Sample Size Calculation in Clinical Studies: Some Common Scenarios

Mohammad Z. I. Chowdhury[1,*], Khokan C. Sikdar[1,2] and Tanvir C. Turin[1,3]

[1]*Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada*

[2]*Cancer Control Alberta at Alberta Health Services, Calgary, AB, Canada*

[3]*Department of Family Medicine, University of Calgary, Calgary, AB, Canada*

Abstract: Determining the optimal sample size is crucial for any scientific investigation. An optimal sample size provides adequate power to detect statistical significant difference between the comparison groups in a study and allows the researcher to control for the risk of reporting a false-negative finding (Type II error). A study with too large a sample is harder to conduct, expensive, time consuming and may expose an unnecessarily large number of subjects to potentially harmful or futile interventions. On the other hand, if the sample size is too small, a best conducted study may fail to answer a research question due to lack of sufficient power. To draw a valid and accurate conclusion, an appropriate sample size must be determined prior to start of any study. This paper covers the essentials in calculating sample size for some common study designs. Formulae along with some worked examples were demonstrated for potential applied health researchers. Although maximum power is desirable, this is not always possible given the resources available for a study. Researchers often needs to choose a sample size that makes a balance between what is desirable and what is feasible.

Keywords: Sample Size Calculation, Power, Hypothesis Test, Level of Significance, Mean, Proportion.

INTRODUCTION

The goal of a research project is to get a scientifically valid answers of the research questions. The validity in research is achieved through rigorous design, proper data collection and appropriate analysis [1]. Use of an accurate data is necessary to validate a research study. Ideally, in order to obtain accurate information to answer a research question, one should study the entire population, which is any entire collection of people, animals, plants, things or subjects from which we may collect data. It is the entire group we are interested in, which we wish to describe or draw conclusions about [2]. While it is often not feasible to examine every member of an entire population, investigating a part of that population (which is referred to as "sample") is a typical practice in research. A sample is a subset of people, items, or events from a large population of interest, from which we collect information and analyze them to make conclusion about a specific unknown value of that population. How many individuals or items should be included in a research study or how large a sample should be, is one of the commonly asked question [3]. What should be the appropriate sample size is an important consideration in any research, including clinical studies. Studies with a small sample size may fail to detect important effects on the outcomes of interest. In consequence, the estimates may be erroneous and be termed as a waste of resources due to their incapability to yield useful results [4, 5]. On the other hand, studies with a larger sample than necessary may sometime be a waste of available resources. An oversized study may have the potential to expose an unnecessarily large number of subjects to potentially harmful or futile treatments [4]. The determination of an appropriate sample size, that is, the number of individuals that should be included for study is crucial part of study design to ensure validity, accuracy, reliability and scientific and ethical integrity of the study. The objective of this article is to review the key factors that determine an appropriate sample size and present methods for sample size calculation in common clinical studies.

MATERIALS AND METHODS

Approaches to Sample Size Calculation

There are two major approaches for sample size calculation: the precision-based approach and the power-based approach.

The precision-based approach estimates an unknown parameter of a population (e.g. the prevalence of a disease in a defined population) with a specific precision. In this case, it is the responsibility of researchers to make sure whether estimates are obtained with required precision, accuracy or level of confidence while the sample size is calculated [6]. This approach limits the confidence interval of the parameter to a certain width.

The power-based approach is related to hypothesis testing and concerned with detecting an effect of intervention on outcome. An ideal application is comparing a new treatment with standard one in

*Address correspondence to this author at the HSC G012G, 3330 Hospital Drive NW, Calgary, AB T2N 4N1, Canada; Tel: +1 (587) 917-2993; Fax: +1 (403) 270-4329; E-mail: mohammad.chowdhury@ucalgary.ca

assessing the effect of the new treatment on patient outcome. The calculation of sample size is performed to ensure whether a clinically meaningful effect exists.

So basically, which approach to consider in sample size calculation depends on the type of study the researcher is conducting. For descriptive/estimation studies (main aim is the estimation of one or more characteristics of the population) precision-based approach is frequently used while for comparative studies (main aim is to establish whether there are statistically significant differences between groups with respect to some key outcome variable) power-based approach is used. In this article, we are mainly focusing on the power-based approach, however we will very briefly discuss about the precision-based approach for some extremely common scenarios.

Power-Based Approach of Sample Size Calculation

Steps Involved in Sample Size Calculation

To calculate appropriate sample size, one need to proceed with the steps that are outlined in the Figure **1**, followed by a detail description.

Step 1. Define Research Objective or Primary Question and Specify a Hypothesis Test

In scientific investigation, investigators need to define what they would like to measure and what change they are hoping to see. One way of evaluating a question is to use the method of hypothesis testing. When a hypothesis test is set to determine the validity of a question, both null hypothesis and alternative hypothesis need to be defined. Typically, the null hypothesis says that nothing new is happening or the comparison groups are the same (i.e., their difference is equal to zero). The word “null” can be thought of as “no change”. The alternative hypothesis is just an alternative to the null, which refers to a difference (or an effect) between the comparison groups anticipated by the researcher. This means the observed pattern of the data is not due to a chance occurrence. Alternative hypotheses can be non-directional or directional. In a one-tailed test, testing for the possibility of the relationship is one directional, completely disregarding the possibility of a relationship in other direction. If a significance level (α) of = 0 .05, is used, a one-tailed test allots all alpha (α) to test the statistical significance in the one direction of interest. This means that, 0.05, is in one tail of the distribution (bell curve) of test statistic. On the other hand, if the alternative hypothesis is non-directional, a two-tailed hypothesis is used. A two tailed hypothesis states there is a difference between groups, but, do not specify the direction of the effect. In a two-tailed test, the possibility of the relationship in both directions are tested. If a significance level (α) of 0.05, is used, a two-tailed test allots half of alpha (α) to test the statistical significance in one direction and half of alpha (α) to test statistical significance in the other direction. This means that .025 is in each tail of the distribution (bell curve) of test statistic (Figure **2**). There is a debate on whether or not it is appropriate to use a one-tailed test. The safest bet is to conduct a two-tailed tests for sample size determination [7]. This is due to the fact that a two-tailed test allows testing for the possibility of an effect in two directions, both the positive and the negative, and thus it is a conservative approach. One-tailed tests, meanwhile, allow testing for the possibility of an effect in only one direction, while not accounting for an impact in the opposite direction. For example, suppose a new drug for cardiovascular disease has been developed, which is less expensive compared to the existing drug. The investigator thinks that this new drug is just as effective as a drug already

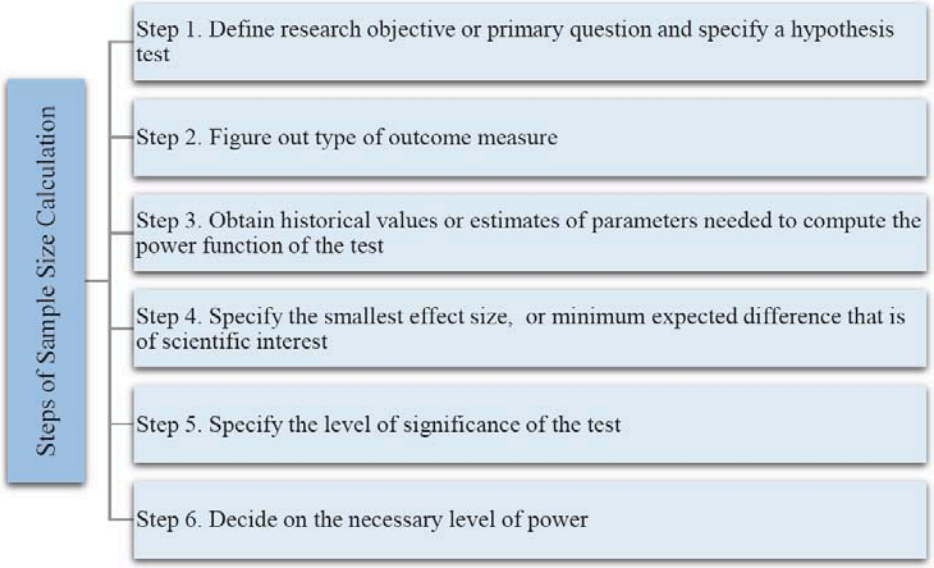

Figure 1: Sample size calculation steps

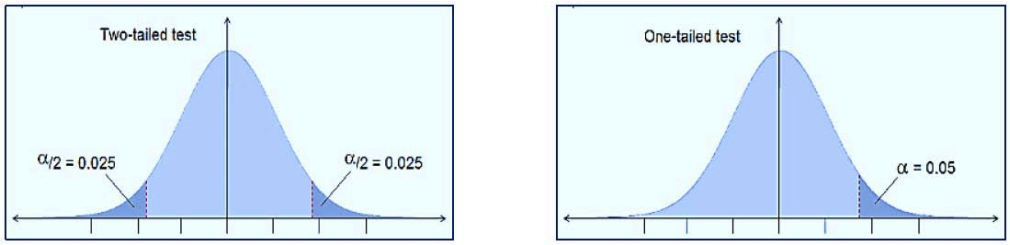

Figure 2: A hypothetical example of area under the curve for one tail and two tail test

in the market. In order to test this hypothesis, the investigator may prefer a two-tailed to one-tailed test. A one tailed test is appropriate when a large difference in one direction would lead to the same action as no difference at all [8]. Considering the same example, one possibility could be that the investigator does not really care about the new drug being more effective than the existing one. She/he would be happy as long as it is not any less effective given that the new drug is relatively cheaper than the existing one. In this type of situation, a one-tailed test, referring whether the new drug is at least as effective as the existing drug, should be of interest. A one tailed test require a smaller sample size to detect minimum expected difference than a two tailed test. The sample size of a one tailed test with a significance level α is equal to the sample size of a two tailed test with a significance level of 2α, while other factors remain unchanged. A study may have multiple hypotheses, but sample size calculation is often performed based on primary hypothesis only.

Step 2. Figure Out Type of Outcome Measure

The methods for sample size determination depends on the type of outcome expected in a study [9]. There are three different categories of outcome that are commonly considered in a study. The first category of outcome is continuous outcome which can take any numerical values within a range. Example includes: weight, height, blood pressure, heart rate etc. Researchers are interested in mean and standard deviation of continuous outcome. The second category of outcome is dichotomous outcome, which have only two categories or levels. Example includes: yes/no, disease/non-disease etc. The third category of outcome is multi-level outcome, which have more than two categories or levels. Example includes: health condition (poor, good, excellent), severity of diseases (early, mild, moderate, severe), blood groups (A, B, AB, O) etc. For binary and multi-level outcomes, researcher's interest lies on proportion, percentage or rates of outcome.

Step 3. Obtain Historical Values or Estimates of other Parameters Needed to Compute the Power Function of the Test

Studies with continuous outcome, require information about the parameter standard deviation to calculate sample size. Standard deviation, denoted by sigma σ, is a measure that is used to quantify the amount of variation or dispersion of a data. It provides a measure of how spread out the numbers in the data. In determining the sample size, an investigator needs to anticipate the variation in the measures that are being studied. The sample size needed to detect the minimum difference increases as the variability (standard deviation) increases [3]. If the study population is homogeneous then standard deviation is small and require a smaller sample size. On the other hand, if the study population is heterogeneous, the standard deviation is large and require a larger sample size. The population standard deviation of the variable of interest is often unknown. Lacking an exact value for the standard deviation can be mitigated by using an estimated value, determined on the basis of previous data collected from a similar study population. If no prior information is available, standard deviation can be estimated on the basis of a pilot study, subjective experience or a range of possible values can be assumed [3]. An estimate of standard deviation is not required when proportions are being compared (in contrast to a mean), because the standard deviation is mathematically derived from the proportion [3].

Step 4. Specify the Smallest Effect Size or Minimum Expected Difference that is of Scientific Interest

Minimum expected difference (also known as effect size) is clinically significant difference one wishes to detect in a study [3]. Investigators needs to decide the smallest difference between the comparison groups, which is clinically significant or important enough to have practical implications. Generally, the bigger the sample size, the greater the chance that the investigator will detect such a difference. Thus, if the clinically significant difference is considered smaller, a larger sample is required to detect it. Investigators often face difficulty in deciding how big a difference between the comparison groups would be regarded as clinically important. It is often based on a clinical judgment and experience with the problem being investigated.

Step 5. Specify the Level of Significance of the Test

The level of significance, usually denoted by alpha α, is the probability of rejecting the null hypothesis

given that it is true (type I error). This is also called the error rate that investigators are willing to accept [10]. Alpha is a threshold value used to judge whether a test statistic is statistically significant. It is chosen by the researcher. In practice, 0.01, 0.05, and 0.1 are the most commonly used values for alpha, representing a 1%, 5%, and 10% chance of a type I error occurring. An alpha of 0.05 means that the investigator wishes the chance of mistakenly designating a difference "significant" (when in fact there is no difference) to be no more than 5%. In other words, we are ready to accept that the probability that the result is observed due to chance (and not due to intervention) is 5%. To put it in different words, we are willing to accept the detection of a difference 5 out of 100 times when actually no difference exists. It is recommended to specify alpha before analyzing data. Specifying alpha after performing an analysis opens one up to the temptation to tailor significance levels to fit the results. As the level of significance decreased sample size needed to detect the minimum expected difference increases.

Step 6. Decide on the Necessary Level of Power

The power of a study is its ability to detect a difference or effect, if it in reality exists. In other words, this is the probability of correctly identifying a difference between the two comparison groups in the study sample when one genuinely exists in the population from which the samples were drawn [11]. Power also refers to the likelihood of avoiding a false negative (positive instances that were erroneously reported as negative) [12]. The investigator needs to set the desired power level. A commonly used power in clinical studies is 80%, which means the study has a 80% chance of ending up with a *p* value of less than 5% in a statistical test (e.g. a statistically significant treatment effect) if there really was an important difference (e.g. 10% versus 5% mortality) between treatments. Sometimes a study may not show a significant difference between groups being studied. The main reasons for this could be: either there was really no significant difference (hence a true negative result) or there was a difference but the study failed to detect it (false negative result) [13]. The latter may arise because the study was poorly designed (e.g. used imprecise measurements) or because the study was lacked power. If the statistical power of a study is low, the study results will be questionable (the study might have been too small to detect any differences). Statistical power depends on the magnitude of the true difference between the study groups (a smaller difference requires more power), the level of significance (the lower the significance level, the lower the power), and the sample size (the larger the sample size, the higher the power) [14].

Table **1** gives an idea how different factors such as power, level of significance, minimum expected difference, standard deviation and test direction affect sample size calculation.

Formulae for Sample Size Calculation under Different Scenarios

The basic formula for sample size calculation is, $n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 2\sigma^2}{d^2}$

where $Z_{1-\frac{\alpha}{2}}$ is the critical value of the Normal distribution at $1-\frac{\alpha}{2}$ (e.g. for a confidence level of 95%, α is 0.05 and the critical value is 1.96), $Z_{1-\beta}$ is the critical value of the Normal distribution at $1-\beta$ (e.g. for a power of 80%, β is 0.2 and the critical value is 0.84), σ^2 is the population variance, and d is the different you would like to detect.

Formulae for sample size calculation varies depending on the type of research designs, type of

Table 1: Impact of Different Factors in Sample Size Calculation

Factors that affect sample size calculations		
Factor	**Magnitude**	**Required sample size**
Power	Low	Small
	High	Large
Level of significance	Small	Large
	Large	Small
Minimum expected difference	Small	Large
	Large	Small
Standard deviation	Small	Small
	Large	Large
Test direction	One sided	Small
	Two sided	Large

Table 2: Meaning of Common Symbols Used in Equations for Sample Size Calculation

Symbol	Meaning
α	The level of significance. 5% or 0.05 is the most commonly used value for α
1-β	The power of the study. 80% or 0.80 is the most commonly used value for 1-β
$Z_{1-\alpha}$	The critical value of standard normal distribution for the specified level of significance (α) for one sided test
$Z_{1-\frac{\alpha}{2}}$	The critical value of standard normal distribution for the specified level of significance (α) for two sided test
$Z_{1-\beta}$	The critical value of standard normal distribution for the specified level of power 1-*β*

outcome measures and other factors as discussed above. Most of the formula contains some common symbols which are provided in Table **2**.

Sample Size Calculation in Testing the Equality of the Mean of a Single Population or Equality of Means between Two Populations

Table **3** presents the formulae for sample size calculation in testing the equality of the mean of a single population or equality of means between two populations, when the primary outcome of interest is a continuous random variable (for which the mean and standard deviation are expression of results or estimates of population characteristics).

Where,

$\mu_A - \mu_0$ = the difference between the null and alternative hypothesis (minimum expected difference).

$\mu_1 - \mu_2$ = mean difference between the two groups (minimum expected difference).

$\sigma = \sqrt{\frac{(n_1-1){s_1}^2+(n_2-1){s_2}^2}{n_1+n_2-2}}$ = the pooled standard deviation (if unknown, an estimate from previous study or pilot study is used). The meaning of other symbols in the equations are provided in Table **2**.

Example of Sample Size Calculation in Testing the Equality of Means between Two Populations

A randomized controlled trial has been planned to evaluate a brief psychological intervention in comparison to usual treatment in the reduction of suicidal ideation amongst patients presenting at hospital with deliberate self-poisoning. Suicidal ideation will be measured on the Beck scale; the standard deviation of the suicidal ideation on the Beck scale in a previous study was 7.7 points, and a difference of 5 points in Beck scale between the intervention group and usual treatment group is considered to be of clinical importance [15]. Investigators wants to determine the required sample size for this study considering a two sided test with the level of significance (α) = 5% and power (1-β) = 80%. To illustrate sample size calculation, we proceed as follows: Here, standard deviation (σ) = 7.7, the expression $(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 = (Z_{0.975} + Z_{0.80})^2 = (1.96 + 0.84)^2 = 7.84$ and $(\mu_1 - \mu_2)^2 = 5^2 = 25$. Substituting these values in the sample size calculation formula,

Table 3: Sample Size Formula to Test Equality of the Mean of a Single Population and Means between Two Populations

	Hypothesis	Sample size formula
Test of a single population mean with respect to a pre-specified mean		
One-tailed test	$H_0: \mu = \mu_0$ vs $H_A: \mu > \mu_0 \; or \; \mu < \mu_0$	$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2}$
Two-tailed test	$H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$	$n = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 \sigma^2}{(\mu_A - \mu_0)^2}$
Test to comparing two population means		
One-tailed test	$H_0: \mu_1 = \mu_2$ vs $H_A: \mu_1 > \mu_2 \; or \; \mu_1 < \mu_2$	$n = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 2\sigma^2}{(\mu_1 - \mu_2)^2}$ $(per\ group)$
Two-tailed test	$H_0: \mu_1 = \mu_2$ vs $H_A: \mu_1 \neq \mu_2$	$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 2\sigma^2}{(\mu_1 - \mu_2)^2}$ $(per\ group)$

$$n = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 2\sigma^2}{(\mu_1 - \mu_2)^2} = \frac{7.84\times2\times(7.7)^2}{25} = 37.186$$
$$\approx 38\ (per\ group)$$

The sample size we determined is always recommended to round-up to the next higher integer. Therefore, the total sample size for this study would be, $2\ groups \times n = 2\times38 = 76$ patients.

Assume the investigator wants to determine the sample size for the study with the level of significance (α) = 1% and power (1-β) = 90%. In this case, the expression $(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 = (Z_{0.995} + Z_{0.90})^2 = (2.58 + 1.28)^2 = 14.89$, and the revised required sample size is:

$$n = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 2\sigma^2}{(\mu_1 - \mu_2)^2} = \frac{14.89\times2\times(7.7)^2}{25} = 70.626$$
$$\approx 71\ (per\ group)$$

With the revised characteristics, the total sample required for the study would be: $2\ groups \times n = 2\times71 = 142$ patients.

Sample size calculation in testing the equality of the proportion of a single population or equality of proportions between two populations

As noted above, the proportion or percentage of an event may be the outcome of interest. The sample size can be estimated using the formulas presented in Table **4**.

Where,

π = proportion or percentage of subjects who have the characteristic (under null hypothesis, this is assumed to be π_0).

$\pi - \pi_0$ = the difference in proportion between the null and alternative hypothesis (minimum expected difference).

$\pi_1 - \pi_2$ = the difference in proportion between the two groups (minimum expected difference). The meaning of other symbols in the equations are provided in Table **2**.

Example of Sample Size Calculation in Testing the Equality of Proportions between Two Populations

A placebo-controlled randomized trial proposes to assess the effectiveness of colony stimulating factors (CESS) in reducing sepsis in premature babies [16]. A previous study has shown the underlying rate of sepsis to be about 50% in such infants around 2 weeks after birth, and a reduction of this rate to 34% would be of clinical importance. With the level of significance (α) = 5% and power (1-β) = 80%, the required sample size considering two-sided test for this study can be estimated as follows: Here, $\pi_1 = 0.50$ and $\pi_2 = 0.34$ Now the expression $(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 = (Z_{0.975} + Z_{0.80})^2 = (1.96 + 0.84)^2 = 7.84$ and $\pi_1(1-\pi_1) + \pi_2(1-\pi_2) = 0.50(1-0.50) + 0.34(1-0.34) = 0.4744$. Also $(\pi_1 - \pi_2)^2 = (0.50 - 0.34)^2 = 0.0256$. Substituting these values in the sample size calculation formula,

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 [\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]}{(\pi_1 - \pi_2)^2}$$
$$= \frac{7.84\times0.4744}{0.0256}$$
$$= 145.285 \approx 146\ (per\ group)$$

Therefore the total sample size for this study would be: $2\ groups \times n = 2\times146 = 292$ babies.

Table 4: Sample Size Formula to Test Equality of the Proportion of a Single Population and Proportions between Two Populations

	Hypothesis	Sample size formula
Test of a single population proportion with respect to a pre-specified proportion		
One-tailed test	$H_0: \pi = \pi_0$ vs $H_A: \pi > \pi_0\ or\ \pi < \pi_0$	$n = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 \pi(1-\pi)}{(\pi - \pi_0)^2}$
Two-tailed test	$H_0: \pi = \pi_0$ vs $H_A: \pi \neq \pi_0$	$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 \pi(1-\pi)}{(\pi - \pi_0)^2}$
Test to comparing two population proportions		
One-tailed test	$H_0: \pi_1 = \pi_2$ vs $H_A: \pi_1 > \pi_2\ or\ \pi_1 < \pi_2$	$n = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 [\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]}{(\pi_1 - \pi_2)^2}$ $(per\ group)$
Two-tailed test	$H_0: \pi_1 = \pi_2$ vs $H_A: \pi_1 \neq \pi_2$	$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 [\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]}{(\pi_1 - \pi_2)^2}$ $(per\ group)$

Sample Size Calculation in Testing the Equality of the Correlation Coefficient of a Single Population or Equality of Correlation Coefficients between Two Populations

Correlation is a statistical measure that can show whether and how strongly pairs of variables are related. The correlation coefficient measures the strength and the direction of relationship between two variables. The most common correlation coefficient, called the Pearson product-moment correlation coefficient**,** measures the strength of the linear association between variables. The sample size needed for correlation study can be obtained using the formulas presented in Table **5**.

Where,

$\rho =$ the correlation coefficient.

$\rho_1 =$ the correlation coefficient in group 1.

$\rho_2 =$ the correlation coefficient in group 2.

and "ln" stands for natural logarithm. The meaning of other symbols in the equations are provided in Table **2**.

Example of Sample Size Calculation in Testing the Equality of the Correlation Coefficient of a Single Population

Suppose, based on the literature, it is assumed that, the correlation between salt intake and systolic blood pressure is 0.30 i.e. $\rho = 0.3$. A study is conducted to attests this correlation in a population, with the level of significance (α) = 1% and power = 90%. The sample size for such a study can be estimated as follows: Here, $\rho = 0.3$. Now the expression $(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2 = (Z_{0.995} + Z_{0.90})^2 = (2.58 + 1.28)^2 = 14.89$ and $[\ln(\frac{1+\rho}{1-\rho})]^2 = [\ln(\frac{1+0.3}{1-0.3})]^2 = 0.383$. Substituting these values in the sample size calculation formula,

$$n = 3 + \frac{4(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{[\ln(\frac{1+\rho}{1-\rho})]^2} = 3 + \frac{4 \times 14.89}{0.383} = 158.51 \approx 159$$

person.

Sample Size Calculations for Odds Ratio

The Odds Ratio is a measure of association between an exposure and an outcome; which represents the odds that an outcome (e.g. disease or disorder) will occur given a particular exposure (e.g. health characteristic, aspect of medical history), compared to the odds of the outcome occurring in the absence of that exposure. The odds ratio is used to assess the risk of a particular outcome (or disease) if a certain factor (or exposure) is present. The odds ratio is a relative measure of risk, tells how much more likely it is that someone who is exposed to the factor under study will develop the outcome as compared to someone who is not exposed.

If the probability of the event (observing an outcome of interest) in the treatment group is p_T and the probability of the event in the control group is p_C, then the odds ratio is:

$$OR = \frac{p_T/(1-p_T)}{p_C/(1-p_C)} = \frac{p_T/(1-p_C)}{p_C/(1-p_T)}$$

Let n_T and n_C are the numbers of individuals/patients in the treatment group and in the control group, respectively. Then under the assumption, $\frac{n_T}{n_C} = k$, the sample size can be estimated using the formulas presented in Table **6**.

If $k = 1$, then, $n = n_C = n_T$.

Example of Sample Size Calculation in Testing the Equality of the Odds Ratio

Suppose in a clinical trial, the relative risk between a new therapy (treatment) and a standard therapy (control) for prevention of relapse in patients with schizophrenia and schizoaffective disorders are being

Table 5: Sample Size Formula to Test Equality of the Correlation Coefficient of a Single Population and Correlation Coefficients between Two Populations

	Hypothesis	Sample size formula
Test of a single population correlation coefficient with respect to a pre-specified value		
One-tailed test	$H_0: \rho = 0$ vs $H_A: \rho > 0 \; or \; \rho < 0$	$n = 3 + \frac{4(Z_{1-\alpha} + Z_{1-\beta})^2}{[\ln(\frac{1+\rho}{1-\rho})]^2}$)
Two-tailed test	$H_0: \rho = 0$ vs $H_A: \rho \neq 0$	$n = 3 + \frac{4(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{[\ln(\frac{1+\rho}{1-\rho})]^2}$
Test to comparing two population correlation coefficients		
One-tailed test	$H_0: \rho_1 - \rho_2 = 0$ vs $H_A: \rho_1 - \rho_2 > 0 \; or \; \rho_1 - \rho_2 < 0$	$n = 3 + \frac{4(Z_{1-\alpha} + Z_{1-\beta})^2}{[\ln(\frac{1+\rho_1}{1-\rho_1}) - \ln(\frac{1+\rho_2}{1-\rho_2})]^2}$
Two-tailed test	$H_0: \rho_1 - \rho_2 = 0$ vs $H_A: \rho_1 - \rho_2 \neq 0$	$n = 3 + \frac{4(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{[\ln(\frac{1+\rho_1}{1-\rho_1}) - \ln(\frac{1+\rho_2}{1-\rho_2})]^2}$

Table 6: Sample Size Formula for Odds Ratio

	Hypothesis	Sample Size Formula
One-tailed test	$H_0: OR = 1$ vs $H_A: OR > 1 \; or \; OR < 1$	$n_C = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{[\log(\mathrm{OR})]^2} \left(\frac{1}{Kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)}\right)$
Two-tailed test	$H_0: OR = 1$ vs $H_A: OR \neq 1$	$n_C = \frac{(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{[\log(\mathrm{OR})]^2} \left(\frac{1}{Kp_T(1-p_T)} + \frac{1}{p_C(1-p_C)}\right)$

studied. Based on the results from a previous study with 365 patients (i.e., 177 patients received the new therapy and 188 received the standard therapy), about 25% (45/177) and 40% (75/188) of patients receiving the new therapy and the standard therapy experienced relapse after the treatment, respectively. The investigator is interested in studying the odds ratio of the new therapy as compared to the standard therapy for prevention of experiencing the first relapse [17]. Assuming the relapse rates in the treatment group and the control group are 25% and 40%, respectively,

$$OR = \frac{0.40(1-0.25)}{0.25(1-0.40)} = 2$$

The sample size, $n = n_C = n_T$ (assuming $k = 1$) needed per group to achieve 80% power at 5% ($\alpha = 0.05$) is given by

$$n = \frac{(Z_{0.975}+Z_{0.80})^2}{[\log(2)]^2}\left(\frac{1}{0.25(1-0.25)} + \frac{1}{0.40(1-0.40)}\right) = \frac{7.84}{0.69^2}(4.17 + 5.33) = 156.4 \approx 157$$ patients (per group).

Precision-Based Approach of Sample Size Calculation

Precision-based approach of sample size calculation is considered in descriptive studies (also known as estimation studies) where concern is with the estimation of one or more characteristics of the populations called parameter(s) (e.g. the prevalence of disease in the population). The sample size calculation is required in such studies to ensure that estimates of the parameters are obtained with required precision/accuracy or level of confidence. Sample size determination for descriptive studies is based on confidence intervals or in other words margin of error; that is, the level of precision required in providing estimates of the rates, proportions and means.

When our interest is in a population mean (i.e. the primary outcome variable is measurement/continuous), the formula for the required sample size (n) is:

$$n = \frac{4Z^2_{1-\frac{\alpha}{2}}\,\sigma^2}{W^2}$$

where, σ is the standard deviation of the variable, W is the width of the confidence interval (equal to twice the margin of error), and $Z_{1-\frac{\alpha}{2}}$ is the value from the standard normal distribution related to and representing the confidence level (equal to 1.96 for 95% confidence). In terms of margin of error (half the width of the confidence interval, W), the formula can be rewritten as:

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}}\,\sigma}{E}\right)^2$$

where, E is the margin of error.

When interest is in a population proportion (i.e. the primary outcome variable is categorical-specifically, binary), the formula for the required sample size (n) is:

$$n = \frac{4Z^2_{1-\frac{\alpha}{2}}\,p(1-p)}{W^2}$$

where, p is the expected proportion who have the characteristic of interest, W, and $Z_{1-\frac{\alpha}{2}}$ is as defined earlier. An estimate for the expected proportion in the study population can be obtained from previous studies conducted in the same population or from a pilot study. In terms of margin of error, the formula can be rewritten as:

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}}}{E}\right)^2 p(1-p)$$

Sample Size Adjustments

Loss-to-Follow-Up

In sample survey design, some of the study subjects may refuse or may not be able to answer a particular question, whereas in studies involving long-term follow-up some individuals may dropout before the end of the study. These are known as attrition in epidemiologic research. It is part of the design stage when the investigator should think about and plan to mitigate possible attrition.

Assume n is the required number of subjects for a study, determined based on an appropriate sample size calculation formula, and a proportion of subjects,

denoted by l, is expected to an estimate of attrition. In this case, the investigator should increase the sample size to n', where

$$n' = \frac{n}{(1-l)}$$

The proportion l, may be unknown at the beginning of a study. An approximate estimates should be obtained using information from similar studies or based on researcher's educated guess.

Unequal Group Size

The sample size calculation formulas in the case of comparison of two groups discussed above assumes that the two comparison groups are of equal sized. Studies with equal numbers of subjects in each group has advantage that they tend to have greater power than would otherwise be the case [18, 19]. However, often in observational studies and in clinical trials unequal group size is desired due to some practical consideration and limitation. In such case the sample size need to be adjusted by a factor dependent on allocation ratio [20]. The required sample size in each group can be estimated in two steps: First, calculate the sample size n (across both groups) assuming that the groups are equal sized (as described above) and then adjust the sample size n according to the actual ratio of the two groups (k). Let n_1 be the sample size in the first group and n_2 be the sample size in the second group, then, n_1 and n_2 are given by

$$n_1 = \frac{1}{2}n(1+\frac{1}{k}) \text{ and } n_2 = \frac{1}{2}n(1+k)$$

Where $k = \frac{n_2}{n_1}$ is the ratio of the two groups (the anticipated degree of imbalance in sample size for a study comparing two independent groups).

DISCUSSION

In this article, we have discussed sample size calculation procedures in some common situations likely to be encountered and how these procedures depends on different factors. Still there are many other sample size calculation procedures which depends on the type of research study such as, qualitative study, descriptive study, time to event study, incidence study, studies with multiple outcome and studies with more than two groups. The procedures of sample size calculations for these studies are more complex and may be referred to in standard statistical textbooks. Sample size calculation is very important and choosing the right formula is crucial in all types of research studies. Sample size determination is an important major step in the design of a research study [9] and best considered early in the planning of a study, when modification in the study design can still be made [3]. For both scientific and ethical reasons, sample size in clinical studies must be carefully planned if the results are to be credible [21]. Currently, there are many user friendly computer software's and online sample size calculation tools are available which can assist researchers to determine appropriate sample size.

REFERENCES

[1] Carneiro AV. Estimating sample size in clinical studies: basic methodological principles. Rev Port Cardiol 2003; 22: 1513-21.

[2] Nayak BK. Understanding the relevance of sample size calculation. Indian J Ophthalmol 2010; 58: 469-70. https://doi.org/10.4103/0301-4738.71673

[3] Eng J. Sample Size Estimation: How Many Individuals Should Be Studied? Radiology 2003; 227: 309-13.

[4] Altman DG. Statistics and ethics in medical research: III How large a sample? Br Med J 1980; 281: 1336-8.

[5] Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994; 272: 122-4. https://doi.org/10.1001/jama.1994.03520020048013

[6] Binu VS, Mayya SS, Dhar M. Some basic aspects of statistical methods and sample size determination in health science research. Ayu 2014; 35: 119-23. https://doi.org/10.4103/0974-8520.146202

[7] Walonick DS. Statistics Calculator software. StatPac Inc.

[8] Bland JM, Altman DG. Statistics Notes: One and two sided tests of significance. BMJ 1994; 309: 248. https://doi.org/10.1136/bmj.309.6949.248

[9] Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci 2012; 5: 7-13. https://doi.org/10.4103/0974-1208.97779

[10] Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis testing, type I and type II errors. Ind Psychiatry J 2009; 18: 127-31. https://doi.org/10.4103/0972-6748.62274

[11] Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care 2002; 6: 335-41. https://doi.org/10.1186/cc1521

[12] Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet 2005; 365: 1348-53. https://doi.org/10.1016/S0140-6736(05)61034-3

[13] Statistical Power [Internet]. Statistical Power. [cited 2015 Dec11]. Retrieved from: http://www.med.uottawa.ca/sim/data/study_design_power_e.htm

[14] Chadha VK. Sample size determination in health studies. NTI bulletin 2006; 42: 55-62.

[15] Guthrie E, Patton GC, Kapur N, Mackway-Jones K, Chew-Graham C, Moorey J, Mendel E, Marino-Francis F, Sanderson S, Turpin C, Boddy G. Randomised controlled trial of brief psychological intervention after deliberate self-poisoning Commentary: Another kind of talk that works? BMJ 2001; 323: 135-38. https://doi.org/10.1136/bmj.323.7305.135

[16] Thabane L. Sample Size Determination in Clinical Trials. Lecture presented at the; McMaster University 2004.

[17] Wang H, Chow SC, Li G. On sample size calculation based on odds ratio in clinical trials. J Biopharm Stat 2002; 12: 471-83. https://doi.org/10.1081/BIP-120016231

[18] Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. BMJ 1995; 311: 1145-48. https://doi.org/10.1136/bmj.311.7013.1145

[19] Rosner B. Fundamentals of biostatistics. Nelson Education 2015.

[20] Woodward M. Formulae for sample size, power and minimum detectable relative risk in medical studies. Statistician 1992; 41: 185-96. https://doi.org/10.2307/2348252

[21] Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. Can Med Assoc J 2001; 165: 1197-202.

Received on 12-06-2017 Accepted on 20-10-2017 Published on 08-12-2017

https://doi.org/10.6000/1929-6029.2017.06.04.3