ORIGINAL RESEARCH PAPER

Volume-6 | Issue-9 | September-2017 | ISSN No 2277 - 8179 | IF : 4.176 | IC Value : 78.46

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

SAMPLE SIZE ESTIMATION FOR CLINICAL RESEARCH STUDIES USING MEAN AND PROPORTION.

Statistics

Prabhaker Mishra	Department of Biostatistics and Health Informatics, SGPGIMS, Lucknow, Uttar Pradesh, India. - Corresponding author
Chandra M Pandey	Department of Biostatistics and Health Informatics, SGPGIMS, Lucknow, Uttar Pradesh, India.
Uttam Singh	Department of Biostatistics and Health Informatics, SGPGIMS, Lucknow, Uttar Pradesh, India.
Vishal Sharma	Department of Community Medicine, ACMS, New Delhi, India
Sachin Singh Yadav	Department of Community Medicine, TMMC&RC, Moradabad, UP, India
Ranjana Kar	Department of Community Medicine, S.C.B. Medical College, Cuttack, India

KEYWORDS

Introduction

One of the most common questions that a researcher should ask when planning a study is "How large a sample size do I need. One of the fundamental phases of planning a study is the calculation of the sample size **[1]**. It is logically neither practical nor feasible to study the entire population in any study. Hence, a set of participants is selected from the population, which is less in number (size) but adequately represents the population from which it is drawn so that true inferences about the population can be made from the results attained from its sample **[2]**. This set of individuals is known as the "sample. If the sample size is too small, even a well-conducted study may fail to answer its research question, may fail to detect important effects or associations, or may estimate those effects or associations too inaccurately **[3]**. Similarly, if the sample size is too large, the study will be more difficult and costly, and may even lead to a loss in accuracy. Hence, optimum sample size is an essential component of any research. Careful consideration of sample size and power analysis during the planning and design stages of clinical research is crucial **[3]**. This article describes the principles and methods used to calculate the sample size for different objectives:

What is a sample size :

A sample is a selection of respondents chosen from the target population in such a way that they represent the total population as good as possible. The number of completed responses / number of participants in a sample in our survey receives is sample size. It's called a sample because it only represents part of the group of people (or population) whose opinions or behavior, we care about. A sample is called Random Sample," when respondents are chosen entirely by chance from the target population at large.

Need for appropriate Sample size: A sample that is too big will lead to the waste of precious resources such as time and money, while a sample that is too small will not allow you to gain reliable insights [4]. Using a correct sample size is crucial for your research.

Concept behind sample size estimation: Main objective of sample size estimation is to find out the adequate group of individuals, those result could be able to predict at ethically and scientifically valid and meaningful outcome with associated precision and error for the target population **[2]**.

Determining the sample size:Factors that influence sample size:

The "right" sample size for a particular application depends on many factors, including : Cost considerations (e.g., maximum budget, desire to minimize cost), Administrative concerns (e.g., complexity of the design, research deadlines), Minimum acceptable level of precision, Confidence level, Variability within the population or sub-population (e.g., stratum, cluster) of interest, prevalence / effect size and Sampling method. Some more factors that can be considered while calculating the final sample size include the expected drop-out rate, equal /unequal allocation ratio, and the objective and design of the study **[5]**.

Inputs for sample size estimation: Prior to calculate sample size for the study, following information require i.e. acceptable level of significance, power of the study, underlying event rate in the population (i.e. prevalence/ percent), relative/margin of error of the estimated value, expected effect size etc. Although type of information/ inputs is varying as per study design and objective of the study.

Level of significance: Prior to starting a study we set an acceptable value for level of significance, is also called "p value". When we say, for example, we will accept a $p<0.05$ as significant, we mean that we are ready to accept that the probability that the result is observed due to chance (and NOT due to our intervention) is 5%. Conventionally, the p value of 5% ($p = 0.05$) or 1% ($p = 0.01$), which means 5% or 1% chance of erroneously reporting a significant effect is accepted.

Confidence level: A measure of how certain you are that your sample result accurately reflect the population, within its margin of error. Common standards used by researchers are 95%, 99% and 99.9%. This is also indicates that, the larger your p value, the smaller your confidence level (because sum of the p value and confidence level is equal to 1 or 100%).

Confidence Interval: The confidence interval calculations assume, you have a genuine random sample of the relevant population. If your sample is not truly random, you cannot rely on the intervals. This reflects the confidence with which you would like to detect a significant difference between the two proportions/means etc. If your confidence level is 95%, then this means you have a 5% probability of incorrectly detecting a significant difference when one does not exist, i.e. a false positive result (Known as type I error).

Power of the study: Statistical power is the probability of detecting a significant difference when one exists. In other words, statistical power explains the generalizability of the study results and its inferential power to explain population variability. Sample size is directly related to power; the bigger a sample, the higher the statistical power. The power of a study increases as the chances of reducing a Type II error (denoted by β). The "power" of the study then is equal to $(1 - \beta)$. Usually most of the studies, we accept a power of minimum 80%. This means that we are accepting that one in five times (that is 20%) we will miss a real difference i.e. when power is 80%, then this means that you have a 20% probability of failing to detect a significant difference when one does exist (false negative result).

Expected Mean/ proportions: In epidemiology, prevalence is a

statistical concept referring to the number of cases of a disease that are present in a particular population at a given time, whereas mean refers to the average per individuals.

The sample proportions/means are what we expect the results to be. This can often be determined by using the results from a previous survey, or by running a small pilot study. If you are unsure, use proportions near to 50%, which is conservative and gives the largest sample size.

Margin/relative error of the estimated value: First of all, the margin of error is a statistic expressing the amount of random sampling error in a sample survey's results. It asserts a likelihood (not a certainty) that the result from a sample is varying with its population. In short, this is the positive and negative deviation you allow on your survey results for the sample w.r.t. population. In other words, the deviation between the opinions of sample respondents and the opinion of the entire population respondents. Suppose in a sample study, 30% of our respondents have anemia. If we set our margin of error on 5%, it means, in our study, our sample prevalence of anemia may be vary between ±5% i.e. 30% ±5% i.e. 25% to 35%. It means we can be 'sure' that prevalence of anemia of the entire population might be between 25% (30%-5%) and 35% (30%+5%). In the above example: Margin of error of the prevalence is 5% of the 30% prevalence. It means margin of error is 16.67% of the estimated value (30%). So this 16.67% is the relative error of estimated value of 30%. The relationship between relative error and margin of error is as follows.

Margin of error (%) = Estimated value X relative error in %
We calculate margin of error of the estimated value at certain level of confidence usually at 95%.

Expected effect size: An effect size is a quantitative measure of the difference between two groups. Suppose, average weight loss following one diet program is 20 kg and following another is 10 kg, the absolute effect size would be 20-10=10 kg. Ex- The difference between the value of the variable in the control group and that in the test drug group is known as effect size. We can estimate the effect size based on previously reported or clinical studies. It is important to note that if the effect size is large between the study groups then the sample size required for the study is less and if the effect size between the study groups is small, the sample size required is large.

Standard Deviation: Standard deviation is the measure of dispersion or variability in the data. While calculating the sample size, a researcher needs to anticipate the variation in the measures that are being studied. With higher standard deviation, need larger sample size to detect a difference between interventions.

One tailed and two tailed test:A one- or two-tailed test is determined by objective of the test. The one-tailed test is performed if the results are interesting whether they turn out in a particular direction **[1]**. The two-tailed test is performed if the results would be interesting in either direction. Because the one-tailed test provides more power to detect an effect, you may be tempted to use a one-tailed test whenever you have a hypothesis about the direction of an effect. Before doing so, consider the consequences of missing an effect in the other direction. Example we developed a new drug that believe is an improvement over an existing drug. You wish to maximize your ability to detect the improvement, so you use a one-tailed test. In doing so, we fail to test for the possibility that the new drug is less effective than the existing drug. The consequences in this example are extreme, but they illustrate a danger of inappropriate use of a one-tailed test. When using a two-tailed test, regardless of the direction of the relationship you hypothesize, you are testing for the possibility of the relationship in both directions. For example, we may wish to compare the mean of a sample to a given value x using one sample t-test. Our null hypothesis is that the mean is equal to x. Two-tailed tests will test both if the samples mean is significantly greater than x and if the sample mean significantly less than x. The mean is considered significantly different from x if the test statistic is in the top 2.5% or bottom 2.5% of its probability distribution, resulting in a p-value less than 0.05. For one sided test and two sided test, all the inputs in sample size formula are same except critical ratio (Z value), which is differ in one and two sided test (table 1) denoted by Z1-œ/2 (at two sided test) and Z1-œ (at one sided test).

Table 1 : Z value

Level of confidence	Two tailed ($Z^{1-œ/2}$)	One tailed ($Z^{1-œ}$)
0.95	1.96	1.65
0.01	2.58	2.33
0.001	3.29	3.09
Power of the test	**Z value ($Z^{1-ß}$)**	
0.80	0.84	
0.90	1.28	
0.95	1.65	
0.99	2.33	

Type of Methods of calculating sample size: To calculate sample size, there are various sample size estimation method, depend upon objective of the study. For one / two sided test, only value of z at desired level of confidence is changeable, rest formula is same in both sided test. Here we are discussed the methods based on single/two means and proportions are given below:

1. Based on proportions:

1A. Sample size to estimate a proportion: If in our study, objective is to calculate sample size for a study, when main outcome of the interest of the study is prevalence / percent of an event of the interest (like disease).Formula is:

Sample Size (n) = $Z_{1-œ/2}^2 pq / d^2$

Where: n is sample size that to be finding out, $Z_{1-œ/2}$ is the critical value of the corresponding level of confidence at two tailed, p is prevalence and q is non-prevalence equal to 1-p, d is the margin of error of the estimated prevalence. As it is for non-comparative study, always use two sided test **[1,6,7,8]**.

Example: In a metropolitan city, a study stated that prevalence of anemia in pregnant women was 40%. At 95% confidence level and 10% relative allowable error in the prevalence, how many individuals (here "pregnant women") required to conduct new study?

In above example: main objective of the study is to estimate prevalence of anemia in the pregnant women. As it is a non-comparative study and formula for sample size is:

Sample Size (n) = $Z_{1-œ/2}^2 pq / d^2$
In this example: p = 40% =0.4, q = 60% =0.6,
d = prevalence* relative error = 40% * 10% =40% * 0.1 = 4% = 0.04
$Z_{1-œ/2}$ =1.96 is value of 95% confidence level at two tailed.
Sample Size (n) = $Z_{1-œ/2}^2 pq, d^2$
$=(1.96)^2 * (0.4) * (0.6) / (0.04*0.04) = 576$

I.e. To conduct the above prevalence study, minimum 576 women required for the study. Although in the study, we have to target to cover more than 576 pregnant women to achieve at least 576 individuals after excluding non-response.

1B. Sample size to detect the difference (effect size) between two groups: If in our study, objective is to calculate sample size for a comparative study, when main outcome of the interest of the study is difference in proportions (effect size) like difference in proportions between treatment and control groups. Formula at two sided test:

Sample Size (n)
$=[Z_{1-œ/2}\sqrt{\{2p(1-p)\}}+Z_{1-ß}\sqrt{\{p_1(1-p_1)+p_2(1-p_2)\}}]^2/(p_1-p_2)^2$
Where: n is sample size that to be finding out, $Z_{1-œ/2}$ and $Z_{1-ß}$ is the critical value of the given level of confidence at two sided test and power of the study, p_1, p_2 are the proportion in treatment and control groups, p is the average value of p_1 and p_2, d is the effect size.As it is for comparative study, we can use either one sided test/two sided test, depend upon our hypothesis **[1,6,7].**

Example: In a study, the improvement in the patients in treatment A and treatment B groups are 40% and 20% respectively. At 95% confidence level and 80% power of the study, to detect the 20% difference would be clinically important, how many patients are required:

Sample Size (n)
$=[Z_{1-œ/2}\sqrt{\{2p(1-p)\}}+Z_{1-ß}\sqrt{\{p_1(1-p_1)+p_2(1-p_2)\}}]^2/(p_1-p_2)^2$
$=[1.96\sqrt{\{2*0.3(1-0.3)\}}+0.842\sqrt{\{0.4(1-0.4)+0.2(1-0.2)\}}]^2/(0.4-0.2)^2$
$=[1.96\sqrt{\{0.42\}}+0.842\sqrt{\{0.24)+0.16\}}]^2/(0.2)^2$

$= [\ 1.96^* 0.648 + 0.842^*0.632\ ]^2 /0.04 = [1.270 + 0.532]^2 / 0.04 = 3.247 / 0.04 = 81$

We have required at least 81 patients in each of the groups (i.e. 81 patients in treatment A and 81 patients in treatment B).

We need to have some estimate of the proportion of events expected. This can often be obtained from routinely collected data or previous studies.

2. Based on means:

2A. Sample size to estimate a mean: In our study, if objective is to calculate sample size for a study, when main outcome of the interest of the study is mean of the variable (example-mean weight of group of the individuals etc.).

Sample size (n) = $Z_{1-\alpha/2}^2 s^2 / d^2$

Where: n is sample size that to be finding out, $Z_{1-\alpha/2}$ is the critical value of the given level of confidence, s is the standard deviation and d is the margin of error in the mean value. As it is for non-comparative study, always use two sided test **[1,6,7,8]**.

Example: In a city, a previous study stated that mean and standard deviation of hemoglobin level in pregnant women were 12.3 and 2.3 g/dl. At 95% confidence level and 10% relative error, how many pregnant women required to conducting a new study in the community?

Sample size (n) = $Z_{1-\alpha/2}^2 s^2 / d^2$

$= (1.96)^2 *(2.3)^2 / (1.23)^2$

$= 3.84 \text{ X } 5.29 \ / 1.51 = 20.31 / 1.51 = 13.45,$

Minimum sample size require for study = 14

In above, d = margin of error = mean X relative error = 12.3 X 0.1 = 1.23

I.e. To conduct the above mean study, minimum 14 women required for the study. Although in the study, we have to target to cover more than 14 pregnant women to achieve at least 14 individuals after excluding non-response.

2B.Sample size to detect the difference in means (effect size) between two groups: In our study, if objective is to calculate sample size for a comparative study, when main outcome of the interest of the study is to detect the difference between two means.

Sample size (n) = $(Z_{1-\alpha/2+} Z_{1-\beta})^2 (s_1^2 + s_2^2) / d^2$

Where: n is sample size that to be finding out, $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ is the critical value of the given level of confidence (two sided) and power of the study, s_1 and s_2 are the standard deviation of treatment A and treatment B groups, d is detected difference in means (effect size).As it is for comparative study, we can use either one sided test/two sided test, depend upon our hypothesis **[1,6,7]**.

Example: In a study of geriatric age group, assuming that mean (standard deviation) of hemoglobin level in rural area and urban area are 11.5 (2.2) and 13.5(2.1) respectively. At 95% confidence level and 80% power of the study , to detect difference 2g/dl in the groups, how many geriatric age group individuals required in the study from rural and urban each.

Sample size (n) = $(Z_{1-\alpha/2+} Z_{1-\beta})2^*(s_1^2 + s_2^2) / (x1\text{-}x2)2$

$= (1.96+0.842)^2(2.22+2.12)/ (2.0)^2$

$= (2.802)2^*(4.84+4.41) / (2.0)2 = (7.851)^*(9.25) / (4.0) = 18.16$

We have required at least 19-19 patients in each of the groups.

Design Effect (DEFF): The above mentioned methods of sample size calculation provide required minimum sample size in case simple random sampling method is used. However, due to some inevitable reasons, sometimes we used cluster random sampling as a result of this, target precision is difficult to achieve as simple random sampling. To overcome this problem, calculated sample size has to be adjusted in terms of design effect (DEFF), which is equal to the ratio of expected variance in cluster random sampling with expected variance in simple random sampling. The sample size calculated by simple random sampling will be multiplied by design effect produced new sample size where DEFF is always ≥1. In most of the immunization and nutritional survey, we usually take DEFF around two. Although if more variation from one cluster to another cluster, we can take DEFF more than two also [1].

Other methods for sample size estimation: There are various methods available to calculate sample size depend upon objective of the study. Instead calculating sample size through mean(s) and proportion(s), we can calculate sample size for other studies those objectives are including survival analysis, ROC curve analysis, Linear/Logistic regression, Correlation coefficient, Clinical trials and group sequential test etc. There are online/offline software's available to calculate sample size. Power and sample size (PASS) software is one of the useful software to estimate sample size **[7]**.

Advantages of the calculated sample size:

1. Minimum sample size requirement is prime need of the significant study. In case of lack of sample size or overestimated sample size mislead the result.
2. Studies without appropriate sample size failed to detect the actual effect of the treatment and result might be gives false interpretation.
3. Through calculated sample size, we can ensure approximate error in the study at desired level of confidence and power of the study (if applicable)[9].
4. If sample size was estimated at minimum level of confidence and power of the study (if applicable), then difference achieved in the study will be called due to treatment effect not due to chance and result can be generalized.

Conclusions: The calculation of the sample size is one of the most important steps in designing a study. Although techniques for sample size calculations are common but performing these calculations can be complicated **[10]**. Because sample size calculations are sensitive to errors and validity of the finding, to performing a study, we recommend to perform the calculations with caution or to ask for statistical advice during the calculation of the sample size.

References:

1. Sundaram KR, Dwivedi SN, Sreenivas V. Medical Statistics: Principles and Practice, 2/e. Lippincott Williams & Wilkins publication, 2014. ISBN: 9789351293217.
2. Kadam P, Bhalerao S. Sample size calculation. International Journal of Ayurveda Research. 2010;1(1):55-57. doi:10.4103/0974-7788.59946.
3. Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. Journal of Human Reproductive Sciences. 2012; 5(1):7-13. doi:10.4103/0974-1208.97779.
4. Altman DG. Statistics and ethics in medical research: III. How large a sample? BMJ 1980; 281: 1336-1338.
5. Whitley E, Ball J. Statistics review 4: Sample size calculations. Critical Care. 2002;6(4):335-341.
6. Charan J, Biswas T. How to Calculate Sample Size for Different Study Designs in Medical Research? Indian Journal of Psychological Medicine. 2013;35(2):121-126. doi:10.4103/0253-7176.116232.
7. Hintze J. PASS 8. NCSS, LLC. Kaysville. Utah, USA. 2011. www.ncss.com.
8. Gang Xu. Estimating sample size for a descriptive study in quantitative research. Published: June 1999 Article ID:19990603.
9. Statistics: An introduction to sample size calculations. Mathematics learning & Support Centre. Rosie Cornish. 2006.
10. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. Nephrol Dial Transplant 2010; 25: 1388–93.