

# Sample Size Estimation for Health and Social Science Researchers: The Principles and Considerations for Different Study Designs

Oladimeji Akeem Bolarinwa

Department of Epidemiology and Community Health, Faculty of Clinical Sciences, University of Ilorin, Ilorin, Nigeria

## Abstract

Sample size is one of the important considerations at the planning phase of a research proposal, but researchers are often faced with challenges of estimating valid sample size. Many researchers frequently use inadequate sample size and this invariably introduces errors into the final findings. Many reviews on sample size estimation have focused more on specific study designs which often present technical equations and formula that are boring to statistically naïve health researchers. Therefore, this compendium reviews all the common sample size estimation formula in social science and health research with the aim of providing basic guidelines and principles to achieve valid sample size estimation. The simplification of the sample size formula and detailed explanation in this review will demystify the difficulties many students as well as some researchers have with statistical formulae for sample size estimation.

**Keywords:** Health, sample size, social science, study design

## BACKGROUND

Every scientific research requires carefully designed methods to produce valid and relevant results. In achieving such results, a scientifically proven sample size estimation must be adopted. In almost all quantitative researches, sample size will be required to provide credible findings. Therefore, sample size estimation is a vital consideration at the concept development and proposal phase in research. One of the key questions health researchers are likely to ask is, how much of a population is needed for valid and reliable study? In some instances, researchers may choose to study all those within a target population. This is possible when the entire population of interest is small and there are resources to study them. This scenario is called exhaustive survey,<sup>[1]</sup> and in this instance, a sample size calculation may not be required or may not be applicable even when estimated. In most instances, it is not feasible to study the entire subjects or respondents in a population of interest. Therefore, a sample or sub-set of the population will be required.<sup>[1,2]</sup> It will be impractical to

study the entire population of interest, when there is large geographical spread of the population, when the subjects within the population are too large and when there are limited resources to study the whole population. In all these situations, a scientific method of selecting representatives of the population will be vital.

In health and social science research, scientists are often faced with challenges of estimating valid sample sizes. Many researchers frequently use inadequate sample size and this invariably introduces errors into the final findings. Taking ‘too much’ or ‘too small’ of a population sample is not only a waste of scarce resources but the researcher is also working with wrong research assumptions<sup>[3]</sup> which could possibly have ethical concerns as well. This will undermine the integrity of the outcome of the study with spurious effects on future researches that may use such outcomes. In essence, sample

**Address for correspondence:** Dr. Oladimeji Akeem Bolarinwa, Department of Epidemiology and Community Health, Faculty of Clinical Sciences, University of Ilorin, Ilorin, Nigeria.  
E-mail: [drdeji@yahoo.com](mailto:drdeji@yahoo.com)

Received: 01-02-2020,

Revised: 29-02-2020,

Accepted: 16-03-2020,

Published: 11-04-2020

### Access this article online

Quick Response Code:



Website:  
[www.npmj.org](http://www.npmj.org)

DOI:  
10.4103/npmj.npmj\_19\_20

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** [reprints@medknow.com](mailto:reprints@medknow.com)

**How to cite this article:** Bolarinwa OA. Sample size estimation for health and social science researchers: The principles and considerations for different study designs. Niger Postgrad Med J 2020;27:67-75.

size should be ‘large enough’ that an effect or precision of such magnitude as to be of scientific or clinical significance will also be statistically significant. Sample size is so important that it has evidential link with previous studies, characteristics of the population of interest, scientific assumptions, allowable study errors, sampling methods, analysis methods and study designs. Available literatures on sample size focused more on specific study designs and often present technical equations and formula that are boring to statistically naïve health researchers. This compendium reviews all the common sample size estimation formulae in social science and health research. In addition, it provides basic guidelines and principles to achieve valid estimation. The simplification of the sample size formula and detailed explanation in this review will demystify statistical formulae in sample size estimation for researchers.

## IMPORTANCE OF SAMPLE SIZE DETERMINATION IN HEALTH RESEARCH

Both internal and external validities of the research are ensured with an accurately estimated sample size that leveraged on previous studies or evidences. When representativeness in a study is accurately determined, it ensures that it measured the population attributes it purports to study. In human and animal experiment, sample size is a pivotal issue for ethical reasons. Inadequate sample size will produce scientific inference with small power. This will expose subjects to potentially harmful treatments without advancing knowledge. On the other hand, oversized experiments will recruit an unnecessarily large number of subjects into the study. This will in turn expose them to unnecessary harmful treatment. The volunteer in the study will be needlessly troubled without the study adding significant contribution to scientific knowledge.

## DYNAMICS OF SAMPLE SIZE DETERMINATION

Some researchers have classified sample size determination into four depending on the aim and procedure involved.<sup>[2]</sup> These are; sample size estimation/determination, sample size justification, sample size adjustment and sample size re-estimation. Sample size estimation/determination requires the actual calculation using scientific assumption and evidence to achieve desired statistical significance of valid and reliable outcome. This is the most common method which requires attributes such as prevalence, proportion and means from previous studies. Predetermined assumptions for validity and reliability such as power of study, level of significance and design effect (Deff) may be needed in sample size estimation.<sup>[2]</sup> Sample size justification is necessary when a sample size is already chosen. It becomes expedient for the researcher to provide a ‘statistical justification’ for the selected sample size.<sup>[2]</sup> Usually, a small size of the population will be recruited initially due to budgetary constraints or for medical consideration. A good example of this is the sample size in the first phase of clinical trials. Various methods for sample size adjustment have been described in the literature.<sup>[1-4]</sup> For reasons like small study population, e.g., for

instance a population <10,000, expected attrition or dropouts, non-response, covariates, e.g., controlling for confounders<sup>[2,3]</sup> and Deff in cluster sampling.<sup>[1,4]</sup> These adjustments are for the purpose of yielding sufficient number of analysable subjects for valid statistical findings of the health research.<sup>[2]</sup> In sample size re-estimation, there is no known or little evidence in the literature about some attributes to be studied, especially past prevalence, incidence and means. In some other instances, certain aspect of the study needs to be monitored for safety and relevance before exposing more participants to the intervention. Therefore, there may be a need for a pilot study or interim study (in clinical trials).<sup>[2]</sup> In these situations, sample size re-estimation is required to adjust for the initial sample size calculated for the pilot study and to confirm the preliminary study assumptions such as power. In this manuscript, sample size estimation, calculation and determination will be used interchangeably. Of all the four methods, sample size estimation will be discussed extensively in this review. A little note will be added towards the end of the review on sample size adjustment.

## GENERAL CONSIDERATIONS IN SAMPLE SIZE DETERMINATION

It is very important to understand the dimensions of the research to be conducted in terms of characteristics of the proposed study population, the appropriate study designs and the intended methods of analysis.<sup>[5]</sup> Characteristics of the population are relevant consideration in sample size determination. These characteristics could be human sociodemography, animal species, human body parts or system to be studied and type of health records available. Study sites’ characteristics should also be considered. Some of the study site characteristics are community setup, household, hospital or institutional-based study sites, geographic spread, confinement and security considerations. The study designs have great influence on analysis methods. As will be shown later, a good idea of the proposed study design that is appropriate for the study concept and analysis method will help define the appropriate sample size estimation for the study. Explicitly, the following study characteristics are essential to the validity of sample size determination.

### Objectives or hypothesis

The objectives, research question and hypothesis are interrelated considerations to choosing the best sample size determination.<sup>[2]</sup> For some studies, these considerations may have more than one attributes (prevalence, incidence and means) which needed to be well thought-out before estimating the sample size. For instance, the prevalence in a study that aimed at assessing the treatment outcomes and health-related quality of life of hypertensive patients attending a local hospital has more than one dependent variables, e.g., clinical outcomes and quality of life, to consider when estimating sample size. Literatures agreed that researchers should calculate for all the attributes and choose the higher or highest sample size.<sup>[2,5]</sup> Another consideration is the direction of the null hypothesis

stated. Is the hypothesis one-tail or two-tail test? This is more relevant in analytical study types, especially experimental studies and some descriptive studies. As would be discussed later, the hypothesis connects sample size and the methods of analysis of the study.

### Study designs

A properly applied study design will need appropriate sample size based on whether the study is descriptive (cross-sectional, surveys or case studies types) or analytical (observational or experimental types).<sup>[2,5]</sup> A good study requires that each of the study design has specific sample size estimation consideration. For example, a cross-sectional study that aimed at assessing the health-care utilisation pattern in a community will need not set power (1-type 2 error) for the sample size estimation. Whereas, a clinical trial that aims at assessing the effectiveness of drug X as against drug Y will be interested in setting a stringent power.

## ELEMENTS REQUIRED FOR SAMPLE SIZE

### DETERMINATION

#### Outcome variable/parameter/endpoints

In health research, units of measuring variables are of two classes. It is either numeric or categorical. These two categories have other sub-types of units of measurement. The unit of measurement in categorical variables is in proportion (percentages and rates) and at times could be in ratio. The numeric variables are presented as means and median mostly (measures of central tendency). In some health researches, odd ratio (OR) and relative risk are also measured as outcome variables. The chosen unit of measurement in sample size estimation should be taken into consideration at all time.<sup>[4,6]</sup> A previous literature that uses the same or similar unit of measurement for the variable should be adopted for the sample size estimation. However, in some instances, a variable could be interpreted in more than one unit of measurement in health research. For example, blood pressure (BP) can be expressed as a mean value in mmHg. It can also be reported as controlled BP or uncontrolled BP. Another classification of BP could be optimal, Grade I, Grade II or Grade III.

#### Variability of the parameter

This is the measure of how spread out or dispersed individual unit in a variable is from the middle. The wider the variability, the more sample size that will be required to achieve a significant effect size if any. The reason is that any two highly dispersed variables being compared will overlap.<sup>[5]</sup> For the numeric parameters, the measures of dispersion for a sample mean is variance (standard deviation), whereas for median is range (interquartile range). These are usually reported by previous literatures and available for the researcher to leverage on to estimate the study sample size. However, for categorical parameter, the variability for sample proportions is based on spread towards 0.5 (or 50%). If a previous study reported a prevalence of 0.5 (50%), the dispersion will also equal 0.5 (that is 1–0.5). A prevalence tending towards 50%

indicates maximum variability.<sup>[7]</sup> The prevalence moving towards extreme of the spectrum 100% (or 1) and 0 will not have as much variability. This simply means that majority of the sample population possess or do not possess the attribute of interest.<sup>[7]</sup>

#### Detectable difference (effect size) of the parameter

This is the smallest clinical effect that is detectable in the finding.<sup>[5,8]</sup> It is a parameter that elicits the difference in the outcome of one arm of study (intervention, experimental or study group) to the other arm (control or comparator). It is the attribute of analytical studies which determines the probability that an independent factor will be strongly associated with an outcome or dependent variable.<sup>[5]</sup> Depending on the unit of measuring the outcome variables, effect size could be mean difference or change in the proportion. It is expedient to mention that effect size is interrelated to the hypothesis set at the beginning of the research, the outcome measurement and clinically detectable difference in the outcome measurement. As a general rule of thumb, a small effect size will require a large sample size to be able to detect a clinically meaningful difference, whereas a large effect size will require a small sample size.<sup>[4,5]</sup> The effect sizes to input in sample size estimation are often obtained from previous research.

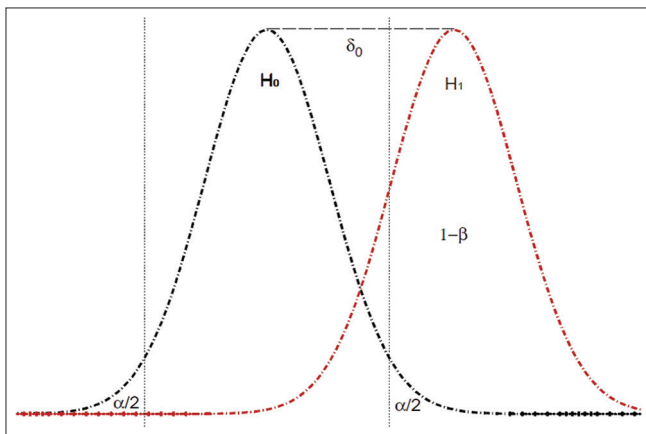
Three variants of detectable difference have been described in the literature.<sup>[2]</sup> Absolute difference means that a clinically acceptable effect size can be presumably set for the study. For instance, a difference of 5 mmHg can be presumed to be clinically acceptable between a new and the existing drug for hypertension treatment. Relative difference requires that researcher set the study to detect certain change in proportion of a clinical outcome. For example, a 10% decrease in systolic BP can be set to be of practical importance (20%–30% is usually taken as clinically acceptable). Cohen, decades ago, established that for an experimental (interventional) study with 2 arms of comparison, a ratio of effect size and standard deviation termed standardised effect size or standard difference can be applied.<sup>[8,9]</sup> The standardised effect size was classified as small, medium or big if this ratio is 0.2, 0.5 and 0.8, respectively.<sup>[8]</sup>

#### Error rates

The concept of error assumption in research stemmed from the hypothesis testing.<sup>[2,5,8]</sup> The type of error committed when researcher wrongly rejects a null hypothesis that is true is called type I or alpha ( $\alpha$ ) error. This is also described as ‘failure to accept a true null hypothesis’.<sup>[2,5,8]</sup> On the other hand, type II or beta ( $\beta$ ) error means to wrongly accept a false null hypothesis. It is also described as ‘failure to reject a false null hypothesis’.<sup>[2,5,8]</sup> The implication of type I error ( $\alpha$ ) is that the researcher has to set an assumption for the level of type I error he/she wishes to allow in the study. This assumption of type I error is also called setting ‘level of significance ( $P$  value)’. It is frequently set at 5% which means the researcher is willing to allow the 5% probability of ‘failure to accept a true null hypothesis’. However, some researches such as clinical trials can set a very small  $\alpha$ -error. The smaller the  $\alpha$ -error, the larger

the sample size required.<sup>[8]</sup> The level of significant thereby means that at less than 5% ( $P = 0.05$ ) or 1% ( $P = 0.01$  in stringent trials) of error, the variations observed in the outcome are due to chance and not due to ‘too much error’.<sup>[10]</sup> An important caution here is that majority of the analysis software like SPSS, set  $P$ -value at 0.05 as a default. Consequently, if there is a need to use  $P$  value lower than 5%, the researcher needs to change this from the software setting to the desired value. Otherwise, the researcher’s assumption of  $P$  value of 1% could be erroneously presenting the result at  $P$  value of 5%. Another note of relevance is that when researcher fails to reject null hypothesis, it does not mean that it is true, it is just that there is not enough evidence to reject the null hypothesis.<sup>[10]</sup>

Type II error ( $\beta$ , beta error) on the other hand gives rise to ‘power’ of the study which is  $1-\beta$ .<sup>[2,5,8,10]</sup> The power of the study therefore means the other proportion left behind after removing the errors committed by wrongly accepting a false null hypothesis [Figure 1]. This connotes a proportion of rightly rejected false null hypothesis.<sup>[2,5]</sup> Power of the study is often assumed or set at the proposal stage similar to the level of significance. For example, suppose a researcher assumes a 20%  $\beta$ -error, the power of the study will be set at 80%.<sup>[2,5,8]</sup> Random values of 0.05 for  $\alpha$  and 0.2 for  $\beta$  (power, 0.8) are often used by researchers, but conventionally,  $\alpha$  values could range from 0.01 to 0.10, whereas  $\beta$  can be set between 0.05 (power, 0.95) and 0.20 (power, 0.80).<sup>[5]</sup> Like the  $\alpha$  error, the lower the  $\beta$  (higher power), the larger the sample size is required to achieve clinically detectable changes in the outcome.<sup>[2,5,8]</sup> As applicable to the actual sample size estimation formula, the values of  $\alpha$  and  $\beta$  cannot be used directly. This required conversion on the standard normal deviate in the Gaussian curve.<sup>[8]</sup> This is called the  $Z$ -scores denoted as  $Z^\alpha$  and  $Z^\beta$  for  $\alpha$  and  $\beta$  errors, respectively [Table 1]. Fianlly, a few clarification need to be stated about the relationship between confidence level and  $\alpha$ -error. Similar to the power of the study, confidence level simply means the other proportion left behind after removing the  $\alpha$ -error ( $1-\alpha$ ) usually set as 0.95 as shown in Figure 1.<sup>[11]</sup> It is the precision of the study which means the confidence of not rejecting a true null hypothesis.<sup>[2]</sup> For analytical studies,



**Figure 1:** The relationship between type 1 and type 2 errors as they relate to the hypothesis<sup>[11]</sup>

setting a confidence interval (CI) means that the interval of the width of the confidence level will be estimated during analysis.<sup>[2]</sup> The CI like the  $P$  value indicates the statistical significance of the study outcomes.

## SAMPLE SIZE ESTIMATION FOR DIFFERENT STUDY DESIGNS AND STATISTICAL ANALYSIS

### Cross-sectional studies and surveys

Prevalence studies and surveys are descriptive in nature. They are employed to show the associations between factors and generated hypothesis for future researches.<sup>[4]</sup> Estimating sample size for these type of research requires outcomes/variables/parameters such as prevalence, incidence, means, rates and ratios. Out of all these, prevalence ( $p$ ) and means ( $\mu$ ) are commonly used for outcomes that are categorical (qualitative) or numeric (quantitative) in nature. The variability for each of  $P$  ( $1-p$ ) and  $\mu$  (variance =  $\sigma$ ), normal standard deviate for  $\alpha$ -error ( $Z^\alpha$ ) and a precision level ( $\delta$ ) usually assumed at 5% (0.05) are all required. The followings depict the formula for both the categorical and numeric outcome variable cross-sectional studies:<sup>[4,6,8,12]</sup>

- a. Categorical outcome (proportion) 
$$\text{Sample size } (N) = \frac{Z^2 Pq}{d^2}$$
- b. Numeric outcome (mean). 
$$N = \frac{Z_\alpha^2 \sigma^2}{\delta^2}$$

### Analytical studies: Independent case-control and cohort studies

In this type of studies, there are comparator groups called ‘controls’ that are weighed against the group with the attributes been studied called ‘cases’. While the case-control study captures the cases with outcome (disease or other health related issue) and search retrospectively to determine the exposed factors, the cohort study starts from exposed factors and follow the cohort prospectively to determine the associated outcomes. Only few studies have extensively documented sample size

**Table 1: Commonly used standard normal deviate for  $\alpha$  and  $\beta$  errors**

	$\alpha$	Z
Direction of $H_0$ testing		
$\alpha$ -error		
Two-tail	0.05	1.960
Two-tail	0.025	2.326
Two-tail	0.01	2.576
One-tail	0.05	1.645
One-tail	0.025	1.960
One-tail	0.01	2.326
$\beta$ -error		
	$1-\beta$ (power)	
0.40	0.60	0.25
0.20	0.80	0.84
0.10	0.90	1.28
0.05	0.95	1.64
0.01	0.99	2.33

formula for case-control and cohort studies.<sup>[6,7,13]</sup> Other study variants' formula (such as matched and paired studies) can be found in some other literature<sup>[7]</sup> and internet sources. Formulae for independent studies are shown in this review.

c. Independent case-control (retrospective study).<sup>[7,13]</sup>

$$N = \frac{[Z_\alpha \sqrt{(1+m)P^*(1-P^*)} + Z_\beta \sqrt{P_1(1+P_1) + mP_0(1-P_0)}]^2}{(p_1 - p_0)^2} \quad (1)$$

$$P^* = \frac{P_1 + p_0 / m}{1 + 1/m} \quad (2)$$

$$P_1 = \frac{p_0 \omega}{1 + P_0(\omega - 1)} \quad (3)$$

$$N_c = \frac{N}{4} \left( 1 + \sqrt{1 + \frac{2(m+1)}{Nm | p_0 - P_1 |}} \right)^2 \quad (4)$$

In equation C (1),  $N$  is the estimated sample size for the independent case-control,  $Z_\alpha$  is the standard normal deviate for  $\alpha$  error and  $Z_\beta$  is the standard normal deviate for power  $(1 - \beta_{\text{error}})$ .  $P^*$  is the average probability of the exposure (similar to pooled variance or proportion) calculated as shown in formula C (2).  $m$  is ratio of control subjects to case subjects desired, while  $P_1$  is the probability of exposure in the control group, calculated in equation C (3) from known prevalence of the exposure from the population ( $P_0$ ) and OR ( $\omega$ ) of the exposure between cases and control.<sup>[7]</sup> As shown in C (4) formula,  $N_c$  is the continuity-adjusted sample size for further analysis such as Chi-square and Fisher's exact, taking into consideration the ratio of control to case, prevalence in the population and probability of the exposure.<sup>[7]</sup> When OR ( $\omega$ ) is not available but only prevalence is available, a more simple alternative formula is prescribed:<sup>[13]</sup>

$$N = \frac{m+1}{m} \frac{\left( Z_\beta - Z_\alpha \frac{1}{2} \right)^2 P^*(1-P^*)}{(P_0 - P_1)} \quad (5)$$

d. Independent cohort (prospective study)<sup>[7,13]</sup>

$$N = \frac{[Z_\alpha \sqrt{(1+1/m)P^*(1-P^*)} + Z_\beta \sqrt{P_0(1-P_0)/m + P_1(1-P_1)}]^2}{(p_0 - p_1)^2} \quad (1)$$

$$P^* = \frac{P_1 + mp_0}{1+m} \quad (2)$$

$$N_c = \frac{N}{4} \left( 1 + \sqrt{1 + \frac{2(m+1)}{Nm | p_0 - P_1 |}} \right)^2 \quad (3)$$

In equation d (1),  $N$  is the estimated sample size for the independent case-control,  $Z_\alpha$  is the standard normal deviate for  $\alpha$  error and  $Z_\beta$  is the standard normal deviate for

power  $(1 - \beta_{\text{error}})$ .  $P^*$  is the average probability of the exposure calculated as shown in formula d (2).  $m$  is the ratio of control subjects to cohort or experimental subjects desired, while  $P_0$  is the probability of event in the control group and  $P_1$  is the probability of the event in the study or experimental group.<sup>[7]</sup> As shown in d (3) formula,  $N_c$  is the continuity-adjusted sample size for further analysis such as Chi-square and Fisher's exact.<sup>[7]</sup>

**Analytical studies: Cross-sectional analytical comparative studies**

These are various types of observational study that compare population proportions ( $P_1$  and  $P_2$ ) and means ( $\mu_1$  and  $\mu_2$ ). It is formerly known as 'comparative study'. In this study, there is no form of intervention or experimentation. For instance, a study that aimed at comparing the cardiovascular risk score between the residents in rural and urban communities. The formula for cross-sectional analytical study can be applied to categorical and numerical variables as shown below:<sup>[4,8,12-14]</sup>

d. Comparing two proportions

$$N = \frac{(Z_\alpha + Z_\beta)^2 [P_1(1-P_1) + P_2(1-P_2)]}{(P_1 - P_2)^2}$$

f. Comparing two means  $N = \frac{(Z_\alpha + Z_\beta)^2 2\sigma^2}{(\mu_1 - \mu_2)^2}$ .

**Analytical studies: Randomised controlled trials**

There are four variants of randomised control trials (RCT) described in the literature<sup>[10,15]</sup> as shown in Table 2:

1. Equality trial: ( $H_0: \mu_T - \mu_S = 0$ ). This trial is designed to hypothesise that there is no clinical difference or effect between the mean of the new treatment/intervention ( $\mu_T$ ) and the mean of the comparator ( $\mu_S$ )
2. Equivalence trial: ( $H_0: |\mu_T - \mu_S| = \delta$ ). This trial hypothesises that both the treatment/intervention and the comparator ( $\mu_T$  and  $\mu_S$ ) are equally effective
3. Non-inferior trial: ( $H_0: \mu_T - \mu_S \geq \delta$ ). It is a design to prove that the treatment/intervention is as effective as the comparator and not necessary better than comparator (standard or usual or placebo)
4. Superiority trial: ( $H_0: \mu_T - \mu_S \leq \delta$ ). The purpose of this design is to prove that the treatment/intervention is more effective (statistically or clinically) than the comparator (standard or usual or placebo).

The trials can also be one-sided (one-tail) hypothesis. This means that the direction of the difference or the effect is stated (more/greater or less/lower than). More commonly, many researchers prefer to adopt two-sided (two-tail) hypothesis which usually do not state the direction of the differences or effects expected. This states that there is no difference between the effect of the treatment/intervention and the comparator (standard/usual/placebo), and the common analysis method is independent  $t$ -test. In addition to the direction of the hypothesis, the design variants of the trials such

**Table 2: Sample size considerations for common types of randomised control trials**

Design	Hypothesis	Sample size estimation rule	
		Numeric	Categorical
One-tail	Equality	$\frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_T - \mu_S)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{(P_T - P_S)^2}$
	Superior	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{(P_T - P_S)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{(P_T - P_S - \delta)^2}$
	Equivalence	$\frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{( \mu_T - \mu_S  - \delta)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{( P_T - P_S  - \delta)^2}$
Two-tail parallel	Equality	$\frac{2(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_T - \mu_S)^2}$	$\frac{2(Z_\alpha + Z_\beta)^2 p(1-p)}{(P_T - P_S)^2}$
	Non-inferior	$\frac{2(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_T - \mu_S - \delta)^2}$	$\frac{2(Z_\alpha + Z_\beta)^2 p(1-p)}{(P_T - P_S - \delta)^2}$
	Superior	$\frac{2(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_T - \mu_S + \delta)^2}$	$\frac{2(Z_\alpha + Z_\beta)^2 p(1-p)}{(P_T - P_S + \delta)^2}$
	Equivalence	$\frac{2(Z_\alpha + Z_\beta)^2 \sigma^2}{( \mu_T - \mu_S  - \delta)^2}$	$\frac{2(Z_\alpha + Z_\beta)^2 p(1-p)}{( P_T - P_S  - \delta)^2}$
Two-tail crossover	Equality	$\frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{2(\mu_T - \mu_S)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{2(P_T - P_S)^2}$
	Non-inferior	$\frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{2(\mu_T - \mu_S - \delta)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{2(P_T - P_S - \delta)^2}$
	Superior	$\frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{2(\mu_T - \mu_S + \delta)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{2(P_T - P_S + \delta)^2}$
	Equivalence	$\frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{2( \mu_T - \mu_S  - \delta)^2}$	$\frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{2( P_T - P_S  - \delta)^2}$

as the parallel, cross-over and cluster RCTs also have effects on the sample size calculation as shown in Table 2.<sup>[2,6,10,15]</sup>

$\sigma^2 = \text{pooled variance} = \frac{(\sigma_T + \sigma_S)^2}{2}$  where  $\sigma_T$  is the variance of the treatment group and the  $\sigma_S$  is the variance of the comparator group or  $\sqrt{\frac{(S_T + S_S)^2}{2}}$  if standard deviation is given for the treatment ( $S_T$ ) and comparator ( $S_S$ ) groups. Alternatively, a more comprehensive pooled standard variation ( $S_{\text{pooled}}$ ) calculation has been suggested<sup>[11]</sup> =  $\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

keeping in view the standard deviations ( $s_1, s_2, \dots$ ) and sample sizes ( $n_1, n_2, \dots$ ) of the groups.  $P$  is also a pooled prevalence and is simply  $P_T + P_S/2$ .  $P_T$  and  $P_S$  are the prevalence of the outcomes in the treatment and the comparators, while  $\mu_T$  and

$\mu_S$  are the mean outcomes in the treatment and the comparator groups. Clinically acceptable margin effect is denoted as  $\delta$  in the above equation.

## OTHER SAMPLE SIZE CONSIDERATION IN RANDOMISED CONTROL TRIALS AND INTERVENTIONAL STUDIES

### Cluster randomised control trials designs

For a detailed explanation on sample size considerations on cluster RCTs, standard reviews should be consulted.<sup>[15,16]</sup> However, a brief and helpful explanation is provided here from existing literature.<sup>[15,16]</sup>

The initial step is to follow the appropriate sample size estimation  $N$  for RCT over individuals as shown in Table 1, and then corrections will be considered for the  $\kappa$  number of clusters in each arm of size  $m$ . This will produce a total number of  $N_c = m\kappa$  individuals in each arm. As a rule of thumb, to

compensate for the selection error inherent in cluster sampling, there is a need to inflate the variance of the difference ( $\delta_c$ ) to be detected by a variance inflation factor (VIF). How well individuals in the clusters are correlated to each other known as the intra-cluster correlation coefficient ( $\rho$ ) is important when multiplying with VIF. This is called Deff.

Therefore,  $VIF = [1 + (m-1)\rho]$ .

There are times that the cluster sizes are not equal, then  $VIF = [1 + ((\delta_v^2 + 1)m^*-1)\rho]$ .

The  $\delta_v$  means the coefficient of variation of the cluster sizes and  $m^*$  represents average cluster size. Substituting the multiplier for VIF in any of the individual RCT formula is:

$N_c = N [1 + (m-1)\rho] = N[VIF]$  – for equal cluster size

$= N [1 + ((\delta_v^2 + 1)m^*-1)\rho]$  – for unequal cluster sizes.

### QUASI-EXPERIMENTAL STUDIES

One good example of quasi-experimental study is pre- and post-test or before and after test. This is also described as repeated measure. Another description of this situation is that each subject is serving as his/her own control. Repeated measures analyses such as paired *t*-test (for numeric) and McNemar test (categorical) are employed for the analysis of these forms of study as shown below:<sup>[11,12]</sup>

$$\text{Numeric: } N = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\delta)^2}$$

$$\text{Categorical } N = \frac{(Z_\alpha + Z_\beta)^2 p(1-p)}{(p_1 + p_2)^2}$$

It looks very similar to the two-sample situation, but with two important changes. First, there is no multiplier of ‘2’. Second, the  $\sigma$  is the standard deviation of the differences within pairs, while  $\delta = \mu_1$  and  $\mu_2$  are the means before and after intervention, respectively.<sup>[11,12]</sup> Similarly,  $p_1$  and  $p_2$  are the proportion/prevalence before and after intervention. The  $P$  is the pooled prevalence of the before and after prevalence. The  $\sigma$  is the variance of the difference in the repeated measure  $= \sigma_1^2 + \sigma_2^2 - \rho\sigma_1\sigma_2$ <sup>[11,12]</sup> where  $\rho$  is the correlation between baseline and post-intervention values on the same group. If only one  $\sigma_1$  is reported, then  $\sigma = 2\sigma_1(1-\rho)$ .

### SURVIVAL ANALYSIS (OUTCOME) STUDY

This type of study is carry out when research subjects are followed up over a time to generate outcome variable that is of time-to-event type.<sup>[12]</sup> A good example of this is in the clinical trial that set out to compare the survival rates of the experimental drug or an intervention group compared to the control (non-experimental) group. One striking feature of survival study is that by design, it is not every research subject that survive to the end of the study.<sup>[12]</sup> Hence, research subjects exit at different points along the follow-up period. Log-rank

test is mostly applied to this type of analysis, thereby making it expedient to take differential total number of events into consideration.<sup>[12]</sup> Therefore, both the sample size estimation and duration of stay in the study are important considerations for this type of study design.<sup>[12]</sup> The first consideration is the number of events ( $d$ ) estimated using the  $\alpha$ -error, the power ( $1-\beta$ ) and effect size or the treatment effect ( $\delta$ ). However, the treatment effect is embodied by the probability of the occurrence of the events in the two study groups.<sup>[12]</sup> This probability is termed ‘hazard ratio’ (HR).

The total number of events can be estimated as:

$$d = (Z_\alpha + Z_\beta)^2 \left( \frac{1 + \delta_0}{1 - \delta_0} \right)^2 \quad (1)$$

$$\text{The } \delta_0 \text{ is equal to } HR = \frac{\log(P_e)}{\log(p_c)} \quad (2)$$

The  $p_e$  and  $p_c$  are the estimated survival probability in the experimental and control groups, respectively.

$$\text{The final sample size required } N = \frac{2d}{2 - p_c - p_e} \quad (3)$$

### SAMPLE SIZE CONSIDERATION IN CORRELATION AND DIAGNOSTIC TESTS

#### Correlational studies

Despite being a common descriptive study, only few literature<sup>[5]</sup> have described sample size estimation in correlational study. In this study type, the main focus is the correlational coefficient ( $r$ ) and the Fisher’s transformation of the correlation coefficient ( $C_r$ ).

$$\text{One sample correlation formula: } N = \left\{ \frac{(Z_\alpha Z_\beta)}{C_r} \right\}^2 + 3.$$

$$\text{where } C_r = \frac{1}{2} \ln \left\{ \frac{1+r}{1-r} \right\}.$$

$$\text{Two sample correlation formula: } N = \left\{ \frac{(Z_\alpha Z_\beta)}{C_{r1} - C_{r2}} \right\}^2 + 3.$$

$$\text{where } C_{r1} = \frac{1}{2} \ln \left\{ \frac{1+r_1}{1-r_1} \right\} \text{ and } C_{r2} = \frac{1}{2} \ln \left\{ \frac{1+r_2}{1-r_2} \right\}.$$

#### Accuracy tests (sensitivity/specificity)

Further detail reading can be found in the literature.<sup>[17]</sup> For the purpose of this review, a simple and an all-purpose formula is given here:<sup>[17]</sup> the sensitivity ( $S_e$ ), specificity ( $S_p$ ), disease prevalence ( $P$ ) and precision ( $\delta$ ) are all required.

Sample size when the aim of the accuracy test is for single sensitivity or specificity:

$$\text{Sensitivity } (S_e) = \frac{Z_\alpha^2 S_e (1 - S_e)}{\delta^2 (P)}$$

$$\text{Specificity } (S_p) = \frac{Z_\alpha^2 S_p (1 - S_p)}{\delta^2 (1 - P)}$$

Sample size for sensitivity (or specificity) of a single diagnostic test in comparison with a standard: The comparison is of the value of the sensitivity/specificity ( $P_1$ ) of a diagnostic test been compared with a predetermined or a gold standard sensitivity/specificity ( $P_0$ ).

$$N = \frac{[Z_{\alpha}\sqrt{P_0(1-P_0)} + Z_{\beta}\sqrt{P_1(1-P_1)}]^2}{(p_0 - p_1)^2}$$

Sample size for a Sensitivity (or specificity) of more than one diagnostic tests: the comparison in this design involves two alternative diagnostic tests ( $P_1$  and  $P_2$ )

$$N = \frac{[Z_{\alpha}\sqrt{2xP^*(1-P^*)} + Z_{\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)}]^2}{(p_1 - p_2)^2}$$

### Sample size adjustments

There are various reasons that can warrant adjustment for an initially estimated sample size.

#### Multiple outcome variables

When there are more than one outcome variables of interest in a study, sample size of each of all these variables should be estimated and the highest of them should be applied for the study.<sup>[8,15]</sup>

#### Unequal comparison group

Some researches have comparison group which may have equal or unequal subjects per group. In this instance that the arms of the study have unequal subjects in the group, it become expedient to adjust the initially calculated sample size ( $N$ ) that assumed that the arms of study are equal,<sup>[8]</sup> using the actual ratio between the unequal arms of the research ( $m$ ).

$$\text{The adjusted sample size} = N^* = \frac{N(1+n)^2}{4n}$$

#### Non-consent, missing response, withdrawal from study and dropout

Sample size is calculated as a minimum number required to achieve research aim. In practice, reasons ranging from incomplete response to loss to follow-up ( $N^*$ ) can adversely affect the final sample size that is useful for the research.<sup>[8,15]</sup> Researcher should have adequate knowledge of these losses and have good idea of the proportion ( $P$ ) that may be lost to any of these in a study.

$$\text{Therefore } N^* = \frac{N}{(1-P)}$$

#### Finite population correction

Logically, searching for a few coloured grains of corn in a large bowl will take longer than finding same coloured grains in a handful scoop of corn. After estimation of sample size for a population of less than 10,000 ( $N_0$ ), need arises for the researcher to correct the sample size ( $N$ ) for the small study population.<sup>[7]</sup>

$$N^* = \frac{N}{1 + \frac{N}{N_0}}$$

#### Design effects

The cluster trials design and the VIF have been discussed in detail in the preceding section. It should be noted that stratified sampling has similarly Deff like cluster randomisation and should be corrected as well.<sup>[8]</sup>

#### Multivariate analysis and covariates

More advanced analysis and modelling are being frequently used in health research nowadays; some of these analyses such as analysis of covariance, log-linear analysis and cox's proportional hazard analysis will require sample size adjustments.<sup>[8]</sup> Proper methods of doing these are still evolving.<sup>[8]</sup>

## CONCLUSION

This review discussed common sample size estimation formula in health research and offers basic guidelines and principles to achieve valid estimation. The simplification of the sample size formula and detail explanation were also provided. Sample size estimation is an important step in conducting a valid and generalisable research. The variable of outcomes, research designs, analysis methods, error assumptions and effect size among other important elements are cardinal to estimating a scientifically correct sample size. Certain situations require adjustment for the sample size and they are to be considered at all times in health research. This compendium will ease the struggles student and young researchers go through to deploy scientifically strong sample size estimation in their studies.

#### Financial support and sponsorship

Nil.

#### Conflicts of interest

There are no conflicts of interest.

## REFERENCES

- Umulisa C. Sampling methods and sample size calculation for the SMART methodology. University; 2012;2:20-30.
- Chow SC, Shao J, Wang H, Lokhnygina Y. Sample Size Calculations in Clinical Research: Chapman and Hall/CRC Biostatistics Series. 3<sup>rd</sup> ed. New York: Taylor and Francis; 2017.
- Lenth RV. Some Practical Guidelines for Effective Sample-Size Determination; 2001. p. 1-11.
- Habib A, Johargy A, Mahmood K, Humma H. Design and determination of the sample size in medical research. IOSR J Dent Med Sci (IOSR-JDMS). 2014;13:21-31.
- Warren SB, Thomas BN, Hulley SB. Estimating Sample Size and Power: Applications and Examples. In: Hulley SB, Cummings SR, Browner WS, Grady DG GN, editors. Designing Clinical Research. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2013. 44-55. Available from: [https://www.academia.edu/36931058/Designing\\_Clinical\\_Research](https://www.academia.edu/36931058/Designing_Clinical_Research). [last Accessed on 2020 Jan 24].
- Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013;35:121-6.



7. Kasiulevicius V, Šapoka V, Filipaviciute R. Sample size calculation in epidemiological studies. *Gerontologija* 2006;7:225-31.
8. Hazra A, Gogtay N. Biostatistics series module 5: Determining sample size. *Indian J Dermatol* 2016;61:496-504.
9. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2<sup>nd</sup> ed. Hillsdale, NJ: Erlbaum; 1988.4-17.
10. Zhong B. How to calculate sample size in randomized controlled trial? *J Thorac Dis* 2009;1:51-4.
11. Shintani A. Sample Size Estimation and Power Computation on Paired or Skewed Continuous Data; 2006. p. 1-15.
12. van der Tweel I. Sample size determination. Intern report. 2006;2-15.
13. Sharma A. Sample Size Calculation for Research Studies in Ophthalmology; 2015. p. 78-81. Available from: <https://pdfs.semanticscholar.org/efd7/00bb717ce68bd9474d43e05d0a7f48920422.pdf>. [last Accessed on 2020 Jan 24].
14. 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.
15. Thabane L. Sample Size Determination in Clinical Trials HRM-733 CLass Notes; 2004. p. 31. Available from: <http://www.lehanathabane.com>. [last Accessed on 2020 Jan 24].
16. Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Med Res Methodol* 2011;11:102.
17. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193-204.