JAYPEE

Basics in Epidemiology & Biostatistics

Waqar H Kazmi
Farida Habib Khan

Foreword
Waris Qidwai

Contents

CHAPTER

3

Sampling Procedure

POPULATION

A major purpose of the research is to infer or generalize findings from a sample to a target population. Population is the term statisticians use to describe a large set or collection of items that have something in common (i.e. all pregnant women, all pregnant women in third trimester, all anemic pregnant women in third trimester, etc.). Target population is a group about which researcher aims to draw conclusion. In medicine, population generally refers to patients or other living organisms, but the term can also be used to denote collections of inanimate objects, such as autopsy reports, X-ray reports, or birth certificates.

Figure 3.1 shows relationship among target population, study population and sample. Target population is a population of ultimate clinical interest about which researcher aims to draw a conclusion. On account of the cost and other practical issues, the entire target population cannot be studied. Study population is a subset of target population that can be studied. Samples are subsets of study populations investigated in clinical research because often not every individual in a study population can be measured.

A "sample" is a subset of population with all its inherent qualities. Studies are conducted on samples but inference is made about target population. That is why it is important that the sample should be a true representative of the target population. Hence, the selected elements should be properly approached, recruited in the study and interviewed. Thus, selection of sample is critical as, otherwise, the research findings might not be valid.

It is vital to have a clear understanding of the terms population and sample; these two terms must not be used interchangeably.

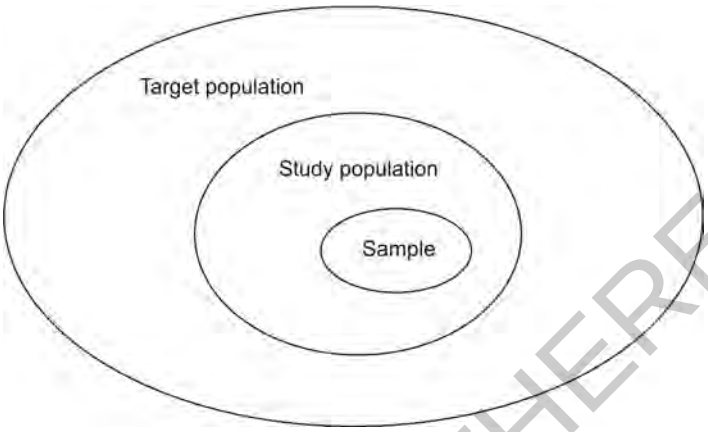


Figure 3.1 Relationship among target population, study population and sample

REASONS FOR SAMPLING

It is reasonable and practical to collect information from sample rather than the whole population. Below are the reasons listed for sampling:

- Samples can be studied more quickly than population
- A study of a sample is less expensive than that of an entire population
- A study of a population is impossible in most situations
- Samples are more often accurate than results based on a population
- If samples are properly selected, probability methods can be used to estimate the error in the resulting statistics
- Samples can be selected to reduce heterogeneity.

SAMPLING TECHNIQUES

Broadly, there are two types of sampling techniques (**Table 3.1**):

1. Probability sampling techniques.
2. Nonprobability sampling techniques.

In a probability sampling technique, each participant in a study population has an equal (or at least a known) chance of being selected. The method protects the research from bias and ensures

Table 3.1: Different sampling techniques

Probability sampling	*Nonprobability sampling*
1. Simple random sampling	1. Consecutive sampling
2. Systematic random sampling	2. Convenience sampling
3. Cluster sampling	3. Purposive sampling
4. Stratified random sampling	4. Quota sampling
	5. Snowball sampling

that the sample is a true representative of a population. Importantly, it helps a researcher to make meaningful statistical estimation while analyzing the results of the research. In a nonprobability technique, each participant does not have an equal chance of being selected.

Probability Sampling Techniques

Simple Random Sampling

Simple random sampling is the simplest method of probability sampling. In this type of sampling technique each individual within the study frame has an equal chance of inclusion in the sample. A common example is sometimes called the 'lottery method' and illustrated in **Figure 3.2**.

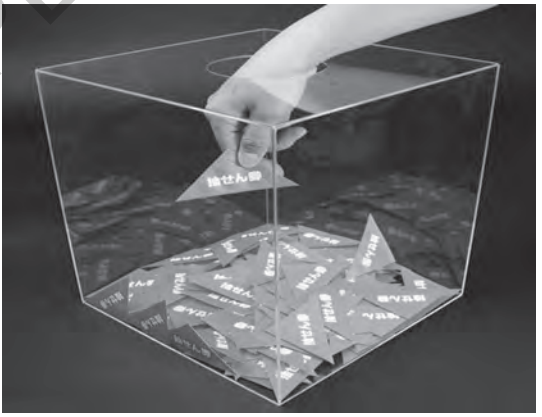


Figure 3.2 Lottery sampling technique

For example in a recruitment for a study there are 100 participants available, of these 25 have to be selected (sample size). The participants to be recruited in the study will be selected randomly by drawing a chit bearing the names/ID number of the 100 individuals. Each individual in the study frame has an equal probability of being selected for the study (i.e. when the first participant is to be selected the probability is 1/100 for all participants, for second participant the probability is 1/99 for all participants, for third participant the probability is 1/98 for all participants and so on). Thus each participant has an equal probability of being selected for the study.

The recommended way to select a simple random sample is to use a table of random numbers, or a computer-generated list of random numbers. For this approach each participants should have an identification number (ID), and a list of ID numbers called a "sampling frame".

The steps of simple random sampling are as follows:

- Prepare the sampling frame (assign a number to each element) of the whole population [Participants are numbered from 1 to 100].
- Determine the sample size [Estimated sample size is 25]
- Randomly select the element [Any 25 numbers are picked from 1 to 100]

OR

- If using computer generated lists to randomly select the participant
  - Enter lowest ID number (i.e. in this case 001)
  - Enter highest ID number (i.e. in this case 100)
  - Enter the estimated sample size as 25
  - Computer generated randomization software will generate a table of randomly selected participants/ID number (**Fig. 3.3**).

Systematic Random Sampling

In systematic sampling technique study participants are selected at regular intervals using a sampling frame (**Fig. 3.4**).

Just estimate the population size (N) and calculate the required sample size (n).

Now divide population size by sample size, i.e. N/n. This will give you the kth number (sampling interval). In the above study example, the number of individuals were 100 and the required sample size

001	002	003	**004**	005	006	007	008	009	**010**
011	012	013	014	**015**	016	017	018	**019**	020
021	022	023	**024**	025	026	**027**	028	029	030
031	032	**033**	034	035	**036**	037	038	039	040
041	042	**043**	044	045	**046**	047	048	049	**050**
051	**052**	053	054	**055**	056	057	**058**	059	060
061	**062**	063	064	065	066	**067**	068	069	**070**
071	072	073	074	**075**	076	077	078	079	080
081	**082**	083	084	085	086	087	**088**	089	090
091	092	093	094	095	**096**	097	098	**099**	100

Figure 3.3 Random selection of 25 participants represented by bold

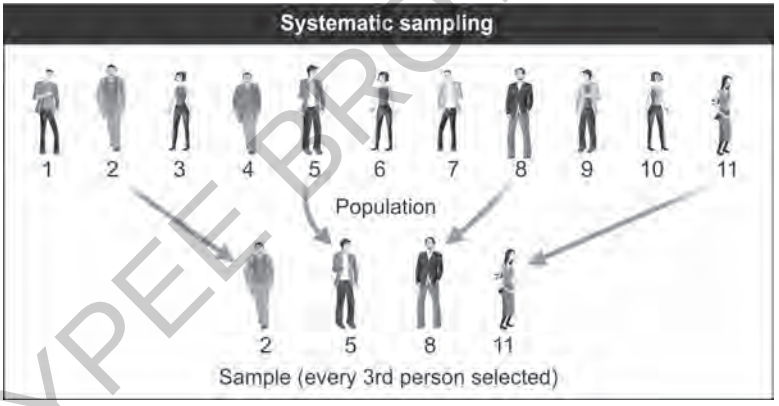


Figure 3.4 Systematic random sampling (Every 3rd selected)

was 25, hence 100/25 would be 4 and so every 4th X-ray should be selected.

First element is selected randomly from 1st to kth element (i.e. in above example from 1 to 4). Then every kth element is selected till the researcher achieves the required sample size. For example in **Figure 3.5** second individual in the study population is selected at random and then every fourth individual is selected (i.e. 6th, 10th, 14th, etc.).

001	**002**	003	004	005	**006**	007	008	009	**010**
011	012	013	**014**	015	016	017	**018**	019	020
021	**022**	023	024	025	**026**	027	028	029	**030**
031	032	033	**034**	035	036	037	**038**	039	040
041	**042**	043	044	045	**046**	047	048	049	**050**
051	052	053	**054**	055	056	057	**058**	059	060
061	**062**	063	064	065	**066**	067	068	069	**070**
071	072	073	**074**	075	076	077	**078**	079	080
081	**082**	083	084	085	**086**	087	088	089	**090**
091	092	093	**094**	095	096	097	**098**	099	100

Figure 3.5 Systematic random sampling (Every 4th selected)

Stratified Random Sampling

Stratified random sampling is a sampling technique that divides the population into various sub-groups, i.e. based on gender, age groups, ethnicity, etc. (**Fig. 3.6**) and then any of the random sampling technique is employed to randomly select participants from each group (**Fig. 3.7**). Suppose a population consisted of more females than males. In spite of the random technique employed, females will constitute a greater proportion of sample than males. Such problem could be overcome by utilizing stratified random sampling.

For example, a population consisted of 60 individuals and the researcher wants to select equal representation of all the strata based on ethnicity. Firstly, the population is stratified according to ethnicity (i.e. Caucasians, African-American and Hispanic-American). There are 30 Caucasians, 20 African-American and 10 Hispanic-American. As the researcher wants to select 15 participants thus each strata must constitute 5 participants. Finally, 5 participants are randomly selected from each strata.

One of the main purposes of stratified sampling is to compare different strata (e.g. males with females, different age groups, etc.) which may not be possible with simple random sampling alone.

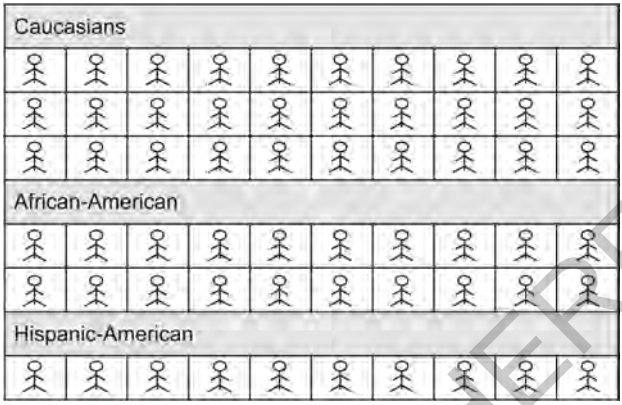


Figure 3.6 Stratified random sampling technique (Individuals in each strata)

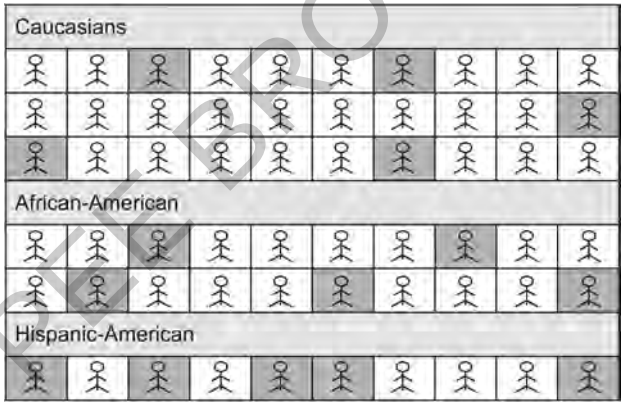
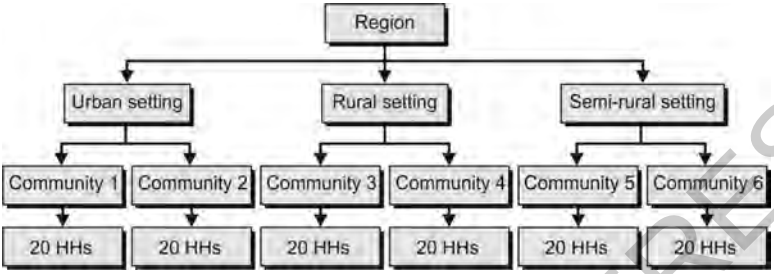


Figure 3.7 Stratified random sampling technique (Participants selected from each strata represented by bold headed stickman)

Cluster Sampling

In clustered sampling technique sub-group of population is used as a sampling unit instead of individuals. It is a probability sampling technique, employed when the researcher aims to select participants from a large geographical area i.e. country, province, state or city (**Flow chart 3.1**). Suppose the city of Karachi consisted of 18 towns and each town consisted of 10 union councils. Initially, 5 towns are

Flow chart 3.1 Cluster random sampling technique



selected by either of the random technique methods. Later, from each town 4 union councils are randomly selected. Finally, from 20 union councils houses are randomly selected. Thus in this type of sampling method households are the sampling unit instead of individual residents.

Nonprobability Sampling Techniques

Consecutive Sampling

It involves sequential selection of all accessible eligible participants that meets the selection criteria. If the study participants are selected in a *consecutive manner,* they might be inherently similar to eligible participants that meets inclusion and exclusion criteria for the study. Suppose, a strategy is devised to recruit 100 patients (the estimated sample size) for a study that satisfies the selection criteria and seen in a Nephrology clinic from Monday to Friday between 9.00 am to 12.00 pm. The first 100 patients who meets the eligibility criteria and attend the outpatient clinic during these days and timings will be recruited in the study. This method is best among nonprobability sampling techniques as it minimizes selection bias by recruiting complete accessible population within the parameters of estimate sample size and selection criteria.

Convenience Sampling

Convenience sampling is presumed to be the most commonly used technique in clinical research. It involves the selection of subjects that are conveniently accessible to the researcher. Suppose, a

researcher working as a professor of nephrology aims to identify the communication skills of postgraduate trainees. The description of "20 postgraduate trainees" is assuredly 20 postgraduate trainees in nephrology ward who volunteered for this study. The participants were selected on account of investigator feasibility to recruit these participants, as working in the nephrology ward. The method is easy, fast and less expensive but not the representative of a larger overall population thus introducing selection bias in the research.

Purposive Sampling

Purposive sampling is also called *judgmental sampling*. The technique is criticized for introducing selection bias in the research as the researcher recruit participants based over pre-existing belief that certain subjects will be more likely benefit, compliant or respond in certain way. Thus, the researcher selects study participants with a 'particular purpose' in mind.

For example, if the researcher wants to check the hypothesis that Pakistani females have better knowledge regarding medical research than American females. Selection of Pakistani females medical students (a group that has better understanding of medical research than other women) and American females who came to the market for shopping were selected. As the two groups are noncomparable, evidently Pakistani females will display a better knowledge regarding medical research which might not be the case. Such deviation from truth is on account of purposeful sampling.

Similarly, while conducting a knowledge survey on the mode of transmission of HIV; selecting participants that are relatives of AIDS patients will demonstrate an excellent knowledge regarding transmission modes of HIV. Evidently the selection of study participants was biased as the sample was not the true representative of the target population.

Snowball Sampling

Snowball sampling method is employed when study participants are difficult to identify, access or locate. The method is commonly employed to recruit participants from hard to reach group (i.e. sex workers, IV drug users, etc.). The sample is built through chain referrals. Suppose, you are investigating the knowledge about

Flow chart 3.2 Snowball sampling technique

contraception among female sex workers. Female sex workers are hard to identify as they are not registered in Pakistan. Thus one female sex worker will be identified and recruited in the study. Later, the participant will be requested to recommend more sex workers. Each of these will recommend more sex workers. In this way, a sizeable sample may be obtained even for hard to reach group (**Flow chart 3.2).**

Quota Sampling

Quota sampling is a nonprobability sampling method that ensured a certain number of study participants from different subgroups constitute the sample so that all these characteristics are represented. Suppose you aim to identify the quality of life among dialysis patients but you think that socioeconomic status has a strong affect on quality of life in these patients. Thus you decide to include 25% of respondents from each socioeconomic groups (i.e. upper, middle, lower middle and lower). If the estimated sample size is 200, each socioeconomic group will include 50 participants. Thus initially a population is divided into different strata and then any nonprobability sampling technique will be applied to select participants.

BIBLIOGRAPHY

1. Beth Dawson-Saunders, Robert G Trapp. Basic and Clinical Biostatistics, 1989.

2. Hulley SB, Newman TB. Choosing the study subjects: specification, sampling, and recruitment. In: Hulley SB, Cummings SR (Eds). Designing clinical research. Philadelphia, PA: Lippincott Williams and Wilkins; 2007.pp. 27-36.
3. Kuzma JW, Bohnenblust SE (Eds). Populations and samples. Basic statistics for the health sciences. Boston: McGraw Hill; 2005.pp. 16-28.
4. Last John M. A dictionary on Epidemiology. Oxford University Press 1983.
5. Morris JN. Uses of Epidemiology. ELBS 3rd edn, 1983.

Basics in Epidemiology & Biostatistics

Salient Features

- Contains text in a very simple and easy-to-understand language
- Introduces the medical/dental students, postgraduates, researchers and clinicians to the study of statistics applied to medicine
- Multiple clinical and nonclinical examples have been used so that the reader can understand the basic concepts better
- Presents the entire text in a pictorial and tabular form
- Incorporates the experiences of the authors associated with teaching epidemiology and biostatistics to develop a comprehensive text covering the traditional topics of biostatistics and epidemiology
- Helpful for undergraduates, postgraduates, clinicians and basic science doctors.

Waqar H Kazmi is Principal, Professor of Nephrology and Director Research, Karachi Medical and Dental College/Abbasi Shaheed Hospital, Karachi, Pakistan. He graduated (1977) from Dow Medical College, Karachi and received his earlier nephrology training from Pakistan and then worked as a Fellow at New England Medical Center, Tufts Medical University, USA. He received Masters in Clinical Research from Tufts Medical University, Boston, USA. He has more than 50 international papers to his credit and is very actively involved in teaching Epidemiology and Biostatistics including Nephrology.

Farida Habib Khan is Professor of Community Medicine, Princess Nora Bint Abdulrahman University, Riyadh, Kingdom of Saudi Arabia. She is a Co-author of *Textbook on Community Medicine*.

Shelving Recommendations
PREVENTIVE & SOCIAL MEDICINE
COMMUNITY MEDICINE
BIOSTATISTICS

ISBN 978-93-5152-631-5

9 789351 526315

Available at all medical bookstores or buy online at www.jaypeebrothers.com

JAYPEE BROTHERS
Medical Publishers (P) Ltd.
www.jaypeebrothers.com

Join us on facebook.com/JaypeeMedicalPublishers