

Basics of Biostatistics

**Ephrem Mannekulih (BSc, MSc)
Biostatistics and Health Informatics**

Course Descriptions

- ❖ **Course title:** Basic Biostatistics
- ❖ **Course code:** PubH601
- ❖ **Credit hour:** 3
- ❖ **Course Policies**
 - Attendance: at least 90% of the course to sit for the exam
 - Absence due to illness: must accompanied by a medical certificate for less than 20% of the class
 - Students must do every assignment given and ²

Course Objectives

- ❖ Discuss the roles of statistics and the main uses of statistical methods in health science
- ❖ Describe statistical methods of data collection, organization, analysis, summarization, presentation and interpretation
- ❖ Explain the context and meaning of statistical inferences (estimation and hypothesis testing)
- ❖ Understand different techniques of sample size determination and sampling techniques

Student's Evaluation

- ❖ Assignments
- ❖ Class Activity
- ❖ Presentation
- ❖ Examination

Introduction

What is statistics?

- ❖ It is a field of study concerned with designing, collection, organization, analysis, summarization, presentation and interpretation of data and;
- ❖ The drawing of inferences /conclusion/ about a population based on observed data taken from a sample
 - Statistics helps us use numbers to communicate ideas

Biostatistics:

- ❖ **Biostatistics:** The application of statistical methods to the fields of biological, public health and medical sciences.
- ❖ Concerned with interpretation of biological/public health/ medical sciences data & the communication of information derived from these data
- ❖ Has central role in medical investigations

Origin and development of statistics in Health Research

- ❖ In 1929 a huge paper on application of statistics was published in Physiology Journal by Dunn.
- ❖ In 1937, 15 articles on statistical methods by Austin Bradford Hill, were published in book form.
- ❖ In 1948, a RCT of Streptomycin for pulmonary TB., was published in which Bradford Hill has a key influence.

Role and Uses of biostatistics

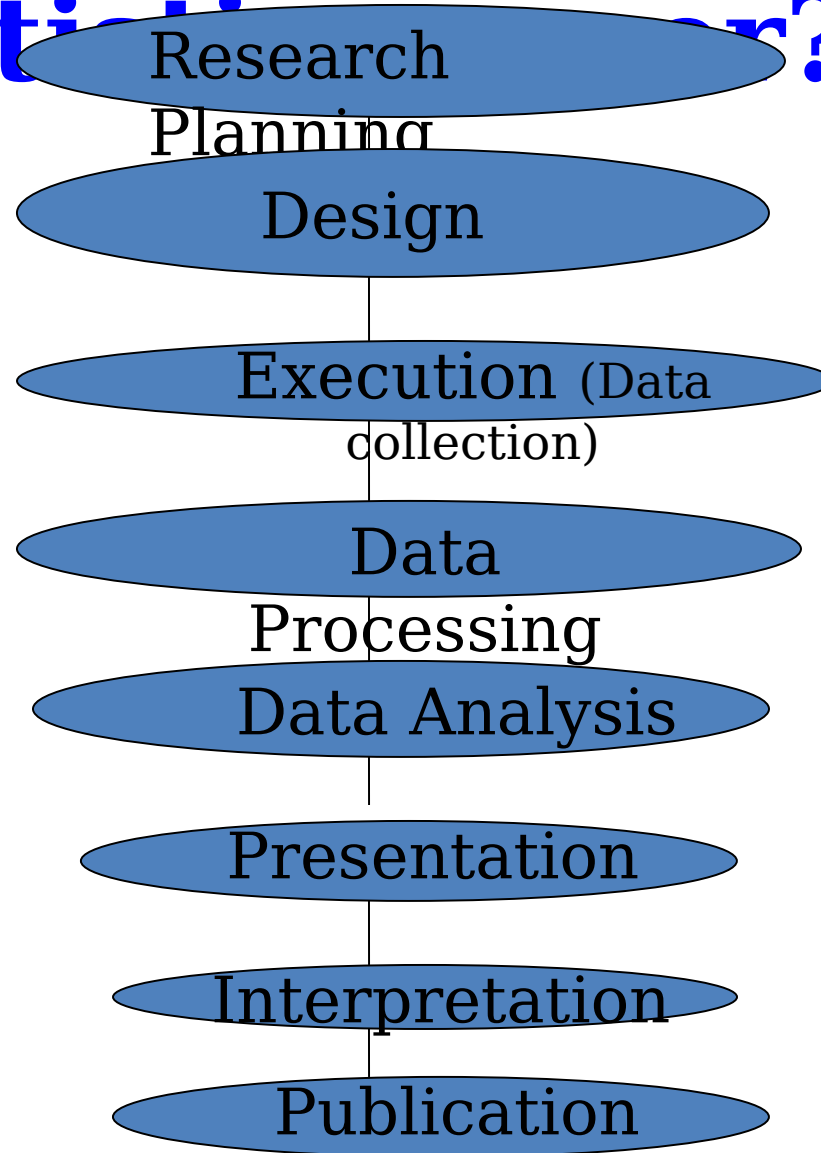
- ❖ Provide methods of organizing information
- ❖ Assessment of health status
- ❖ Health program evaluation
- ❖ Resource allocation
- ❖ Magnitude of association
 - Strong vs weak association between exposure and outcome
- ❖ Assessing risk factors
 - Cause & effect relationship

Cont'd...

- ❖ Evaluation of a new vaccine or drug
 - What can be concluded if the proportion of people free from the disease is greater among the vaccinated than the unvaccinated?
 - How effective is the vaccine (drug)?
 - Is the effect due to chance or some bias?
- ❖ Drawing of inferences
 - Information from sample to population

What does biostatistics do?

Biostatistical
thinking
contribute in
every step in
a research



The best way
to learn about
biostatistics is
to follow the
flow of a
research from
inception to
the final
publication

Research Design

- ❖ We can not study all subjects (all pregnant women, or all people) living in a given geographical area
 - Sampling technique
 - Inclusion/exclusion criteria
 - Sample size calculation
 - Study design
 - Method of data collection
 - Etc

Analysis

- ❖ Analysis part is the major part of learning about biostatistics
 - There are dozens of different methods of analysis, which makes difficult the choice of the correct method for a particular case
 - It is necessary to consider the philosophy that underlies all methods of analysis:
 - Use data from a sample to draw inference about a wider population

Interpretation

- ❖ Interpretation of results of statistical analysis is not always straightforward, but is simpler when the study has a clearer aim
- ❖ If the study has been well designed and correctly analyzed the interpretation of results can be fairly simple

Types of Statistics

1. Descriptive statistics:

- ❖ Ways of designing, collection, organization, analysis, summarization, presentation and interpretation of data
- ❖ It help to identify the general features and trends in a set of data and extracting useful information
- ❖ Also very important in conveying the final results of a study
- ❖ Example: tables, graphs, numerical summary measures

Cont'd...

2. Inferential statistics:

- ❖ Methods used for drawing of inferences /conclusion/ about a population based on observed data taken from a sample
- ❖ Example: Principles of probability, estimation, confidence interval, comparison of two or more means or proportions, hypothesis testing, etc.

Population and Sample

❖ Population:

- The largest collection of entities for which we have an interest to study at a particular time
- It could be people, animals, machines, places, or cells

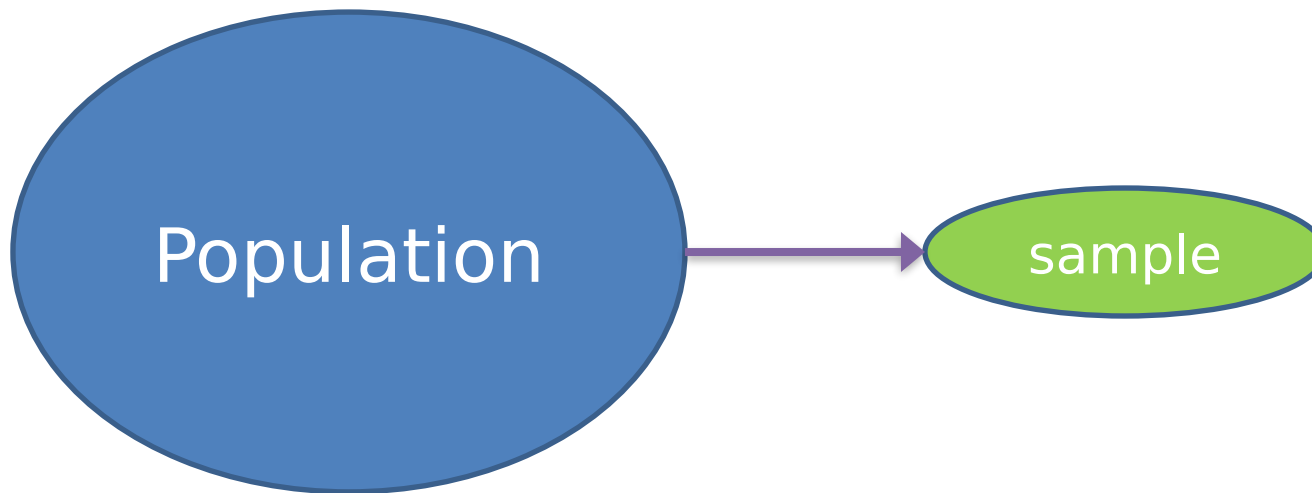
❖ Sample

- It is a group of subjects selected from a population.
- They are assumed as representative of population they are selected from

Population and Sample...

Cont'd

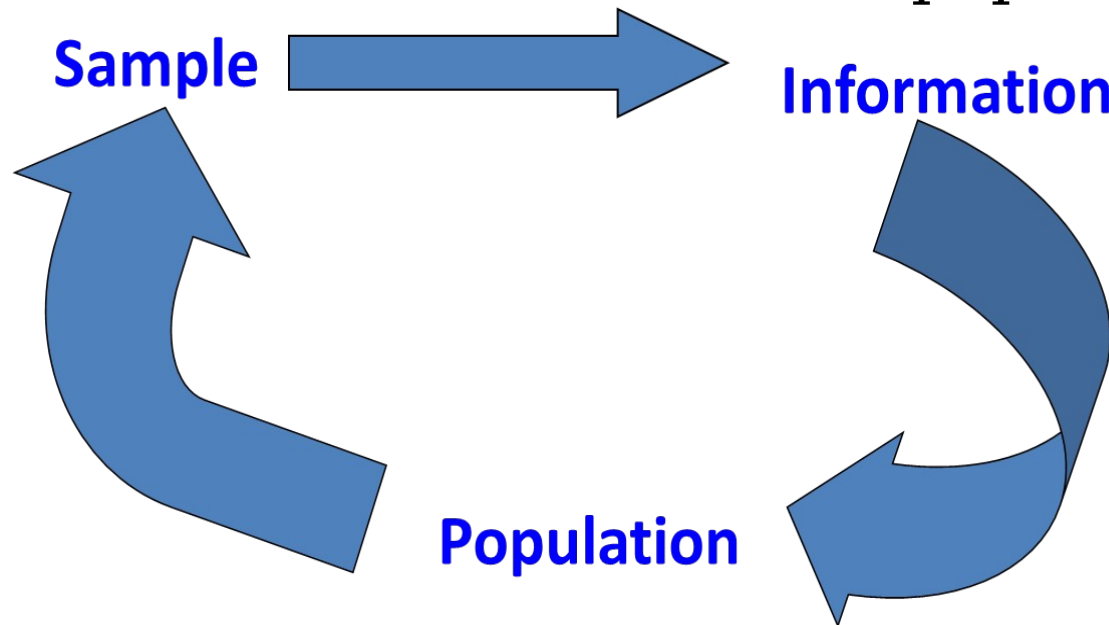
- ❖ The population is too large to study, instead we select a sample of individuals hoping that they are **representative** of the whole population.



Population and Sample...

Cont'd

- ❖ Researchers are not interested in the sample itself
- ❖ The interest is to learn from the sample to generalize/conclude about entire population.



Population and Sample...Cont'd

- ❖ **Inferences/conclusions** about the population are based on the information taken from the sample
- ❖ The **accuracy of the conclusions** depend on how well the sample is **representative**.
- ❖ How to ensure **representativeness**;
 - o **Sample size**
 - o **Sampling techniques**

Population and Sample...

Cont'd

❖ Points to be considered while taking a sample,

- A. Defining about the population from which sample is drawn? (**Target, Source and sample population.....**)
- B. Determine **Sample size**
- C. Select appropriate **sampling techniques**

Population and Sample...

Cont'd

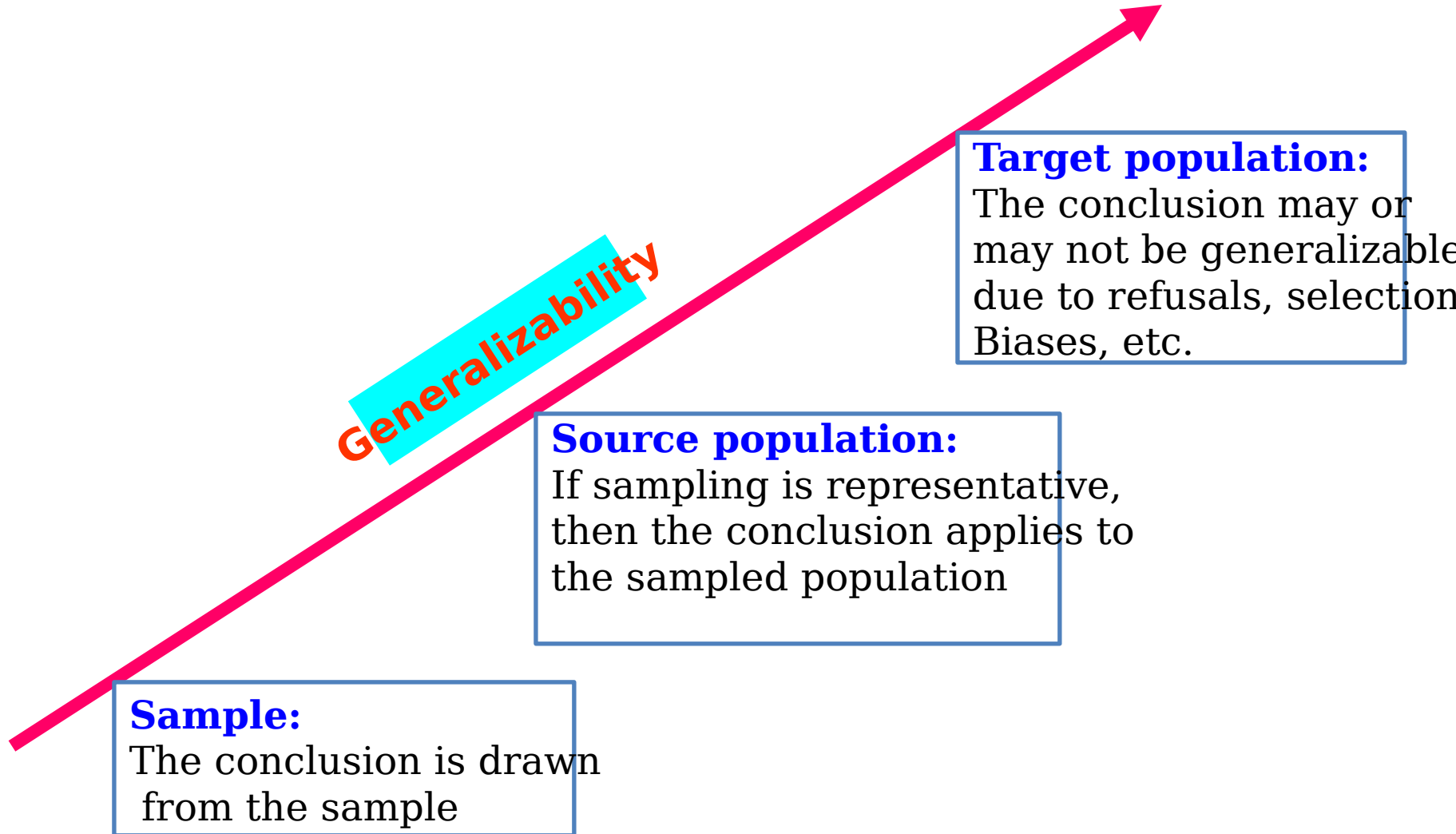
- ❖ **Reference population (or target population):** the population of interest to whom the researchers would like to make generalizations.
- ❖ **Source population:** the subset of the target population from which a sample will be drawn.
- ❖ **Study population:** the actual group in which the study is conducted sometimes similar with sample
- ❖ **Sampling unit:** the units from which we select study subject/ (house hold)
- ❖ **Study unit:** the units on which information will be collected: (persons)

Parameter and Statistic

- ❖ **Parameter:** A descriptive measure computed from the data of a population.
 - E.g., the mean (μ) age of the target population
- ❖ **Statistic:** A descriptive measure computed from the data of a sample.
 - E.g., sample mean age ()

Generalizability

- ❖ The role of statistics is using information from a sample to make inferences/generalization about the population
- ❖ Procedurally we need to be able to generalize from:
 - The study to the source population, &
 - Then from the source to the target population
- ❖ If the sample is not representative of the population, the conclusions are restricted to the sample & don't have general applicability



What is Data?

❖ **Data:** is collection of facts and evidences from which we can extract information and draw conclusions.

❖ **Types of data**

- **Primary data:** data collected directly from individuals or subjects or respondents for the purpose of certain study.
- **Secondary data:** data which had been originally collected by certain people or agency, and then statistically treated and the information contained in it is used for other purpose

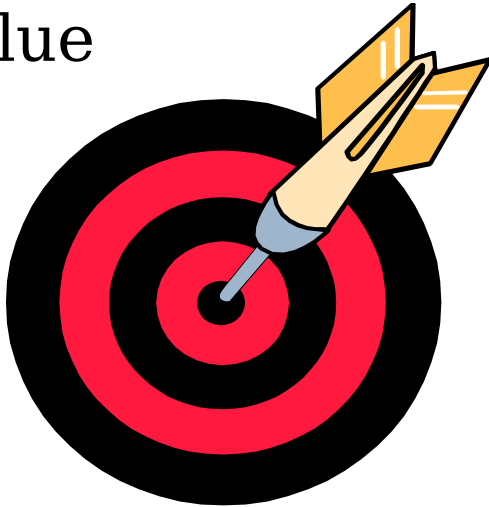
Sources of data

- ❖ Routinely kept records
- ❖ literatures
- ❖ Surveys
- ❖ Experiments
- ❖ Reports
- ❖ Observation, etc.

Accuracy vs. Precision

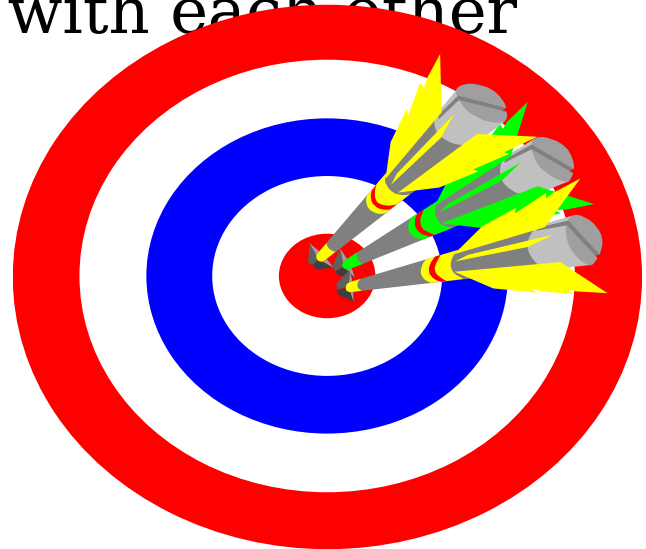
Validity

- How well a measurement agrees with an accepted value



Reliability

- How well a series of measurements agree with each other



Stages of data collection

❖ **Three Stages in the Data Collection Process**

- Stage 1: Permission to proceed
- Stage 2: Data collection
- Stage 3: Data handling

Stage 1: Permission to proceed

- Ethically approved and consent must be obtained from the relevant authorities.

Cont'd...

Stage 2: Data collection

- ❖ When collecting our data, we have to consider:
 - **Logistics:** who will collect what, when and with what resources
 - **Quality control**
- ❖ **Measures that helps to ensure good quality of data:**
 - Prepare a field work manual for the research team as a whole,
 - Train research assistants (data collectors, supervisors) carefully
 - Pre-test research instruments

Cont'd....

❖ **Stage 3: Data handling**

- A clear procedure should be developed for handling and storing them.

Data Collection Techniques

- ❖ In the collection of data, we have to be systematic.
- ❖ Data collection techniques allow us to systematically collect data about study subjects and the settings in which they occur.
- ❖ If data are collected haphazardly, it will be difficult to answer research questions in any conclusive way.
- ❖ Depending on the **type of variables and the objective of the study** different data collections

Data collection Techniques

1. Self administered Questionnaires
2. Interviews
 - a. Face to face/Telephone
 - b. In-depth interview
 - c. Focus group interviews
3. Observations
4. Documentary sources (for secondary data)

1. Self-administered Questionnaire:

❖ It is a data collection tool in which written questions are presented to be answered by the respondents in written form.

❖ **Advantages:-**

- Can cover a large number of people or organizations
- Relatively cheap
- No prior arrangements are needed
- No interviewer bias

Types of Questionnaire

- ❖ **Structured:** They offer a list of possible options or answers from which the respondents must choose.
- ❖ **Semi-structured:** offer a list of possible options or answers from which the respondents choose and it have a space to permit extra answer
- ❖ **Unstructured/ in-depth interview:** permit free responses that would be recorded in the respondent's own words

Cont'd....

❖ **Disadvantages:-**

- Difficult to design and often require many rewrites before an acceptable questionnaire is produced.
- Questions have to be relatively simple
- Time delay for waiting response
- Assume no literacy problem
- Historically low response rate
- Not possible to give assistance if required

Cont'd....

- ❖ A self-administered questionnaire can be administered in different ways
 1. Through mailing to respondents
 2. Gathering all or part of respondents, giving oral or written instructions, and letting them fill out the questionnaires;
 3. Hand-delivering questionnaires to respondents and collecting them later
- ❖ The questions can be either open ended or closed **(with pre -categorized answers)**

2. Interview Method

- ❖ It involves oral questioning of respondents, either individually or as group
- ❖ A technique used to gain an understanding of the underlining reasons and motivations for peoples' attitudes, preferences or behavior.
- ❖ Types of interviews:-
 - Structured
 - Semi-structured
 - Unstructured(in-depth interview or FGD)

Cont'd...

❖ **It can be of low degree flexibility such as:**

- Face to face,
- Phone interview

OR

❖ **It can be of high degree flexibility such as:**

- In-depth interview
- Focused Group Discussion(FGD)

Cont'd...

Low degree flexibility

- ❖ They are useful;
 - When the researcher is relatively knowledgeable about expected answers or
 - When the number of respondents being interviewed is relatively large

Cont'd...

High degree flexibility

- ❖ Can be used for interviewing individuals as well as groups of key informants.
- ❖ Useful if a researcher has as yet little understanding of the problem or situation under investigation.
- ❖ It is frequently applied in **exploratory studies** and also used during **case studies**

Face to Face Interview

❖ Advantages :-

- Good response rate
- Completed and immediate
- Possible in-depth questions
- Interviewer in control and can give help if there is a problem

Face to Face Interview...Cont'd

❖ Disadvantages:-

- o Time consuming
- o Need to set up interviews
- o Geographic limitations
- o Can be expensive

In-depth interview

- ❖ Data collection technique characterized by extensive **probing and open-ended questions**,
- ❖ Conducted on a one-on-one basis between the respondent and a highly skilled interviewer

In-depth interview...Cont'd

❖ Goals:

- To get narrative, stories
- To elicit potential cognitive/cultural domains
- To assess reported behavior
- To obtain broader context of your research problem

❖ It Lasts 30-90 minutes

❖ It requires a minimum sample size 20-30 respondents

In-depth interview...Cont'd

When to use In-depth interview?

- When the subject matter is highly sensitive.
- For example, conducting a study among women who have had an abortion, regarding their feelings about sexuality and family planning.

Focus group discussion(FGD)

- ❖ A qualitative data collection technique involving in-depth, guided discussions among a group of participants facilitated by a trained moderator
 - Moderator- leads the discussion
 - Last 1-2 hours
 - Are recorded; helper takes notes
 - Take place in a private, quiet setting

FGD....Cont'd

- ❖ The purpose of a FGD is to obtain in-depth information on concepts, perceptions, and ideas of the group
- ❖ Good way to get a sense of:
 - Social norms related to a given topic
 - How participants talk to each other about a topic
 - Explicit use of group interaction to produce data and insights that would be hard to get otherwise

FGD....Cont'd



FGD....Cont'd

❖ Advantages:-

- Quick result and cost-effective
- Groups may generate important issues
- Ideas as how to proceed with the study may be generated.

❖ Disadvantages:-

- Topic of discussion may be missed
- The discussion may be manipulated by the moderator.
- Needs well trained professionals

Observation

- ❖ **Observation** is a technique which involves systematically selecting, watching and recording behaviors and characteristics of living beings, objects or phenomena.
- ❖ It is much used data collection technique
- ❖ Observations are usually **complementary to other data** collection techniques.
- ❖ They can give additional, more accurate information on behavior or people than interviews or questionnaires

Observation...Cont'd

❖ Advantages:-

- Gives more accurate data on behavior and activities
- Collection of information on facts
- Can be used in any type of study

❖ Disadvantages:-

- Investigators or observers own bias
- Need more resource & skills
- They are time consuming
- Often used in small -scale studies

Observation...Cont'd

❖ It can be undertaken in two different ways:

- Participant observation:

- Non-participant observation:

OR

- Structured observation

- Unstructured observation

Observation...Cont'd

❖ Participant observation

- The researcher is part of the group that is being investigated
- The researchers would live in tribal villages attempting to understand the customs and practices of that culture.

❖ Non-participant

- The observer(he or she) does not normally question or communicate with the people being observed. or
- The observer watches the situation, openly or concealed, but does not participate

Observation...Cont'd

❖ **Structured observation**

- The researcher specifies what is to be observed and how the measurements are recorded.
- Appropriate when the problem is clearly defined

❖ **Unstructured observation:**

- The researcher monitors all aspects of the phenomenon that seems relevant.
- It's appropriate when the problem has yet to be formulated precisely
- The potential for bias is high

Review of Documents

- ❖ Mainly used for secondary data
- ❖ There is a large amount of data that has already been collected by others.
- ❖ Locating these sources and retrieving the information is a good starting point in any data collection effort.

Review of Documents...Cont'd

- ❖ Some sources of such data are:
 - Mortality reports
 - Morbidity reports
 - Epidemic reports
 - Reports of laboratory utilization (including laboratory test results)
 - Reports of individual case investigations
 - Reports of epidemic investigations
 - Special surveys (e.g., hospital admissions, disease registers, and serologic surveys)
 - Demographic data

Review of Documents...Cont'd

❖ Advantages:

- Relatively easy
- The best means of studying **past events**.

❖ Disadvantages:

- Problems of **reliability and validity** due to:
 - The records are maintained not for research purposes
 - Presence of Incomplete data

Selection of data collection method

❖ **Selection of data collection method is based on:**

- The resource required to apply the method
- Acceptability of the method
- Accuracy of the method
- Relevance
- The number of participants aimed to be covered by the study.
- Familiarization of the procedure

Problems in gathering data

- ❖ Language barriers
- ❖ Lack of adequate time
- ❖ Expense
- ❖ Inadequately trained and experienced staff
- ❖ Cultural norms

Bias in Data Collection

- ❖ **BIAS** in information collection is a distortion which results in the information not being representative of the true situation
- ❖ Possible sources of bias during data collection:-
 - Defective instruments
 - Observer Bias
 - Information bias
 - Effect of the Interview on the Informant

Validity and reliability of data

❖ Reliability

- **Reliability** refers to the repeatability of scientific observations under identical conditions
- If repeated measurements of a characteristic in the same individual under identical conditions produce similar results, we would say that the measurement is reliable.
- A study result is said to be reliable if the same result is obtained when the study is repeated

Validity and reliability...Cont'd

❖ Validity

- Validity refers to the degree to which scientific observations actually measure or record what they allege to measure
- A measurement is said to be valid if it measures what it is supposed to measure
- We would seriously doubt the answers from interviewing on sensitive subjects, because they are generally lacks validity

Ways to Make Data More Reliable

- ❖ Training
- ❖ Use of different sources of data
 - Combining Different Data Collection Techniques
- ❖ Pre-testing
- ❖ Supervision