Ann Ibd. Pg. Med 2025. Vol.23, No.3 8-17

A NARRATIVE REVIEW OF SAMPLE SIZE ESTIMATION FOR QUANTITATIVE EPIDEMIOLOGICAL AND CLINICAL STUDY DESIGNS

T.M. Ipinnimo

Department of Community Medicine, Federal Teaching Hospital, Ido-Ekiti, Nigeria.
Department of Community Medicine, Afe Babalola University, Ado-Ekiti, Nigeria.

Correspondence:
Dr. T.M. Ipinnimo
Dept. of Community Medicine,
Federal Teaching Hospital,
Ido-Ekiti, Nigeria
E-mail: abbeymagnus@yahoo.com

Submission Date: 13th Aug., 2025
Date of Acceptance: 26th Dec, 2025
Publication Date: 31st December., 2025

ABSTRACT

Introduction: **Sample size estimation is a crucial step in conducting valid and reliable research. It is the most common question researchers ask biostatisticians, as these investigators often face the problem of estimating the correct sample size for their research.**

Aim: **This article aims to discuss the methods of sample size estimation in quantitative epidemiological and clinical research, presenting key principles and commonly used techniques in a simplified manner.**

Discussion: **Unlike the previously published reviews, this article presents simple, straightforward, and easy-to-understand sample size formulas as well as basic principles on sample size estimation of common study designs in epidemiological and clinical research. In addition, explanations were provided on different statistical theories and concepts, such as outcome variable, confidence level, power, and effect size, which are important considerations in estimating scientifically appropriate sample size.**

Conclusion: **Based on the descriptions and simplified formulas in this article, researchers can easily estimate a sample size that aligns with their study design and outcome measures, enabling more robust and reliable estimates and ensuring that the findings can be generalised to the wider population.**

Keywords: Determination, Estimation, Formula, Research, Sample size, Study design

INTRODUCTION

Sample size estimation is an important aspect of the design stage of scientific research, but researchers, including students, are faced with the problem of determining a valid sample size for their study. Manuscripts are often rejected by journals because of insufficient sample size[1] and many studies published in international publications revealed that sample size computations were not reported appropriately.[2] Based on my professional experience providing methodological support for research projects, questions regarding sample size estimation are among the most frequently raised by researchers. This observation is consistent with previously published literature.[3] In order to obtain accurate and relevant results, a researcher must know the number of subjects to include in their study. If the sample size is too small, even the most meticulously implemented study may fail to achieve its objectives. Apart from a too-small sample size, choosing a too-large sample than needed is also a waste of valuable scarce resources.

Sample size determination is a combination of both practical and scientific considerations. For instance, the upper limit of the sample size is set by the availability of resources, while the lower limit is set by the required accuracy.[4] The sizes of the population sample should be enough to allow for appropriate bivariate and multivariate analysis, provide the desired level of accuracy, and allow for the validity of the significance test[4] When the entire population of interest is not large, researchers may sometimes decide to conduct an exhaustive survey and study the whole population, especially when the time and resources are available.[5] It is common to find researchers apply a sample size formula intended for one study design to a different study design. Investigators need to understand that sample size estimation methods vary by study design and no one formula for calculating sample size can be used for all study designs.[2] Furthermore, several articles have been published discussing sample size estimation, but most of them use statistical equations and challenging illustrations that may be confusing and difficult to understand by non-statisticians. In addition, most of the past literature typically discusses the sample size estimation of a single study design.[5] This article, therefore, aims to bridge the highlighted gaps by discussing the sample size estimation methods of various quantitative study designs in epidemiological

and clinical research as well as simplifying them for researchers who may not be biostatistics experts, including students, for their easy comprehension. In addition, links to user-friendly online tools and sample size calculators are provided to support practical application. With the combination of the new tips and explanations of basic statistical knowledge in this paper, researchers should be able to utilize the tools provided to accurately estimate the sample size of their studies.

METHODS

This narrative review explores the sample size estimation technique in epidemiological and clinical research. A review of published literature was conducted after searching and identifying relevant articles using PubMed, Google Scholar, and Google. Keywords used in these search engines were, but not limited to, sample size estimation, sample size calculation, sample size determination and sample size formula. Relevant research articles discussing aspects of sample size estimation in epidemiological and clinical research were included for review. Printed literature and textbooks found appropriate were also incorporated. A total of 23 articles and books were used in this review. Data on statistical theories and concepts, as well as sample size formulae for the various study designs, were extracted and presented in a simple and easy-to-understand format, drawing on the author's professional experience in supporting multiple research projects in epidemiology and public health.

Statistical theories, concepts and considerations of sample size determination

The following statistical theories and concepts are essential and must be taken into consideration when estimating the sample size required for the validity of a study:[3–10]

Study designs

Sample size estimation is based on the study design. These study designs can be broadly categorized into two types. The first type is observational studies where no intervention is given, such as descriptive cross-sectional, comparative cross-sectional, case-control and cohort studies. The second type is experimental or interventional studies, which involve deliberate manipulations of variables to evaluate their effects, exemplified by randomized controlled trials (RCTs) and quasi-experimental studies. The descriptive cross-sectional measures the parameter of a group in the population at one point in time "snapshot", while the analytical studies (comparative cross-sectional, case-control and cohort) compare parameters from two groups. The diagnostic accuracy studies, a type of observational study, which evaluates the validity of a screening test in discriminating between diseased from healthy individuals, have been reported in medical literature. It is important to note that each of these study designs has a specific population sample size estimation method.

Null and alternative hypotheses

Experimental and observational analytical studies use the null (H_0) and alternative (H_A) hypotheses to state predictions. These hypotheses are formulated in line with the research questions and objectives. The H_0 is the statistical hypothesis under test and it is stated in a way that does not prejudge the phenomenon of interest, such as "there is no significant difference" or "there is no association". The H_A on the other hand, contradicts the H_0 and it is the conception of the researcher that led to the research work. It is usually stated opposite to the H_0, such as "there is a significant difference" or "there is a significant association" and it is the conclusion drawn when the data provide sufficient evidence to reject the H_0. The sidedness of the hypothesis, which could be directional as seen in a one-tailed hypothesis that looks for either "increase" or "decrease" in outcome, or non-directional as seen in a two-tailed hypothesis that looks for a change in outcome (that is effect in both directions) is relevant in the sample size determination as well as in statistical analysis of RCT and analytical studies.

Quantitative and qualitative outcome variable

The outcome variable, also called the dependent variable, consists of the parameters that the researcher intends to measure in a study. This primary outcome variable, which could be numerical/quantitative or categorical/qualitative, is summarized with mean/median and rates/ratios/proportions, respectively, and they form the key components of the sample size formulas. The sample size determination formula for a qualitative outcome variable is different from that of a quantitative outcome variable. It is therefore vital for researchers to adopt the right formula and search the literature for similar outcome variable type measures before commencing their sample size calculation.

Additionally, a study may have more than one outcome variable of interest to take into consideration when calculating sample size. In this case, the researcher is expected to calculate for all the outcome variables and select the largest sample size. This is because this minimum sample size will ensure the study has adequate power to detect a significant effect for all the outcome variables under consideration.

Population estimates (prevalence and means)

It is also important to consider the parameters, such as the prevalence rate and expected mean value of the

outcome of interest in the study population, when calculating sample size. This could be gotten from pilot or previous studies; however, if this is not available, the investigator can try to make an informed assumption or assume a 50% in case of prevalence rate. The assumed 50% maximizes the expected variance and gives a sample size that is sure to be large enough. The more the variability or dispersion of a parameter, the larger the sample size needed to get a significant effect size. For the prevalence rate, the dispersion is calculated by multiplying the population estimate by 100% minus the population estimate, that is, p(1–p), while the dispersion for the mean and median is standard deviation (SD) and interquartile range, respectively.

Alpha error, confidence level and degree of accuracy (precision)

Alpha error (á-error), otherwise known as type I error, is committed when a true H_0 is wrongly rejected. In the presence of enough evidence, a false H_0 should be rejected, and a true H_0 should not be rejected, that is, accepted (although the word "accepted" is not conventionally used in statistics) to avoid á-error [Table 1]. This error is described as the false positive rate (one minus specificity) in the evaluation of a screening test [Table 1]. Alpha error is inversely related to the precision or degree of accuracy, such that a lower á-error contributes to increased precision; however, true precision is more closely tied to the width of the confidence intervals, which depend not only on á-error but also on variability and sample size.

A researcher must make assumptions about the level of type I error they will tolerate in their study. This is known as the level of significance and is usually set at 5% (0.05), which means that the researcher is willing to allow that there is 5% probability that the results observed may be due to chance (or 5% probability of failure to "accept" a true H_0). If a researcher seeks a higher degree of accuracy, one would have to assume a lower level of significance, such as 1% (0.01), which will subsequently raise the sample size higher than if a low degree of accuracy were sought. For example, a very small α-error is usually recommended for therapeutic trials compared with a survey on perception because the cost of error in the therapeutic trial could be great, leading to morbidity and mortality. Furthermore, researchers need to decide how confident they would like to be about the accuracy of the sample estimates using the appropriate level of significance. This confidence level is simply 100% minus α-error (1 - α). Therefore, 5% and 1% levels of significance will give 95% and 99% confidence levels (CI), respectively. Similarly to the á-error, the narrower the width of the CI, the higher the sample size. Both á-error and CI are used for statistical significance of the study outcomes during data analysis and the values set during the design phase must be in line with those used for statistical interpretation.

Table 1: Relationship between null hypothesis and type I and II errors

	Gold standard/Real world (Presence of Enough Evidence)	
Experiment (Statistical Test)	False null hypothesis (Effect Present)	True null hypothesis (Effect Absent)
Null hypothesis rejected (Positive Test)	**RIGHT DECISION (TRUE POSITIVE)**	**TYPE I ERROR (α) (FALSE POSITIVE)**
Null hypothesis not rejected i.e "accepted" (Negative Test)	**TYPE II ERROR (β) (FALSE NEGATIVE)**	**RIGHT DECISION (TRUE NEGATIVE)**

Table 2: Z-value or critical value for commonly set values of α-error and power

α-error	**One-tailed test**	**Two-tailed test**
0.1% (0.001)	3.074	3.291
1% (0.01)	2.326	2.576
2.5% (0.025)	1.960	2.240
5% (0.05)	1.645	1.960
10% (0.10)	1.282	1.645
Power (1 - β)		
70% (0.70)	0.524	0.524*
80% (0.80)	0.842	0.842*
90% (0.90)	1.282	1.282*
95% (0.95)	1.645	1.645*
99% (0.99)	2.326	2.326*

**: The Z-value for power (Z1-) is the same for both one-tailed and two-tailed tests*

Beta error and power

Beta error (β-error) or type II error is committed when a false H_0 is not rejected or when a false H_0 is wrongly "accepted". It is described as the false negative rate (one minus sensitivity) in the evaluation of a screening test [Table 1]. β-error gives rise to the power of a study, defined as 100% minus β-error (1 - β). Study power is the ability to reject a false $H_{0,}$ that is, how often a true difference can be detected when it exists. The higher the power, the lower the â error, the easier to detect a smaller difference between groups and the higher the sample size. Similar to the level of significance, power is often assumed at the design phase and usually set at 80% (20% β error).

Z-value/ standard normal deviate

The z-score of a value is the number of SDs away from the mean that an observation lies. It can be found by subtracting the observation from the mean and dividing the result by the SD. Z-value or standard normal deviate (SND) is the value obtained from the standard normal distribution table or Z-table (a normal distribution with mean = 0, SD = 1) for a given probability or significance level. It represents how many SDs a point lies from the mean on the standard normal curve and is commonly used as a critical value in hypothesis testing and sample size estimation. The critical value or SND for the chosen α-error and power (1 – β) is obtained from the Z-table before using it in the sample size formulas. The critical value for á-error (Z_α in one-tailed and $Z_{\alpha/2}$ in two-tailed) and power ($Z_{1-\beta}$) at 5% level of significance for a two-tailed test and 80% power are 1.960 and 0.84, respectively.[11] Critical values for commonly set values of α-error and power are presented in Table 2.

Effect size (minimum difference) of the parameter

This is the magnitude of the difference a study aims to detect between the measured effects of two groups of a population. This difference between the measured effects is the effect size and it is usually chosen because it is clinically meaningful and relevant to the research question. Effect size is used in studies comparing two groups and could be the mean difference for quantitative outcome studies or the prevalence/ proportion difference for qualitative outcome studies. Generally, the smaller the effect size or the difference one expects to be significant, the larger the sample size and vice versa. The effect size used for calculating sample size could be obtained from pilot and previous studies. The Cohen guideline could also be employed when deciding on the effect size.

Adjustments of sample size

For obvious and practical reasons, there are situations where the calculated sample size using the prescribed formula will not hold good and, in these cases, will need adjustment. The common reasons why we adjust sample size include: when studying a small or finite population (<10,000), compensation for non-response or loss to follow-up, and in the situation of unequal comparison groups. Adjustments are also made for multiple outcome variables, design effect as well as advanced statistical analysis and modelling.

The smaller the total population, especially if <10,000, the lower the sample size required. An adjusted sample size (*na*) is usually calculated for a finite population (*N*) after the initial sample size estimate (*n*) using the formula:

$$na = \frac{n}{1 + n/N}$$

To compensate for non-response or drop out, researchers need to assume an anticipated loss to follow-up or non-response rate (*v*) based on the nature of the study or previous research. The adjusted sample size (*na*) in this case will be determined from the initial sample size estimate (*n*) using the formula:

$$na = \frac{n}{1 - v}$$

To adjust for sample size (*na*) for comparison groups with unequal subjects per arm, the ratio between the unequal arms or allocation ratio (*k*) and the initially estimated sample size (*n*) is required using the formula:

$$na = \frac{n(1 + k)^2}{4k}$$

The formulas provided in this article assume a simple random sampling; deviations from this, such as cluster sampling, increase sampling error. To compensate for the increased sampling error, an adjustment (*na*) of the initially calculated sample size (*n*) with a design effect (*D*) is usually done:

$$na = n \times D$$

Design effect should ideally be estimated from pilot data or previous studies; it is generally e"1, and we usually make a conservative assumption of 2 for cluster design. Also, design effect may be determined using the average size of the cluster (c_s) and the intraclass correlation coefficient, ICC (ñ), using the formula:

$$D = 1 + \rho(c_s - 1)$$

Sample size for the various study designs

Sample size for descriptive study: cross-sectional study

The sample size (*n*) for a cross-sectional study with a qualitative outcome, such as assessing the prevalence

or proportion of obesity among adolescents in a population, is calculated using the formula:[2,4,5,7,12]

$$n = \frac{(z_{\alpha/2})^2 \times p \times q}{d^2}$$

n = desired sample size
$Z_{\alpha/2}$ = SND or Z-value of the assumed á-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]
p = the expected proportion of the outcome variable in the target population based on pilot or previous study
q = variability of proportion of the outcome variable = 1 – *p*
d = precision or degree of accuracy or margin of error desired. This is usually set at 5% (0.05)
Similarly, the sample size (*n*) for a cross-sectional study with a quantitative outcome, such as assessing the average body mass index (BMI) among adolescents in a population, is calculated using the formula: [2,5,7,13]

$$n = \frac{(z_{\alpha/2})^2 \times \sigma^2}{d^2}$$

σ = standard deviation (SD) of the outcome variable in the target population based on pilot or previous study
All other parameters are as previously defined.
For example, the sample size for a study that aims to assess the average BMI of adolescents at 5% á-error, precision of 1kg/m^2 and SD of BMI among adolescents from a previous study of 5kg/m^2 will be

$$n = \frac{1 \cdot 96^2 \times 5^2}{1^2} = 96$$

If need be, this 96 could be adjusted for non-response or finite population as the case may be, using any of the adjustment formulas above.

Sample size for descriptive study: correlational study

In a correlational study, the researcher looks at the statistical relationship linking two variables, for example, is the recent increase in the importation of fat in a community related to obesity? The sample size (*n*) of a study design to detect the correlation coefficient (*r*) between two variables is estimated using:[5,14–16]
For one-sample correlation:

$$n = \left[(Z_{\alpha/2} + Z_{1-\beta}) \div C\right]^2 + 3$$

n = desired sample size
$Z_{\alpha/2}$ = SND or Z-score of the assumed á-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]
$Z_{1-\beta}$ = SND or Z-score of the assumed power, which is usually set at 80%, and this equals 0.84 [Table 2]
C = Fisher's transformation of the correlation coefficient (*r*):

$$C = 0 \cdot 5 \times \ln[(1 + r) \div (1 - r)]$$

In = Log_e
r = expected correlation coefficient between the two variables
For the difference between 2 correlations (is r_1 different from r_2):

$$n = \left[(Z_{\alpha/2} + Z_{1-\beta}) \div (C_1 - C_2)\right]^2 + 3$$

C_1 and C_2 = Fisher's transformation of the correlation coefficient for r_1 and r_2

Sample size for analytical study: comparative cross-sectional study

The sample size (*n*) for a comparative cross-sectional study comparing two proportions, such as comparing the prevalence of obesity among adolescents in rural and urban communities, is estimated using the formula:[3,7,16]

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

n = desired sample size per arm or group
$Z_{\alpha/2}$ = SND or Z-score of the assumed á-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]
$Z_{1-\beta}$ = SND or Z-score of the assumed power, which is usually set at 80%, and this equals 0.84 [Table 2]
P_1 and P_2 = proportion/prevalence of the outcome variable in the two groups based on pilot or previous studies. (P_1 - P_2 = effect size)
The sample size (*n*) for a comparative cross-sectional study comparing two means, such as comparing the average BMI of adolescents in rural and urban communities, is estimated using the formula:[3,7,13]

$$n = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2 \times 2 \times \sigma^2}{(u_1 - u_2)^2}$$

n = desired sample size per arm or group
σ = SD of the outcome of interest in the general population
μ_1 and μ_2 = mean of the outcome variable in the two groups based on pilot or previous studies. (μ_1 - μ_2 = effect size)
All other parameters are as previously defined.

Sample size for analytical study: case-control study

A case-control study compares cases (a group with the disease or health outcome) with controls (another group without the disease or health outcome) for exposure to the risk factor of interest. The sample size (*n*) for this type of study can be estimated using:[2,7]

For qualitative outcome variable:

$$n = \frac{(k+1)}{k} \times \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2 P(1-P)}{(P_1 - P_2)^2}$$

n = desired sample size per arm or group
$Z_{\alpha/2}$ = SND or Z-score of the assumed α-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]
$Z_{1-\beta}$ = SND or Z-score of the assumed power, which is usually set at 80%, and this equals 0.84 [Table 2]
k = ratio of control to cases desired, this is 1 if an equal number of subjects is in the case and control group
P_1 and P_2 = Proportion of exposure in cases and in controls, respectively, based on pilot or previous studies. (P_1 - P_2 = effect size)
P = Average proportion of exposure = (P1+P2)/2
For quantitative outcome variable:

$$n = \frac{(k+1)}{k} \times \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2 \times \sigma^2}{(u_1 - u_2)^2}$$

n = desired sample size per arm
σ = SD of the exposure based on pilot or previous study
μ_1 and μ_2 = mean of the exposure in the cases and controls based on pilot or previous studies. (μ_1 - μ_2 = effect size)
All other parameters are as previously defined.

Sample size for analytical study: cohort study

A cohort is a group of people with similar attributes. Cohort studies follow up subjects without an outcome of interest who are with and without an exposure to a particular risk factor over a period to see the rates of outcome among the exposed and non-exposed groups. The sample size (n) in a cohort study can be determined using.:[2,7,13,17]

$$n = \frac{\left[Z_{\alpha/2}\sqrt{\left(1+\frac{1}{k}\right) \times P \times (1-P)} + Z_{1-\beta}\sqrt{\frac{P_1(1-P_1)}{k} + P_2(1-P_2)}\right]^2}{(P_1 - P_2)^2}$$

n = desired sample size per arm
$Z_{\alpha/2}$ = SND or Z-score of the assumed á-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]
$Z_{1-\beta}$ = SND or Z-score of the assumed power, which is usually set at 80%, and this equals 0.84 [Table 2]
k = ratio of control (unexposed) to study/experimental (exposed) subjects desired
P_1 = Probability of outcome in control (unexposed) group
P_2 = Probability of outcome in study/experimental (exposed) group
P = Average probability of outcome in control (unexposed) and study/experimental (exposed) group:

$$P = \frac{P_2 + (k \times P_1)}{1+k}$$

If the number of controls per study subject is 1, then the formula subsequently becomes:

$$n = \frac{\left[Z_{\alpha/2}\sqrt{2P \times (1-P)} + Z_{1-\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)}\right]^2}{(P_1 - P_2)^2}$$

$$P = \frac{P_1 + P_2}{2}$$

Sample size for interventional study: randomized controlled trial (parallel design)

Randomized controlled trials are scientific experiments characterized by the process of randomization, the presence of a control group and deliberate or artificial manipulation of the experimental group. The researcher provides interventions and compares the effects in control and experimental groups. A parallel RCT design that randomizes participants into two (most common) or more groups and treats them concurrently is the most common design.[6] Other variants include cross-over, sequential and cluster designs.[3] Also, different types of comparison can be made in an RCT and they include:[6,7,18]

a) Equivalence trial: This trial aims to show that the novel/new treatment and the control/standard treatment are equally effective.
b) Non-inferiority trial: This trial aims to verify that the novel/new treatment is as effective but now more effective than the control/standard treatment.
c) Superiority trial: This trial aims to ascertain that the novel/new treatment is more effective than the control/standard treatment.

Generally, one-tailed tests are used in both superiority and non-inferiority trials, while two-tailed tests are used in equivalence trials.[6] However, the appropriateness of one or two-tailed hypothesis tests in RCT is controversial. A two-tailed test could sometimes be used when comparing two active treatments in a superiority trial.[19] Irrespective of how the hypothesis is being formulated, the sample size should be estimated to reflect the sidedness [Table 2]. Nonetheless, the two-tailed hypothesis, which does not indicate direction, is used by most researchers. Also, regulatory authorities, such as the FDA and EMA, generally recommend using two-tailed tests even in superiority trials to minimize potential bias and ensure robust, unbiased inference.
The sample size for RCT, assuming two groups, can be estimated using:[6,7,18]
For equivalence trial with a qualitative outcome variable:

$$n_1 = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2}{\left[\Delta - |d|\right]^2} \times \left[\frac{P_1(1-P_1)}{k} + P_2(1-P_2)\right],\ n_2 = k \times n_1$$

For equivalence trial with a quantitative outcome variable:

$$n_1 = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2}{\left[\Delta - |d|\right]^2} \times \sigma^2 \times \left(1 + \frac{1}{k}\right),\ n_2 = k \times n_1$$

For non-inferiority and superiority trial with a qualitative outcome variable:

$$n_1 = \frac{\left[Z_{\alpha} + Z_{1-\beta}\right]^2}{\left[d - \Delta\right]^2} \times \left[\frac{P_1(1-P_1)}{k} + P_2(1-P_2)\right],\ n_2 = k \times n_1$$

For non-inferiority and superiority trial with a quantitative outcome variable:

$$n_1 = \frac{\left[Z_{\alpha} + Z_{1-\beta}\right]^2}{\left[d - \Delta\right]^2} \times \sigma^2 \times \left(1 + \frac{1}{k}\right),\ n_2 = k \times n_1$$

n_1 = sample size for new treatment group
n_2 = sample size for control/ standard treatment group = $k \times n_1$
The total sample size for the trial = $(n_1 + n_2)$
Z_{α} and $Z_{\alpha/2}$ = SND or Z-score of the assumed á-error, which is commonly set at 5%, and this equals 1.65 and 1.96 for a one and two-tailed tests, respectively [Table 1]
$Z_{1-\beta}$ = SND or Z-score of the assumed power, which is usually set at 80%, and this equals 0.84 [Table 2]
k = ratio of subjects in the new treatment to the control group desired
P_1 = the response rate (proportion of participants with outcome) in the new treatment group
P_2 = the response rate (proportion of participants with outcome) in the control/standard treatment group
d = effect size $(P_1 - P_2)$
Δ = clinically acceptable margin effect or clinically meaningful difference, which may be obtained from the literature
σ = SD of the general population from the previous study or the pooled SD of both comparison groups:

$$\sigma = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

S_1 and S_2 = SD from the previous study of the new treatment and control groups
n_1 *and* n_2 = sample size from the previous study of the new treatment and control groups

Sample size for interventional study: cross-over randomized controlled trial

Cross-over design is commonly used for a study for which the available treatment is palliative rather than curative.[3] It is a within-subject comparison in which the treatment group also serves as the control group at different points in time. The types of comparison in cross-over RCT are similar to parallel design and their sample size is estimated using:[5]
For equivalence trial with qualitative outcome variable:

$$n = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2}{2\left[\Delta - |d|\right]^2} \times P \times (1-P)$$

For equivalence trial with quantitative outcome variable:

$$n = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2}{2\left[\Delta - |d|\right]^2} \times \sigma^2$$

For non-inferiority and superiority trial with a qualitative outcome variable:

$$n = \frac{\left[Z_{\alpha} + Z_{1-\beta}\right]^2}{2\left[d - \Delta\right]^2} \times P \times (1-P)$$

For non-inferiority and superiority trial with a quantitative outcome variable:

$$n = \frac{\left[Z_{\alpha} + Z_{1-\beta}\right]^2}{2\left[d - \Delta\right]^2} \times \sigma^2$$

P = pooled response rate (proportion of participants with outcome) in both comparison groups $(P_1 + P_2 / 2)$

All other parameters are as previously defined.

Sample size for interventional study: cluster randomized controlled trials

A cluster RCT is a study in which pre-existing groups called clusters are randomly assigned to treatment arms. Generally, the sample size (n) for cluster RCT is estimated using the appropriate RCT formula and then adjusted with the design effect (D) as discussed under the adjustment of sample size. This adjustment using the design effect is a result of increased sampling error associated with cluster sampling. Hemming *et al* give a detailed description of the sample size considerations of cluster RCT.[20]

Sample size for interventional study: quasi experimental studies

Quasi-experimental studies are similar to RCTs as they aim to assess the outcome of an intervention, but without conducting the process of randomization. Examples of this study include the uncontrolled trial, natural experiment, before and after comparison study, as well as pre- and post-test study. In most of these studies, the subjects serve as their own control and the sample size can be determined using: [5,16]
For a qualitative outcome variable

$$n = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2}{\left[P_1 - P_2\right]^2} \times P \times (1-P)$$

For a quantitative outcome variable

$$n = \frac{\left[Z_{\alpha/2} + Z_{1-\beta}\right]^2}{\left[d\right]^2} \times \sigma_c^2$$

$Z_{\alpha/2}$ = SND or Z-score of the assumed α-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]

$Z_{1-\beta}$ = SND or Z-score of the assumed power, which is usually set at 80%, and this equals 0.84 [Table 2]

P_1 and P_2 = proportion of outcome of interest before and after the intervention

P = pooled proportion of outcome of interest from both groups

d = difference in mean of the outcome of interest before and after the intervention. ($\mu_1 - \mu_2$ = effect size)

σ_c = SD of difference or change within pair

$$\sigma_c = \sigma\sqrt{2 \times (1 - r)}$$

σ = SD of the outcome in the population

r = within-subject correlation of the outcome or correlation between before and after values in the same group

Screening studies: sensitivity and specificity

Sensitivity and specificity are common analyses used to assess screening and diagnostic studies. These studies aim to determine how sensitive a screening or diagnostic test is in predicting an outcome.[21] The sample size for assessing either the sensitivity or specificity of a screening or diagnostic test is estimated using:[5,21,22]

For sensitivity:

$$n = \frac{[Z_{\alpha/2}]^2 \times Sn(1 - Sn)}{P \times [d]^2}$$

For specificity:

$$n = \frac{[Z_{\alpha/2}]^2 \times Sp(1 - Sp)}{(1 - P) \times [d]^2}$$

$Z_{\alpha/2}$ = SND or Z-score of the assumed α-error, which is commonly set at 5%, and this equals 1.96 for a two-tailed test [Table 1]

Sn = Sensitivity from pilot or previous study

Sp = Specificity from pilot or previous study

P = Prevalence of the disease or health outcome

d = degree of accuracy or precision assumed

A larger sample size is required for evaluating sensitivity when the prevalence of the disease is lower, while a larger sample size is required for specificity when the prevalence is higher and vice versa in both cases. In research that seeks to determine both sensitivity and specificity of a screening/diagnostic test, it is recommended that the larger sample size be used. However, a sum of the two may be used as demonstrated by Negida *et al.*[22] but this approach is subject to debate.

Animal studies

Animals are studied for various reasons, and one of such important reasons in medical literature is testing the efficacy of preventive (such as vaccines) and therapeutic (such as drugs) measures. The standard sample size formula in animal experiments is the same as that of the RCT. This formal power-based sample size estimation method remains the preferred and more rigorous approach whenever feasible. However, the parameters, such as SD and effect size needed, may not be readily available in the literature, and one may resort to using the "resource equation method"(REM).[2] REM is an approximate method intended for use only as a last resort, applicable when effect size and SD cannot be reliably estimated or when multiple endpoints are being evaluated. It uses a rule of thumb and the first step is to decide the total number of comparison groups in the study and then apply the formula:[2]

$$n = REV + \text{Total number of groups}$$

n = total number of animals

REV = resource equation value, set between 10 and 20.

The rule of thumb here is that REV (total number of animals minus total number of groups) should lie between 10 and 20 for optimum sample size. If REV is <10, then we should increase the sample size, and if >20, we should reduce the sample size.

Electronic Resource to sample size determination

Examples of such resources are:[23]

1. Epi Info™
   Available from: https://www.cdc.gov/epiinfo/index.html
2. G*Power
   Available from: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower
3. Select Statistical Services
   Available from: https://select-statistics.co.uk/calculators/
4. Cleveland Clinic I Department of Quantitative Health Sciences
   Available from: https://riskcalc.org/samplesize/
5. UCSF Clinical and Translational Science Institute
   Available from: https://sample-size.net/

General recommendations on sample size estimation

After estimating the sample size, investigators should always assess the sample and evaluate whether it is adequate and reasonable enough for the set research question(s).[13] Generally, qualitative outcome research requires more sample sizes than a quantitative outcome study. Health outcome with low prevalence requires larger sample sizes.[5,13] Also, if the research proposal involved analysis of subgroups, the sample size may need to be expanded to show significant findings if present. For example, to study the characteristics of diabetes patients, a researcher may only need 100

samples; however, if the researcher chooses to conduct subgroup analysis on those with diabetic foot ulcer or any specific complication, this subgroup sample size may be too small to show significant findings, except that the total sample size is increased.[4]

CONCLUSION

Sample size determination is an important step in carrying out valid and reliable research. This article highlighted the principles and methods of sample size estimation of common epidemiological and clinical research designs. Statistical theories and concepts such as the direction of hypothesis, type of outcome variables, level of confidence, power, effect size assumed and situations requiring adjustments, such as non-response, are important considerations in calculating scientifically appropriate sample size. Based on the explanations, simplified formulas, and suggested easy-to-use electronic resources in this article, researchers should be able to determine the correct sample size for their study in order to provide more reliable results and improve the generalization of their findings.

Competing interests: The author declares that there is no conflict of interest

Funding: Nil

Author contributions: TMI conceptualized, designed, and conducted this study. He drafted the manuscript, reviewed it, as well as read and approved the final version.

Acknowledgment(s): Nil

REFERENCES

1. **Süt N,** Ajredani M, Koçak Z. Importance of sample size calculation and power analysis in scientific studies: an example from the Balkan Medical Journal. Balkan Med J. 2022;39(6):384–385.
2. **Charan J,** Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med. 2013;35(2):121–126.
3. **Bamgboye AE.** Medical Statistics. 2nd ed. Ibadan: Folbam Publisher; 2014.
4. **Araoye MO.** Research Methodology with Statistics for Health and Social Sciences. Ilorin: Nathadex Publishers; 2004.
5. **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(2):67–75.
6. **Zhong B.** How to calculate sample size in randomized controlled trial? J Thorac Dis. 2009;1(1):51–54.
7. **Sharma SK,** Mudgal SK, Thakur K, Gaur R. How to calculate sample size for observational and experimental nursing research studies? Natl J Physiol Pharm Pharmacol. 2020;10(1):1–8.
8. **Wassertheil-Smoller S.** Biostatistics and Epidemiology: A Primer for Health and Biomedical Professionals. 3rd ed. New York: Springer; 2004.
9. **Parks K.** Park's Preventive & Social Medicine. 23rd ed. Jabalpur, India: M/s Banarsidas Bhanot Publishing; 2015.
10. **Das PK.** Monitoring & evaluation: adjusting sample size [Internet]. 2023. Available from: http://monitoringevaluation.weebly.com/sample-size-adjustments.html. Accessed April 4, 2023.
11. Appendix - z-score percentile for normal distribution [Internet]. 2023. Available from: https://www.pindling.org/Math/Learning/Statistics/z_scores_table.htm Accessed April 4, 2023.
12. **Hazra A,** Gogtay N. Biostatistics series module 5: determining sample size. Indian J Dermatol. 2016;61(5):496–504.
13. **Althubaiti A.** Sample size determination: a practical guide for health researchers. J Gen Fam Med. 2022;24(2):72–78.
14. **Bujang MA,** Baharum N. Sample size guideline for correlation analysis. World J Soc Sci Res. 2016; 3(1):37–46.
15. **Negida A.** Sample size calculation guide - part 7: how to calculate the sample size based on a correlation. Adv J Emerg Med. 2020;4(2):e34.
16. **Hulley SB,** Cummings SR, Browner WS, Grady D, Newman TB. Designing Clinical Research. 4th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2013.
17. **Khaled Fahim N,** Negida A. Sample size calculation guide - part 2: how to calculate the sample size for an independent cohort study. Adv J Emerg Med. 2018;3(1):e12. doi:10.22114/AJEM.v0i0.104s
18. **Kishore K,** Mahajan R. Understanding superiority, noninferiority, and equivalence for clinical trials. Indian Dermatol Online J. 2020;11(6):890–894.
19. **Ghosh K.** Two-sided vs. one-sided tests: this should not be "controversial" [Internet]. 2020. Available from: https://inferenceinc.com/two-sided-vs-one-sided-tests-this-should-not-be-controversial/#:~:text=For%20superiority%20trials%2C Accessed April 5, 2023.
20. **Hemming K,** Girling AJ, Sitch AJ, *et al.* Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. BMC Med

Res Methodol. 2011;11:102. doi:10.1186/1471-2288-11-102. Erratum in: BMC Med Res Methodol. 2017;17(1):8.
21. **Bujang MA,** Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. J Clin Diagn Res. 2016;10(10):YE01–YE06.
22. **Negida A,** Fahim NK, Negida Y. Sample size calculation guide - part 4: how to calculate the sample size for a diagnostic test accuracy study based on sensitivity, specificity, and the area under the ROC curve. Adv J Emerg Med. 2019;3(3):e33.
23. **Egbuchulem K.I.** The basics of sample size estimation: An editor's view. Annals of Ibadan Postgraduate Medicine. 2023;21(1):5–10.