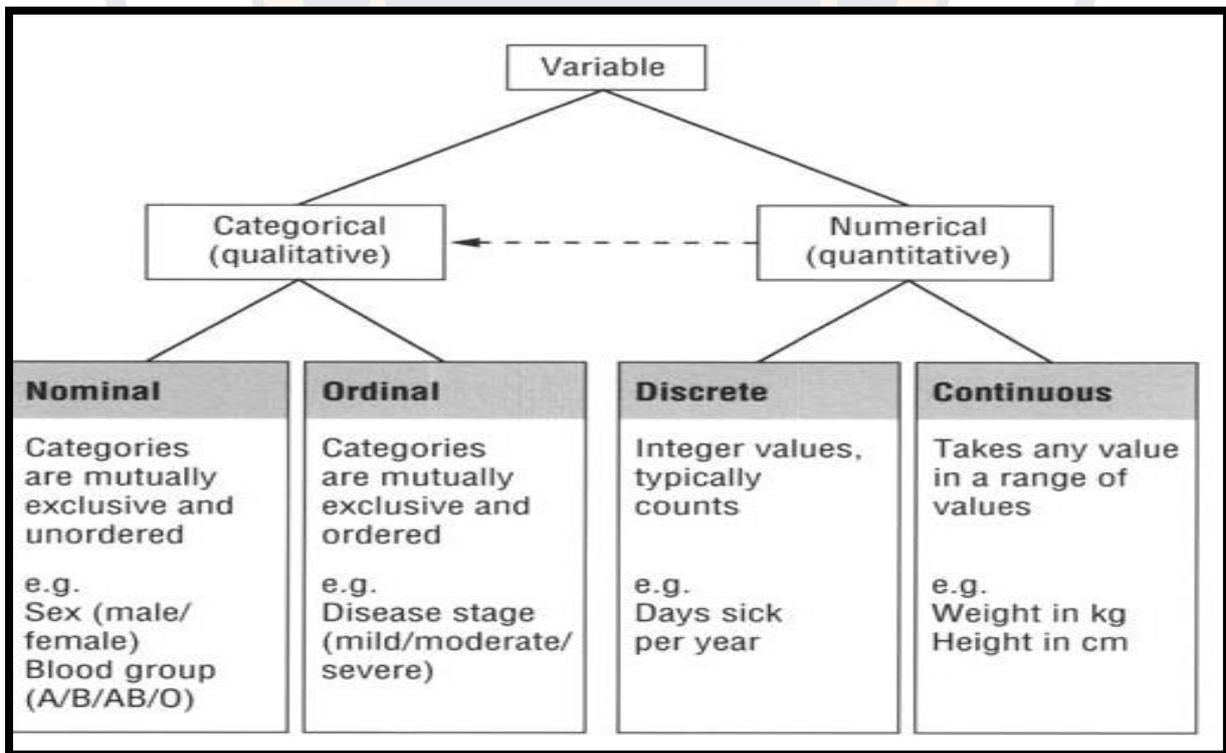
### **Data & Information**

- ☒ Data consist of discrete observations of variables that carry no or little meaning when considered alone.
- ☒ Data need to be transformed (manually or by computer programs) into information by reducing them and adjusting them for variations in age and sex and others. Information support decision-makers, policy makers and planners to take proper action in their works.

## ✚ Sources of Data:

- ☒ Population Census
- ☒ Registration of vital events e.g. Births and deaths, marriage
- ☒ Notification of diseases (Disease Registers) Communicable and non-communicable diseases.
- ☒ Hospital Records
- ☒ Epidemiological surveillance
- ☒ Health Service records
- ☒ Environmental Health data
- ☒ Health Surveys
- ☒ Published articles and reports

✚ **Definition of variable:** Is a characteristic or attribute that vary from person to person, from time to time and from person to person





## **Analysis of quantitative data**

### **☒ Measures of central tendency or averages.**

- Mean
- Median
- Mode

### **☒ Measures of dispersion (spread)**

- Range
- Mean deviation
- Variance
- Standard deviation

### **☒ Measures of location**

- Percentile
- Quartile

## **Measures of central tendency or averages.**

### **☒ Mean (Average):**

- Is obtained as sum of all values divided by the no. of values  $Q$ .
- $\text{Mean} = \sum x/n$
- **Advantages:** used in quantitative continuous data (normally distributed)
- **Disadvantages:**
  - affected by extreme values
  - it should not be used for non parametric or skewed data
- **Importance:** Best summarizing value for normally distributed data

### **☒ Median:**

- The median is the value that lies in the middle of the ordered observations.  
**A. When sample size is odd number:**
  - The observations are ordered according to an ascending or descending magnitude.
  - Determine the rank of the median given by  $((n + 1)/2)$
  - Using the obtained rank and referring back to the ordered or arranged observations and find the value of median.



### B. When sample size is even:

- In a distribution with even no. of total values: Such a distribution has 2 middlemost values; median is the average of two middlemost values when arranged in an ascending or descending order of values.
- Median = Mean (average) of  $(n/2)$ th and  $(n/2 + 1)$ th value in ascending order
- **Advantages :**
  - It can be used with quantitative & qualitative ordinal variables (e.g. median number of patients in cancer stages).
  - It is useful for summarizing data with extreme values as it is not affected by extreme values.
- **Disadvantages**
  - It cannot be used with qualitative nominal variables.
  - 2- It is not easy to be used in statistical analysis

### ☒ Mode:

- Most frequent or most commonly occurring value in a distribution
- This is done by finding the observation which has the highest frequency.
  - e.g. weight of five children as follows : 9, 8, 12, 7, 8 kg.
  - It is seen that eight is the observation of highest frequency.
  - The mode = 8 kg
- A similar procedure can be used for finding the mode from qualitative data.
- **Advantages:**
  - It can be used in all types of variables
  - It is not affected by extremes or out-lying observation
- **Disadvantages:**
  - Sometimes the mode cannot be determined, this happens when all observation have the same frequency (i.e. uniform distribution).
  - Sometimes we may obtain two modes (bimodal) or more (multimodal) from the same group of data.
  - e.g. 22, 24, 26, 28, 24, 26, Mode= 24 & 26



## **Measures of dispersion (spread)**

- Using measures of central tendency is not enough to describe completely a mass of data.
  - If we have five persons with age 30, 34, 32, 36 and 28 years, the mean age is 32 years.
  - We get the same mean age of 32 years for other five persons have their ages as 12, 30, 8, 62 and 48 years but the two groups are totally different.

### **Measure of dispersion include :**

#### **☒ Range:**

- It is a simple measure of dispersion and by definition range is difference between the biggest and smallest observation.
- From the above two examples range for first group = 36 - 28 = 8 years and for second group = 62 - 8 = 54 years.

#### **☒ Mean deviation**

$$\frac{\sum |x - \bar{x}|}{n}$$

#### **☒ Variance**

$$\frac{\sum (x - \bar{x})^2}{n - 1}$$

#### **☒ Standard deviation**

- It is the commonly used measure of dispersion and generally the best.
- It measures the deviation of observations from the arithmetic me
  - obtaining the deviation of each value from the arithmetic me
  - square the deviation from the mean.
  - The squared deviations are summed and divided by the number of observations minus one (n-1) to get the variance (S<sup>2</sup>)



- The square root of variance (S<sup>2</sup>) gives us the standard deviation(S).

$$= \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

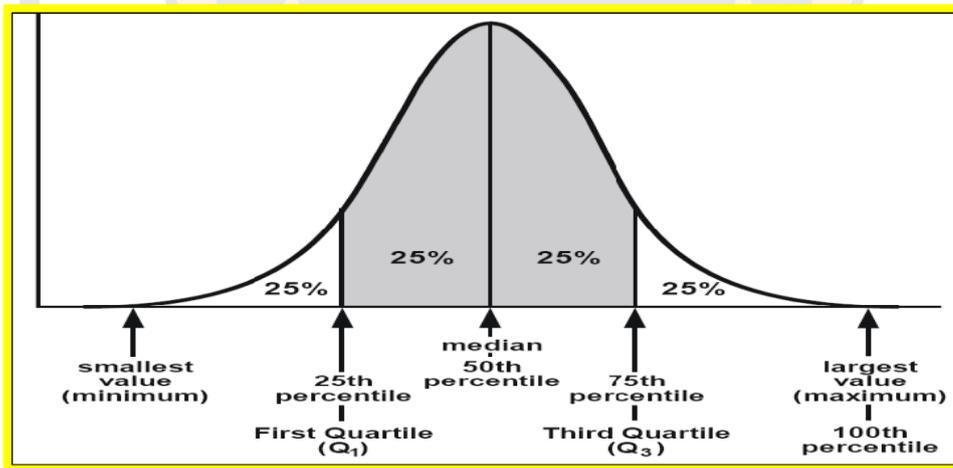
### Measures of location

#### ☒ Quartile:

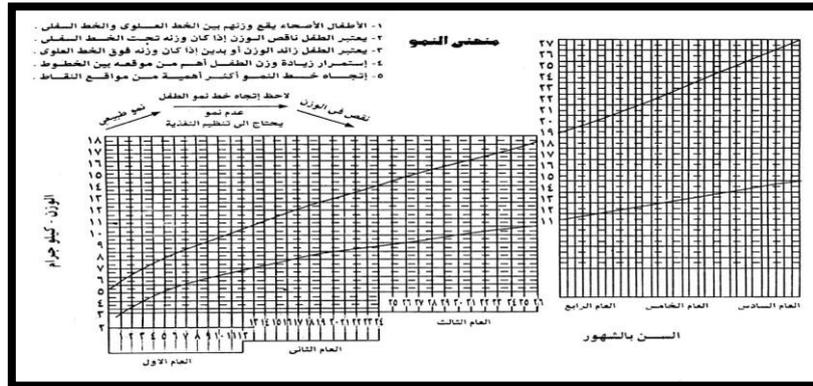
Divides a distribution into 4 equal parts, so the number of intercepts required will be 3, i.e. Q<sub>1</sub>, Q<sub>2</sub>, Q<sub>3</sub>

#### ☒ Percentile:

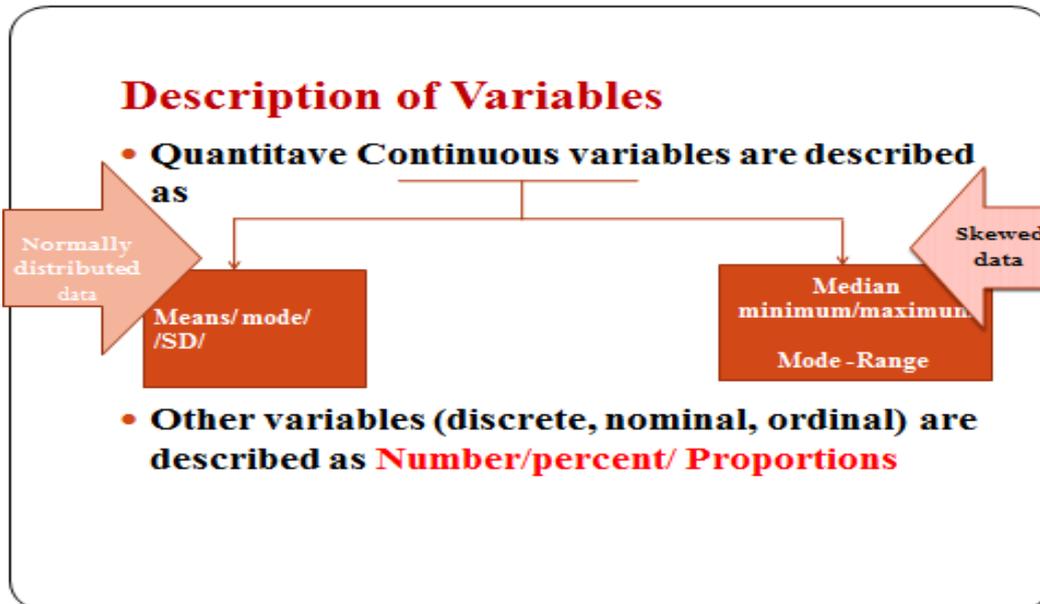
Divides a distribution into 100 equal parts, AFTER arranging in an ascending order such that each part/segment has equal number (n/100) of subjects.



Quartiles



### Percentiles



## Medical statistics II:

### Data distribution & presentation

#### Data presentation

✚ Data presentation can be either tabular or graphical

#### I . Tabular

##### ☒ Requirements of tabulation:

- Use a clear and concise title that describes the content of the data in the table.
- Precede the title with a table number.
- Label each row and each column and include the units of measurement for the data (for example, years, mm Hg, mg/dl).
- Show totals for rows and columns, where appropriate.
- Explain any codes, abbreviations, or symbols in a footnote.

##### ☒ Types of Tabular presentation

#### A. Descriptive

#### 1. Descriptive Table for Quantitative data (frequency distribution table)

Classes (height in cm)	Frequency (no of children)
100-	10
110-	15
120-	25
130-140	10
Total	60

## 2. Descriptive table summarizing Qualitative data (simple frequency table)

Barriers	Frequency (no)	%
Lack of facilities	20	20
Priority to patients needs	30	30
Fear of dry hands	10	10
Forgetfulness	40	40
<b>Total</b>	<b>100</b>	<b>100</b>

### B. Analytic

Table 1: Demographic and Occupational Characteristics of injured construction workers versus noninjured workers

Variable	Construction workers		p Value	Crude OR 95% CI
	With injuries (n = 100)	No injuries (n = 90)		
Age (mean ± SD) Min-max	33.2 ± 10.7 (15–80)	35.02 ± 11.3 (17–64)	0.2	–
Residence				
Urban (r)	16 (16)	35 (38.9)	≤0.001	3.3 (1.6–6.6)
Rural	84 (84)	55 (61.1)		
Education				
Read and write (r)	10 (10)	13 (14.4)	–	1
Basic education	43 (43)	19 (21.2)	0.05	2.94 (0.9–8.8)
Secondary and above	47 (47)	58 (64.4)	0.7	1.17 (0.43–3.2)
Marital status				
Single (r)	24 (23.8)	24 (26.7)	0.7	1.15 (0.57–2.33)
Married	76 (75.2)	66 (73.3)		
Smoking				
Non-smoker (r)	37 (37)	26 (28.9)	0.3	1.45 (0.75–2.79)
Smoker	63 (63)	64 (71.1)		
History of cannabis use				
Negative (r)	87 (87)	81 (90)	0.5	1.3 (0.5–3.3)
Positive	13 (13)	9 (10)		
Job category				
Installers of roof and floor (r)	30 (30)	64 (71.1)	–	1
Carpenters	21 (21)	7 (7.7)	≤0.001	6.4 (2.2–18.7)
Painters	37 (37)	11 (12.3)	≤0.001	7.2 (3.02–17.1)
Electrician	5 (5)	4 (4.4)	0.1 <sup>a</sup>	2.6 (0.6–12.9)
Demolition workers	7 (7)	4 (4.4)	0.04 <sup>b</sup>	3.7 (1.01–13.1)
Type of shift				
Day (r)	90 (90)	84 (93.3)	–	1
Night and alternating	10 (10)	6 (6.6)	0.5	1.56 (0.49–5.0)
Duration of employment in years				
≤9 (r)	35 (35)	42 (46.7)	–	1
10–20	42 (42)	26 (28.8)	0.05	1.9 (0.9–3.9)
>20	23 (23)	22 (24.5)	0.6	1.25 (0.56–2.8)
Past history of injury				
Negative (r)	67 (67)	85 (94.4)	<0.001	1
Positive	33 (33)	5 (5.6)		8.3 (3.1–22.6)

SD: standard deviation; OR: odds ratios; r: reference group.  
<sup>a</sup>Chi-square test.  
<sup>b</sup>Fisher's exact test.

## ☒ Tests of Hypotheses and Significance

- An investigator conducting a study usually has a research idea in mind
  - Research Question ? What are the risk factors ( age – Job category –past hx of injuries –drug abuse for occupational injuries among construction workers?
- **Null hypothesis:** there is no association between risk factor and injuries

## ☒ Alternative hypothesis:

There is association between risk factor and injuries

## ☒ Test of significance:

Quantitive data	Qualitative data
t test	Chi square test
To compare between different means	To compare frequencies of categorical variable in different groups

☒ The result of the statistical test either supports or rejects the Null hypothesis.

## ☒ Probability value (p-value)

- The p-value is the probability of obtaining the effect observed in the study (or one stronger) if the null hypothesis of no effect is actually true.
- The p value gives the probability of any observed difference having happened by chance.
- It is the cutoff for rejecting null hypothesis ( $p=0.05$ ).
- Interpretation of results:

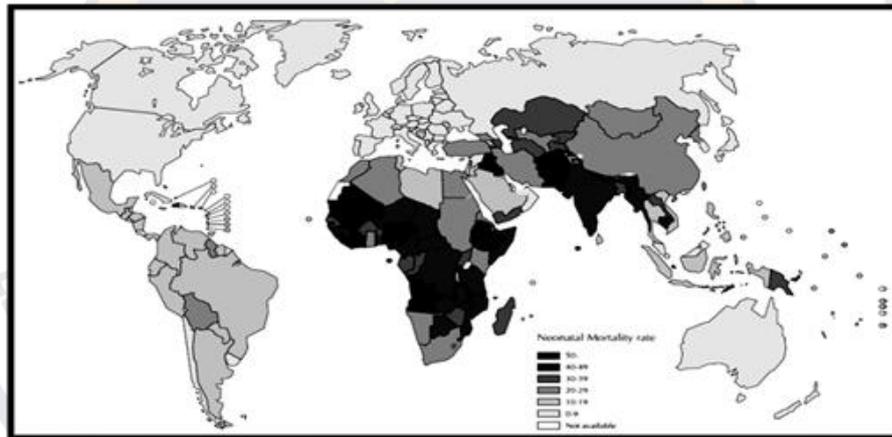
If p value less than 0.05	If p value more than 0.05
Reject null hypothesis	Accept null hypothesis
Accept alternative hypothesis	Reject Alternative hypothesis

## II . Graphical presentation:

- ☒ Visual display of data using plots and charts
- ☒ Not substitute for tables
- ☒ Stress on certain information ,quick idea about situation
- ☒ Should be as simple as possible, self-explained without reference to text

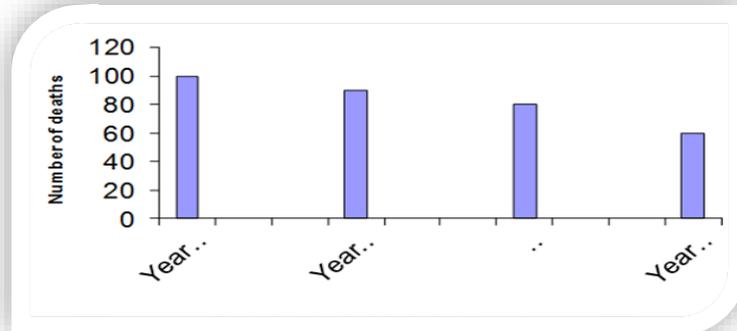
### 1. Map diagram/cartogram:

- is a map which demonstrates geographic distribution of a particular characteristic or variable
- e.g. prevalence of certain disease or infant mortality. The following Cartogram demonstrates Neonatal mortality rate, by country, 2000.

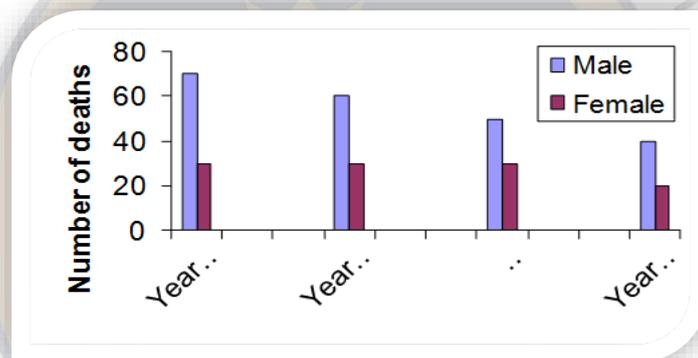


### 2. Bar chart

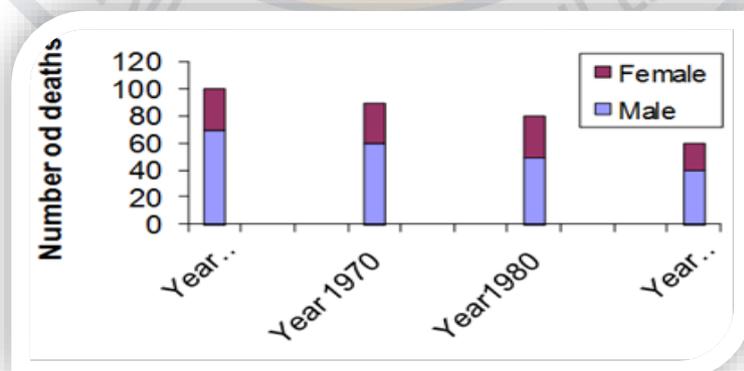
- **Simple bar chart:** Different values of a particular variable are illustrated by vertical or horizontal bars to show simple comparison of size. The base line forms the time scale.
- **Multiple bar charts:** 2 components of data represented by one chart.
- **Component bar chart:** Each bar stands for a number of components according to their relative proportion.



Simple bar chart showing number of deaths among patients admitted to the hospital (X) in years (1960-1990).



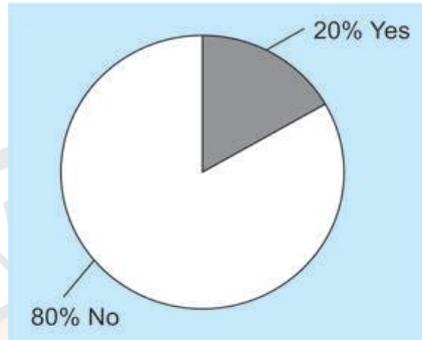
Multiple bar chart showing males & females deaths among patients admitted to the hospital (X) in years (1960-1990)



Component bar chart showing males & females deaths among patients admitted to the hospital (X) in years (1960-1990)

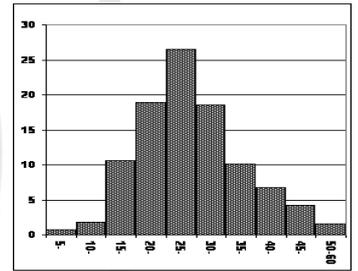
### 3. Pie or circular charts:

- Is for 'presentation of discrete data or qualitative characteristics'
- All pie categories are mutually exclusive, with a total of 100% (360°).



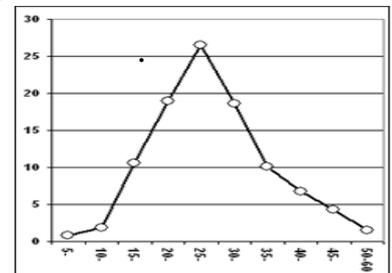
### 4. Histogram:

- Observations of frequency distribution table are illustrated on arithmetic paper as rectangles drawn side by side with no spaces between to get a block diagram (histogram). Classes of the frequency distribution are plotted along
- the X axis & height of constructed rectangles is corresponding to the frequency.

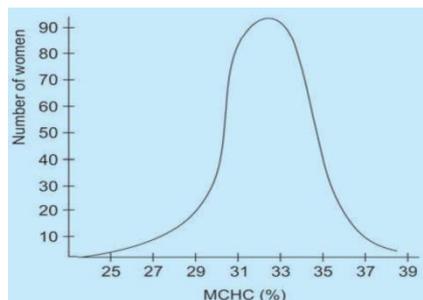


### 5. Frequency polygon:

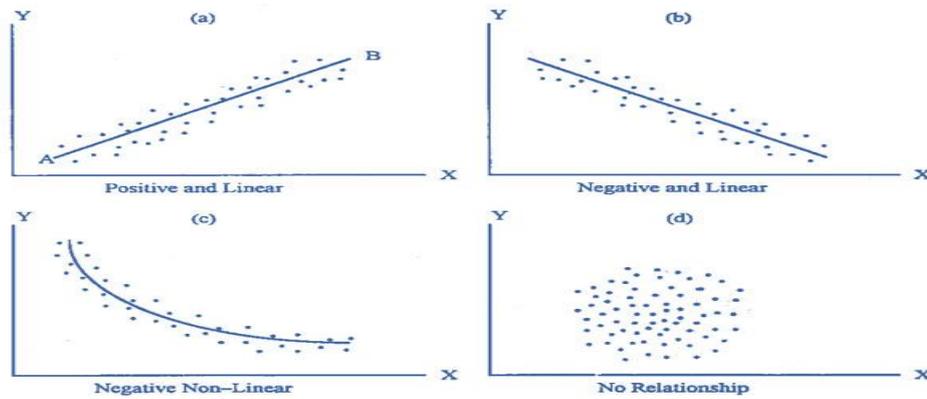
- Mid points of upper bases of rectangles are connected together by series of straight lines



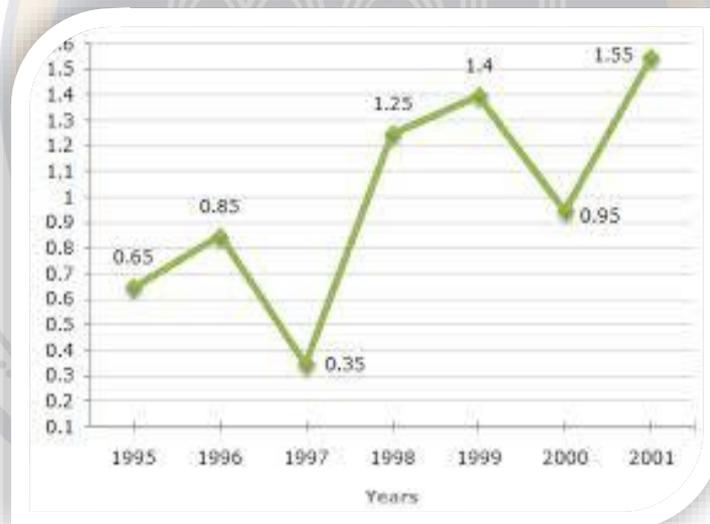
### 6. Smooth curves:



7. **Scatter diagram:** Illustrate the relationship between two continuous variables



8. **Line graph:** Is a frequency polygon presenting variations by line– It shows the trend of an event over a period of time.



Display data over time for example:  
prevalence of diseases by time

### Summary

- \* The most common types of graphical presentation for **discrete data** & qualitative data are Bar & Pie charts and map diagrams.
- \* Histogram, frequency polygon & smooth curves for **continuous data** or grouped data arranged in frequency distribution.

## Data distribution

### ✚ Symmetric distribution curves: Normal (Bell- shaped) curve & U shape of death rate according to age

#### I. Normal Distribution

- ☒ It is bell shaped and symmetric curve
- ☒ The curve rises to its peak at the mean where mean = median = mode and it is located at the midpoint of the base
- ☒ The area under normal curve unity = 100%, each half = 50%
- ☒ The area starts from -ve to +ve and the two edges of curve do not meet X line except at infinity
- ☒ The X axis is divided according to standard deviation into approximately 3 standard deviations
- ☒ Mean  $\pm$  1 standard deviation = 68.2% and Mean  $\pm$  2 standard deviations = 95.45% and mean  $\pm$  3 standard deviations = 99.73%

#### ☒ Example

- One thousand randomly selected men have a mean systolic blood pressure of 120 mmHg with a SD of 10 mmHg. The population is normally distributed with respect to systolic blood pressure.
- About 68% (680) of the men have systolic blood pressure of  $120 \pm 10$  (mean  $\pm$  1 SD) or systolic blood pressure ranging from 110-130 mmHg.
- 95% (950) of men have systolic blood pressure of  $120 \pm 2(10)$  (mean  $\pm$  2 SD) or systolic blood pressure ranging from 100 – 140 mmHg and 99.7% their systolic blood pressure is  $120 \pm 3(10)$  (mean  $\pm$  3 SD) or men systolic blood pressure ranging from 90 -150mmHg.

#### II. Skewed (asymmetric) distribution curves:

- ☒ Skewed to the right: mean>median>mode
- ☒ Skewed to the left: mean<median<mode

## Normal versus skewed data(non normal)

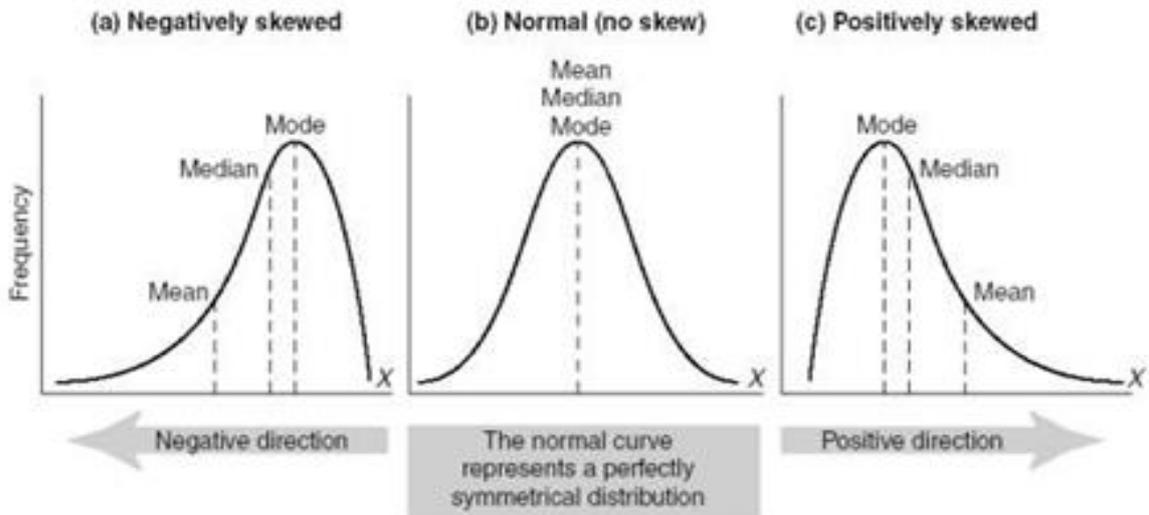


FIGURE 15.8 Examples of normal and skewed distributions

## Role of screening tests in disease diagnosis

### + Definition:

- ☒ Screening is the investigation of apparently healthy individuals to detect unrecognized cases or individuals with high risk of developing a disease.
- ☒ Therefore, intervention can be done to prevent occurrence of the disease or to improve its prognosis when it develops.

### + Objective of a screening test:

- ☒ **Immediate objective:** Simple test applied on large number to exclude those free from the disease and pick up those possibly suffering of the disease and subjected to detailed investigation to prove or disprove the diagnosis
- ☒ **Ultimate objective:** to reduce mortality and morbidity

+ **Screening test:** a simple test applied on large number to exclude those free from disease & to pick up those possibly suffering from disease & subjected to detailed investigation to prove or disprove the diagnosis (i.e. reference test).

### + Difference between screening test & diagnostic test:

Screening test	Diagnostic test
Done on apparently healthy.	Done on those with disease indication.
Used on groups.	Used on an individual basis.
Less accurate.	More accurate.
Less expensive.	More expensive.
Not a basis for treatment.	Used as a basis for treatment.



#### + Nature of screening tests: Screening tests may be:

- ☒ A clinical step (e.g., breast palpation),
- ☒ A laboratory (e.g., glucose tolerance test for diabetes mellitus)
- ☒ Other investigation (e.g., mammography).

#### + Types of Screening:

- ☒ **Mass Screening** offered to all individuals, irrespective of the presence of particular risk to the disease in question. This is not a useful preventive measure unless it is backed-up by treatment & follow-up facilities for positive screening.
- ☒ **High Risk Screening** offered to those with special risk, e.g., screening of close relative of known diabetics (a greater number of cases can be identified at less cost).
- ☒ **Multiphase screening** for a variety of diseases at one time. This is a well-established procedure in antenatal care & school examinations.

#### + Requirements for a screening program:

- ☒ Suitable disease
- ☒ Suitable test
- ☒ The population to be screened

#### + Requirements Of Screening Program regarding (The disease):

##### 1. Importance of the disease:

The disease should be an important health problem, i.e., high frequency and/or bad sequelae, e.g., congenital hypothyroidism, although rare, should be detected early because of its serious sequelae if untreated and because it is treatable.

##### 2. Adequate understanding of the natural history of the disease:

to identify the points at which the disease can be detected by screening with effective intervention before irreversible damage, to evaluate the effectiveness of any intervention





3. A recognized latent period or asymptomatic stage.
4. Can be detected before onset of symptoms and signs
5. At risk individuals can be identified and screened
6. Available facilities for diagnosis and treatment.
7. Agreed policy on whom to treat as patients
8. An effective treatment, available , effective and acceptable
9. Benefits of early detection exceeds risks and costs (money, manpower and equipment).

#### + Requirements of a screening test:

- Valid
- Reliable
- Cheap, easily and quickly done.
- Safe, not painful.
- Objective rather than subjective.
- Acceptable by the population

#### + Validity and reliability of screening test:

- It is the capacity of a test to give **true results**. Therefore, a valid test is the test which **correctly** detects the presence or absence of a condition.
- e.g., glucosuria as a test to detect diabetes mellitus has poor validity compared to glucose tolerance test.

#### Validity includes:

- **Sensitivity:** The ability of the test to identify correctly those who have the disease, i.e., it gives few false negative results.
- **Specificity:** The ability of the test to identify correctly those who do not have the disease, i.e., it gives few false positive results.
- **Positive predictive value:** abbreviated PPV or PV+, is the proportion of all people with positive tests who truly have the condition –  $a / (a+b)$  in the above table.



- **Negative predictive value:** (NPV or NP-) is the proportion of all people with negative tests who truly do not have the condition –  $d / (c+d)$  in the above table.

#### ✚ Results of screening test & the true diagnosis:

Screening test	Disease		Total
	Present	Absent	
Positive	(TP) A	(FP)B	A+b
Negative	C(FN)	D(TN)	C+d
Total	A+C	B+D	A+B+C+D

- ☒ **Sensitivity:** = probability of a positive test in people with the disease =  $a / (a+c)$
- ☒ **Specificity:** = probability of a negative test in people without the disease =  $d / (b+d)$
- ☒ **Positive predictive value:** probability of the person having the disease when the test is positive =  $a / (a+b)$
- ☒ **Negative predictive value:** probability of the person not having the disease when the test is negative =  $d / (c+d)$ .
- ☒ **Accuracy:** =  $(a+d) / (a+b+c+d)$
- ☒ **Reliability (Repeatability):**
  - It is the level of agreement between repeated measurements; therefore, a technique will give the same values on repeated application on the same individual.



## Morbidity and mortality Statistics

### + Definition of Morbidity Statistics:

- Statistics that enumerate the extent, frequency, or severity of disease in a community.

### + Types of Morbidity Statistics:

#### ☒ Incidence rate

- **Definition:**

The rate of occurrence of new cases in a specified population.

- **Calculation:**

$$\frac{\text{Number of reported new cases of a disease in certain } Y / L}{\text{At risk population in the same } Y/L} \times 10^n$$

#### ☒ prevalence rate

- **Definition:**

Frequency of existing cases (old and new) in a defined population.

- **Calculation:**

$$\frac{\text{Number of people with a disease or condition (old + new cases)}}{\text{Total number of examined population at same locality and time}} \times 10^n$$

**Constant is  $10^n$ , where  $n = 1$  or  $2$ , or  $3$  etc.**





## Types of Mortality Statistics:

### A. Crude (death) mortality rate

$$\frac{\text{Total number of deaths a certain Y \& L}}{\text{Estimated mid year population at same Y \& L}} \times 1000$$

### B. Sex Specific Death Rate

$$\frac{\text{Total number of deaths of a certain sex in a certain Y \& L}}{\text{Total number of same sex at same Y \& L}} \times 1000$$

### C. Cause-specific mortality rate

$$\frac{\text{Total number of deaths of a specific cause in a certain Y \& L}}{\text{Estimated mid - year population at same Y \& L}} \times 100000$$

### D. Proportionate mortality rates

$$\frac{\text{Total number of deaths of a specific cause in a certain Y \& L}}{\text{Total deaths from all causes at same Y \& L}} \times 100$$

### E. Case fatality rate

$$\frac{\text{Total number of deaths of a specific cause in a certain Y \& L}}{\text{Total number of cases at same Y \& L}} \times 100$$

### F. Maternal Mortality

- **Definition:**

Maternal mortality means death among mothers due to causes related to and/or aggravated by pregnancy, labor & puerperium.

- **Rate:**

$$\frac{\text{Total number of maternal deaths in a certain Y \& L}}{\text{No. of female in childbearing period (15 - 49 y) at same Y \& L}} \times 100000$$



- **Causes**

1. Hemorrhage: May occur during pregnancy, labor or puerperium. It forms the most important causes in Egypt.
2. Hypertensive disease of pregnancy (eclampsia & preeclampsia).
3. Puerperal sepsis: most preventable cause.
4. Pre-existing diseases aggravated by pregnancy, labor & puerperium e.g.
  - Rheumatic heart disease.
  - Chronic glomerulonephritis complicated by renal failure.
  - Uncontrolled D.M

- **Age specific mortality rates:**

- A. Still birth**

- **Definition:**

- Stillbirth is the delivery, after the 28th week of pregnancy, of a baby who has died.
      - Loss of a baby before the 28th week of pregnancy is called a miscarriage.

- **Rate:**

$$\frac{\text{Total number of still births in a certain Y \& L}}{\text{No. of total births at same Y \& L}} \times 1000$$

- B. Infant Mortality Rate**

$$\frac{\text{Number of deaths less than 1 year of age in a certain Y \& L}}{\text{No. of live births at same Y \& L}} \times 1000$$

*Infant mortality rate is divided into neonatal and post neonatal mortality rates.*

### C. Neonatal mortality rate

$$\frac{\text{Number of deaths less than 28 days of age in a certain Y \& L}}{\text{No. of live births at same Y \& L}} \times 1000$$

#### o Causes:

- Prematurity (preterm & LBW)
- Congenital malformations & Rh incompatibility
- Birth injuries
- Asphyxia neonatorum.
- Infections as congenital infection, tetanus neonatorum, acute respiratory disease & diarrhea.

### D. Post-neonatal mortality rate

$$\frac{\text{Number of deaths from 28 days to 1 year in a certain Y \& L}}{\text{No. of live births at same Y \& L}} \times 1000$$

#### o Causes:

- Infections in the form of:
  - \* Acute respiratory diseases
  - \* Infective diarrheal diseases.
- Other infections e.g. tetanus neonatorum, pertussis & measles.
- Sever nutritional deficiency e.g. PEM.
- Accidents.
- The remaining section of prematurity & congenital malformations



### E. Child (1-4years) mortality rate

$$\frac{\text{Number of child (1 – 4 years of age) in a certain Y \& L}}{\text{No. of total children in same age group at same Y \& L}} \times 1000$$

#### ○ Causes:

- Infectious diseases as acute respiratory diseases, diarrhea, pertussis, measles & meningitis.
- Accidents.
- Severe nutritional deficiency.
- Other causes as rheumatic heart disease

#### Best wishes





## Descriptive + Case control studies

### Training 1:

#### **Novel Coronavirus (SARS-CoV-2) in Febrile Infants in New York**

**Introduction:** The 2019 novel coronavirus (SARS-CoV-2) has spread rapidly across the globe since it was identified in January 2020. The Chinese Center for Disease Control and Prevention reported that only 1% of cases (416 out of 72 314) whether confirmed or suspected cases were found in children <10 years of age.

**Case presentation:** Three febrile infants <2 months of age admitted to a large tertiary care children's hospital in New York. All 3 patients presented with fever, feeding difficulty, lymphopenia, thrombocytosis on laboratory evaluation and 2 out of the 3 sick infants were found to have neutropenia. All 3 patients subsequently found to be infected with severe acute respiratory syndrome coronavirus.

### Questions & Answers:

#### **1- Which of the following represents this type of the study design?**

- A) Case report.
- B) Case series.
- C) Cohort study.
- D) Case control study.
- E) Cross sectional study.





## 2- Which of the following is among the application of this study design?

- A) Study multiple risk factors.
- B) Estimate the prevalence rate.
- C) Describe previously un-described disease.
- D) Describe association between exposure and outcome.
- E) Describe a number of similar cases with a given disease in one report.

### **Training 2:**

#### **Diet and Alzheimer's Disease**

A multi-country study links between diet and Alzheimer's disease have been conducted using the national survey prevalence of Alzheimer's disease in 11 countries along with dietary supply factors. It found that total fat and total caloric supply were strongly correlated with prevalence, while fish and cereals/grains were inversely correlated. Diet is now considered an important risk-modifying factor for Alzheimer's disease.

#### **Questions:**

- 1- What is the type of this epidemiologic study?
- 2- What are the applications of this study design?
- 3- Mention the other types of studies which belong to that category?





### **Training 3:**

#### **Prevalence of Obesity in a Saudi Obstetric Population**

A study done by El-Gilany & El-Wehedy involved 791 women registered for antenatal care in primary health care centers in Saudi Arabia. The objectives were to measure body mass index (BMI) & assess the prevalence of obesity. Weight & height were abstracted from maternal records. BMI was calculated as weight (kg)/height<sup>2</sup> (m). They found that 67, 311, 187 & 226 women were underweight, normal weight, overweight and obese, respectively.

#### **Questions:**

- 1- Mention the study design?
- 2- Calculate the prevalence of obesity among the study population?
- 3- What are the disadvantages of this study design?

### **Training 4:**

#### **The natural history of primary dysmenorrhoea:**

**Background:** Dysmenorrhoea is a common condition among women.

**Objective:** To describe the course of primary dysmenorrhoea in women of all reproductive ages.

**Population:** We analyzed data from 404 women who responded to both surveys, but denied endometriosis, pelvic inflammatory disease or uterine fibroids.





**Methods:** Participants were surveyed twice at an interval of six years regarding menstrual cycle characteristics.

**Questions:**

- 1- What is the type of this epidemiologic study?
- 2- What is the outcome of this study design?
- 3- Mention the other types of studies which belong to that category.

**Training 5:**

**A study to identify risk factors for acute hepatitis C virus infection in Egypt:**

**Background:** Identification of risk factors of acute hepatitis C virus (HCV) infection in Egypt is crucial to develop appropriate prevention strategies.

**Methods:** A study was carried out from June 2007-September 2008 to investigate risk factors for acute HCV infection in Egypt among 86 patients and 287 age and gender matched controls identified in two infectious disease hospitals in Cairo and Alexandria. Case-patients were defined as: any patient with symptoms of acute hepatitis; lab tested positive for HCV antibodies and negative for HBsAg, HBc IgM, HAV IgM; and 7-fold increase in the upper limit of transaminase levels. Controls were selected from patients' visitors with negative viral hepatitis markers. Subjects were interviewed about previous exposures within six months, including community-acquired and healthcare associated practices. Case-patients were more likely than controls to have received injection with a reused syringe (OR= 23.1).





### Questions:

- 1- What is the type of this epidemiologic study?
- 2- What is the outcome of this study design?
- 3- Which of the following characterize this study design?
  - A) It is a retrospective study.
  - B) Relative risk can be measured.
  - C) Incidence rates can be calculated.
  - D) It proceeds from cause to outcome.
  - E) It is an observational descriptive study.
- 4- The underlined number indicates that:
  - A) Syringe injection is not a risk factor for HCV at all.
  - B) Syringe injection is a protective factor for acute HCV.
  - C) No association between acute HCV and syringe injection.
  - D) Increased exposure to syringe injection leads to increased acute HCV infection.
  - E) Odd of exposure to syringe infection among controls/Odd of exposure to syringe infection among cases= 23.1
- 5- Advantage of this study design includes:
  - A) Study rare diseases.
  - B) Useful in rare exposure.
  - C) Incidence rate can be calculated.
  - D) Less bias in selection of the control.
  - E) It is an expensive and time-consuming study.





### Training 6:

Study the relation between smoking and lung cancer:

Smoking	Lung cancer	Controls
Smokers	16	8
Non smokers	2	24
Total	18	32

### Questions:

- 1- Calculate proportion of exposure among lung cancer cases and controls.
- 2- Calculate odd of exposure among lung cancer cases and controls.
- 3- Calculate odds ratio, interpret the result.

### Training 7:

#### Questions & Answers:

- 1- To determine an odds ratio, which of the following studies a researcher can follow?
  - A) Cohort study.
  - B) Case-control study.
  - C) Longitudinal study.
  - D) Cross-sectional study.
  - E) Randomized clinical trial.





**3- Researchers want to know whether low birth weight (LBW) is associated with the maternal smoking during pregnancy. The researchers obtained smoking history from mothers with LBW infants and comparing it with smoking histories of an equal number of mothers with normal LBW. What type of this study?**

- A) Cohort study.
- B) Prospective study.
- C) Case-control study.
- D) Cross-sectional study.
- E) Randomized Clinical trial





## Analytic studies (cohort study)

### Training 1.a

Match each study with the correct definition:

#### A. Cross-Sectional

#### B. Cohort

#### C. Case-Control

1. Diabetics are compared with non-diabetics to assess its relationship with junk food.

2. A study of women aged 50–60 years in a community located close to a nuclear power facility.

3. Subjects who exercise twice a week are compared with subjects who do not to investigate the effect of exercise on obesity.

### Training 1.b

• Which of the following characterize cohort study?

- It is an observational analytical study
- It proceeds from outcome to cause
- Prevalence rates can be calculated
- Odds ratio can be measured





### Training 1.c

- Which study design is best for rare exposure?
  - a) Cohort study
  - b) Case-control study
  - c) Randomised controlled trial
  - d) Case-crossover study

### Training 2

**You want to estimate the incidence of depression following accidental trauma. Which the following study design would be most suitable for your research question?**

- a) Cohort study
- b) Case-control study
- c) Randomised controlled trial
- d) Case-crossover study





### Training 3

A research team follows 2000 HBsAg positive patients and a similar group of HBsAg negative at outpatient clinics over 15 years to determine the rate of occurrence of hepatocellular carcinoma.

This study is example of:

- A. Case control study.
- B. Prospective Cohort study.
- C. Cross section study.
- D. Experimental study.

### Training 4

- Over a period of 20 years, lung cancer occurred in 20 people among 2000 smokers and 5 persons out of 5000 non-smokers.
- 1. construct a 2x2 table for the above data
- 2. calculate the relative risk of developing carcinoma lung in smokers

### Training 5

A study of Heart disease and smoking among British physicians reported these results.

**Table: Annual death rates per 100000 persons**





	Heart Diseases	No Heart Diseases	Total
Heavy smokers	599	99401	100000
Non-smokers	422	99578	100000

Calculate and interpret relative risk of dying from heart disease





## Training 6

**A study examined the relationship between body mass index (BMI) and all-cause mortality in US cohort who were 50 to 71 years old for a year.**

Relative risk was calculated to measure the association between abnormal BMI and the risk of all-cause mortality (the abnormal BMI groups versus the normal BMI group)

Table: Relative risk of death in men aged 50 to 71 years across BMI categories.

BMI at age 50	Relative Risk
<18.5	1.29
18.5 to 24.9	Comparison group
25.0 to 29.9	1.31
30.0 to 34.9	1.96
35.0 to 39.9	2.46
≥ 40	3.82

**Interpret relative risk in the table across BMI categories**





## Role of screening in diagnosis of disease

### Training (1)

On evaluation of tuberculin skin test as screening test for tuberculosis, a cut off value of 10 mm was considered to be positive. The following results were found:

Tuberculin skin test	Tuberculosis disease	
	Present	Absent
$\geq 10$ mm	40	20
$< 10$ mm	10	130

Calculate:

1. Sensitivity
2. Specificity
3. Positive predictive value
4. Negative predictive value
5. Accuracy
6. Prevalence of tuberculosis in the study population

### Training (2)

You apply a new test to serum taken from a 320 newborns to detect infection with a virus. Test results were positive in 100 of these newborns but only 70 of them were found to harbour the virus when viral cultures were performed (viral isolation is considered to be the Gold Standard). When samples were cultured from test negative newborns, 20 returned a positive culture. Assume the 320 newborns were representative of the population that exists in your practice region. You want use this test to screen all newborns in your practice.

Answer the following questions:

1. Construct a 2 x 2 table of the results of the study.
2. Assess validity of the studied screening test.
3. Calculate the positive predictive and negative predictive values of the test.





### Training (3)

Breast mammography which has a PPV of 93% and a NPV of 99.9% is used to screen 4,000 women for breast cancer. Five percent of the women (200 women) have positive test results and 95% (3800 women) have negative test results.

1. Construct the 2x2 table for the results
2. Assess validity of the studied screening test.

### Training (4):

The findings of a test using a screening level of 130mg% of blood sugar (random) are shown in the following table:

Screening test	Disease status	
	Diabetic	Nondiabetic
Positive	86	98
Negative	84	9752

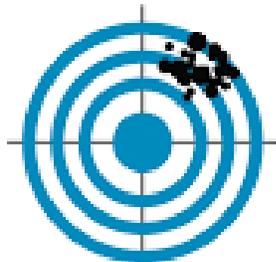
1. The ability of the screening test to correctly identify those who have diabetes is:
  - A. 40%
  - B. 51%
  - C. 60%
  - D. 75%
  - E. 99%
2. The ability of the screening test to correctly identify those who have the disease from all those who test positives:
  - A. 39%
  - B. 52%
  - C. 66%
  - D. 88%
  - E. 90%





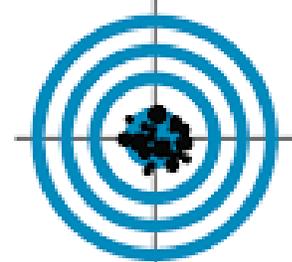
**Training (5):**

**On assessment of three new screening tests, the following diagram represent the results of 100 trials for each test. Describe and comment on each diagram.**



(A)

(B)



(C)



### Training (6): Multiple Choice questions

**1- The proportion of people without disease who are correctly classified by a screening test as negative is called:**

- a) Positive predictive value
- b) Sensitivity
- c) Prevalence
- d) Specificity
- e) Incidence

**2- Which of the following characterize screening test:**

- a) More accurate
- b) A basis for treatment
- c) Used on community basis
- d) More expensive
- e) Confirm the diagnosis

**3- The property of a test to identify the proportion of free persons (not diseased) in a population who are identified as negative by a screening test.**

- a) Sensitivity
- b) Specificity
- c) Positive predictive value
- d) Negative predictive value

**4- In a population of 1000 persons tested for hepatitis C, 5 positives are identified using a perfect test. What is the sensitivity of the test being used?**

- a) 0.5%
- b) 99.5%
- c) 100%
- d) Cannot be determined

**5- In screening test, of reliability, repeatability, validity and accuracy which pair is analogous?**

- a) Validity is analogous to repeatability.
- b) Reliability is analogous to accuracy.
- c) Validity is analogous to reliability.
- d) Validity is analogous to accuracy.

**6- In a communicable disease with high mortality, tests must be:**

- a) Highly sensitive
- b) Highly specific
- c) Easy to perform
- d) Economical

**7. A screening test is negative in the majority of free individuals but false negatives is much higher than true negatives, this indicates:**

- a) Low PPV
- b) Low NPV
- c) Low accuracy
- d) High validity
- e) High specificity

## medical statistics1

### Training 1:

State the type of variable & **summarizing value** for each of the following:

Biological variables	Type of variable
Blood pressure in mmHg	
Cholesterol in Mmol/L	
diabetes (YES/NO)	
BMI (Kg/m <sup>2</sup> )	

Blood pressure (mmHg) is a ..... variable, but when expressed as hypotensive, normotensive, and hypertensive is considered as ..... variable.

The number of patients with HCV is .....

## Training (2): MCO

### 1. Which of the following techniques yields a simple random sample?

- Choosing volunteers from 4<sup>th</sup> year medical students to participate
- Listing the individuals by blood group and choosing a proportion from within each blood group at random.
- Numbering all the elements of a sampling frame and then using a random number table to pick cases from the table.
- Randomly selecting schools, and then sampling everyone within the school.
- Not all the individuals have equal chance to be in the sample

## Training (3):

### Match each description of a sampling procedure with the correct term.

- Systematic sampling
- Simple random sampling
- Stratified sampling
- Cluster sampling

1. Each individual of the total group has an equal chance of being selected.

2. Households are selected at random, and every person in each household is included in the sample.

3. First individual is selected randomly, and every fourth subject is selected.

4. Individuals are divided into subgroups on the basis of specified characteristics and then random samples are selected from each subgroup

**Training (4):**

**Calculate Median- Mean for the following set of data:**

Serial	Weight (Kg)	Height (cm)
1	70	160
2	75	168
3	67	179
4	80	158
5	79	160

**Weight**

Median=                      mean=

**Height**

median=                      mean=

**Training (5):**

1-What are the types of the three variables?

Sex, residence

BMI



### Training (6):

Match the measurement with the data type. Drag the data type in the correct box:

- |                                 |             |
|---------------------------------|-------------|
| 1- color of shirt               | a- Ordinal  |
| 2- Temperatures in F            | b- Interval |
| 3- weight of students           | c- Nominal  |
| 4- Grade of students ( A, B ,D) | d- Ratio    |



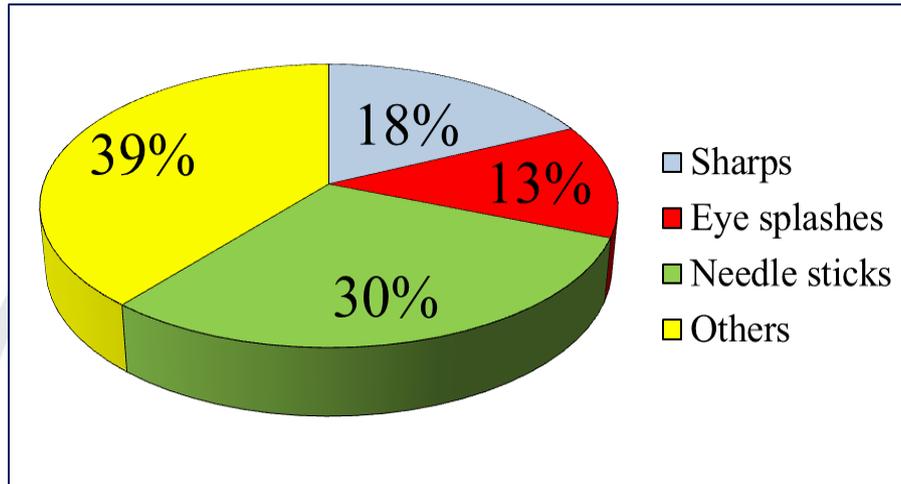
2:



## medical statistics 2

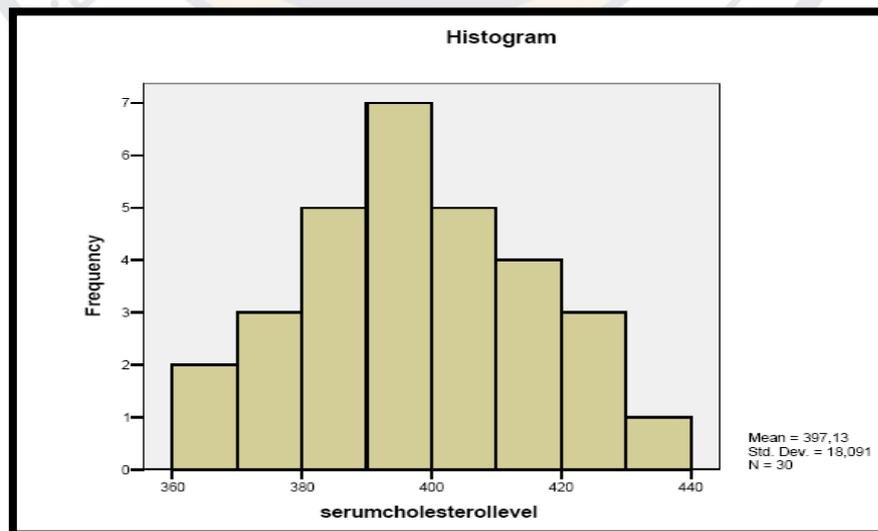
### Training (1):

Look at the following figure then answer:



1. What is the type of this graph?
2. What is the type of variable presented by this graph?
3. Could other types of graphs present the same type of data?

### Training (2)



- 1.
- 2.
2. Figure1. Serum cholesterol (mg/dL) of 30 diabetic patients at MUH, 2010



1. What is the type of this graph?
2. What is the type of variable presented by this graph?
3. Could other types of graphs present the same type of data?

### Training (3):

In a hospital with 3000 healthcare workers, 50% are physicians, 30% nurses, 10% technicians, and 10% service workers.

1. What is the type of this variable? **"Job title "**
2. How to present these data?
3. If we want to plot the job title together with the gender of HCWs in that hospital, what are the suitable graphs?

### Training (4)(Activity):

Complete the following table

Variable	Presentation
Age in years	
Gender	
Blood group	
Blood glucose in mg/dl	
Marital status	
Stage of cancer (I, II, III, IV)	
Number of students	

### Training (5):

The following table shows the age and sex differences between the intervention and control groups in a clinical trial. Read carefully then answer the questions:

2:



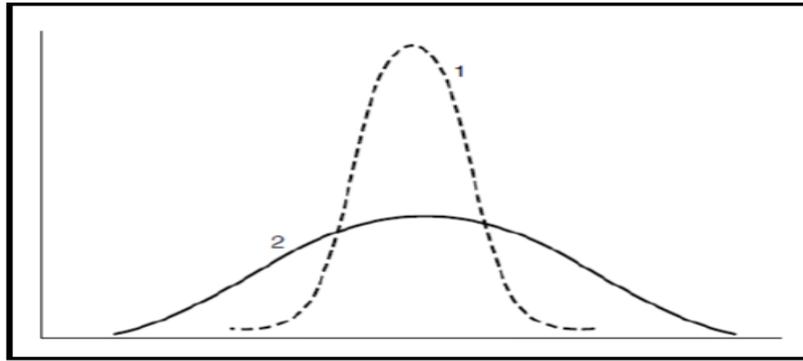
Variables	Intervention (n=19)		Control (n=25)		Test of significance	P-value
	No	%	No	%		
<b>Age/years</b>						
Mean ± SD	11.84 ± 3.43		12.14 ± 3.03		t = 0.30	0.76
Min-Max	3-16		6-17			
≤12	12	63.2	16	64.0	χ <sup>2</sup> =0.003	0.95
>12	7	36.8	9	36.0		
<b>Sex</b>						
Male	12	63.2	11	44.0	χ <sup>2</sup> =1.58	0.20
Female	7	36.8	14	56.0		

1. What are the types of variables in the above table?
2. What are the measures of central tendency and dispersion available in the above table?
4. What is the interpretation of p value?

### Training (6): MCQ Activity

- 1) measure of dispersion of a set of observations in which it is calculated by the difference between the highest and lowest values produced is called:
  - a) standard deviation
  - b) variance
  - c) range
  - d) mode
  - e) mean

3) Consider the following two distribution curves. Which numerical summary measure would allow you to discriminate between the two distributions?



a. Median    b. Mean    c. Mode    d. Standard deviation    e. Range

4) The normal distribution curve is determined by

- a) Mean and sample size
- b) Range and sample size
- c) Range and standard deviation
- d) Mean and standard deviation

(5) In a population of 100 females in the age group of 15-45, the mean systolic BP was found to be 120. In a normal curve distribution, the number of people who would have an average BP above 120 will be:

- A. 25
- B. 50
- C. 75
- D. 95
- E. 85



## Morbidity and Mortality statistics

### Training 1

In 1990, there was 80.000 individuals aged 30 years and more with history of coronary heart disease (CHD), while an estimated 50.000 new cases of CHD was reported in the same year and age group in that community. If the total population of those aged 30 and more was 20.000.000. **Calculate the incidence and prevalence rate of CHD in this community.**

### Training 2

A study was done to investigate an outbreak of meningococcal meningitis in a city with 80,000 populations. A total of 721 cases were recorded at the local hospital during a period of one week. The daily number of cases is: 50, 70, 100, 200, 150, 100, and 51. Fifty-eight deaths were recorded among the admitted patients.

Calculate case fatality rate.

### Training 3

In 2018, in a community the estimated midyear population was one million (450.000 males and 550000 females). The number of deaths was 30.000 death among them 14000 were males.

Calculate the following:

The crude death rate of the same year.

The female death rate for this country. 21





#### Training 4

In 2019, a total of 15,555 homicide deaths occurred among males and 4,753 homicide deaths occurred among females. The estimated 2001 mid-year populations for males and females were 139,813,000 and 144,984,000, respectively.

Calculate the homicide-related death rates for males and for females.

#### Training 5

In an epidemic of Hepatitis A traced to green onions from a restaurant, 555 cases were identified. Three of the case patients died as a result of their infections.

Calculate the case-fatality rate.

#### Training 6

1- If the number of deaths from tuberculosis is expressed in relation to the total mid-year population, it is:

- a) Case fatality rate
- b) Age specific death rate
- c) Proportionate mortality rate
- d) Crude death rate
- e) Cause specific death rate

2- Which of the following mortality rates use the estimated mid-year population as its denominator?

- a) Age-specific death rate
- b) Sex-specific death rate





- c) Case fatality rate
- d) Crude death rate
- e) Proportionate mortality rate

**3- In a country, estimated mid-year population is 1 million and total number of deaths is 15,000. Crude death rate is:**

- a) 15 per 1000 persons
- b) 15 per 100 persons
- c) 150 per 1000 persons
- d) 150 per 100 persons
- e) 15 per 10000 persons

**4- Denominator of maternal mortality rate:**

- a) Estimated mid-year population
- b) Live birth
- c) Female in childbearing period (15-49)
- d) Married females in childbearing period (15-49)
- e) Females in a specific age group

**5- If the number of deaths from road traffic accidents is expressed in relation to the total number of deaths from all causes, it is.....**

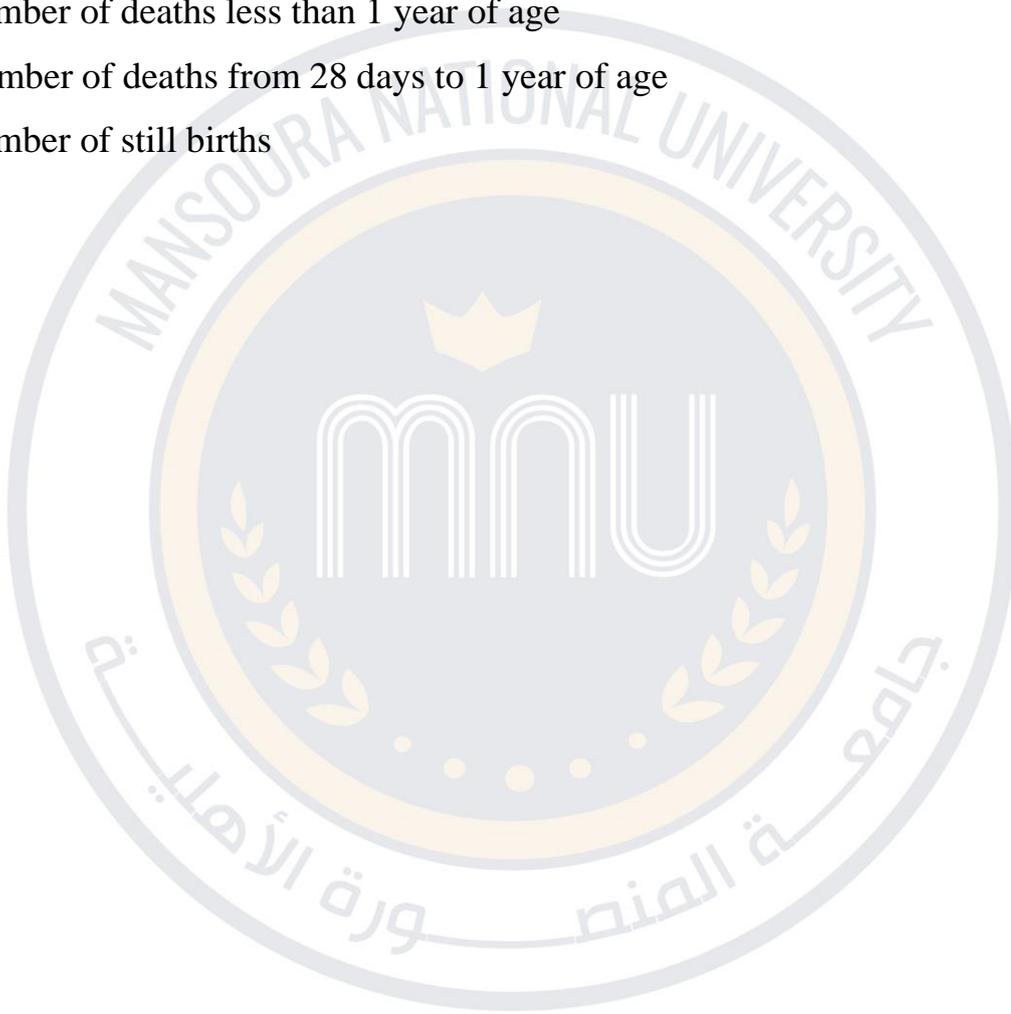
- a) Case fatality rate
- b) Age specific death rate
- c) Proportionate mortality rate
- d) Crude death rate
- e) Cause specific death rate





**6- Which of the following is the numerator of neonatal mortality rate?**

- a) Number of deaths less than 28 days of age
- b) Number of total births
- c) Number of deaths less than 1 year of age
- d) Number of deaths from 28 days to 1 year of age
- e) Number of still births



## Training 7

The underlying table show number of reported cases and death from GIT cancer in a given country where estimated number of population is 30,000,000 and total number of deaths was 400,000 in 2021.

Site of cancer	No. of reported cases	No. of death
Esophagus	5,000	3,500
Stomach	8,000	6,300
Duodenum	2,000	870
Jejunum and ileum	1,500	340
Colon	14,000	3,000
Rectum	9,000	5,400
<b>Total</b>	<b>39,500</b>	<b>19,410</b>

Calculate the following:

1. Cause specific mortality rate for GIT cancer.
2. Proportionate mortality rate for GIT cancer.
3. Case fatality rate for stomach cancer.

30



كلية الطب - جامعة المنصورة الأهلية  
Faculty Of Medicine - MANSOURA NATIONAL UNIVERSITY



3:

M N U

الطريق الدولي الساحلي - منطقة 15 مايو - مدينة جمصة - محافظة الدقهلية  
International Coastal Road - 15<sup>th</sup> of May District - Gamsa City - Dakahlia Governorate  
✉ [medic@mansnu.edu.eg](mailto:medic@mansnu.edu.eg)



الصفحة الرسمية للجامعة



الموقع الرسمي للجامعة