Educational Review

A beginner's Guide to Sampling Methods in Medical Research

Moslem Basti [1] (ID), Farzan Madadizadeh [2]* (ID)

Abstract

Background: Sampling methods are one of the main components of each research. Familiarity with a variety of sampling methods is essential for researchers.

Objective: The main purpose of this study was to teach different probabilistic and non-probabilistic sampling methods to improve the knowledge of researchers in conducting more accurate research.

Methods: In this tutorial article, useful information about each sampling method, as well as how to properly use each method and its strengths and weaknesses are provided.

Results: Five cases of probabilistic sampling methods and four cases of non-probabilistic sampling methods that are common are mentioned. Probabilistic sampling included simple random sampling, stratified random sampling, cluster sampling, systematic random sampling, and multi-stage random sampling. In addition to introducing each method, its strengths and weaknesses are also mentioned.

Conclusion: Probabilistic sampling methods despite limiting assumptions provide more reliable results. Therefore, if it is possible, researchers should use probabilistic sampling methods to increase the accuracy of the study.

Keywords: Sampling, Probability Samples, Sampling Study

Introduction

It is costly, time consuming and beyond the ability of a researcher to collect information from the whole population; therefore, the best solution is to follow efficient sampling procedures. Sampling has several advantages and the most important one is saving time and money. It also requires less equipment and facilities. It can be said that the most important problem of sampling is the possibility of bias [1].

To achieve an ideal and suitable sample, the following points are need to be considered:

1. A clear definition of the statistical population

2. The format and framework of sampling

3- Characteristics of the statistical population to generalize the findings

4- Logical sample size

5- Adoption of appropriate sampling techniques

6- The rate of non-response according to the sample size [2].

Before introducing the types of probabilistic sampling, statistical population and the sample are elaborated below.

Statistical population is a term used by epidemiologists and statisticians to describe a large set of objects or people who share a common characteristic.

Target population: This informal term as a whole means a subset of the statistical population considered only for study, but sometimes in the literature this term is mistakenly used instead of the term statistical population [3].

A sample is a subset of the statistical population, which is selected to represent the statistical population.

External validity: External validity means the ability to generalize the results of the sample to the statistical population from which the sample was taken. This criterion is highly sensitive to participant selection [3]. In addition to the concept of external validity, there is another concept called internal validity, indicating that the results and relationships discovered are reliable and are not affected by confounding factors.

Researchers choose the type of sampling method based on the purpose of their research and the capabilities of the statistical population. If the purpose of sampling is to estimate population parameters, it is necessary to use probabilistic sampling [3].

In general, sampling techniques are divided into probabilistic and non-probabilistic methods. Given

sampling is an important issue in medical research, it is elaborated in plain English. Basic statistics terms are initially explained followed by four types of probabilistic sampling, namely, simple random sampling, stratified random sampling, cluster sampling, and Systematic sampling. In the following, the subject of non-probabilistic sampling and its types are dealt with. Figure 1 reveals the types of sampling methods and Table 1 shows the strengths and weaknesses of each method.
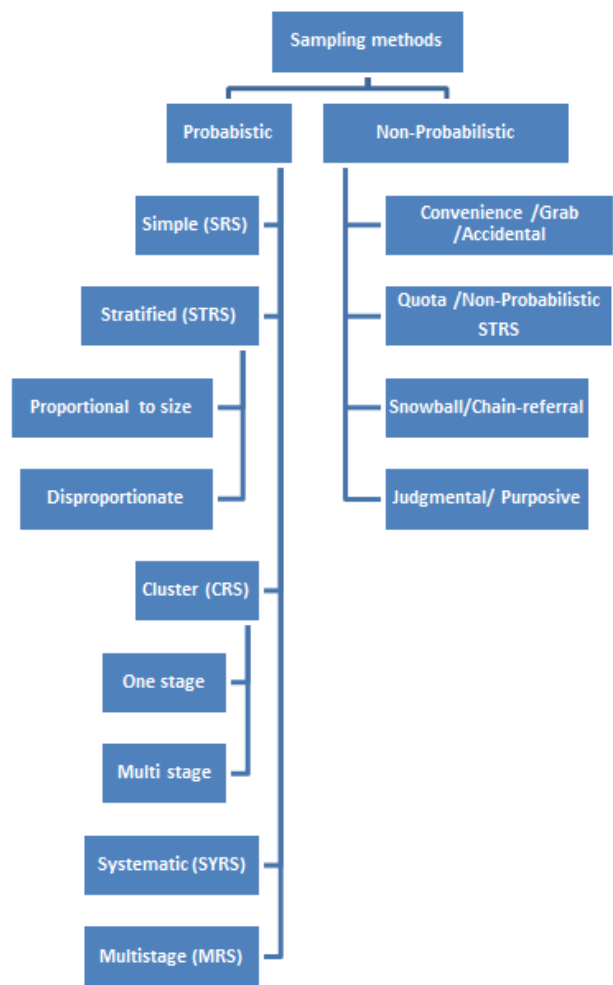
Figure 1. All types of sampling methods

1- Probabilistic Sampling methods

As a reliable sampling technique, it offers equal chance for individuals and objects to represent the whole population. [4]. When the base population is homogeneous, the chances of selecting each member in the sample are quite high [1].

1-1 Simple random sampling (SRS)

In this sampling method, each member of the statistical population has an equal chance to enter the sample or be selected in the study. The correct method recommended for selecting a simple random sample includes the use of a random number table or generated random numbers using statistical software [1, 4].

There are two types of simple random sampling [1]:

1. Simple random sampling with replacement (SRS+R)

2. Simple random sampling without replacement (SRS-R)

In the SRS with replacement, each member has the chance to be re-elected more than once; while in random sampling without replacement, each member has a chance to be selected only once.

Example: The researcher intends to study the clinical outcomes of COVID-19 in patients hospitalized in Shiraz hospitals, Iran. Therefore, the target population is patients hospitalized with COVID-19. The number of these patients is 5,000 and the sample size obtained is 500 patients. Therefore, the chance of choosing each patient is one-tenth. If each patient in the population is returned to the statistical population, they will have a chance to be re-selected, which is called SRS with replacement (SRS+R). Also, if any patient in the sample is removed from the statistical population after selection, the patient will not have a chance to be re-selected, which is called simple random sampling without replacement (SRS-R).

Advantages of SRS method include minimization of the selection bias and simplification of the analysis. It is easy to estimate the accuracy obtained from the sample, but care needs to be taken to avoid sampling error; since random selection may lead to a sample that does not represent the statistical population. Moreover, this sampling method is not suitable for research aimed at subgroup analysis and may provide biased estimators [4].

1-2 Stratified random sampling (STRS)

Stratified random sampling is a probabilistic sampling method in which the population is first divided into independent subgroups (strata) that have members with identical characteristics in each strata, and then a random sample is selected from each strata [1, 5].

This method is suitable in studies where the researcher intends to compare different subgroups. Therefore, in this type of classification, the quantities based on which they are classified (such as age group) should be related to the desired characteristics and attributes to ensure sampling efficiency. In addition, it is more dynamic than simple random sampling; since it requires fewer samples [4].

STRS advantages

1- This method is more efficient than other sampling methods,

2- It is easier to use

3- It needs a smaller sample size [1].

Example: Suppose the heterogeneous statistical population of a university students. Sampling can be performed based on the faculty as strata. If sampling is done as a proportion of the of students of each faculty from all university students, then this type of stratified sampling is called probability proportional to size (PPS) sampling; otherwise, sampling is called disproportionate.

1-3 Cluster random sampling (CRS)

Cluster random sampling is the result of dividing a population into similar clusters and heterogeneous populations within each cluster and then randomly selecting several clusters or groups. This sampling method is a two-step process [4]. In cluster sampling, the unit of measurement is a group of individuals (or clusters) that are naturally formed. This method is used when there is a massive population. Therefore, to save time and money, we can focus on clusters instead of people [6].

CRS is divided into one-stage and multi-stage according to the method of implementation. In one-stage cluster sampling, clusters are selected only

randomly among all elements. Multi-stage cluster sampling is an extended form of single-stage cluster sampling, so that it is carried out in several stages.

Example: Suppose we want to examine the level of education of the employees of a city because this statistical population is heterogeneous and scattered. First, the employees of each organization are considered as a cluster, and then among all organizations, several organizations (clusters) are randomly selected. Several units are extracted from each organization and then from the organizational units, the sample is randomly selected from each unit.

1-4 Systematic random sampling (SYRS)

The main idea of SYRS is to regularly select a number of members on a list. Suppose we want to select names from a long list. A simple way is to select the appropriate distance (systematic sampling interval or sampling fraction) and select the names at equal intervals along the list. Systematic random sampling is often easier to perform than simple random sampling and also reduces the possibility of error [7]. This sampling method is more advanced than the simple sampling method and gives the researcher more detailed information about the statistical population.

Steps of sampling in SYRS method are as follows: selecting the statistical population, selecting the sample size, assigning a value to each member of the sample, determining appropriate distance between the sample (sampling fraction), choosing a random starting point, and finally starting the sampling systematically [1].

For example, suppose we want to systematically select a random sample with size 8 from a statistical population with size 24.

First all the population units are numbered, then the sampling fraction (SF) is calculated by dividing preset sample size (n) by the size of the population (N),(SF=8/24=3). In the next step, the number of the first unit of sample (random starting point) will

be randomly selected from 1 to 24, then the number of the next units is obtained by adding SF. For example, if the number of the first unit of the sample is 2, the number of the next units will be 5,8,11,14,17,20,23.

1-5 Multi-stage random sampling (MSRS)

Depending on the characteristics of the community and the purpose of the research, the desired sample is sometimes selected in several stages. In this method, a sample of n1 is selected from a community of N people. Then, a smaller group is selected from n1; hence two-stage, three-stage, ... and m stage can be define accordingly. This random sampling is suitable when the statistical population is too large [1].

Example: It may be necessary to sample the population of a city in a study. By considering multi-stage sampling, one can first select several regions from among urban areas through cluster sampling. Then, within each region select several neighborhoods by systematic random sampling method. Finally, through simple random sampling, one can select several building blocks and homes.

2- Non-probabilistic sampling (NPS) methods

In NPS, the chances of selecting people or objects are not the same. In this type of sampling, there is a selection bias because the selection of individuals by the researcher is conscious and arbitrary. Therefore, estimating the sample error will be difficult due to non-random selection of individuals and it is impossible to obtain an accurate estimate of the sample error. Moreover, the results obtained from the sample cannot be generalized to the total population, because the obtained sample does not represent the community. NPS is cheaper and easier than probabilistic sampling, but the results are of lower validity than probabilistic sampling [4, 8].

2-1 Convenience sampling (CS) (Grab /accidental / opportunity sampling)

Convenience sampling is a type of NPS in which the researcher studies the part of the statistical

population that is more accessible. The results of this type of study cannot be attributed to the statistical population; since the sample size is not a true representative of the population. It is also suitable for the initial pilot study *[4]*. In general, this type of sampling method is preferred by researchers, because it is easier and cheaper than other sampling techniques *[5]*, though it has some weaknesses, such as the selection bias. Therefore, samples do not represent the statistical population, and it is not recommended for descriptive and causal studies *[9]*.

2-2 Quota sampling (QS) (Non-probabilistic version of STRS)

In the QS, first the population is divided into separate subgroups, and then individuals are selected from the subgroups based on the proportion to size sampling. In this type of sampling, sample selection is non-random and often unreliable. Moreover, there is a selection bias error; since everyone in the community does not the same selection chance, making it the least popular sampling method. Quota sampling is usually used when time is short, there is no sampling framework, study budget is limited, and sampling accuracy is not a matter. In QS, it is possible to determine how many people are selected from each subgroup. The researcher places people in demographic groups based on age and gender variables. When the quota of a demographic group is saturated, sampling is completed*[4]*. In terms of strengths, it should be noted that in this type of sampling method, the sample can be controlled with certain characteristics. Selection bias is also its main weakness *[9]*.

2-3 Snowball sampling (SS) (Cold-calling, chain/chain-referral / referral sampling)

In statistical and epidemiological research, SS is a technique, in which the study group introduce other people, the study group grows like a rolling snowball. This sampling technique is often used in secret populations that are difficult to access, like drug addicts. This method has many limitations, because it does not have a specific framework and is time consuming *[9, 10]*.

2-4 Judgmental sampling (JS) (purposive/ authoritative sampling)

JS is also called purposive sampling and is selected non-randomly to achieve a specific goal. In this method, the researcher preferably uses people who have the necessary experience and knowledge in the field, so in JS, the researcher studies people who thinks are more suitable for study. This method is used when the number of people with a particular experience is small. It is also used in pilot studies. This type of sampling method has almost the same disadvantages as the convenience sampling method; therefore, the results are not of high validity *[1, 4]*. In terms of strength, it should be noted that JS has low cost, is easy, and is not time consuming, hence suitable for exploratory research projects. In this method, sampling is carried out usually by experienced researchers, and the results obtained in this method cannot be generalized to the statistical population*[1, 9]*.

Table 1: Pros and cons of sampling methods

Sampling methods	Pros	Cons
Probabilistic Sampling		
Simple random sampling	• Minimizing the selection bias • Simplification of the analysis • Providing representative sample of population • Consider equal chance of selection for every unit of population	• Not suitable for heterogeneous population • Providing full list of population units could be very difficult and impossible
Stratified random sampling	• Easy to use • Need a smaller sample size, • More accurate than other sampling methods	• Only used in population with exist underlying classification • Finding list of population units may very difficult and impossible • Risk of overlapping with common member in two strata
Cluster random sampling	• Saving time and money • Managing variability • Feasible method with no need to sampling frame	• Only is used in a massive, heterogeneous population with subgroup structure • Risk of selection bias and sampling error • requiring equality of size clusters, so it decreases precision
Systematic random sampling	• Reducing the possibility of error • More advanced than the simple sampling method • Providing more detailed information about the target population without full list of units.	• Risk of bias related to hidden periodic features
Multi-stage sampling	• Suitable when the statistical population is too large • Flexible	• Time-consuming, cost consuming • Not suitable for small population • High risk of subjectivity and arbitrariness • Requiring group-level information • Risk of missing data
Non-probabilistic Sampling		
Convenience sampling	• Suitable for the initial pilot study • Easier and cheaper than other sampling techniques	• Cannot be attributed to the target population • Selection bias • It is not recommended for descriptive and causal studies • Does not provide representative sample • Hard to replicate results • High risk of sampling error • Low level of credibility
Quota sampling	• Time and cost consuming • Easy to do	• Risk of selection bias, • Not reliable results
Snowball sampling	• Used in secret and hidden populations that are difficult to access	• Risk of bias, time-consuming • High level of sampling error • Does not provide representative sample • Time-consuming
Judgment sampling	• Low cost, is easy, • Not time consuming, • Suitable for exploratory research projects	• Low level of validity results • Does not provide representative sample • Impossible reliability evaluation in sampling process

Conclusion

Determining the type of sampling method is one of the important components of medical studies. Therefore, researchers are expected to improve knowledge of sampling methods and how to use them (random and non-random). As a result, they have more mastery in choosing the type of sampling method in their studies. The sampling method and its logic is directly related to external validity, so that the external validity will decrease if the researcher makes a mistake in choosing the sampling frame or selecting participants. Therefore, researchers should also pay attention to external validity. In sampling, both internal and external validity should be high. External validity is highly sensitive to selection bias, so in non-probabilistic sampling the external validity is much lower than in probabilistic sampling [3]. Furthermore, choosing the right sampling method can reduce time and cost and enhances quality of research.

Authors' contribution

Conceptualization: Madadizadeh F
Formal analysis: Basti M, Madadizadeh F
Investigation: Basti M, Madadizadeh F
Methodology: Madadizadeh F
Supervision: Madadizadeh F
Writing – original draft: Basti M, Madadizadeh F
Writing – review & editing: Basti M, Madadizadeh F

References

[1] Bhardwaj PJJotPoCS. **Types of sampling in research**. 2019;5:157-63
[2] Taherdoost HJHtCaSTfR. **Sampling methods in research methodology; how to choose a sampling technique for research**. 2016;5:18-27
[3] McEwan B. **Sampling and validity**. *Annals of the International Communication Association*. 2020;44: 235-47
[4] Adèr HJ. **Advising on research methods: A consultant's companion**: *Johannes van Kessel Publishing.*; 2008
[5] Ackoff R. **The design of social research. chicago: Universidad de Chicago**. AMERICAN; 1953
[6] Etikan I, Bala KJB, Journal BI. **Sampling and sampling methods**. 2017;5:00149
[7] Mweshi GK, Sakyi KJAoBRV. **Application Of Sampling Methods For The Research Design**. 2020;8:180-93
[8] Yin RK. **Case study research: Design and methods (Vol. 5)**. 2003
[9] Malhotra NK, Birks DF. **Marketing research: An applied approach**: *Pearson education*; 2007
[10] Brewerton PM, Millward LJ. **Organizational research methods: A guide for students and researchers**: *Sage*; 2001