

5

Sampling

LEARNING OBJECTIVES

After the completion of the chapter, the readers will be able to:

- Define the terminology related to the sampling.
- Enumerate the steps of sampling process.
- Understand the sampling methods and their application in research.
- Identify sample size for nursing research.
- Differentiate probability and nonprobability sampling.
- Enumerate the bias in sampling.

CHAPTER OUTLINE

- Introduction
- Common Terms
- Purposes of Sampling in Research
- Characteristics of a Good Sample
- Sampling Procedure/Process
- Eligibility Criteria for Sampling (Inclusion and Exclusion Criteria)
- Factors Influencing Sampling Process
- Sampling Methods
- Sample Size and its Determination
- Sampling Error
- Problems of Sampling

Nursing Knowledge

An Initiative by CBS Nursing

INTRODUCTION

Along with research design, sampling design is nothing but selection of participants to be observed or studied in the research study. Selection of participants, done with the help of different methods or techniques is known as sampling. Each participant is termed a sample or study subject. It is widely used in everyday life and in other professional research too. In medicine and health sciences generalization is always done on the basis of samples. The **main purpose** of sampling is to draw conclusions about the given population from samples by directly observing a portion of the population. Sampling is a procedure by which some members of a given population are selected as a true representative of the entire population.

COMMON TERMS

- **Population:** The universe in totality or aggregate, the entire set of all individuals, subjects or objects with specified common characteristics in a given study area. The population may be homogeneous (all elements have similar characteristics) or heterogeneous (all elements have different characteristics). It is important to be clear that the population is not only restricted to human beings. It also includes hospital records, specimens like blood and urine or laboratory animals, etc.
- **Target population:** The aggregate of individuals from which the researcher wants to generalize the findings of research study. In simple words, a population is the group of people or things to whom a researcher wants to apply or generalize the study findings. For example, diabetes mellitus patients are target population.
- **Accessible population:** The aggregate of individuals who are accessible to the researcher from which the samples are chosen for the research study. In simple words, it is a population is the group to whom a researcher wants to generalize the findings. For example, diabetes mellitus patients who are admitted to a medical ward are accessible population.
- **Finite population:** Also known as a countable population, it refers to the samples or elements within a population that can be measured in numbers (e.g., the number of patients who visited the OBGY outpatient department).
- **Infinite population:** Also known as an uncountable population, it refers to the samples or elements within a population that cannot be measured in numbers (e.g., the number of microbes in a patient's body).
- **Sample:** A subset or a small portion of a population selected for the research study, whose properties are to be generalized to the larger population or set. Sample must be less than the total population.

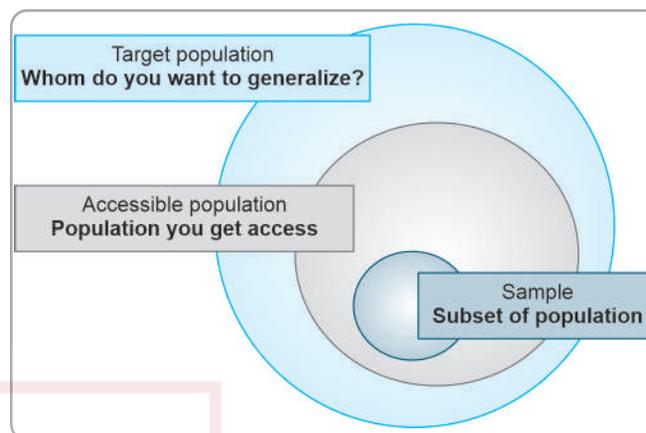


Fig. 5.1: Target population, accessible population and sample

- **Sampling unit:** The basic unit chosen in the sampling process, such as a person, household, village, etc., is considered a sampling unit.
- **Sampling frame:** It is a set of all sampling units that comprise the study population from which a sample is drawn, otherwise called working population. For example, a list of all villages and a list of all mothers, etc.
- **Sample size:** The number of individuals or observations included in a sample is called the sample size. A sample is considered small if it has <30 observations, and large if it has >30.
- **Representative sample:** The sample which has most of the characteristics of the population is termed representative sample which is essential for the generalization of the findings.
- **Random sample:** A sample which is selected by giving every sampling unit the same chance of being included in the study.

Target population, accessible population and sample are given in Figure 5.1

PURPOSES OF SAMPLING IN RESEARCH

- **Economical:** A significant purpose for using sampling in research is to get information/data from the large populations (infinite) who are scattered at different geographical locations with reduced costs, time and with increased accuracy of results.
- **True representativeness:** Another purpose is to have true representativeness (sample must possess all the characteristics of a population) of sample to the entire population. Representativeness is very much essential to generalize the results to the entire population/universe.
- **Quick results:** A study on the entire population takes enormous time and it is a laborious process, where it is difficult to conduct the research on it is a large population,

however, with a sample it is possible to have results faster and quick.

- **Precision and accuracy of data:** Sampling involves collection of data from a portion of the entire population, where it helps to generate precise and accurate data in a given research study.

CHARACTERISTICS OF A GOOD SAMPLE

A sample is a basic element or unit of the entire population. Given below are the salient features of a sample where each and every sample must possess the following elements:

- **Truly representative:** The representative sample is the one which possesses the characteristics which are similar to the entire population, only then the generalization of findings is possible. In simple terms, the sample should have all characteristics (homogeneous) that are present in the population.
- **Free from sampling error and bias:** Each sample of the population should have equal opportunity to be recruited to the research study. Use of probability techniques avoids the sampling bias. The sampling bias leads to over or under representation of the entire population, wherein it can be minimized when the samples are homogeneous in nature.
- **Optimum size:** The sample size must be large enough (adequate number) to provide sufficient precision of results as well as to draw the inference accurately. The sample size should be identified via scientific method rather than arbitrary method.
- **No substitution and incompleteness:** The sample should be good. Once a participant is selected for the study, they should not be replaced till the researcher has obtained the complete data required for the research study.

SAMPLING PROCEDURE/PROCESS

Sampling is a process of selection of a sample from the available population. Selection of a sample is an important step in the research process, as the results are validated and generalized based on sampling methods and procedures. The above mentioned points must be considered before sampling procedure. Following are the steps of sampling procedure (Table 5.1 and Fig. 5.2):

1. **Identify and define the target population:** Once the statement of problem is finalized, the researcher identifies and determines the target population on whom he/she wants to generalize the findings. The target population is the complete group of specific population elements or aggregate of cases related to the research project.
2. **Identify the accessible population:** Here, the researcher identifies the portion of the target population which is

Table 5.1: Steps of sampling procedure with example

Sl. no.	Steps	Example
1.	Identify and define the target population	Study on diabetes mellitus patients
2.	Identify the accessible population	Diabetes mellitus patients admitted at hospitals
3.	Describe the sampling frame	List of diabetes mellitus patients admitted in all medical wards of hospitals
4.	Determine the selection criteria of sampling	Diabetes mellitus patients above 45 years of age, and who give written informed consent
5.	Decide the sample size	Sample size: 90 (with 5–10% attrition rate)
6.	Specify the sampling methods	Probability method, systematic random sampling technique
7.	Select the sample and conduct field work	Collect data and identify the results

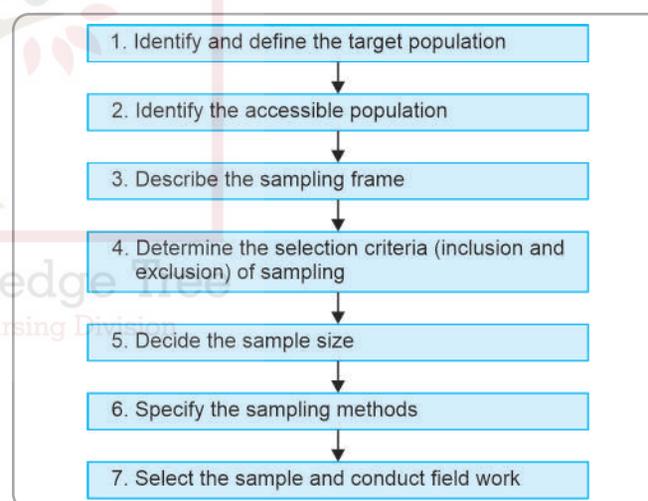


Fig. 5.2: Steps of sampling procedure/process

more accessible to them, and they are recruited for the research work. In simple term, it is the population which is readily available for the researcher.

3. **Describe the sampling frame:** The researcher prepares the complete list of all units or elements, called sampling frame, from which the samples are drawn for the research study.
4. **Determine the selection criteria (inclusion and exclusion) of sampling:** The researcher must clearly lay down the eligibility criteria for the selection of a sample. It is always carried out by considering the inclusion and exclusion criteria for enrolling the participants in the study.

5. **Decide the sample size:** The sample size has to be determined based on the type of study with the help of power analysis or statistical software packages (e.g., OpenEpi software). It should not be done in an arbitrary method. The attrition rate (dropout) must always be kept in mind while determining the sample size. The sample size helps in the planning of implementation of sampling process.
6. **Specify the sampling methods:** Sampling method is decided based on the size of population and its geographical concentration. The researcher must choose either the probability (random selection) or nonprobability (purposive) method for the selection of sample based on type and nature of the research study.
7. **Select the sample and conduct field work:** Select the sample as per the sampling plan, decide techniques, and collect the data (research work) to achieve the objectives.

ELIGIBILITY CRITERIA FOR SAMPLING (INCLUSION AND EXCLUSION CRITERIA)

The terms, inclusion criteria, eligibility criteria, exclusion criteria or delimitations are used to define the attributes of a sample for the given population to ensure that the selected samples are homogeneous in nature. The inclusion criteria are called eligibility whereas exclusion criteria are called delimitations of the study. The population characteristics serve as a base for eligibility criteria of sample, and importantly both sample and population should be similar. Here, higher the similarity or congruence higher the representativeness of the sample.

The inclusion and exclusion criteria are made to have control over the extraneous variables of study or the bias that will influence the outcome of the study. Further, it is very important that the inclusion and exclusion criteria should have valid reasons to include or exclude them. The careful development of inclusion or exclusion criteria for study sample will increase the research precision and strength of evidence, thereby contributing to the accuracy and generalizability of the study findings.

FACTORS INFLUENCING SAMPLING PROCESS

The systematic process of recruiting the sample in the study is known as sampling. Following are the factors that need to be considered while selecting a sample:

- **Researchers' experience:** The research is a scientific activity where the researchers need to have variety of qualities. They should have awareness, training and skill in the sampling process as well as recruitment of sample. The lack of awareness, interest inadequate training, time

and extensive workload influence the selection of samples for the study. A novice researcher may find difficulty in performing research work and sampling process.

- **Characteristics of sample:** The nature of subjects to be recruited for the sample plays a significant role in the sampling process. The sample characteristics may include age, gender, education, economic status and diagnosis, etc. At times, the researcher finds difficulty to identify, locate and recruit subjects in the study.
- **Sample size:** If the sample does not have homogeneity of characteristics of the entire population, generalization is not possible. The sample size should be determined via scientific method and arbitrary selection otherwise there will be a scope for sampling bias which will influence the study results. The generalization of findings is not possible if the sample size is small.
- **Sampling frame:** Faulty sampling frame leads to defective sampling process and flawed research work, thus the researcher must have awareness on population characteristics thereby correct sampling frame can be made for the research study.
- **Sampling technique:** Sampling method is a crucial factor which influences the sampling process. The selection of sample depends on the type of study, availability of sample, cost, logistic resources and time availability. The random sampling is a preferred method for any research, however, nonprobability sampling is used in qualitative research studies.
- **Sample attrition:** The chance of dropout rate among the recruited samples for study also influences the sampling process especially in experimental research. Usually 5–10% attrition rate shall be kept to achieve the predetermined sample size even with dropouts during the study.
- **Other logistical factors:** Time availability, geographical distribution of sample, natural calamities and lack of cooperation from the participants influence the sampling process.

Points to be kept in mind while sampling

- Population to be studied
- Population size and geographical distribution
- Heterogeneity of population with respect to variables
- Availability of list of sampling units
- Level of precision required
- Resources available

SAMPLING METHODS

The sampling methods are broadly classified into two types, i.e., probability and nonprobability sampling methods; they are used in research as per the type and nature of study. Following are the commonly used sampling methods in nursing research (Fig. 5.3):

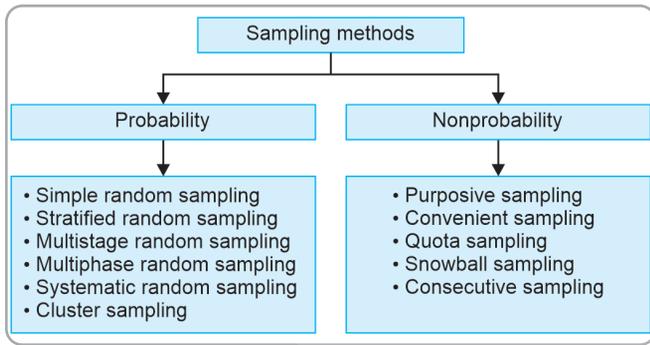


Fig. 5.3: Classification of sampling methods

Probability sampling (scientific):

- Simple random sampling
- Stratified random sampling
- Multistage random sampling
- Multiphase random sampling
- Systematic random sampling
- Cluster sampling

Nonprobability sampling (nonscientific/nonrandom):

- Purposive sampling
- Convenient sampling
- Quota sampling
- Snowball sampling
- Consecutive sampling

Probability Sampling

Probability sampling involves the random selection procedures to ensure that each sample, element and individual in the population is selected on the basis of chance. Here, all the units of study population should have equal and fair or at least a known chance of being included in the sample for the research study. The basic principle of random sampling is ‘equal chance/probability of drawing each unit’ which enables the researcher to have representativeness of the whole population. These methods of sampling are commonly used in quantitative and experimental research studies.

Simple Random Sampling

Simple random sampling is a sampling method where each sampling unit or element in the sampling frame has the same and equal chance of being selected in the sample. It is applicable when the population is small, homogeneous (similar characteristics) and readily available. However, a clear sampling frame is must for drawing the sample and it will not be possible if there is no clear sampling frame. The commonly used methods are: (a) Lottery Method and (b) Random Number Tables.

- **Lottery method:** In this method, the samples of total population are listed in a paper (chit) which is similar size, shape, texture and color. The chits, folded in the same form are placed in a bowl. These chits are handpicked randomly (blindly) and recorded by the researcher. The selected chit’s numbers constitute the sample. In this sample, each chit is eligible to be a part of the research work.
- **Random number table:** In this method, the samples of total population are listed in the form of a table. The samples are selected randomly (blindly) from the table numbers for research work. There are a number of readymade random tables available which can be used as per the total number of the population. Examples of random table numbers are Tippets (1927) random number table, Fisher and Yates (1938) table and Rand corporation (1953) random number table, etc. Alongside there are standardized readymade random number tables available on various websites and in statistical packages (For example, OpenEpi) for ready reference.

Principle: Equal and independent chance/probability of drawing each unit

Steps of simple random sampling:

1. Identify sampling (accessible) population
2. List all sampling units (sampling frame)
3. Number all the units
4. Randomly draw the units (by table of random numbers or lottery method)
5. Samples are drawn with/without replacement

Random number table for the population of 100

(Required sample size: 25)

01	11	21	31	41	51	61	71	81	91
02	12	22	32	42	52	62	72	82	92
03	13	23	33	43	53	63	73	83	93
04	14	24	34	44	54	64	74	84	94
05	15	25	35	45	55	65	75	85	95
06	16	26	36	46	56	66	76	86	96
07	17	27	37	47	57	67	77	87	97
08	18	28	38	48	58	68	78	88	98
09	19	29	39	49	59	69	79	89	99
10	20	30	40	50	60	70	80	90	100

A total of 25 samples are required from the population of 100; the above table shows the randomly selected samples for the scientific study which are: 04, 09, 13, 17, 19, 21, 26, 30, 34, 39, 43, 47, 52, 55, 59, 65, 68, 73, 76, 80, 84, 87, 91, 96, 99. The data is collected from the chosen samples.

Alongside, the random number table can be generated with the help of 'OpenEpi' software which is depicted below. Here, the researcher needs to enter the minimum number, maximum number and required size in the drop down menu. After the entry of this requirement, the software produces and displays the random number table which can be used for the study.

Random Number Generator

Minimum number: 1

Maximum number: 150

Required sample size: 25

25 Random Numbers from 1 to 150				
Generated by the OpenEpi Random Program				
www.openepi.com				
4	35	44	107	118
77	65	19	40	18
99	146	49	146	47
120	13	98	101	95
89	86	10	114	68

The numbers are generated by the JavaScript Math random function

Source: Results from OpenEpi, Version 3, open source calculator-Random

Print from the browser with ctrl-P or select text to copy and paste to other programs.

Example: Prevalence of tooth decay among the children attending a school

- Enlist the children attending the school (i.e., 450 students)
- Children are numerated from 1 to 450
- Determined sample size is = 30 children
- Selecting sample randomly 30 in number between 1 and 450

Advantages

- It shows the high representativeness of sample characteristics of the population.
- It is free from sampling errors and one of the most unbiased sampling methods.
- It is applicable when population is small, homogeneous and readily available.
- It is more scientific, simple and easy method of recruiting the samples.

Limitations

- If there is no sampling frame, the simple random sampling is not possible as it needs the complete list of population.
- If the sample size is small, the sample will not be a true representative of the population.
- If samples are widely distributed, then it is an expensive and a time consuming technique.

Stratified Random Sampling

This method is used when the population is not homogeneous (does not possess the similar characteristics with each other). The population under the study is divided in homogeneous

groups which are called 'strata'. However, preventive measures should be taken to avoid the overlapping in the strata. In research, geographical, sociological and economical characteristics are generally used for stratification. The examples of strata are *Religion*—Hindu, Muslim, Christian; *Workers*—Skilled, unskilled, clerical and non-clerical; *Gender*—Male and Female.

In stratified random sampling, (a) First of all, the population is divided in homogeneous strata or groups according to a characteristic of interest, i.e., age, gender, religion, residence etc., (b) Prepare a sampling frame, and (c) Then the samples are drawn by using a Simple Random Sampling technique the each stratum/group in proportion to its size. The strata are as homogeneous as possible within and as heterogeneous as possible between. This technique provides more representative sample than Simple Random Sampling technique in a given large population. The size of the sample from each stratum can be *proportionate* (equal proportion/percent) and *disproportionate* (unequal proportion).

Proportionate: In this, a given percentage of samples is selected from each strata. For example, in case 50% of selection, there will be a selection of 5 out of 10 samples or 10 out of 20 samples (equal fraction/percent) depending on the size of the stratum (Table 5.2).

Disproportionate: A different proportion of selection is recommended if the different strata reveal different levels within strata variability. For example, in case of a large group researcher may select only less percentage of samples and in small groups researcher may select increased percentage as sample. If the researcher wants to study the haemoglobin level of adolescents, among available adolescents 100 from urban area, 200 from semi urban area and 300 from rural area. Here, the researcher uses different sampling fraction/percent and randomly select the sample of 50 from each stratum (Table 5.3).

Table: 5.2: Example of proportionate stratified random sampling

Strata	A	B	C
Population size	100	200	300
Sampling fraction/percent	1/2	1/2	1/2
Final sample size	50	100	150

Table: 5.3: Example of disproportionate stratified random sampling

Strata	A	B	C
Population size	100	200	300
Sampling fraction/percent	1/2	1/4	1/6
Final sample size	50	50	50

Principle: Divide the sampling frame in homogeneous sub groups and draw the samples as per the sample size.

Procedure:

1. Draw random sample in each strata (simple or systematic random method).
2. Size of the sample from each strata can be proportionate or disproportionate.

Example: Assess the psychosocial problems among senior citizens.

Divide the population in different groups as strata I (Institutionalized elderly) and strata II (Elderly at homes) and chose the sampling fraction or percent followed by selecting the determined size of samples.

Advantages

- It is applicable when population is diverse/heterogeneous and finite in nature.
- It has more representation of sample characteristics of population.
- It provides greater accuracy with increased precision of sampling and results.
- It has administrative convenience as compared to other random sampling methods.
- It is more scientific, simple and easy method of recruiting the samples.

Limitations

- It is difficult to divide and stratify the samples in a homogeneous group.
- There is a chance for the overlapping of sampling which may leads to not a true representativeness.

Systematic Random Sampling

In this method, the samples are drawn from a fixed interval from the population. It involves the selection of every Kth unit (house/person/patient) in the population or from sampling

frame, where K refers to the sampling interval, which is calculated by $K = N/n$

$$K = \text{Total population (N)}/\text{Sample size (n)}$$

After calculating the sample interval (K), the first unit is selected by random method from among the first K units followed by total the first sample unit plus sample interval is second sample and so on. It is the most convenient technique used for community level or clinic based studies, and the time and labor involved in collection of sample is relatively small. The basic principle involved is selecting the sample at regular interval based on sampling frame.

Example: The total population is 1000 and required sample size is 100. Thus, the samples drawn are:

$$K = N/n$$

$$K = \text{Sample interval } N = \text{total population } n = \text{sample size}$$

$$K = 1000/100 = 10$$

The **first sample is drawn randomly within the limit of sample interval** say it is 3, the second sample will be 10+3, i.e., 13 and third sample is 23 and so on... [3, 13, 23, 33, 43 and so on] till the researcher the determined sample size.

Principle: Select sample at regular intervals from sampling frame based on sampling fraction

Procedure:

1. First obtain the list of total population (N)
2. Determine the sample size (n)
3. Identify the sample interval (k) [$k = N/n$]
4. Select every kth person until the required sample size is acquired

Example: Assessment of the impact of counseling on depression among cancer patients receiving chemotherapy. In this the sample size is 60, wherein the cancer patients are selected in a specific interval till the researcher gets their predetermined sample size.

Advantages

- It is a simple and more convenient method for sample selection
- The time and labor involved in the selection of sample is relatively small
- There is always a scope for the evenly distribution of samples from the entire population.

Limitations

- It cannot be employed when the population size is small
- If the first sample is not selected randomly, afterward it becomes nonrandom sampling method.
- The variance increases if there is periodicity in the numbering of elements, leading to bias in the selection process.

Cluster Sampling

The basic and the smallest unit of population is elements, and the collection of elements is known as clusters. Usually, the clusters are internally heterogeneous in nature. In cluster sampling the population is divided in clusters and samples are drawn randomly either from all clusters or selected clusters of specific geographical areas. The examples of common clusters are wards, units, villages, cities, slums, schools, etc., however the clusters are internally heterogeneous in nature. This type of sampling is commonly employed in national surveys.

In community based studies, 30 cluster sampling is the commonly used method adopted by the World Health Organization for field studies. The basic principle involved in cluster sampling is sample units are not identified independently but in a group (cluster). Cluster sampling has greater logistical advantage for conducting the field work.

Identification of clusters: The cluster sampling is done by considering the following *steps*:

1. Firstly, list all the cities, towns, villages and wards of cities with their population falling in the target area under study.
2. Calculate the cumulative population and divide by 30, which gives the researcher the sampling interval.
3. Select a random number less than or equal to sampling interval having same number of digits – this will form the first cluster.
4. Add random number plus sampling interval, it will give you second cluster.
5. Add second cluster plus sampling interval, it will give the third cluster population and the last (30th) cluster will have 29th cluster plus sampling interval.

The cluster sampling has **two types**:

1. **One stage sampling:** All of the elements within the selected clusters are included in the sample.
2. **Two stage sampling:** A subset of elements within selected clusters are randomly selected for inclusion in the sample. The cluster sampling is commonly used when the population is large, sampling frame is difficult and population is located geographically diverse in nature.

Principle: It involves selection of clusters (group) rather than sampling units, individual or elements.

Procedure:

1. Divide the population in clusters
2. Randomly sample the clusters
3. Measure all units within sampled clusters

Example: Study of knowledge, attitude and practice on family planning among eligible couples in rural community.

Advantages

- Useful when the population is infinite and diverse, and helpful in carry out national or regional surveys.
- The sampling frame (listing of all the units or elements) is not required.

- It is more economical and used during logistical difficulties.

Limitations

- There is a scope for minimal or higher representation of the population.
- There is a possibility of sampling errors when less or inadequate number of clusters involved in the study.

Multistage Random Sampling

Multistage random sampling involves sampling at various stages by using random techniques; however, the samples at the final stage (household, person, etc.) are studied. The stages are considered primary (first stage), secondary (second) and tertiary level (final). Results of the final stage sampling units represent the whole population. Usually, this method is commonly employed in macro level surveys, and community based studies where it has flexibility in sampling which lacks in other techniques. The basic principle involved is consecutive sampling (sampling unit is household: First stage—draw neighborhoods, Second stage—draw buildings and Third stage—draw households). For example, Panchayat system in the villages and a study on nutritional status of under five year children in India, etc.

Principle: Sampling is selected in various stages, but only the last sample of subject is studied. The investigation is restricted to small number of units which represents the whole population.

Example: 'Assess the nutritional status of under five children' In this study, three states are selected randomly from four regions of India (primary stage-I), followed by randomly selecting 3 districts from each state (intermediate stage-II) and later selecting 3 villages randomly from each four districts. Thus, finally, 27 villages are selected as a representative of the entire India (Fig. 5.4). The data is collected from these 27 villages and the findings will be generalized all over India.

Advantages

- It is used for a large geographical area and diverse population, where the sampling frame is not required.
- Less expensive and less time consuming.

Limitation

There is a possibility of minimal or excessive representation of the population.

Multiphase Random Sampling

In this method, sampling is done in different phases, where portion of information is collected from sample and part of information from the subsamples. The each sample is adequately studied before another sample is drawn from it.

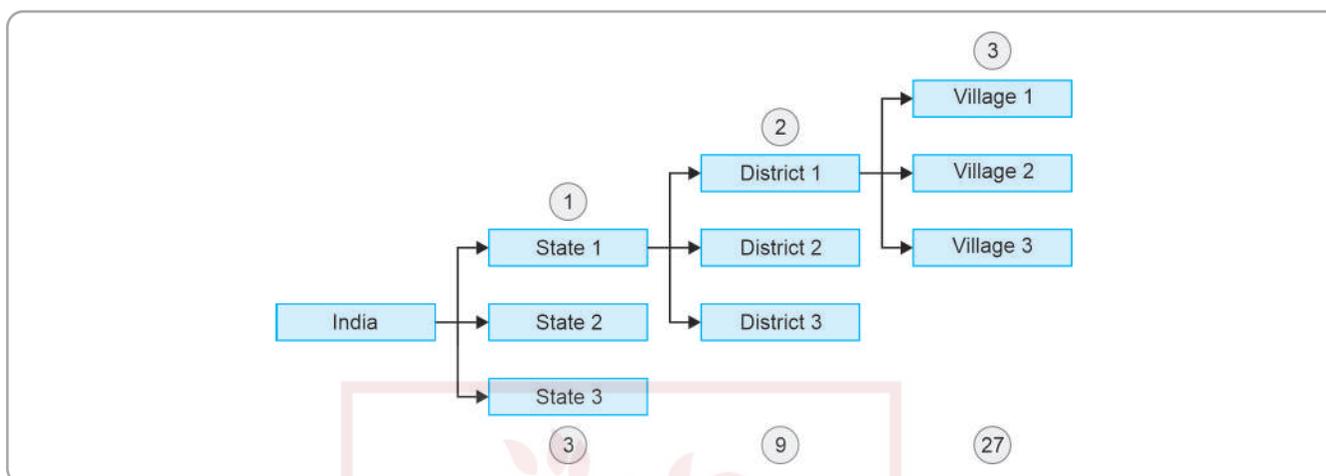


Fig. 5.4: Stages of multistage random sampling

The sample size becomes smaller and smaller from sample to subsamples at different phases, and it is less costly, less laborious and more purposeful for large scale studies. The difference between multistage and multiphase samplings is that *in multistage*—sampling frame is different at each stage whereas in *multiphase*—samples are drawn from the same frame.

Principle: Sampling is done in different phases. Part of information is collected from whole sample and part of information is from subsample.

Example: Prevalence of Tuberculosis, wherein **Phase I:** Mantoux skin sensitivity test was done to all cases, **Phase II:** Those who are positive are screened for X-ray exam, **Phase III:** Those who are positive only their sputum is examined for confirmation and diagnosis of tuberculosis, thus those who need sputum exam will be smaller. However, all the cases are adequately studied before another sample is drawn.

Advantages

- It is less costly, less laborious and more powerful sampling method.
- It reduces the error.

Limitation

It is time consuming and planning and administration is complicated.

Advantages of probability sampling:

- It is a scientific method of choosing samples for study.
- There is always more representativeness.
- Greater accuracy of results.
- Large geographical areas can be covered.
- More economical one.

Limitations of probability sampling:

- Needs complete list (sampling frame) of population.
- Sometimes difficulty in stratification.
- No true representativeness in case of smaller sample size.
- Time consuming and costly.

Nonprobability Sampling

Nonprobability sampling is a nonscientific (nonrandom) sampling method in which not every member, element or unit of the population has an equal chance of being included in the study sample. The choice of selection of sample is generally depends on the judgment of researcher. These sampling methods are routinely used in qualitative and nonexperimental type of studies, and when the random sampling is not feasible and realistic. The major drawback of these sampling is that it lacks of generalization and lack of true representativeness of entire population as it is selected purposefully.

Nonprobability sampling is often used in nursing as it is simple, easy and more convenient method. In case of testing of tool, pilot testing or use of qualitative research the nonprobability method of sampling is commonly employed. Most of the scientists do believe that these techniques should not be used in scientific work except for qualitative type of studies. Nonprobability sampling is most commonly associated with non-experimental and qualitative type of studies that contribute to lower level of evidence.

Features of Nonprobability Sampling

- The nonprobability sampling is a technique where the sample does not have the equal chance to be selected as sample of the research study.

- The samples in nonprobability sampling are usually chosen on the basis of their availability, accessibility and the researcher's judgment whether to include them or not in the study.
- Nonprobability sampling does not represent the true representativeness of the entire population as it is not selected by random sampling methods.
- The nonprobability sampling does not allow the researcher to calculate sampling statistics that provide information about the precision of the results.
- Nonprobability samples tend to be less complicated and less time consuming than probability samples.

Uses of Nonprobability Sampling

The following are the uses of nonprobability sampling in the research:

- Nonprobability sampling is useful when the researcher test the reliability of tools and techniques or during the pilot testing of a major study.
- It is useful when the researcher finds a unique case with special features and faces difficulty in getting the adequate number of samples in a given time and context.
- This sampling is useful if the researcher does not want to generalize the findings to the entire population, and only wants to learn the research methodology and research process as a novice researcher.
- When the randomization is not possible where the samples are minimum and inadequate this sampling can be widely used.

Types of Nonprobability Sampling

This sampling is categorized as purposive sampling, convenient sampling, quota sampling, accidental sampling, sequential sampling and snowball sampling These are discussed below:

Convenience Sampling

Samples are selected based on their availability and accessibility to the researcher or data collectors, and the study units that happen to be present during data collection are chosen as the sample for the study. This technique of sampling is commonly used in nursing research, wherein the availability of samples is of paramount importance and considered easy and less time consuming. The selection of samples are made from available sources like registers, directory and data bases where getting the sample is more convenient. The primary purpose of convenient sampling is to get basic information of the study participants as well as to conduct pilot studies. For example, if the researcher wants to assess the prevalence of diabetes mellitus among elderly residing in a specific village. Here, the researcher has frequent contact with elderly as he resides in

that village, thus the samples are conveniently available and accessible for the research study.

Advantage

It is considered to be the cheapest, easiest, cost effective and least time consuming technique.

Limitations

- There is always a possibility of sampling bias, the sample may not be a true representative of the entire population.
- The results obtained are less reliable as the samples are not the true representative.

Judgment Sampling

Judgment sampling is also called purposive/deliberate sampling technique. The researcher chooses the samples based on the specific purpose in mind that will be fulfilled as per the judgment. In this, samples are chosen based on definite reason rather than easy accessibility of participants. Here the researcher requires basic and advance knowledge about the population characteristics requires to choose the participants without complexity. It is very useful in testing the tools/intervention and pilot testing and utilized generally in qualitative research studies. Alongside, it helps in obtaining information from a specific group of people.

Example: The researcher wants to study the lived experience of myocardial infarction among its patients. Here the person who had myocardial infarction will be selected deliberately to assess the lived experience (variable under study). The researcher should have advance knowledge of participants who have myocardial infarction and they are selected purposefully.

Advantages

- It is a useful method for the exploration of specific variables under study.
- It is a simple method and requires less field work to select the samples.

Limitations

- It requires advance knowledge on the samples.
- There is always a greater chance of sampling bias.
- It limits the generalization of results to entire population.

Quota Sampling

The samples are selected non-randomly from fixed quota/strata. It involves stratification without randomization. It ensures that a certain number of sample units from different categories with specified characteristics appear in the sample so that these characteristics are represented. The basis of dividing the population in sub groups or quota is similar to stratified sampling method. The basis may be age, gender, occupation, education, income, religion, residence, etc., however, the samples are selected conveniently or purposively.

Steps:

1. Divide the population in subgroups or quota
2. Identify the proportion of subgroup's population
3. Choose the subjects from subgroups considering the proportion.

Example: Researcher wants to assess the effectiveness of educative intervention on awareness regarding prevention of anemia among nursing students. The determined sample size was 120, where the researcher selects 30 students from I, II, III and IV years respectively.

Advantages:

- It is economically cheap and easy method for sample selection.
- It is commonly used in field work and social studies.

Limitation:

It does not provide equal opportunity to every individual, and there is always a chance of sampling bias.

Accidental Sampling

Accidental sampling is otherwise called *incidental sampling* technique, where the people might have assembled for certain purpose and the researcher may find it suitable for his/her study. Thus, the people are enrolled as participants of study incidentally without any prior insight by the researcher. In this, neither the participants nor the researcher have any previous knowledge and contact with each other. The common areas to carry out the accidental sampling are educational institute, convention or meeting areas, shopping malls, etc.

Example: Researcher wants to assess the risk factors of myocardial infarction among adults. Here, the researcher goes to a party or a meeting and finds a large number of people above 18 years of age and tries to assess the risk factors among them.

Advantage: It is the simplest and easy method for sample selection.

Limitation: It does not represent true population and increase probability of sampling bias and credibility of results.

Sequential Sampling

Sequential sampling is a consecutive sampling, where the samples are drawn one after another from the population, i.e., data is collected from a group of participants followed by another group, if required. The samples are drawn one after another from a universe depends on the results of earlier samples. The sampling schedule depends on a second sample which can only be obtained after the experiment is conducted on the initial group. Because of repetitive nature of sampling method, there is a scope for modification or changes in research methods. It is commonly used in quality control process.

Example: Assessment of effectiveness of massage on sleep quality among cancer patients. In this, the researcher tries

to assess the effectiveness of massage on sleep quality on one group patients and draw inference or test hypothesis, if not achieved or tested, researcher select another group of patients and test the interventions.

Advantages:

- It is cost effective, simple and easy method.
- In sequential sampling, there is no limit for sample size (can be relatively small or large) and has flexible sampling process.

Limitation:

It lacks true representation of population.

Snowball Sampling

Snowball sampling is also known as network or chain sampling. It is employed based on the advantage of social networking, where first chosen sample identifies other participants through friends or other referrals. It is used to locate the potential subjects who are difficult to trace out. After the selection of first sample, researcher asks them regarding the availability of other individual with similar problems then researcher enrolls them in the study. This technique is very useful where the population size is very small, difficult to locate and helps to identify rare cases.

Types of Snowball Sampling Technique

1. Linear snowball sampling (initial sample provides reference of one similar subject) (Fig. 5.5).
2. Exponential nondiscriminative snowball sampling (initial sample provides reference of at least two similar subjects) (Fig. 5.6).

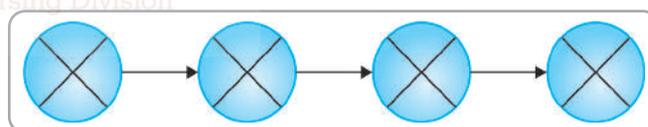


Fig. 5.5: Linear snowball sampling

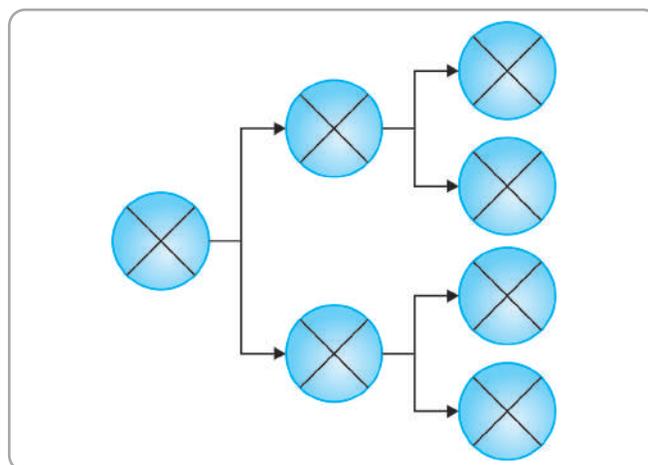


Fig. 5.6: Exponential nondiscriminative snowball sampling

sample size, degree of precision and relative frequency, attrition rate, etc. are considered for calculating the sample size for research studies. It is always advisable that the sample size must be determined before the commencement of study, i.e., during the preparation of research proposal time only.

A general rule is to always use a larger sample size for greater representative of the population, generalization of findings and to have minimal or no bias in the results. The statistical procedures are commonly used for estimating the sample size, and the best method is called power analysis. Power analysis is an advanced statistical technique that is commonly used by researchers and is a basic prerequisite for funding agencies to grant funds based on scientific sampling procedure and sample size. No matter how better the quality of the research design is, the findings of a study and its generalizability are weakened when the power analysis is not evaluated for calculating the sample size to determine the effectiveness of the intervention.

Sample Size in Qualitative Research

There are no specific guidelines for determining the sample size for qualitative research design. The sample size in qualitative studies mainly depends on the aim of the study, type and quality and quantity of information required and type of sampling procedure employed. If researcher is interested to collect more and variety of information the sample size should be more or in case of less number of characteristics are interested then the sample size should be smaller in nature.

When the quantity of data is sufficient and saturation is reached, the researcher does not need to collect further information, which ultimately results in a smaller sample size. In qualitative studies, the saturation of data is the indicator for deciding the sample size. The types of qualitative research like ethnography, phenomenology, and grounded theory need more sample size but within 15–30 samples. However, the case studies, case report and historical research can be carried out with minimum of one sample also.

Sample Size in Quantitative Research

Quantitative research needs to have scientific method of determining the sample size rather than arbitrary method of selection which is usually carried out in qualitative research. The scientific determination of sample size and scientific method of sampling (probability) helps in generalization of research findings to the entire population or universe. Thus, sample size determination is of paramount importance in scientific research. The larger the sample size, better and accurate results whereas smaller the sample size greater the sampling errors and sampling bias. The sample size in quantitative research varies based on the research design employed to get answer to the questions.

Number of Respondents in a Sample

Research can be carried out even with single sample (as a case study) or with larger sample (as survey). However, sample of 30 and less than that is considered small sample size and sample of more than 30 is considered large sample size. It is important to have accurate sample size, if the size is large it may lead to increase of cost, time and laborious, whereas if it is small and not calculated scientifically may lead to inaccurate results. Thus, determining sample size is very important in research. The following assumptions must be kept in mind while determining the sample size, they are: Alpha error (α) usually 5%, Beta error (β) usually 20%, hypothesis (one sided or two sided), provision for loss to follow up / attrition (usually 5%). The sample size should be determined scientifically by using power analysis and use of statistical software [www.openepi.com]

The standard formula for calculating the sample size is:

$$\frac{n}{1 + n(e^2)}$$

n = Total number of population and e = Error (95%), i.e., 0.05 level

Prerequisites for the sample size calculation by any statistician:

- Study design
- Estimates of primary variables
- Alpha error
- Beta error
- Precision expected
- Hypothesis (one sided or two sided)
- Z value

Adequacy of Sample Size

The following are the factors to be considered for determining the adequacy of sample size, they are:

- **Degree of difference:** It is the difference between two groups, example: The difference between two means or propositions (control value and the expected test value).
- **Type I error:** It is otherwise called alpha error (α), the chances of detecting a difference which does not exist (false positive). It is always set at 0.05 or 5% which is called level of significance.
- **Confidence level (CL):** The probability that an estimate of a population parameter is within certain specified limits of the true value (denoted by $1 - \alpha$) \times 100. When α is decided the confidence level is automatically fixed.

α	CL
5%	95%
10%	90%

Confidence level is directly proportional to the sample size which means higher the confidence level higher the sample size.

- **Type II error:** It is otherwise called beta error (β), the chances of not detecting a difference which actually exists (false negative). It is always set at 0.2 or 20%. $1 - \beta$ is called power ($0.2 = 0.8$)
- **Power of the test:** It is the probability of correctly rejecting a false null hypothesis. It is denoted by $1 - \beta$. In terms of percentage, it is $100 \times (1 - \beta)$ when β is selected, the power of test is automatically fixed.

Beta	Power
10%	90%
20%	80%

- **Variation of results:** Standard Deviation (SD) of control mean and the test mean.
- **Dropout:** The expected dropout has to be determined and appropriate adjustment is done.
- **Noncompliance:** It will increase the variation and hence sample size has to be adjusted to the degree of expected non-compliance.
- **Z value (standard normal distribution):** $Z = (1 - \alpha)/2$, $Z(1 - \alpha)$ and $Z(1 - \beta)$ represent the number of standard deviations away from mean. $Z(1 - \alpha)/2$ and $Z(1 - \alpha)$ are the functions of the confidence level, while $Z(1 - \beta)$ is the function of the power of the test. Following are some of the commonly used Z values at 95% confidence interval.
 - Two sided test: $Z(1 - \alpha)/2 = 1.96$ and one sided test: $Z(1 - \alpha) = 1.65$
 - At 90% confidence level: two tailed test: $Z(1 - \alpha)/2 = 1.65$ and one sided test: $Z(1 - \alpha) = 1.28$
 - At 90% power: $Z(1 - \beta) = 1.28$ and at 80% power: $Z(1 - \beta) = 0.84$

The sample size also depends on the aims of the study, the estimated outcome, the required level of precision, operational constraints, and the availability of relevant information.

Methods to Determine the Sample Size

There are a number of methods to determine the sample size like: Arbitrary method (not a recommended and scientific method), from the result of the previous studies, nomograms, tables, can be used, formulas and computer programs (easy method usually based on formulas).

Statistical Determination of Sample Size

- **One sample situation:** For example, to estimate the true immunization coverage in a community of under five children, the previous studies showed that immunization coverage is around 80% where absolute precision is within 4% of the true value at confidence level = 95%
Sample size ' n ' = $Z^2 \cdot p \cdot (1 - p)/d^2$
 d = absolute precision = 0.04

p = expected proportion in the population = 0.80

$Z(1 - \alpha)/2 = 1.96$

Hence ' n ' = $Z^2 \cdot p \cdot (1 - p)/d^2 = (1.96)^2 \cdot (.80) \cdot (.20)/(0.04)^2 = 384$

- **Case control study:** For example, estimation of sample size based on odds ratio with specified relative precision
 - Two of the following should be known:
 - Anticipated probability of 'exposure for people with disease' = $a/(a + b) = P_1$
 - Anticipated probability of 'exposure of people without disease' = $c/(c + d) = P_2$
 - Anticipated odds ratio = OR
 - Confidence interval = $100(1 - \alpha)\%$
 - Relative precision = e
Sample size ' n ' = $Z^2(1 - \alpha)/2 \{1/[P_1(1 - P_1)] + 1/[P_2(1 - P_2)]\}/\log_e(1 - e)^2$
- **Cohort study:** For example, estimation of sample size based on relative risk with specified relative precision
 - Two of the following should be known:
 - Anticipated probability of disease in people exposed to factor = P_1
 - Anticipated probability of disease in people not exposed to factor = P_2
 - Anticipated relative risk = RR
 - Confidence interval = $100(1 - \alpha)\%$
 - Relative precision = e
Sample size ' n ' = $Z^2(1 - \alpha)/2 [(1 - P_1)/P_1 + (1 - P_2)/P_2 \log_e(1 - e)]^2$

- **Difference in two mean (single group):** The following small letters in the formulae used below represent the following:

n = sample size

s = standard deviation

r = rate

e = required size of standard error

p = percentage

Sample size ' n ' = $s_1^2 + s_2^2/e^2$

- **Difference in two rates (single group):**
Sample size ' n ' = $r_1 + r_2 / e^2$
- **Difference in two proportions (single group):**
Sample size ' n ' = $P_1(100 - P_1) + P_2(100 - P_2) / e^2$
- **Sample size calculation for significant difference between two groups:**

The following small letters in the formulae used below represent the following:

n = sample size

s = standard deviation

r = rate

e = required size of standard error

p = percentage

u = one sided percentage point of normal distribution

Comparison of two means:

Sample size ' n ' = $(u + v)^2 (s_1^2 + s_2^2)/(m_1 - m_2)^2$

- **Comparison of two rates:**

Sample size ' n ' = $(u + v)^2 (r_1 + r_2)/(r_1 - r_2)^2$

- **Comparison of two proportions:**

Sample size ' n ' = $(u + v)^2 \{p_1(100 - p_1) + p_2(100 - p_2)\}/(p_1 - p_2)^2$

Determination of Sample Size with Computers

Other than statistical calculation, the sample size can be determined with the help of openepi software in sample size menu (www.openepi.com) is presented with appropriate examples for better understanding of readers. The statistical packages and computer softwares are estimated by the sample size using the formulas. Here the data is required (number or percentage of specific items/variables) to compute the sample size.

The sample size is determined based on the type (design) of studies involved. They are as follows:

Example 1: Sample size determination for proportion or descriptive study: This module calculates sample size by determining the frequency factor in a given population and the sample size is given for confidence levels from 90% to 99.99%. Enter the population size (150000 or any) and outcome factor 50% (always), and then it generates the sample size with open source calculator. The nearest integer to be considered a sample size at 95%, wherein the calculated size is 384 further it can be rounded as 390, thus the final size will be 390.

Sample Size for Frequency in a Population		
Population size (for finite population correction factor or fpc) (N):	150000	
Hypothesized % frequency of outcome factor in the population (p):	50% +/-5	
Confidence limits as % of 100(absolute +/- %) (d):	5%	
Design effect (for cluster surveys-DEFF):	1	
Sample Size (n) for Various Confidence Levels		
	Confidence Level (%)	Sample Size
	95%	384
	80%	165
	90%	271
	97%	470
	99%	661
	99.9%	1075
	99.99%	1500
Equation		
Sample size $n = [DEFF * N * p(1-p)] / [(d^2 / Z_{1-\alpha/2}^2 * (N-1) + p(1-p)]$		

Source: Results from OpenEpi, Version 3, open source calculator – SS Propor Print from the browser with ctrl-P or select text to copy and paste to other programs.

Example 2: Sample size determination for Unmatched Case-Control Study: This module calculates sample size for an unmatched case control study. Here you need to enter the desired confidence level (always 95), power (always 80), a hypothetical percentage of exposure among the controls (mean value of control group of previous study, e.g., 47), and either an odds ratio (always 1) or a hypothetical percentage of exposure among the cases (mean value of experimental group of previous study, e.g., 73). Results are presented using methods of Kelsey, Fleiss, and Fleiss with a continuity correction. Always consider the higher the values (62) and make it rounded with nearest integer 65, thus the final sample size for control group is 65 and experimental group is 65.

Sample Size for Unmatched Case-Control Study		
For:	Two-sided confidence level (1-alpha)	95
	Power (% chance of detecting)	80
	Ratio of Controls to Cases	1
	Hypothetical proportion of controls with exposure	47
	Hypothetical proportion of cases with exposure:	73
	Least extreme Odds Ratio to be detected:	3.05

Contd...

	Kelsey	Fleiss	Fleiss with CC
Sample Size – Cases	56	55	62
Sample Size – Controls	56	55	62
Total sample size:	112	110	124

References

Source: Kelsey et al., Methods in Observational Epidemiology 2nd Edition, Table 12-15

Fleiss, Statistical Methods for Rates and Proportions, formulas 3.18 and 3.19

CC = continuity correction

Results are rounded up to the nearest integer.

Print from the browser menu or select, copy, and paste to other programs.

Example 3: Sample Size for Comparing Two Means: This module calculates sample sizes for comparing two means. Enter a desired confidence interval (95%), power (80%), ratio of sample size of group 2 (e.g., 37.14 ± 9.21) to group 1 (e.g., 31.76 ± 9.76), mean (or mean difference) and standard deviation (or variance), and the sample size will be calculated for each group. The calculated number must be made rounded with nearest integer, i.e., 50 and 50 and total of 100.

Sample Size For Comparing Two Means			
Input Data			
Confidence Interval (2-sided)	95%		
Power	80%		
Ratio of sample size (Group 2/Group 1)	1		
	Group 1	Group 2	Difference*
Mean	31.76	37.14	-5.38
Standard deviation	9.76	9.21	
Variance	95.2576	84.8241	
Sample size of Group 1	49		
Sample size of Group 2	49		
Total sample size	98		
*Difference between the means			

Source: Results from OpenEpi, Version 3, open source calculator – SS Mean Print from the browser with ctrl-P or select text to copy and paste to other programs.

Example 4: Sample Size for Cross sectional, Cohort and Randomized Control Trials: This module calculates sample size for unmatched cross sectional and cohort studies, including clinical trials. Enter the desired confidence level (always 95), power (always 80), ratio of exposed to unexposed samples (always 1), and a hypothetical percentage of outcome among the controls (based on previous literature). Then enter one of four parameters to be detected, and the others will be calculated. Results are presented using methods of Kelsey, Fleiss, and Fleiss with correction factor. Here the final sample size is for control and study group is 466 and make it rounded, i.e., 470 for control and study group, and the final sample size is 940.

Sample Size for Cross Sectional, Cohort and Randomized Clinical Trials	
Two-sided significance level (1-alpha):	95
Power (1-beta, % chance of detecting):	80
Ratio of sample size, Unexposed/Exposed:	1
Percent of Unexposed with Outcome:	10
Percent of Exposed with Outcome:	5
Odds Ratio:	0.47
Risk/Prevalence Ratio:	0.5
Risk/Prevalence difference:	-5

Contd...

	Kelsey	Fleiss	Fleiss with CC
Sample Size - Exposed	429	428	466
Sample Size-Non exposed	429	428	466
Total sample size:	858	856	932
References			
Source: Kelsey et al., Methods in Observational Epidemiology 2nd Edition, Table 12-15			
Fleiss, Statistical Methods for Rates and Proportions, formulas 3.18 & 3.19			
CC = continuity correction			
Results are rounded up to the nearest integer.			
Print from the browser menu or select, copy, and paste to other programs.			

Online software for sample size calculation:

Rollin Brant's sample size calculators: <http://newton.stat.ubc.ca/~rollin/stats/ssize/>

MGH Biostatistics Centre, USA: http://hedwig.mgh.harvard.edu/sample_size/size.html

University of IOWA's Java-applets: <http://www.stat.uiowa.edu/~rlenth/Power/index.html>

University of Texas: For many other online statistical tools: <http://www.utexas.edu/its/rc/world/stat/online.html>

Factors Affecting Sample Size

Sample size and its adequacy plays a significant role to ensure the scientific nature of research work, as well as generalization of its findings. However, there are numerous factors that are influencing the sample size, which are described as follows.

Factors influencing sample size:

- Design effect
 - Availability of resources
 - Nature of study
 - Sampling methods
 - Homogeneity of sample
 - Effect between variables
 - Attrition rate
 - Non-compliance
 - Degree of confidence
- **Design effect:** The study design influence the sample size, wherein the number of variables under study and the type of design is an important factor for the determination of sample size. Quantitative research needs larger sample size while qualitative research needs smaller size, the sample variation is considered for estimating the sample size. The longitudinal studies and experimental studies may require more samples than descriptive studies as the experimental and longitudinal studies does have higher attrition rate (dropout).
 - **Availability of resources:** Though the larger sample size and random selection of samples are produce highly precise and accurate results; however this process requires cost, time, data collectors and other logistic resources.
 - **Nature of study:** The sample size depends on the type of study carried out wherein the descriptive studies and qualitative studies require smaller sample size and interestingly case studies can be carried out even with one participant, whereas the experimental and extensive type of studies may require larger samples.
 - **Methods of sampling:** Though the samples are small but selected scientifically proved to be better than the larger samples that are selected by nonprobability methods.
 - **Homogeneity of sample:** The sample which is similar to the population and possesses the parallel characteristics though the samples are small still it is sufficient and scientific one than the selection of samples who are heterogeneous in nature.
 - **Effect between variables:** If the relationship between independent variable and dependent variable is strong and tested or proved statistically then the small sample size is enough for the research. If the causal relationship is not established between variables under study, then it requires larger sample size.
 - **Attrition rate:** The attrition rate (dropout) is a common problem that the researcher faces during the course of study. This is commonly noticed in experimental studies as well as in longitudinal type of studies where the participants need to be contacted and followed for a long period. In case of determining sample size the researcher should keep or add 5–10% of attrition rate in the final calculated sample size. If the calculated sample size is 100 and add 10% of attrition and make it 110 as sample size, here even though 10 participants dropout from study still the researcher will be able to have a predetermined samples for study.

- **Degree of confidence:** Confidence is nothing but the researcher’s wish for keeping the percentage of error on sampling results. Here, the higher the confidence level higher the sample size and lower the degree of confidence lower the sample size. Usually, 0.05 or 95% is the degree of confidence, the researcher prefers which indicates the

chance of error is only 5% and the results are 95% precise and correct; which is shown below.

- **Noncompliance:** It will increase the variation in the results and hence sample size has to be adjusted to the degree of expected noncompliance.

Example: Sample size at different confidence levels

Sample Size for Frequency in a Population			
Population size (for finite population correction factor or fpc) (N):			150000
Hypothesized % frequency of outcome factor in the population (p):			50% +/-5
Confidence limits as % of 100(absolute +/- %) (d):			5%
Design effect (for cluster surveys-DEFF):			1
Sample Size (n) for Various Confidence Levels			
Confidence Level (%)	Sample Size		
95%	384		
80%	165		
90%	271		
97%	470		
99%	661		
99.9%	1075		
99.99%	1500		
Equation			
Sample size n = [DEFF*Np(1-p)]/[(d2/Z21-α/2*(N-1) + p*(1-p)]			

SAMPLING ERROR

Sampling error is the deviation of the selected samples from the true characteristics, traits, features, behaviors, qualities of the entire population. It is a systematic error in sampling procedures that leads to a distortion in the results of the study, and it may be related to improper sampling techniques and samples are not a representative of the entire population. Following are the reasons for possible sampling bias:

- **Improper sampling procedures** (improper/faulty sampling technique): If the researcher follows improper sampling methods which are not in conjunction with research design leads to sampling bias and lack of representativeness.
- **Nonresponsiveness** (no or incomplete response from participants): Though the researcher follows scientific sampling techniques, the participants provide only incomplete response or do not respond properly to the questions/interview, then there will be a chance for sampling bias.
- **Studying volunteers only** (Sampling of registered patients only): Studying the volunteers only will not impose the true representativeness. For example, a study on the opinion of government schemes through telephonic

interview will not have true participation of population. Here the samples who do not have telephone is excluded, thus, there will be sampling bias.

- **Seasonal bias** (does not provide true representativeness of population): The data collected during a specific season will not be applicable to other seasons, which leads to sampling bias.

Types of Sampling Error

Faulty selection of sample, substitution of sample, faulty demarcation of sample units, response error, errors in coverage, compiling error and publication error

Measures to Minimize Sampling Bias

The best way to avoid sampling bias is to collect the data/test of each participant of the entire population. Unfortunately, it cannot be practiced because of varied reasons. The sampling bias is controlled and avoided mainly by:

- Employ random method of sample selection from the population.
- Determine the sample size by scientific methods rather than arbitrary way.

- Try to have larger sample size for the study.
- Avoid the use of nonprobability (non scientific) sampling methods like purposive and convenient sampling techniques, etc.
- Ensure that the sampling frame is properly devised without leaving any of the elements or units, etc.

PROBLEMS OF SAMPLING

Research is itself a problem oriented activity to find the answers to questions. Similarly, in sampling process there are numerous technical and practical problems. They are mentioned below with examples:

Sampling error: It is nothing but deviation of samples from the normal characteristics or traits of the entire population. It usually occurs when the samples are selected heterogeneous which is not homogeneous with the population characteristics.

Inadequate sample representativeness: If the samples are heterogeneous and do not possess the homogeneous characteristics of population, the results cannot be generalized to the universe. Making the samples with homogeneous characteristics is difficult, however, researcher uses stratification and randomization to make homogeneity of sample characteristics.

Problem in estimation of sample size: Like medical research, nursing research also lacks adequate population parameters

for determining the sample size. Research carried out with an undetermined sample size cannot be generalized with others.

Awareness of the researcher: If the researcher does not have awareness and training on sampling process and selection of samples based on probability and nonprobability methods, the sampling process will be a failure, ultimately leading to faulty and distorted results.

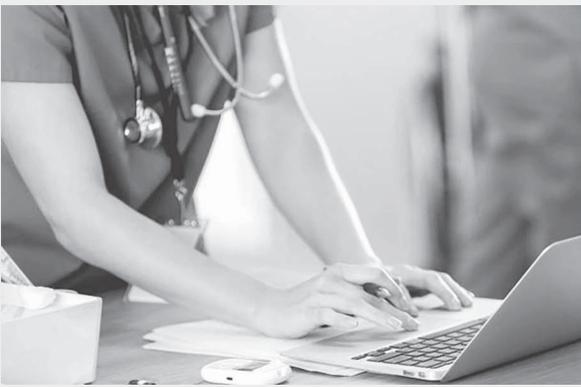
Lack of resources and time: Availability of resources for sampling frame and sampling procedure and the time factor is another crucial factor which influences the sampling in research. It will be difficult to carry out research without adequate resources and availability of sufficient time.

Critical review of sampling:

- Is the sample characteristic completely described?
- Is the sample a true representative of the population?
- Is the inclusion/exclusion criteria for the sample clearly identified?
- Is there a possibility to replicate the study on larger population?
- Is the sample selection method used appropriate as per study design?
- Is there a possibility of bias in sampling, and if so, is there safeguard method to resolve the bias?
- Is the sample size adequate and scientific?
- Are the rights of samples and ethical aspects considered during sampling?

SUMMARY

- Sampling is a process of choosing the units/elements/samples from a given population by using a probability or non-probability sampling method to represent the whole population of study. A sample is a subset of a population selected for a research study. The population can be categorized into two types: The target population and the accessible population. Importantly, the sample must have true representativeness, homogenous, free from sampling error, and must have optimal size that are calculated scientifically by using power analysis, statistical methods or computer-based software.
- The probability sampling is otherwise called quantitative sampling or scientific sampling, the common types are simple random, stratified random, cluster, multi-stage, multi-phase and systematic random sampling. The non-probability sampling is called qualitative sampling or non-scientific sampling, and the classifications are purposive sampling, convenient sampling, quota sampling, snow ball sampling and universal sampling. It is the prerogative of researcher to choose the sampling methods that are appropriate for conduct of study; however, probability sampling methods are more preferable for better generalization rather than the other types.



STUDENT ASSIGNMENT

LONG ANSWER QUESTIONS

1. Define sampling and explain in detail on probability sampling methods with suitable examples.
2. Enlist the characteristics of a good sample, and elaborate in detail on nonprobability sampling methods.
3. Define sample and elaborate in detail on sampling procedure with appropriate examples.
4. Enlist the inputs required for calculating sample size, and explain in detail with appropriate examples.
5. Discuss various software used for sample size calculation and sampling procedures.
6. How do inclusion and exclusion criteria contribute to increasing the strength of evidence provided by the sampling strategy of a research study?
7. Why is it important for a researcher to use power analysis to calculate sample size?
8. Write the classification of scientific sampling and elaborate in detail commonly used sampling methods in community based research.
9. Explain the problems faced by researchers on sampling in research studies and its preventive measures.

SHORT ANSWER QUESTIONS

1. Compare and contrast accessible population and target population.
2. Define sample and write the sampling process with example.
3. Write a note on simple random sampling.
4. Write in brief the factors to be considered while determining sample size.
5. What is the difference between probability and nonprobability sampling?
6. Write a note on sampling methods in qualitative research.
7. Enumerate the advantages and disadvantages of probability sampling.
8. Enumerate the advantages and disadvantages of nonprobability sampling.
9. Differentiate multi stage sampling and multiphase sampling method.
10. Enumerate the determinants of sample size calculation.
11. Write a short note on power analysis and its significance in research.
12. Write a short note on how to minimize the sampling bias in research.
13. What role does the sample size play in the accuracy of a statistical inference?
14. Why is bias more important than variance in research?
15. Write the importance of cluster sampling in epidemiological research.
16. Write a short note on universal sampling and its significance in research work.
17. Write in brief the difficulties faced by the researcher in sampling procedure.
18. Write in brief on snowball sampling and its significance in medical research.
19. Write a short note on factors influencing sample size.

MULTIPLE CHOICE QUESTIONS

1. Which of the following is an example of probability sampling technique?
 - a. Sequential sampling
 - b. Consecutive sampling
 - c. Stratified sampling
 - d. Cluster sampling
2. The process of drawing a sample from a population is known as:
 - a. Sampling
 - b. Census
 - c. Survey
 - d. Universe
3. If the population on which a study is conducted is such that all its units can be reached, it is termed:
 - a. Target population
 - b. Accessible population
 - c. Homogeneous population
 - d. Heterogeneous population
4. Which of the following would generally require the largest sample size?
 - a. Cluster sampling
 - b. Simple random sampling
 - c. Systematic sampling
 - d. Proportional stratified sampling
5. Data collection about everyone or everything in group or population has the advantage of accuracy and detail:
 - a. Census
 - b. Survey
 - c. Probability sampling
 - d. Cluster sampling
6. Participants who are available, volunteer or can be easily recruited, used in the sampling method are called:
 - a. Simple random sampling
 - b. Cluster sampling
 - c. Systematic sampling
 - d. Convenience sampling
7. A sampling method which involves a random start and then proceeds with the selection of every kth element from then onwards is called:
 - a. Simple random sampling
 - b. Stratified random sampling
 - c. Systematic sampling
 - d. Snowball sampling
8. When the researcher is unable to locate the participants, he is likely to use:
 - a. Convenience sampling
 - b. Snowball sampling
 - c. Cluster sampling
 - d. Purposive sampling
9. Which of the following is not a characteristic of a good sample?
 - a. True representativeness
 - b. Free from sampling error
 - c. Minimal sample size
 - d. Free from sampling bias
10. When sampling is done sequentially across two or more hierarchical levels the sampling method is called:
 - a. Stratified sampling
 - b. Multistage sampling
 - c. Multiphase sampling
 - d. Systematic sampling
11. The commonly used level of confidence for determining the sample size is:
 - a. 80%
 - b. 90%
 - c. 95%
 - d. 99%

CRITICAL THINKING QUESTIONS

Activity A: Write the definition of following terms and explain the differences between each set of terms:

1. Sample: -----
 Population: -----
 Differences: -----
2. Target population: -----
 Accessible population: -----
 Differences: -----
3. Inclusion criteria: -----
 Exclusion criteria: -----
 Differences: -----

Activity B: Write the difference between probability and nonprobability sampling methods.

Probability sampling	Nonprobability sampling
1. 2. 3. 4.	

Activity C: Identify the category of sampling method for each of the sampling techniques mentioned below, and indicate the abbreviations from the key provided:

Key: P = Probability sampling N = Nonprobability sampling

1. ----- Convenient sampling
2. ----- Cluster sampling
3. ----- Snowball sampling
4. ----- Quota sampling
5. ----- Multistage sampling

Activity D: Using critical thinking, indicate whether the following statements are **true or false**:

1. ----- Probability sampling enables to have generalization of findings to the larger population.
2. ----- Nonprobability sampling techniques are usually time-consuming and have logistical difficulties.
3. ----- Probability sampling has the greater risk of sampling bias, and its lacks true representativeness.
4. ----- Multiphase sampling uses a random selection procedure for obtaining sample subjects.
5. ----- Purposive sampling technique is commonly used in qualitative research studies.

ANSWER KEY

1. d 2. a 3. b 4. a 5. a 6. d 7. c 8. b 9. c 10. b
 11. c
-