Journal of Young Pharmacists, 2024; 16(2):204-215.
https://www.jyoungpharm.org

Review Article

Types of Sampling and Sample Size Determination in Health and Social Science Research

Mohammad Daud Ali[1,*], Ebrahim Ahmed Jaber Al Hatef[2]

[1]Department of Pharmacy, Mohammed Al-Mana College for Medical Sciences, Al Safa, Dammam, SAUDI ARABIA.
[2]Department of Information Technology, Mohammed Al-Mana College for Medical Sciences, Al Safa, Dammam, SAUDI ARABIA.

ABSTRACT

While determining suitable sample size and sampling is one of the most crucial factors to take into account during the planning stage of a research project, researchers sometimes encounter difficulties in this regard. Inadequate sample sizes are commonly used by researchers, which always results in inaccuracies in the conclusions. The purpose of this methodological review is to highlight the importance of sample size estimation and sampling techniques and offer suggestions for describing sample sizes in an easy-to-understand way. A brief overview of sample size calculation in health sciences research and the terminology used are provided. Techniques for estimating the sample size for health research are described with examples depending on the kind of study design. To estimate sample size, researchers must: Be aware of the statistical analysis that will be used; Decide on acceptable precision levels; Decide on study power; and specify the confidence that will be used. This method review's displays a thorough explanation and simplicity of the sampling technique and the sample size formula that will assist in clarifying the mystery surrounding statistical equations for sampling technique and sample size estimates in healthcare and social science research.

Keywords: Sampling, Sample Size, Health, Study design, Observational Studies, Case-Control, Cohort Studies, Experimental Studies.

Correspondence:
Dr. Mohammad Daud Ali
Department of Pharmacy, Mohammed Al-Mana College for Medical Sciences, Abdulrazaq Bin Hammam Street, Al Safa, Dammam-34222, SAUDI ARABIA.
Email: dali.niper@gmail.com
ORCID: 0000-0003-3752-2331

Received: 23-01-2024;
Revised: 05-02-2024;
Accepted: 14-03-2024.

INTRODUCTION

Every scientific study needs well-thought-out procedures to yield reliable and pertinent data. Such outcomes need the use of a sample size estimate that has been scientifically validated.[1] To provide reliable results, nearly all quantitative studies need a sufficient sample size.[1] As a result, while developing research concepts and proposals, sample size estimation is crucial. If the sample size is too low, the research outcome may not be reproducible. While informal guidelines based on researchers' experience may be sufficient for pilot research studies, they may not be enough for studies requiring funding or institutional review board approval.[2,3] Review committees may expect a proper justification for the sample size. In addition, many academic journals now require evidence of sample size calculation or specific requirements to be provided in the method section of a manuscript. The calculation can also be part of a checklist before submitting the manuscript to a journal.[4,5] When the sample size calculation is not mentioned, reviewers may question whether the sample size is adequate.[5] Under some circumstances, scientists may choose to analyze every individual in a target population. It becomes possible to research the population of interest when the necessary resources are accessible. This is known as a thorough survey, and while it is possible to estimate the sample size, it may not be required in this instance.[6] It is usually not possible to study all of the subjects or responders in a population of interest. It will therefore be necessary to use a sample or subset of the population.[6,7] The selection of population representatives in each of these scenarios will require a scientific approach.

The authors of this method review conducted a thorough search and analysis of statistical concepts related to calculating the appropriate sample size determination and sampling technique determination. This review also explained different sampling techniques, a practical and straightforward method for determining sample size, which can be applied to various types of parameters and variables in observational studies, randomized controlled trials, and diagnostic studies that can be conducted in health and social science research.

SAMPLING

A group of individuals from a population is chosen to estimate the characteristics of the entire population. This process is known as sampling. Cheaper costs and quicker data collection are the two primary benefits of sampling.[8,9] One or more characteristics

DOI: 10.5530/jyp.2024.16.27

of observable subjects that are identified as distinct persons are measured by each observation. To gather information about a population, sampling is frequently employed in social science and medical research.[8,9]

SAMPLING METHODS AND CATEGORIES

A description of the procedure for selecting the subjects of the information has been published in the literature.[8] When choosing people, the following factors need to be taken into account (Figure 1 Explains about types of sampling).

- The organization as a whole may be the subject of an investigation or a representative member may be chosen.
- The heterogeneity within the group should be taken into account while choosing the samples, and the proper sampling method should be applied.

Several common sample designs have been covered in the literature, including quota, random, and purposeful sampling. There exist alternative techniques for random sampling.[8-10]

Purposive Sampling

This method chooses sampling units based on the intended use. Statistical recognition of purposeful sampling is lacking, and it yields a skewed estimate. There are only a few particular uses for this method.[10]

Random Sampling

There will be a predetermined likelihood of inclusion for every unit in the sample under this sampling technique.[10] As opposed to purposive sampling; this sampling yields a more accurate assessment of the study's parameters. A known and non-zero possibility exists for every individual in the sampling frame to be chosen for the sample. That single-stage random sampling is the most recognized and optimum method.[8-10]

Through the Use of Random Number Tables

Most statistics handbooks and publications on research procedures include an appendix with a table of random numbers. A random number table can include 10,000 randomly chosen digits, ranging from 0 to 9, and is often displayed in rows of five. No matter whatever digit comes before it in the table, each digit has the same chance of occurring.[8-11]

Basic random sampling

Every unit in the sample has an equal probability of being chosen when using the basic random sampling procedure. This method offers a more impartial and precise evaluation of the parameters, especially in cases when the population is homogenous.[11]

Random Sampling Stratified

Random Sampling Stratified is a valuable data collection method when the population is diverse. This approach divides the whole diverse population into several homogeneous groupings, or strata, each of which is homogenous within itself. Units are then randomly sampled from each of these strata.[10] The relative relevance of each stratum in the population determines the sample size for each stratum. Stratified sampling is the process of creating this stratified sample. Stated differently, stratification refers to the process of dividing a population into subgroups or strata. Then, sampling in each stratum will be done independently. Strata or subgroups are selected based on available data indicating their relationship to the outcome.[11] The choice of stratum will differ depending on the region and climate. Following stratification, sampling is carried out in each stratum independently. The sampling error in a stratified sample is determined by the population variance inside the stratum, not by the population variance between strata. Another definition of stratified random sampling is a population that falls into multiple unique groups, allowing the frame to be divided into different "strata" according to these classifications. Subsequently, a distinct subgroup within each stratum is selected, from which individual constituents may be randomly selected.[8-10]

Samples by Cluster

The whole population, which has been divided into groups or clusters, is selected for samples by cluster using a random sampling approach. All observations fall inside the sample within the selected clusters.[13] Once "natural" then comparatively consistent clusters remain apparent in a statistical population, cluster sampling is a sampling technique that is employed. Typically, cluster sampling is used once a researcher has a complete list of groups, or "clusters," within a population but not an ample list of the elements of the population they want to investigate.[12] This sample technique might be more cost-effective and practicable than stratified or random sampling alone.[12]

Random Sampling in a Systematic Way

This sampling technique begins by choosing the first unit of the sample randomly. Then, the remaining units are selected systematically. To calculate the sampling interval (R), you can use the formula R=N/n, where “N” is the overall number of items in the population and “n” is the desired sample magnitude. To select the next units, you can choose the first number randomly from the remaining sampling interval between the initially chosen unit and the end of the interval.[11,12]

Multistage Arbitrary/Random Sampling

Multistage arbitrary sampling makes use of several item selection segments. The sample designs may alter or remain the same at each level. Samples that are slightly grouped are used in the multistage sampling technique, sometimes referred to as cluster

sampling. The main benefit of this sampling strategy is that it lets resources be intensive on a small number of frame units; however, sampling errors will rise as a result of this sampling technique.[12,13]

Sampling quota-wise

Sampling quota-wise is a method that involves dividing a population into mutually exclusive subgroups. After this division, subjects or units are selectively chosen from each of these subgroups using judgment, by a predetermined proportion.[12]

Sampling quota-wise is classified as a non-probability sampling methodology because of its second stage. The sample is not chosen at random when using quota sampling. Interviewers, for example, may be biased toward selecting people who appear to be more helpful. The problem with this approach is that since not every candidate is given the opportunity to be selected, the samples obtained may be prejudiced.[8-12]

Spatial Sampling

Survey sampling that involves sampling in two or more dimensions is known as spatial sampling.[9]

Independent Sampling

Sampling from the same population or distinct populations that are independent of each other is known as an independent sample. Put differently, there isn't any association between the samples.[9]

Convenience sampling

Convenience sampling is a non-probability sampling technique in which the units chosen for the sample are the ones that the researcher can access most easily.[9]

Snowball sampling

Researchers invite study participants to help them find more possible subjects as part of a recruiting strategy called snowball sampling.[9]

Importance of sampling and determining sample size in social science and health research

A study's internal and external validity can be guaranteed when the sample size is appropriately assessed and based on prior research or supporting data.[9] To obtain accurate measurements of the population being studied, it is important to establish that the sample being used is representative. It is also crucial to consider the sample size in both human and animal experiments for ethical reasons. Using a small sample size can result in weak scientific inference and potentially harmful treatments for participants, without any new information being gained.[9] On the other hand, studies with excessively large sample sizes can result in an unnecessarily high number of participants being subjected to harmful medical care if the study does not yield a substantial advance in scientific understanding. Therefore, it is important to carefully consider the sample size to avoid needless suffering for study volunteers.[8-12]

Calculating the parameters' sample size

Five research design characteristics often influence sample size decisions: estimated measurement variability, significance criterion, intended statistical power, minimum expected difference, or effect size; then whether a one or two-tailed statistical investigation is being deliberate.[12]

The base of Size-Effects or Least Probable Variance

When conducting a study, the required sample size to identify statistical significance increases as the expected minimal difference decreases.[12] This parameter is usually determined based on the researcher's medical decision or skill with the delinquent being studied. For instance, if research aims to evaluate a new process that may offer higher accuracy but with some uncertainty than an established diagnostic procedure with 85% accuracy, the researcher may set a minimum expected difference of 10% (0.10) if they believe that achieving 95% accuracy with the new method would represent a significant advancement.[11,12]

Estimate of Measurement Variability

The standard deviation of the selected measurements is a parameter for each group being compared. A greater statistical variation necessitates a larger sample size for identifying even the smallest difference. To ensure efficiency, it is important to assess measurement variability using initial data from a similar research population.[11,12]

This value represents the power required for the research. Sample size rises with increasing power. Although high power is greatly preferred, there is a clear trade-off with the number of subjects that could be studied realistically, considering time and resources that are often fixed for research or investigative studies. While many clinical trial specialists currently propose a power of 0.90, statistical power in RCTs is frequently set to a value more than or equal to 0.80.[13]

Implication Criterion (Probability/*p*-value)

This constraint indicates the highest *p*-value at which an alteration qualifies as statistically significant. The minimum difference must be detected with a larger sample size when the significance criterion is lowered. Typically, a 5% threshold is used for statistical significance.[11,12]

Statistical Analysis: One- or Two-tailed

In some rare situations, it is possible to determine before conducting research that changes between the experimental and comparison groups can only occur in one direction. In such cases, it may be appropriate to consider a one-tailed statistical

analysis. When all other parameters are the same, the sample size of a one-tailed research design, through an assumed statistical significance threshold (such as α), is equivalent to the sample size of a two-tailed design with an implication principle of 2α.[11,12,14]

Essential Statistics Ideas for Determining Sample Size

The following data must be available to a researcher to estimate the sample size for a given study.[12]

THE ALTERNATIVE HYPOTHESIS AND THE NULL HYPOTHESIS

As the null hypothesis suggests that here is no discernible alteration, it is used in both experimental and correlation observational research. It is always employed with a different hypothesis (significant difference) that is predicated on rejection. Sometimes it is impossible to reject the null hypothesis; nevertheless, this does not mean that it is false; rather, it simply indicates that there is insufficient data to back rejecting the null hypothesis as put forth.[14]

REASONABLE DEGREE OF RELEVANCE

After the previous sentence, the text explains that an acceptable level of significance is reached when a true null hypothesis is rejected. This means that α (which represents the probability of a Type-I error) should be low. In biology, studies are typically evaluated based on their level of significance, which is either 5% (α/*p*=0.05) or 1% (α/*p*=0.01).[15] This indicates that there is a small probability (either 5% or 1%) that the observed results may have occurred by chance rather than due to human intervention. Depending on the level of significance, a different Confidence Interval (CI) is required. The proper Confidence Interval (CI) for a 5% (α/*p*=0.05) level of significance is CI-95%, whereas a 1% (α/*p*=0.01) level of significance is CI-99%.[15]

RESEARCH POWER

It denotes the ability to rule out a major difference when one is present. Stated differently, it represents the likelihood of extrapolating research results to the entire population. Increasing statistical power lowers the likelihood of Type-II error (β) occurring, hence lowering the potential of hypocritical results. The situation is therefore represented by "1-β". A power of 0.8 (80%) or higher is generally regarded as further acceptable towards determining a statistically substantial difference in clinical trials.[15] With a power of 80%, there is a 20% possibility that we won't notice a meaningful change even if one is present.[16]

ANTICIPATED IMPACT SIZE

The term "effect size" states the extent of the association between two variables in a populace. Such as, if one type of exercise increases blood pressure by an average of 5 mmHg, and another type increase it by 2 mmHg, the out-and-out result size would be 5-2=3 mmHg. Result size refers to the alteration amongst the restrained things of the interventional and rheostat groups. It can be estimated using pilot studies and previously reported data. Social scientists usually rely on Cohen's guide to determine effect size. According to this guide, an effect size of less than 0.1 is considered small; an effect size between 0.3 and 0.5 is intermediate, and an effect size greater than 0.5 is considered a large variance outcome. An effect size of 0.5 is usually cast to reflect moderate to large differences. The situation important to note that the sample magnitude and effect magnitude are reciprocal proportions. Once the effect magnitude is big, a smaller sample size is enough for research, and vice-versa.[16]

ERROR MARGIN

It is important to note that when conducting a study, the sample may not perfectly represent the larger population due to random sampling error. If a study sample has a prevalence of COVID-19 infection of 50% with a margin of error of 5%, the prevalence of COVID-19 infection in the population would range between 45% and 55%.[16]

THE POPULATION'S STANDARD DEVIATION, OR SD

When calculating an outcome variable, such as the effect of a workout driver on blood glucose, we need to consider the population variance. This can be done by finding the mean or the Standard Deviation (SD). If we embrace a group whose blood glucose levels range between 150-350 mg/dL, we will need a larger sample size to identify differences among therapies, as the SD in this group will be greater. However, we can produce a more homogeneous group by considering a trial from a populace

Table 1: Z-values for sample size calculation.

Value	Variance
α-Value (Level of confidence (%))	$Z_{1-\alpha/2}$ (Two-tail)
0.10 (10)	1.64
0.05 (5)	1.96
0.01 (1)	2.58
α-Value (Level of confidence (%))	$Z_{1-\alpha/2}$ (One-tail)
0.10 (10)	1.28
0.05 (5)	1.65
0.01 (1)	2.33
β-Value (Power %)	$Z_{1-\beta}$
0.40 (60)	0.25
0.20 (80)	0.84
0.10 (90)	1.28
0.05 (95)	1.65
0.01 (99)	2.33

whose blood sugar analysis falls between 150-250 mg/dL. This will decrease the study's standard deviation and sample count.[16]

STATISTICAL TESTS FOR INFERENCE WITH ONE AND TWO-TAILS

The study's objective is to determine which one- or two-tail test is better. If a new drug is thought to lower blood pressure better than an existing one, a one-tail test may be all that is needed to assess the hypothesis. If it is unclear if the new medication would drop blood pressure more or less than the current medication, a two-tail test is always preferred. The critical ratio (Z-value) differs between one-tail and two-tail tests as indicated in Table 1.[16,17]

EFFECT OF DESIGN (DEFF)

When a straightforward random sampling approach is presumed to be employed in the study, the sample size estimation techniques offered in this review assist in estimating the appropriate trial magnitude. But only if basic random sampling is feasible can a computed sample be sufficient. The computed sample magnitude needs to remain changed about DEFF to solve this issue. This remains equivalent to the proportion of the predicted discrepancy in cluster arbitrary selection to that of simple random sampling. In most cases, the Design Effect (DEFF) is greater than or equal to 1, so the researcher adopts DEFF=2 in a cluster design.[16,17]

$$DEFF=1+\delta(n-1)$$

(Where, n is the cluster's common size and δ denotes interclass correlation).

ESTIMATING SAMPLE SIZES FOR VARIOUS RESEARCH DESIGNS AND STATISTICAL ANALYSIS

Descriptive or cross-sectional study sample size estimation (prevalence studies)

The purpose of these surveys and studies is to collect information, observe, and article-specific features of a condition as it occurs obviously. [19] They are non-cast to recognize the root cause of a problem, like the source of a prevalent. Investigators collect data on the incidence of liver disease and past alcohol consumption habits using a cross-sectional approach, without manipulating any variables.[17-24]

Scenario 1: If the facts are a nominal or ordinal scale and a percentage stays a criterion, then sample size should be determined.[17]

$$\text{Sample size}(n)=\frac{(z_{1-\alpha/2})^{2*}(p)(q)}{(d)^2}$$

n is the intended sample size.
For each degree of confidence, $Z_{1-\alpha/2}$ = Critical value and a standard value.
(It is 1.96 at a 95% confidence interval or 5% level of significance (type-I error), and 2.58 at a 99% confidence interval.)
P stands for predicted prevalence or based on earlier studies.
q = 1 - p
d = Error or accuracy margin

An investigator intends to carry out a descriptive study to determine the percentage of the elderly population in a city who have hypertension. A previous study revealed that 30% of adults in the community had hypertension. The researcher needs to ascertain the sample size required to carry out further research with a 95% confidence level and a 5% margin of error.[20]

On applying:

$$\text{Sample size}(n)=\frac{(z_{1-\alpha/2})^{2*}(p)(q)}{(d)^2}$$

$$(n)=\frac{(1.96)^{2*}(0.30)(0.70)(0.05)}{(0.05)^{2*}}$$

(n) = 368.79 (n) = 369+37 (taking into account a 10% research participant dropout rate)

Z1-α/2 = 1.96, P = 40% = 0.4, q = 1−0.4 = 0.6, d = 5% = 0.05, and sample size (n) = 406.

For this study, a minimum of 406 people is required to conduct novel cross-sectional research to determine the incidence of diabetes mellitus amongst affected roles.

Scenario 2: When the mean of the study/data is on an interval or ratio scale, the following equation is used to determine sample size estimation.

$$\text{Sample size }(n)=\frac{(Z_{1-\alpha/2})^{2*}(\sigma)^2}{(d)^2}$$

n = The required number of samples
The standardized value for the applicable confidence level is denoted by $Z_{1-\alpha/2}$.
(It is 2.58 at CI- 99%, or 1% type I error, and 1.96 at CI- 95%.)
d = Precision rate or error margin
The standard deviation, or σ, is a number that comes from past studies or experimental projects.

Assume an investigator desires to carry out a new study to estimate the average glycated haemoglobin (HbA_{1c}) level among urban residents. The researcher has a 95% Confidence Interval (CI) of 2.5 g/dL and knows from a prior study that the standard

deviation of adult haemoglobin levels was 5.5 g/dL. The researcher needs to govern the required number of populations to start the new study.

On applying:

$$\text{Sample size } (n) = \frac{(Z_{1-\alpha/2})^{2} * (\sigma)^{2}}{(d)^{2}}$$

$$(n) = \frac{(1.96)^{2} * (5.5)}{(2.5)^{2}}$$

21.13 (n) = 21+2 (n) (Taking into account a 10% research participant dropout rate) $Z_{1-\alpha/2}$=1.96, σ = 5.5 g/dl, and d = 2.5 g/dl are the sample sizes (n) of 23.

To assess the average HbA_{1c} level among adults, new cross-sectional research with a minimum of 23 individuals will be mandatory.

Scenario 3: Sample size for fixed sample size.

$$\text{Sample required } (n) = N/1+N*d^{2}$$

N=Total populace,

d=Edge of error or accuracy.

An investigator desires to determine the prevalence of tachycardia among people with hypertension in a town through a survey. For instance, if there are 3,500 adults in the population who have hypertension, and the researcher aims to operate with a 95% confidence interval and a 5% Edge of error or accuracy rate chosen through the investigator, what would be the required size of the survey sample?

On applying the above formula,

$$(n) = \frac{3500}{1+3500*(0.05)^{2}}$$

$(n) \approx 399.89 = 400$

(n) = 400+40 (considering 10% dropout of study participants) Sample size

(n) = 440, N = 3500, d = 5% = 0.05

Thus, novel cross-sectional research near ascertaining the frequency of tachycardia amongst adults through hypertension will require a minimum of 440 participants.

CALCULATING SAMPLE SIZES FOR CASE-CONTROL RESEARCH

This study aims to determine if there is a relationship between exposure and a particular outcome. It identifies the correlation between exposure and a disease or condition of interest, establishing cause and effect. This observational study examines the widely accepted theory of causality between two groups with different outcomes. For instance, it may involve a case-control study towards investigating the connection between alcohol consumption and liver ailment.[18,21]

Scenario 4: When the data are on a nominal or ordinal scale, the sample size is a research measure related to proportion.

$$n = \frac{(r+1)}{r} * \frac{p(1-p)\left(Z_{1-\beta}+Z_{1-\alpha/2}\right)^{2}}{\left(p_{1}-p_{2}\right)^{2}}$$

n = Number of samples desired
r = Control relative to cases (1 if both groups have the same number of subjects)
p = (P1 + P2)/2 is the population proportion.
The desired power is Z1-β, which is 0.84 for 80% power and 1.28 for 90% power.
Z1-α/2 represents the critical value, together with a standard value for the associated confidence level.
(It is 1.96 at 95% CI or 5% type I error, and 2.58 at 99% CI or 1% type I error.)
P1 = Case proportion
P2 = Share of controls

A final-year pharmacy student aims to investigate the relationship between pulmonary embolism and deep vein thrombosis (DVT) using a case-control study design. The student decides to operate at 80% research power and 95% confidence interval. Assuming a 30% proportion of cases and 20% for controls, the aim is to balance the number of cases between the two groups.[18,21] Can you determine the ideal sample size for each study group?

On applying:

$$n = \frac{(r+1)}{r} * \frac{p(1-p)\left(Z_{1-\beta}+Z_{1-\alpha/2}\right)^{2}}{\left(p_{1}-p_{2}\right)^{2}}$$

$$(n) = \frac{(1+1)}{1} * \frac{0.25(1-0.25)(0.84+1.96)^{2}}{(0.30-0.20)^{2}}$$

(n)=2*147,

(n)=294, (n)=294+30 (considering 10% dropout of study participants).

Sample size (n)=324, p_1=30%=0.3, p_2=20%=0.2, r=1, $Z_{1-\beta=}$0.84, $Z_{1-\alpha/2=}$1.96, p=0.3+0.2/2=0.25.

Thus, an investigator is fictional to take at least 324 volunteers in the case as well as in the control set.

Scenario 5: When conducting a study with mean as a parameter, using data on an interval or ratio scale, the sample size is important.

$$\text{Sample required}(n) = \frac{(r+1)}{r} * \frac{\sigma^{2}\left(Z_{1-\beta}+Z_{1-\alpha/2}\right)^{2}}{d^{2}}$$

n = The number of samples we must examine in order to
r = Control relative to cases
p (population proportion) = P1 + P2 /2
The desired power is $Z_{1-\beta}$, which is 0.84 for 80% power and 1.28 for 90% power.
$Z_{1-\alpha/2}$ represents the critical value and a standard value for the associated confidence level.
(It is 1.96 at 95% CI and 2.58 at 99% CI or 1% type I error.)
Based on a prior study or pilot research, σ = SD
d = Effect size (difference between pilot study or prior research mean differences)

Assume that a researcher wishes to investigate whether there is a correlation between alcohol intake levels and liver cirrhosis. The researcher has concluded that in the previous study, there was a mean difference of 8 mL/day in alcohol drinking between the case and control groups, with a standard deviation of 15 mL/day. To ensure sufficient statistical power of 80% and a confidence interval of 95%, the researcher has decided to have an equal number of participants in both groups. Now, the task is to determine the sample size.

On applying:

$$\text{Sample required}(n)=\frac{(r+1)}{r} * \frac{\sigma^2\left(Z_{1-\beta}+Z_{1-\alpha/2}\right)^2}{d^2}$$

$$(n)=\frac{(1+1)}{1} * \frac{(15)^2(0.84+1.96)^2}{(8)^2}$$

(n) = 2*27.56,

(n) = 55.13+6 (considering 10% dropout of study participants).

Sample size (n) = 61, r = 1 $Z_{1-\beta}$ = 0.84 $Z_{1-\alpha/2}$ = 1.96 σ = 18 d = 10.

So, an investigator will necessitate a minimum (of 61 volunteers) in the control as well as in the case group.

Estimating the Sample Magnitude for Cohort Research

This is a type of research study that follows a group of individuals with similar characteristics who have experienced a common event within a specific time frame, like a particular illness or schooling. Unlike randomized controlled trials, this study does not utilize a control group or patient interventions; instead, participants track their eating habits over time and compare them to their sleep patterns. Unlike randomized controlled trials, observational studies lack patient intervention or a control group. For instance, researchers may ask participants to monitor their eating habits and compare them with their sleep patterns over time.[17,18,22]

Scenario 6: Estimating the sample size for independent cohort research.

$$n=\frac{[Z_{1-\alpha/2}\sqrt{\{(1+1/m)p*(1-p)\}}+Z_{1-\beta}\sqrt{\{p_0*(1-p_0/m)\ p_1(1-p_1)\}}]^2}{(p_0-p_1)^2}$$

n = The total planned study subjects (cases) required to calculate the true relative risk using a two-sided Type of error I, m = The quantity of each experimental subject's control subjects
$Z_{1-\beta}$, or 0.84 for 80% power and 1.28 for 90% power, is the required power.
The critical value is represented by $Z_{1-\alpha/2}$, along with a standard value for the corresponding confidence level.
(It is 2.58 at 99% CI, or 1% type I error, and 1.96 at 95% CI.)
p_0= Possibility of Event in Controls
The experimental probability (p) is equal to $(p_1+m*p_0)/(m+1)$.

An investigator seeks to determine how smoking affects lung cancer risk. According to a prior study, the case group had a 30% lung cancer rate, while the control group had a 20% lung cancer rate. Determine the sample size needed to do the study at 80% power and 95% confidence interval (CI) with an equivalent quantity of case and control individuals.

On applying:

$$n=\frac{[Z_{1-\alpha/2}\sqrt{\{(1+1/m)p*(1-p)\}}+Z_{1-\beta}\sqrt{\{p_0*(1-p_0/m)\ p_1(1-p_1)\}}]^2}{(p_0-p_1)^2}$$

(n) = 183.17 (n) = 183+18 = 201 (taking into account a 10% research participant dropout) Sample size (n) = 201, r = 1, $Z_{1-\beta}$ = 0.84, $Z_{1-\alpha/2}$ = 1.96, p_0 = 0.2, p_1 = 0.3, m = 1, P = (0.3+1*0.2)/(1+!) = 0.25

Consequently, a minimum of 201 volunteers will be needed for the study, according to the researcher.

Determining the sample size needed for comparative research

This type of study design involves comparing two or more groups based on predetermined characteristics such as knowledge, attitude and perception. An integrative tactic is often useful when conducting such studies. Comparative research usually involves the use of quantitative secondary analysis data.[23]

Scenario 7: Sample size where the percentage is the study's parameter and data are on a nominal or ordinal scale.

$$\text{Sample size}(n)=\frac{p_1(1-p_1)+p_2(1-p_2)}{(p_1-p_2)^2} * C$$

n = Sample size required to determine for a single group
Proportion of the two groups, p1 and p2.
C = Standard value for the matching α and β level chosen for the research.

It is as follows:

	$z_{1-\alpha/2}$	0.05	0.01
$Z_{1-\beta}$	0.8	7.85	11.68
	0.9	10.51	14.88

The goal of the study is to compare Group A's and Group B's expertise. The proportions of the two groups-50% and 25%, respectively- are drawn from the earlier study. For the study, the researcher seeks 80% power and 95% CI. Determine how many volunteers are needed in total for new research.

On applying:

$$\text{Sample size}(n)=\frac{p_1(1-p_1)+p_2(1-p_2)}{(p_1-p_2)^2}*C$$

$$(n)=\frac{0.5(1-0.5)+0.25(1-0.25)}{(0.5-0.25)^2}*7.84$$

(n)=7.33*7.85 (n)=57.56+6 (taking into account a 10% research participant dropout rate).

At 95% confidence interval and 80% power, sample size (n)=63.56≈64, p1=50%=0.5, p2=25%=0.25, and C=7.85.

For the study in each group, a researcher will thus need a minimum of 64 people.

Scenario 8: Sample size determination, when research parameters are mean and data are on an interval/ratio scale.

$$\text{Sample size } (n)=\frac{(\sigma_1^2+\sigma_2^2)^2\left(Z_{1-\beta}+Z_{1-\alpha/2}\right)^2}{(d)^2}$$

d = the two groups' mean differences (effect size)
σ_1 = Group 1's SD σ2 = Group 2's SD
$Z_{1-\beta}$ = The intended power
$Z_{1-\alpha/2}$ represents the critical value, together with a standard value for the associated confidence level.
(It is 1.96 at 95% CI and 2.58 at 99% CI, or 1% type I error.)

Based on the prior research, it can be inferred that Group A's and Group B's HbA_{1c} levels were, respectively, 5.5 (3.2) and 7.5 (3.0) mean (SD). Determine the sample size for the new research at CI-95% and the power of the investigation is 80%.

On applying:

$$\text{Sample size } (n)=\frac{(\sigma_1^2+\sigma_2^2)^2\left(Z_{1-\beta}+Z_{1-\alpha/2}\right)^2}{d^2}$$

$$(n)=\frac{(3.2^2+3^2)^2(0.84+1.96)^2}{2^2}$$

(n)=162.20+16 (taking into account of 10% dropout).

Sampling size(n)=78.

Hence, an investigator will need at least (178 volunteers) for the study in each group which means an overall of 356 study volunteers is mandatory.

Scenario 9: The sample size determination for continuous variables, and group comparison.

$$\text{Sample size } (n) = 1+2C\ (SD/d)^2$$

n = The sample size that we must determine for one group.
d = Difference between two groups' means found (effect size)
σ = Standard Deviation
c = Constant value is contingent upon the values of α and β chosen for the investigation.

It is as follows:

	$z_{1-\alpha/2}$	0.05	0.01
$Z_{1-\beta}$	0.8	7.85	11.68
	0.9	10.51	14.88

A scientist is interested in determining the impact of medication X and contrasting it with an inactive drug/placebo. He deliberates that uncertainty of drug X reduces sugar in the blood through 25 mg/dL in comparison to an inactive drug/placebo, and then the situation would be considered clinically relevant. Let us assume that the SD was 50 mg/dL in prior research. He chose to do the study at 80% power with a 95% confidence interval/CI.

Exploit:

Sample size (n)=1+2C $(SD/d)^2$
(n)=1+2*7.85$(60/20)^2$
(n)=1+282.6

(n)= 283.6+28 (taking 10% dropout into account)
At 95% confidence interval and 80% power, sample size (n) = 311 d = 20 mg/dl, SD = 60 mg/dl, and C = 7.85.

The study requires a total of 622 subjects, with a minimum of 311 participants in each group.

ESTIMATING THE SAMPLE SIZE FOR EXPERIMENTAL RESEARCH

Randomized controlled trials, also known as experimental studies, involve researchers intentionally manipulating study factors. Control groups and randomization are essential elements of many types of research. In this type of study, the researcher introduces an intervention, measures its impact, and compares the results of the intervention group with those of the control group.[22-24]

The following categories of comparison exist (details are available in Table 2)
Scenario 10: To calculate sample size for comparing two groups, exclude the effect size and determine based on nominal/ordinal or proportion variables.

Table 2: Types of trial and their description.

Types of trials	Description
Superiority trial	The goal of this kind of study is to prove that a novel medication, therapy, or intervention is better than a control treatment.
Equivalence trial	The goal of this kind of study is to prove that a novel medication, therapy, or intervention is better than a control treatment.
Non-inferiority trial	This kind of study seeks to demonstrate that a newbie intervention or therapy is beneficial, but it need not be better than the control treatment.

n = Sample size for each group
d = Difference in means of two treatment effect
Zx = Standard value for a one or two-tailed
σ_2 = SD of Group 2
δ_0 = Acceptable margin of error
S_2 = Pooled SD (both comparison groups)
p = Response rate of standard intervention

The purpose of the study is to compare the efficacy of treatment B (standard intervention) with treatment A (novel intervention) in treating stroke patients over two months.

d=0.16; δ=0.21, δ0 =0.10; p = 0.40; p0 =0.55; a=0.05 (95% CI); β=0.20 (80% power).

Non-inferiority trial

$$\text{Sample size}(n) = 2 * \left\{ \frac{Z_{1-\beta} + Z_{1-\alpha}{}^2 p}{\delta_0} \right\} * (1-p)$$

n = 260.40+26 (taking 10% dropout into account)

Equivalence trial

$$\text{Sample size}(n) = 2 * \left\{ \frac{Z_{1-\beta} + Z_{1-\alpha/2}{}^2}{\delta_0} \right\} * p * (1-p)$$

n=330.45+33 (considering 10% dropout)
Sample size (n) = 363

Superiority clinical trial

$$\text{Sample size}(n) = 2 * \left\{ \frac{Z_{1-\beta} + Z_{1-\alpha}{}^2}{\delta - \delta_0} \right\} * p * (1-p)$$

(n) = 214+21 (taking 10% dropout into consideration)

Scenario 11: Determining the Sample Size for Comparing Two Groups.

A researcher is conducting a study to compare the efficacy of drug A (the test drug) and drug B (the control drug) in treating diabetes. The study aims to measure the change in glucose level (in mg/dL) of blood from the baseline; with a 95% confidence interval and 80% study power. The researcher assumes all parameters to be equal. The aim is to determine the sample size needed to detect any change (outcome extent) between the 2 groups based on the variance in the nasty or incessant variables.[24]

$$Z_{1-\alpha} = 1.645,\ Z_{1-\beta} = 0.84,\ Z_{1-\alpha/2} = 1.96$$

The mean change in blood glucose of the group receiving new medication therapy is 20 mg/dl, while the group receiving conventional therapy has a mean change of 16 mg/dl. The pooled standard deviation (S) of both comparison groups is 10 mg/dl. The real margin between the two interventions is 4, and the clinically allowed difference is 2.

Non-inferiority trial

$$\text{Sample size}(n) = 2 * \left\{ \frac{Z_{1-\beta} + Z_{1-\alpha}{}^2}{\delta_0} \right\} * SD^2$$

$$(n) = 2 * \left\{ \frac{0.84 + 1.645^2}{2} \right\} * 10^2$$

(n) = 308.7+30 (taking 10% dropout into account)

339 is the sample size (n).

Equivalence trial

$$\text{Sample size}(n) = 2 * \left\{ \frac{Z_{1-\beta} + Z_{1-\alpha/2}{}^2}{\delta_0} \right\} * SD^2$$

$$n = 2 * \left\{ \frac{0.845 + 1.96^2}{2} \right\} * 10^2$$

n = 280.5+28 (considering 10% dropout)
Sample size (n) = 308

Table 3: List of online resources for calculating sample sizes.

Software	URL Link
Raosoft, Inc.	http://www.raosoft.com/samplesize.html
Creative Research Systems	https://www.surveysystem.com/sscalc.htm
G*Power Software finder	https://stats.oarc.ucla.edu/other/gpower/ Free download: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html
PS Power and Sample Size Calculation	https://biostat.app.vumc.org/wiki/Main/PowerSampleSize
Java applets	https://homepage.divms.uiowa.edu/~rlenth/Power/index.html
Growth Analyzer (RCT - Research Calculation Tools)	https://growthanalyser.org/programs/growth-analyser-research-calculation-tools/
Epi Info	Sample Size:X-Sectional, Cohort, and Randomized Clinical: https://www.openepi.com/SampleSize/SSCohort.htm OpenEpi - Sample Size for Unmatched Case-Control Studies: https://www.openepi.com/SampleSize/SSCohort.htm
PASS Sample Size Software	https://www.ncss.com/software/pass/
IBM SPSS	https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=features-power-analysis
SurveyMonkey	https://www.surveymonkey.com/mp/sample-size-calculator/
Qualtrics™	https://www.qualtrics.com/blog/calculating-sample-size/
Eval Academy	https://www.evalacademy.com/articles/finding-the-right-sample-size-the-easy-way
Sigma Magic Analysis software	https://www.sigmamagic.com/blogs/online-sample-size-calculators/
CheckMarket	https://www.checkmarket.com/sample-size-calculator/
OvitionMR	https://www.ovationmr.com/sample-size-calculator/
hotjar	https://www.hotjar.com/poll-survey-sample-size-calculator/
Classgist	https://www.classgist.com/sample-size-calculator.aspx
SurveySparrow	https://surveysparrow.com/blog/sample-size-calculator/
QuestionPro	https://www.questionpro.com/sample-size-calculator/
Biomath	http://www.biomath.info/
Statistical considerations for clinical trials and scientific experiments	https://hedwig.mgh.harvard.edu/sample_size/size.html
Statistics calculators	http://danielsoper.com/statcalc3/default.aspx
Australian Bureau of Statistics	https://www.abs.gov.au/websitedbs/d3310114.nsf/home/sample+size+calculator
WINPEPI (PEPI for windows)	http://www.brixtonhealth.com/pepi4windows.html
Laboratory Animal Service Centre	http://www.lasec.cuhk.edu.hk/sample-size-calculation.html

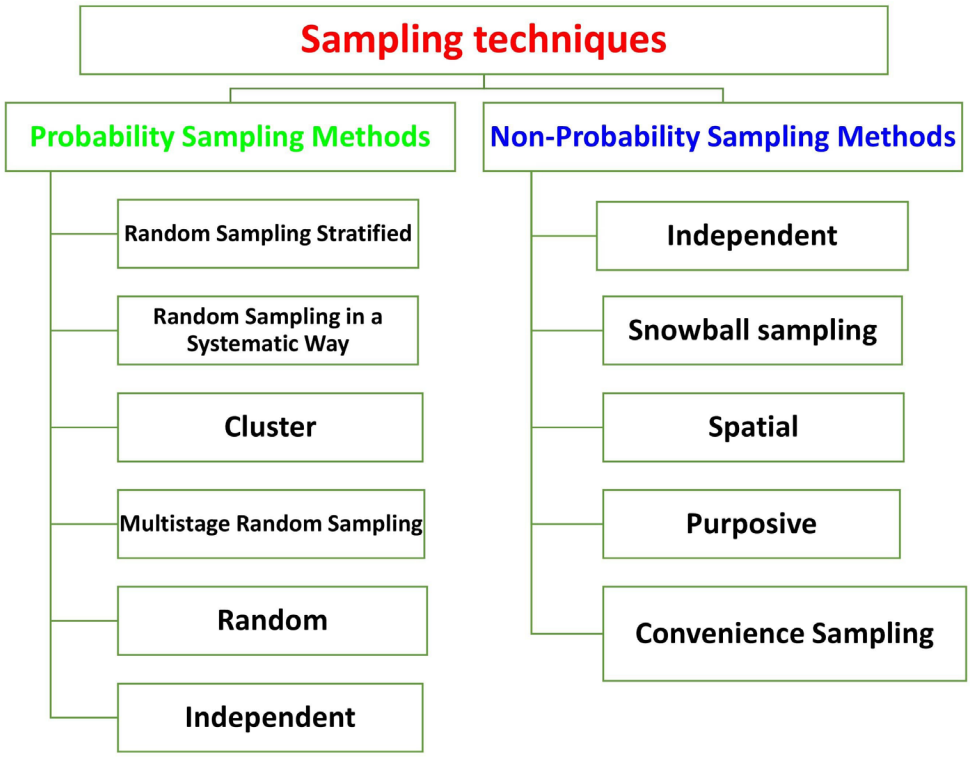

Figure 1: Classification of Sampling Techniques.

Superiority Clinical trial

$$\text{Sample size}(n) = 2 * \left\{ \frac{Z_{1-\beta} + Z_{1-\alpha}{}^{2}}{\delta - \delta_0} \right\} * SD^2$$

(n) = 248.5+24.8 (considering 10% dropout)
Sample size (n)=273.

Calculating Sample Size from Programmes

It is now easy to find online resources for calculating sample sizes. Here are a few commonly used tools in Table 3.[24]

CONCLUSION

This review article offers a comprehensive discussion of various types of samples, sampling techniques, and sample size estimation formulas for observational and experimental research in health and social sciences. This information will assist health and social science researchers in identifying a suitable sample size for their research to generate more accurate and consistent results, thereby improving the generality of research results.

ACKNOWLEDGEMENT

The authors would like to express their heartfelt gratitude and appreciation to Dr. Jenan Al-Matouq, Vice Dean of Academic Affairs at Mohammad Al-Mana College of Medical Science, for allowing us to collaborate and complete this project on time.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

α: Type-I Error; **β:** Type-II Error; **CI:** Confidence Interval; **COVID-19:** Coronavirus Disease; **DEFF:** Effect of design; **R:** Sampling Interval; **N:** Overall Number; **n:** Desired Sample Magnitude; **mmHg:** Millimetres of Mercury; **P:** Probability; **RCTs:** Randomised Controlled Trials; **SD:** Standard Deviation

REFERENCES

1. Button KS, Ioannidis JPA, Mokrysz C, *et al.* Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365-76.
2. Wilson Vanvoorhis CR, Morgan BL. Understanding power and rules of thumb for determining sample sizes. Tutor Quant Methods Psychol. 2007;3(2):43-50.
3. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Medica. 2021;31(1):1-27.
4. Curtis MJ, Bond RA, Spina D, *et al.* Experimental design and analysis and their reporting: new guidance for publication in BJP. Br J Pharmacol. 2015;172(14):3461.
5. Bartlett JE, Kotrlik JW, Higgins CC. Organizational research: determining appropriate sample size in survey research. Inf Technol Learn Perform J. 2001;19(1):43-50.
6. Umulisa C. Sampling Methods and Sample Size Calculation for the SMART Methodology June 2012 Table of Contents. University [Internet]. 2012:20-30.
7. Chow SC, Shao J, Wang H, Lokhnygina Y. Sample Size Calculations in Clinical Research: Chapman and Hall/CRC Biostatistics Series. 3rd ed. New York: Taylor and Francis; 2017.
8. Fox N., Hunn A., and Mathers N. Sampling and sample size calculation The NIHR RDS for the East Midlands/Yorkshire and the Humber 2007. Available from: https://www.bdct.nhs.uk/wp-content/uploads/2019/04/Sampling-and-Sample-Size-Calculation.pdf
9. Sampling methods types, techniques and examples. (2022). Accessed: March 4, 2024: https://www.scribbr.com/methodology/sampling-methods/

10. Kandi V: Research process, study variables, statistical validations, and sampling methods in public health related research: an update. Am J Public Health Res. 2022;10:1-8.
11. Rao NSN (1985): Elements of Health Statistics, First edition, R. Publication, Varanasi, India.
12. Singh, Ajay S., and Micah B. Masuku. "Sampling techniques and determination of sample size in applied statistics research: An overview." International Journal of economics, commerce and management. 2014;2(11):1-22.
13. Zhong B. How to calculate sample size in randomized controlled trial? J Thorac Dis. 2009;1(1):51-4.
14. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: Guided by information power. Qual Health Res. 2016;26: 1753-60.
15. Kirby A, Gebski V, Keech AC. Determining the sample size in a clinical trial. Med J Aust. 2002;177:256-7.
16. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. London: Academic Press; 1977.
17. Kadam P, Bhalerao S. Sample size calculation. Int J Ayurveda Res. 2010;1:55-7.
18. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ, *et al.* Sample size calculations: Basic principles and common pitfalls. Nephrol Dial Transplant. 2010;25(5):1388-93.
19. Ebrahim Valojerdi A, Tanha K, Janani L. Important considerations in calculating and reporting of sample size in randomized controlled trials. Med J Islam Repub Iran. 2017 Dec 25;31:127.
20. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. J Clin Diagn Res. 2016;10(10):YE01-6.
21. Singh AS, Masuku MB. Sampling techniques and determination of sample size in applied statistics research: An overview. International Journal of economics, commerce and management. 2014;2(11):1-22.
22. Mishra P, Pandey MP, Singh U, Sharma V, Yadav SS, Kar R. Sample size estimation for clinical research studies using mean and proportion. Int J Sci Res. 2017;6:587-79.
23. Charan J, Biswas T. How to calculate sample size for different study designs in medical research?. Indian J Psychol Med. 2013;35:121-6.
24. Sharma SK, Mudgal SK, Thakur K, Gaur R. How to calculate sample size for observational and experimental nursing research studies. National Journal of Physiology, Pharmacy and Pharmacology. 2020;10(1): 1-8.

Cite this article: Ali MD, Hatef EAJ. A Review Article: Types of Sampling and Sample Size Determination in Health and Social Science Research. J Young Pharm. 2024;16(2):204-15.