Brief Report

Sample Size Calculation in Clinical Trial and Animal Studies

Alireza Pakgohar[1*], Hossein Mehrannia[2]

[1]Departmment of Statistics, Payame Noor University (PNU). Tehran, Iran.
[2]Department of Statistics, Behbahan Branch, Islamic Azad University, Behbahan, Iran.

Abstract

Objective: This review aims to address the critical role of sample size calculation in designing and conducting animal studies related to diabetes and obesity, particularly focusing on experimental comparison methods like ANOVA design.

Materials and Methods: We discuss the factors influencing sample size decisions, including type I and type II errors, effect size, and standard deviation. We emphasize the importance of avoiding common pitfalls, such as using rules of thumb or arbitrary choices, and advocate for utilizing established formulas to ensure accurate and reliable sample size determination.

Results: This review presents relevant equations for calculating sample size in animal studies with an ANOVA design, providing researchers with a framework for determining the appropriate number of animals needed to achieve robust and ethical research.

Conclusion: Accurate sample size calculation is essential for achieving powerful and statistically sound animal studies in diabetes and obesity research. Utilizing established formulas and avoiding arbitrary choices ensures reliable and ethical research practices while minimizing resource waste and maximizing the validity of collected data.

Keywords: Sample size calculation, Clinical trials, Animal studies, Resource Equation Method, Randomized Controlled Trials Method

Corresponding Author:

Alireza Pakgohar, Department of Statistics, Payame Noor University, Tehran, Iran.

Tel: (98) 913 318 0781
Email: a_pakgohar@pnu.ac.ir
Orcid ID: 0000-0003-2512-1581

Introduction

Sample size calculation is an important step in planning a clinical or trial study, as it affects the validity, reliability, and ethicality of the research. If sample size is too small, it may not accurately represent the population being studied, leading to biased results and limited generalizability. On the other hand it is clear that if sample size is too large, it may be a waste of resources and time. Sample size calculation depends on various factors, such as the type of study design, the type of outcome, the expected effect size, the variability of the data, and the significance level and power chosen by the researchers. For more details see (1).

In clinical or animal experimental studies, it is important to determine the appropriate sample size to ensure reliable and generalizable results. If sample size is too small, significant differences between treatment and control groups may be missed, and the findings may not be applicable to the larger population. Conversely, if sample size is too large, it may lead to the detection of insignificant differences and waste resources and time. For more information see (2,3).

Dell et al. (2002) highlight the importance of determining the appropriate sample size for research studies (4). They discuss two methods for doing this: estimating sample size based on previous experience and using power calculations with formulae. The authors suggest that estimating sample size based on previous experience can be helpful when there is enough reliable data from similar studies, and when the effect size and variability are well-established and consistent. However, they caution against relying on previous experience when there is a lack of data, outdated or irrelevant data, or when the effect size and variability are uncertain or vary. The authors emphasize the need to calculate sample size or power using various potential parameters to ensure a study design that is sensitive and robust, and to avoid studies that are either too weak or too powerful. Additionally, they

provide guidance on how to conduct and report sample size calculations for different types of studies and outcomes.

Wittes (2002) emphasizes that sample size calculation is an important and complex step in designing and conducting randomized controlled trials, as it affects the validity, reliability, and ethicality of the research. The author argues that sample size calculation requires precise specification of the primary hypothesis and the method of analysis, as well as careful consideration of the available data and evidence. The author also suggests that sample size calculation should be viewed as an approximation rather than a precise value, and that researcher should consult with statisticians and report their methods and results transparently (5).

Who to calculate sample size

The findings of this article have been obtained using a review method and by choosing a systematic review method. The main purpose of the upcoming review study was to summarize new findings and materials in the field of sample size in the field of animal studies and clinical trials and critically evaluate it.

To find documents related to writing a review article using keywords such as "sample size", "sample size calculation", "sample size method", " research methods of animal study" were searched for content in specialized science databases such as Google Scholar, PubMed, Base Search, SID.ir, and Elmnet.ir in the last 20 years.

To select the used documents, the titles found by the search engine were first checked in terms of thematic relevance. The found materials were divided into three groups: internet portal, article and book. The criterion for choosing internet portals after thematic connection was having an academic (.ac) or educational (.edu) extension. After reviewing these portals, the ones that were more complete than the others were selected as

references. After examining the title, in the next step, the articles were evaluated in terms of the relevance of the keywords of the article and the abstract with the purpose of the research. The cases that did not cover the subject of size and included the subject of sample size techniques but were not mentioned in other subjects or were not addressed in the field of animal studies and clinical trials care were removed from the list of reviewed references. In the third stage, selected documents were scanned. In the end, if necessary, the content was criticized by the author.

There are the main approaches or formulas for sample size calculation in clinical and trial studies, depending on the specific situation and objective of the study. Some of the common approaches or formulas are: power analysis, precision-based sample size and non-inferiority or equivalence trials significance level, validity, reliability and controlling. In this section we discuss the mention concepts and obtain some formals for sample size calculation in clinical trials and animal studies.

Power analysis

Power analysis is a crucial statistical method employed to ascertain the sample size necessary for a research study to detect a statistically significant effect or difference between groups with a specified level of confidence. This concept takes into account several key factors, including the significance level, types I and II errors. By carefully considering these factors, power analysis aids researchers in determining the optimal sample size required to achieve sufficient statistical power. The desired power, typically set at 80% or 90%, represents the probability of correctly detecting a true effect or difference if it exists. Meanwhile, the significance level, commonly established at 0.05, denotes the threshold for determining statistical significance (6). By increasing sample size we expect to increase the power of test. However, it is important to remember that statistical significance does not necessarily equate to practical significance or

real-world importance. While a large sample size may increase the likelihood of detecting a statistically significant result, it does not guarantee that the observed effect is meaningful or impactful in practical terms. In fact, focusing solely on statistical significance without considering effect size can lead to misleading conclusions in which the effect size refers to the magnitude or strength of the relationship between variables being studied. Therefore, even if a claim is statistically significant due to a large sample size, it is crucial to assess its effect size before drawing any definitive conclusions. Because a small effect size suggests that while there may be a statistically significant relationship between variables, it might not have much practical relevance or impact in real-world scenarios (3).

Precision-based sample size calculation

In some studies, the objective may be to estimate a parameter with a certain level of precision or accuracy. This approach involves calculating sample size required to achieve a specific margin of error or confidence interval width around the estimated parameter (7). To perform a precision-based sample size calculation, researchers use statistical formulas or software tools specifically designed for this purpose. These calculations take into account the desired margin of error, confidence level, standard deviation (equivalently variance), effect size, and statistical power to determine the optimal sample size.

Non-inferiority trials

Non-inferiority and equivalence trials compare a new treatment to a standard treatment in clinical research. They are used when using a placebo is not possible or ethical.

In a non-inferiority trial, the goal is to prove that the new treatment is not worse than the standard treatment by a certain amount called the non-inferiority margin. If the new treatment falls within this margin, it is considered non-inferior. Equivalence trials aim to show that two treatments are essentially the

same in terms of effectiveness or safety. These trials require a larger sample size to detect smaller differences between the treatments. Both trials randomly assign patients to either the new treatment group or the standard treatment group. The primary outcomes are compared to determine if the new treatment meets the non-inferiority or equivalence criteria. These trials provide evidence for introducing new treatments that offer advantages without sacrificing effectiveness or safety, but it is important to carefully select the non-inferiority margin and calculate sample size accurately (8).

Statistical significance

Statistical significance and clinical significance are two different concepts. Statistical significance determines if a difference observed in a study is likely due to chance, while clinical significance assesses the practical importance of the findings. It is important to note that even if a study does not show statistically significant results, there could still be meaningful differences in the population (9).

Variability and reliability

Variability within the data can affect the validity and reliability of study results. Higher variability makes it harder to detect significant differences as random fluctuations can mask true effects. Researchers should consider sources of variability, such as individual differences or measurement errors, and aim to minimize their impact. For more details see: Martic-Kehl et al. (10).

One way to address variability is by determining an appropriate sample size. A larger sample size reduces random fluctuations and provides a more accurate representation of the population, improving statistical power.

Controlling:

Controlling for potential confounding variables is another strategy. Confounders are factors associated with both the independent and outcome variables, making it difficult to determine true effects. Researchers can address confounders by carefully selecting participants or using statistical techniques to account for them and minimize their impact on study results (11). Controlling sample size in animal studies is important for accurate and reliable results. Adequate sample size allows for the detection of significant differences between groups and reduces the chances of false-positive or false-negative findings. Factors affecting sample size determination include statistical power, effect size, variability, and type I error rate. Various methods, such as power analysis, sample size tables, and online calculators, can be used to determine the appropriate sample size for animal studies.

Sample size Methods

This subsection delves into the intricacies of sample size calculation methods for randomized controlled trials (RCTs), the precision-based formula and the simulation-based method, besides, high lighting their theoretical underpinnings, practical applications, and considerations for optimal study design.

The standard formula (RCT Methods)

The standard formula for calculating sample size in randomized controlled trials (RCTs) takes into account factors like the likelihood of wrongly concluding a significant difference between groups (type I error), the likelihood of failing to detect a significant difference if it exists (type II error), the proportion of events observed in the control group (control event rate), and the meaningful difference between intervention and control groups (desired treatment effect). By considering these factors, researchers can determine sample size needed to ensure the study has enough power to find meaningful effects and reduce false conclusions (12,13).

The precision-based formula

The formula used in this context takes into account factors like type I and type II errors, control event rate, and desired treatment effect. It focuses on estimating effect size with precision rather than just statistical significance. The confidence level determines the probability that the calculated interval will contain the true effect, usually set at 95%. The margin of error indicates how much deviation from the estimated effect size is allowed, with smaller margins indicating better precision but needing larger sample sizes (1). The control event rate helps estimate the baseline risk and provides context for evaluating treatment effects, and the desired treatment effect acts as a benchmark for comparing actual effects. Determining an appropriate sample size is important to ensure statistical power and detect meaningful differences between groups. If the observed treatment effect is greater than the desired effect, it suggests high effectiveness and potential additional benefits. However, if it falls short, researchers can assess various intervention factors to identify areas for improvement (14).

The simulation-based method

The described method helps researchers determine the appropriate sample size for studies with complex designs or advanced analysis methods. Researchers generate data based on the study's assumptions and parameters, using the planned analysis method to determine the effect estimate and its uncertainty. Simulating data based on the assumptions of these designs provides more accurate estimates of the necessary sample size (15,16).

In research, we can simulate count data using Poisson or negative binomial regression models, simulate event times using Cox proportional hazards models, and generate simulated data for ordinal scales using ordered logistic regression or proportional odds models. Simulated data helps researchers understand the impact of predictor variables, test hypotheses, evaluate findings, and perform sensitivity analyses. By manipulating predictor variables, they can observe how changes affect the outcome. Simulated data also allows them to explore different scenarios and examine how variations in predictors or sample characteristics affect the predicted probabilities of different outcome categories (17). Simulated data is also valuable for testing statistical models and algorithms, exploring hypothetical scenarios, and improving algorithms by identifying weaknesses and confirming results. It provides advantages in reproducibility and transparency, as synthetic datasets can be shared without revealing sensitive information. The method can be implemented using existing software packages, such as Stata or R. The method can be found in (12,18).

Sample Size Formulas

This section discusses various methods for calculating sample sizes in animal studies. The paper introduces two main approaches: sample size calculation based on power and infection proportion, and sample size calculation using the resource equation method.

Sample size based on Power and Infection Proportion

In the first approach, the formula calculates the number of animals needed to detect the presence of a pathogen. The formula takes into account the chosen power (β) and the proportion of infected animals (p). The formula is (Dell, 2002):

$$n = \frac{\log \beta}{\log q} \qquad [1]$$

Where q= 1- p. This formula helps determine the appropriate sample size based on the desired power and infection proportion.

Example 1: If 30% of animals are infected and the investigator wants a 95% chance of detecting the infection, the formula suggests sampling 9 animals. If the infection prevalence is lower, like 10%, the formula recommends sampling around 30 animals.

Sample Size on Resource Equation Methods

The second approach, known as the resource equation method, is useful for complex experiments with multiple treatment groups. It involves analyzing the results using analysis of variance (ANOVA) and considers the degrees of freedom in the analysis. The formula for calculating the degrees of freedom (E) is (Charan and Kantharia, 2013):

$$E = N_t - N_g, \qquad [2]$$

Where N_t is the total number of animals and N_g is the total number of groups. The aim is to keep E between 10 and 20 for an adequate sample size.

Example 2: A researcher is studying the effects of a drug on rats and has divided them into five groups, each with 10 rats. The equation is used to calculate E, which is (10× 5)- 5= 45. This means sample size is more than necessary. However, if each group is reduced to 5 rats, E will equal 20, which is considered adequate.

Sample Size calculation in Randomize Block ANOVA design

In randomized block ANOVA designs, a useful method for determining sample size is the resource equation method. It involves the formula (Festing, 2006):

$$E = N_t - Tr - B + 1, \qquad [3]$$

Where N represents the total number of observations, Tr stands for the number of treatments, B indicates the number of blocks, and E is the error degrees of freedom, ideally falling between 10 and 20.

Example 4: In a within-litter experiment with 3 treatments, an average litter size of 6, and a proposal to use 5 litters, the equation would be

$$E= (6× 5)- 3- 5+ 1= 23. \qquad [4]$$

Sample Size calculation in one-way ANOVA design

When conducting a one-way ANOVA, the within-subject error degrees of freedom (DFw) can be calculated using the following formula:

$$DFw= N_t- k= k.n- k= k (n- 1). \qquad [5]$$

Here, N_t represents the number of subjects, k represents the number of groups, and n represents the number of subjects per group.

To determine the value of n, the formula can be rearranged as follows:

$$n= DFw/ k+ 1. \qquad [6]$$

To determine the minimum and maximum numbers of animals per group, the range of DFw is taken into consideration, and the formulas are modified with minimum= 10 and maximum= 20 values. Finally, to calculate the minimum and maximum numbers of animals required, the formulas

$$\text{Min } (N_t) = \text{ Min } (n) \times k = k(10/k + 1)$$
$$\text{Max } (N_t) = \text{ Max } (n) \times k = k(20/k + 1) \qquad [7]$$

Example 5: In a study comparing continuous variable X between k=4 groups, the sample sizes per group are as follows:

$$\text{Min } (n) = \frac{10}{4} + 1 = 3.5 = \text{ rounded up to 4,}$$

$$\text{Max } (n) = \frac{20}{4} + 1 = 6 = \text{ (rounded down to) 6.} \qquad [8]$$

Then

$$\text{Min}(N_t) = \text{Min } (n) \times k = 4 \times 4 = 16$$
$$\text{Max}(N_t) = \text{Max } (n) \times k = 6 \times 4 = 24. \qquad [9]$$

In conclusion, for this study, it is recommended to have between 4 and 5 animals per group, resulting in a total of 16 to 24 animals. This ensures that the DF remains within the range of 12 to 20.

Sample Size calculation in two independent samples T-test

The provided formula is for a one-way ANOVA, but it can also be used for an independent T-test with two groups by setting the value of k to 2. Both tests will yield the same p-value. To keep the error degrees of freedom (DFw) within the range of 12 to 20, the formula is modified. The minimum number of animals required is calculated using Min (n)= 10/k+ 2 (rounded up), and the maximum number is calculated using Max (n) =20/k+ 1 (rounded down). Finally, the minimum and maximum numbers of animals required are calculated using:

Min (N_t)= k (10/k+ 2) and Max (N_t)= k (20/k+ 1).   [10]

Example 7: In a study comparing continuous variable X between 2 groups (two independent sample T-test), the sample sizes per group are as follows:

Min (n)= 10/2+ 2= 7= rounded up to 6 animals/ group

Max (n)= 20/2+ 1= 11= rounded up to 6 animals/ group   [11]

Then

$Min(N_t) = Min(n) \times k = 7 \times 2 = 14$

$Max(N_t) = Max(n) \times k = 11 \times 2 = 22$.   [12]

In conclusion, for this study, it is recommended to have 7-11 animals per group, resulting in a total sample size between 14- 22 animals and the remains within the range of 12 to 20.

Sample Size calculation in repeated measure ANOVA design

In a repeated-measures ANOVA design, there is only one group of subjects. The within-subject error degrees of freedom (DFw) can be calculated using the formula DFw= (N_t− 1) (r− 1), where N_t is the total number of subjects and r is the number of repeated measurements. The calculation of n is not applicable in this case. To determine N_t you can use the formula N_t = DFw/(r-1)+ 1. To find the minimum and maximum numbers of animals required, replace the DFs in the formulas with 10 and 20, respectively. The formulas become:

Min (N_t)= 10/ (r-1)+ 1= rounded up to integer num.animals/ group

Max (N_t)= 20/ (r-1)+ 1=(rounded down to) integer num.animals/group   [13]

If the experiment involves sacrificing the animals at each repetition, N_t must be multiplied by r.

Example 8: In an experiment to investigate the impact of a drug on tumor sizes in a group of animals at 3 different time points (pre-treatment, post-treatment 1, 2, and 3), the sample sizes are determined as follows:

$Min(N_t) = \frac{10}{3-1} + 1 = 6$ animals/group

$Max(N_t) = \frac{20}{3-1} + 1 = 11$ animals/group.   [14]

It is important to note that the minimum and maximum numbers of animals are rounded up and down, respectively, in order to maintain the degrees of freedom (DF) for each sample size within the range of 10 and 20. Therefore, the minimum total sample size is 6 animals multiplied by 3 measurements, resulting in 18 animals. Similarly, the maximum total sample size is 11 animals multiplied by 3 measurements, resulting in 33 animals.

Remark: If the animals need to be sacrificed at each measurement, the total sample sizes are calculated by multiplying the minimum and maximum sample sizes by the number of measurements (r).

Sample Size calculation in paired samples T-test

The formula provided is intended for a repeated measure ANOVA, but can also be used for a paired samples T-test with two steps. In this case, k is set to 2. The formula is effective because the error degrees of freedom (DF) for both tests are equal when comparing two groups.

To calculate the within-subject error degrees of freedom (DFw), the formula

DFw= (N_t− 1) (2− 1).   [15]

Then

$N_t = \frac{DF_w}{2-1} + 1 = DF_w + 1$ [16]

To determine the minimum and maximum number of animals required, the DFs in the formulas are replaced with 10 and 20, respectively, resulting in the expression

$11 \leq N_t \leq 21$ [17]

Sample Size calculation in repeated-measures ANOVA with one between-subject factor

This formula provided calculates the error degrees of freedom (DF) for a repeated-

measures ANOVA with one between-subject factor. The between-subject error DF (DFb) is equal to the total number of subjects (N) minus the number of groups (k). The within-subject error DF (DFw) is calculated as (N-k) (r-1), where r represents the number of repeated measurements within each group. The total degrees of freedom (DFt) is the sum of DFb and DFw, which is equal to k (n-1) (r-1). By rearranging the terms, the number of subjects per group (n) can be determined as DFt/ (k×r) +1. The formulas also provide minimum and maximum numbers of animals needed for the experiment by substituting specific values for n and N into the DF calculations. If the animals are sacrificed at each repetition, n and N need to be multiplied by r.

Example 9: Consider a study that aims to compare three treatment groups with four repeated measurements of continuous variable X in an animal study. To determine the sample sizes per group, we need to consider the error degrees of freedom (DF). The minimum sample size per group is calculated as follows: 10/ (3× 4)+ 1= 1.8, which is rounded up to 2 animals per group. The maximum sample size per group is calculated as 20/ (3× 4)+ 1= 2.6, which is rounded down to 2 animals per group. This results in equal sample sizes for the minimum and maximum N. To calculate the minimum and maximum N (total sample size), we multiply the minimum and maximum n (sample size per group) by 3 (number of treatment groups). Therefore, the minimum (maximum) N= 2× 3= 6 animals. If the animals must be sacrificed at each measurement, the total sample size is calculated by multiplying the minimum (maximum) N by the number of repeated measurements (r), which is 4 in this case. Therefore, the minimum (maximum) N× r= 6× 4= 24 animals.

Ethical considerations

The current study is approved by the ethical committee of Payame Noor University (PNU). Tehran, Iran.

Conclusions

Sample size calculation is crucial in research studies, particularly in clinical trials and animal studies, as it affects the statistical power, reliability, and ethicality of the findings. Researchers must consider factors such as type I and type II errors, effect size, standard deviation, study design, outcome type, data variability, and significance level when determining the optimal sample size. In animal studies, additional considerations are required to minimize sample size. By following these guidelines and consulting with statisticians, researchers can conduct high-quality studies that adhere to ethical principles including the 3R ethical approach and produce accurate and reliable results.

Conflict of Interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Authors' contributions

A. P: conceived and designed the analysis, collected the data, contributed data, performed the analysis, prepared first drafts of the manuscript.

H. M: conceived of the presented idea, planning methodology to reach the conclusion, contributed data and performed the analysis.

All authors have accepted responsibility for the entire content of this manuscript and agreed to be accountable for all aspects of the

work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved and approved the version to be published.

References

1. 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. Biochemia medica. 2021;31(1):27-53.
2. Hickey GL, Grant SW, Dunning J, Siepe M. Statistical primer: sample size and power calculations—why, when and how?. European journal of cardio-thoracic surgery. 2018;54(1):4-9.
3. Pakgohar A. The Role of Sample Size on Interpretation of the Result in Applied Research A Study on the Analysis of Regression Models. Methodology of Social Sciences and Humanities. 2023;29(114):19-34.(in Persian)
4. Dell RB, Holleran S, Ramakrishnan R. Sample size determination. ILAR journal. 2002;43(4):207-13.
5. Wittes J. Sample size calculations for randomized controlled trials. Epidemiologic reviews. 2002;24(1):39-53.
6. Hintze JL. Quick start manual. PASS power analysis and sample size system. Kaysville, UT: NCSS. Safarkhani and Moerbeek. 2008.
7. Zorich Jr JN. Reasons for No Longer Teaching [or Using] the Normal Approximation Confidence Interval.2021.
8. Hahn S. Understanding noninferiority trials. Korean journal of pediatrics. 2012;55(11):403.
9. Zhang W. p-value based statistical significance tests: Concepts, misuses, critiques, solutions and beyond. Computational Ecology and Software. 2022;12(3):80.
10. Ko MJ, Lim CY. General considerations for sample size estimation in animal study. Korean Journal of Anesthesiology. 2021;74(1):23-9.
11. Martic-Kehl MI, Ametamey SM, Alf MF, Schubiger PA, Honer M. Impact of inherent variability and experimental parameters on the reliability of small animal PET data. EJNMMI research. 2012;2:1-0.
12. Wang, D., & Ji, L. (2018). Sample size estimation in clinical research. Annals of Translational Medicine, 6(18), 358.https://doi.org/10.1016/j.chest.2020.03.010
13. Zhong B. How to calculate sample size in randomized controlled trial?. Journal of thoracic disease. 2009;1(1):51.
14. Bell ML. New guidance to improve sample size calculations for trials: eliciting the target difference. Trials.2018;19:1-4.
15. Arnold BF, Hogan DR, Colford JM, Hubbard AE. Simulation methods to estimate design power: an overview for applied research. BMC medical research methodology. 2011;11(1):1-0..
16. Charan J, Kantharia N. How to calculate sample size in animal studies?. Journal of Pharmacology and Pharmacotherapeutics. 2013;4(4):303-6.
17. Rahnenführer J, De Bin R, Benner A, Ambrogi F, Lusa L, Boulesteix AL, et al. Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. BMC medicine. 2023;21(1):182.