Journal of Advanced Research in Medical Science & Technology
Volume 7, Issue 2 - 2020, Pg. No. 19-21
Peer Reviewed & Open Access Journal

Review Article

Sampling in Research Series 2: Basic Concepts in Estimating Sample Size

SC Mohapatra[1], Badrinarayan Mishra[2]

[1]Former HOD, Community Medicine, BHU, Varanasi & Former Dean FMHS and Dean Academic Affairs SGT, University. Presently Advisor and Consultant SGT University, Gurgaon, Haryana, India.
[2]Professor, Community Medicine, Ruxmaniben Deepchand Gardi Medical College, Ujjain, Madhya Pradesh, India.
DOI: https://doi.org/10.24321/2394.6539.202008

I N F O

Corresponding Author:
SC Mohapatra, Community Medicine, BHU, Varanasi & Former Dean FMHS and Dean Academic Affairs SGT, University. Presently Advisor and Consultant SGT University, Gurgaon, Haryana, India.
E-mail Id:
vishwamegh@gmail.com
Orcid Id:
https://orcid.org/0000-0002-9605-0867
How to cite this article:
Mohapatra SC, Mishra B. Sampling in Research Series 2: Basic Concepts in Estimating Sample Size. *J Adv Res Med Sci Tech* 2020; 7(2): 19-21.

Date of Submission: 2020-06-11
Date of Acceptance: 2020-07-04

A B S T R A C T

Sampling process/ method has been like an examination to the researchers' botheration from generation to generation from the time of inception towards drawing a representative group of units or cases from a particular population. In series 1 the most popular and frequently used sampling based on probability theory had been depicted. In this series, sampling for special types of studies will be discussed. It has already been emphatically stated earlier '*no thumb rule of 100' (or any other fixed number) of samples are permitted for ideal study or research.* Most research publications and MD theses are being seen to have been accepted in spite of a wrong sample size. Even many journals today publish papers without proper sample size, although it is the responsibility of the reviewer who does not know statistics or epidemiology; the journal concerned is also equally responsible for a cherry-picking reviewer with partial knowledge. In fact, such journals should be stopped publishing or be blacklisted since they dispense the wrong knowledge to the researchers and the young scientist might learn misguidedly. In most of the population studies, $4pq / L2$ where p=Prevalence; q=100 or 1 - Prevalence; and L = permissible limit of error has to be <10 % of the prevalence. L or permissible error should not be taken as absolute term. It's easy to understand that if the error is 10% admitted before a study then the study stands to be null and void. There could be more than 100 types of errors while implementing the study like human error, equipment error, interviewer error, respondent error and so on. If sampling itself has a huge error of 10%; the study will become a Pandora's Box of errors only. Thus, it was imperative to highlight the value of sample size calculation based on probability theory to determine the result of chance and types of different studies.

Keywords: Sampling, Sample Size, Sampling Methods, Sample Size of Unit Variable, Snowball Sample, Sample Size of Two Mean, Rate and Proportion

Introduction

Accurate sample selection of a subset of individuals from within a statistical population is an unavoidable and vital procedure for quality assurance in epidemiological and managerial studies, studies related to Health System Research (HSR), survey any kind of research methodology. However, today and unfortunately for last few years, the majority of researchers or supervisors of postgraduate studies do not follow in most of their researches rendering wrong concepts to the young scientists. In Series 1, probability and nonprobability sampling; qualitative and quantitative samplings; pilot study sampling and census has already been discussed. In this section, special situational sampling will be considered.

Snow Ball Sampling

These are of two types: Hospital-based snow-ball and community-based snowball sample. Case-control design is usually conducted to study the epidemiology of rare diseases and undertaken in the hospital setup. In this case, the sample size cannot be prefixed and the sample is taken as hospital-based snow-ball sampling. Since we cannot produce the disease, for example, gall bladder carcinoma (CA GB), snow-balling is undertaken. It is undertaken from a particular ward (or wards) of surgical oncology unit. Suppose the study is decided to be undertaken for 1 year; then all cases diagnosed as CA GB will be included daily and all the information has to be completed after the consent form is explained to and then signed by the patient before inclusion of the case into the study sample, as per Helsinki declaration.[1] However for each case, at least one matched control needs to be selected. Normally Age and Sex matching are good enough, but care should be taken when "no probable confounding parameter" is matched, which may confuse the study output. The statistical packages cannot check sample size or probable confounders. Thus, the investigator has to have a clear understanding of this. Statistical packages are nonliving substances to find whatever you need from even a profane and foul set of data.

Community-based snowball samplings are equally done like this but controls may not be needed every time. Here also, data is collected for a fixed period, say 1 year, and assembled to make the snow-ball. It is undertaken for diseases/conditions, which are not discovered/disclosed easily (like rare disease in hospital), for example, tuberculosis, HIV/AIDS or substances abuse. In this case (in case of injection drug users), one case if found or traced from hospital record; can be contacted and his known partners and their partners (partners of partners) are added to the sample one by one so as to make the snowball. Another method of snowballing is by 'data triangulation' commonly employed in HIV/AIDS case detection.[2]

Fisher[3] had suggested a special cross-sectional study where a group can be nested and followed for a certain period. For example, suppose a researcher studies ophthalmologic surgical elderly cases and after registration of cases he finds there are many cases at close proximity; hence, their financial implication due to the surgery can be studied. These cases can, therefore, be nested and followed up for the desired period. Such a study can be said as a "cross-sectional nested case follow-up study". Many researchers do not know such studies. However, in such a case the sample size cannot be prefixed and like snowball sampling, any sizable number above 50 can be taken as a good sample.

A cohort study is one of the observational study designs which is used to evaluate the association between exposure and disease. It's a long study involving long-time huge data and large staff support. To reduce cost, several alternative study designs have been proposed. Some of them are 'nested case-control' and 'case-cohort study' designs which are particularly practical in studying rare diseases. A nested case-control study design involves the selection of several healthy controls for each case, typically from those still under observation at the time when the case developed the disease form the studied cohort.[4] Nested case-control studies have some limitations; thus, case-cohort study designs were proposed as an alternative to the nested case-control study design. Such studies require only the selection of a random sample.

Feasibility or Convenient Sample

The feasibility study is usually carried for project implementation or similar studies. It is basically a health management process that estimates the project completion time, cost and other issues in order to complete the project successfully. There are four types of feasibility studies such as operational feasibility study, technical feasibility study, administrative feasibility study and economic feasibility study.[5] For example, aiming to expand a hospital facility, i.e., add an extension to any component like modular operation theatre/ telecommunication, etc, the organization may perform a feasibility study. The study will determine whether the project should be undertaken ahead or not. A feasibility study is *an assessment of the practicality and utility of a proposed plan or method*.

But some other kinds of studies also use this method with the meaning of "convenient sample." This method, as the name suggests, is used as per the convenience or practicality (feasibility) factors of the investigator of small research with the paucity of time or facility. For example, suppose an investigator wants to study on diabetes and related issues in the community the sample size will be around 4000 to 9000 depending upon the prevalence, but the investigator has a paucity of time or difficulty in covering such a large population, a feasibility or convenient sample

ISSN: 2394-6539
DOI: https://doi.org/10.24321/2394.6539.202008

may be chosen.[6] Such a situation is not infrequently seen in case of time limit studies like MD thesis. The feasibility sample in the above study may be calculated as:

Sample size=to cover 20 houses per day X 5 members/ family

X 21 days in a month X 6 months of data collection=1260 population.

However, this method should not be used if other methods are available. This is always a compromised method and investigator's bias remains inbuilt and cannot be removed and therefore this method should be discouraged.

Sample Size in Two Series

Sample Size in two different series with different means, rates or proportions in comparative studies of two populations is calculated in different methods.

Example with two different means is calculated by the formula of sample size=n =s_1^2+s_2^2/e^2 where s_1 is standard deviation of 1st group, s_2 is standard deviation of the second group and e is the standard error of the difference at 95% confidence interval.

Given two different rates for two groups of the population, the sample size is calculated by the formula n=(where the given rates (say birth rate) of both populations are r_1 and r_2 and e is the standard of the difference on them at a 95% confidence interval. In case proportions, the sample size is calculated as n= {P_1 (100-P_1) +P_2(100-P_2)} / e^2 where e is the standard error of difference.

The sample size of two population proportions: For example, we need to estimate the sample size for two districts in order to estimate vaccine drop out coverage of pregnant women from two adjacent districts in a state. Let us assume that the proportion of non-vaccination of pregnant mothers(for tetanus) from district one is 30%, and from district two is 15% per year; thereby yielding a difference of 15%. Assuming the desired 95% CI (confidence interval) for this difference is 5% to 25% with a standard error of 5%, we can calculate the sample size adopting the formula

n=P1(100-P1)+P2(100-P2)/e^2 = 30x70+15x85/5^2 =135 for each districts.[7]

Conflict of Interest: None

References

1. WMA declaration of Helsinki-ethical principles for medical research involving human subjects, 35th WMA General Assembly, Venice, Italy, October 1983, uploaded July, 19, 2018.
2. WHO; Synthesis of result from multiple data sources and decision making, 2009; downloaded May 2020.
3. Fisher, R A, Statistical Method for Research workers, 14th Edition, Edinburgh University Press, 1970.
4. Langholz B, Thomas D. Nested case-control and case-cohort methods of sampling from a cohort: A critical comparison. *American Journal of Epidemiology* 1990; 31: 169-176.
5. Mohapatra SC, Mohapatra M, Mohapatra V. A treatise on health Management, 1st Edition, JP Brother Publication, 2016.
6. Greenwald P, Cullen JW. The scientific approach to cancer control. *CA Cancer J Clin* 1984; 34: 328-332.
7. RC Goyal. Research Methodology for Healthcare Professionals; ISBN:978-93-5025101-0; 1st Edition, 2013,127; JAYPEE BROTHERS Medical Publishers, New Delhi.

ISSN: 2394-6539
DOI: https://doi.org/10.24321/2394.6539.202008