Innovations

Sample Size Estimation in Epidemiological Studies: Review Article

R. A. Eltayeb; R. A. Farah; F. A. Almostafa
Department of MIS-Stat and Quant. Methods unit, Faculty of Business and Economic, University of Qassim, Buraidah, KSA

Corresponding Author: **R. A. Eltayeb**

Abstract:
Background: *Estimating sample size is an essential step in the planning stages of epidemiological studies. Although there are many online calculators available to find out sample size for most epidemiological studies, their use needs to understand the underlying factors that determine sample size.* ***Objective:*** *To provide an overview of the factors that determine sample size in epidemiological studies, to explain sample-size estimation for the most common study designs in epidemiological research and to discuss a few sample size reduction techniques.* ***Discussion:*** *If investigators understand key parameters—Type I error, power, effect size, Variability and drop rate—they can estimate the basic sample size. These factors are used throughout this paper to calculate the sample sizes for the following study designs: descriptive cross-sectional study (for continuous and dichotomous data), analytical cross-sectional study, case-control study, cohort study and randomized clinical trial.* ***Conclusion:*** *An investigator can determine the minimum sample size needed with enoTable1: Type I error, Type II error and powerugh power to identify significant effects by carefully choosing the parameter values to include in sample size calculations.*

Keywords: *Sample size, power, effect size, study designs, Type 1 Error*

1. Introduction

When performing any study, the first thing that needs to be determined is how large of a sample is needed to achieve the study's objectives [(1)].Clinical study sample sizes must be carefully designed, neither increasing nor decreasing. A smaller sample size might make it impossible for the study to accurately discriminate between the study groups, which would make it unethical. Moreover, the conditions in the original population are not widely applicable. However, a bigger sample size exposes a larger group of people to the potential harm of the treatment, which makes the study unethical. Also, it squanders money, staff time,

and research resources.[2]. The calculation of sample size depends on many factors, e.g., type of study (descriptive or analytical), variable under study (dichotomous or continuous), variability of the target population in case of continuous variables, the significance level, the power test, the effect size to be detected and drop rate. Researchers can calculate sample sizes using a variety of online calculators for popular clinical study designs, but they must first understand the factors that affect sample size. In this paper we discuss the components to estimate the sample size as well as a few formulas that are applicable with various study design types.

2 Factors determining sample size

2.1 Type I error

Rejecting the null hypothesis (H0) when it is true is a type I error. A false positive is another term for a type I error. A common choice for the type I error rate in statistical tests is 5%, which means that the chance of wrongly rejecting the null hypothesis is acceptable.[3]. The probability of type I error has been called the level of statistical significance or the alpha (α) level.

2.2 Type II error

However, a type I error is not the only type of error that can occur in a hypothesis test. There is also a type II error, which is also known as a false negative, and it occurs when you fail to reject the null hypothesis when it is false. Type II error is commonly represented by β.

2.3 Power

The power of a statistical test is its ability to correctly reject the null hypothesis when it is false, thus avoiding a type II error (which occurs when the null hypothesis is false but not rejected). Mathematically, power is often defined as $1 - \Pr(\text{type II error})$, or $\text{Power} = 1 - \beta$. In other words, it's the probability of correctly detecting a true effect or difference in the population.

There is a trade-off between the probabilities of a Type I and a Type II error. Alpha level adjustment affects the balance between type I and Type II errors. Reducing the alpha level reduces the probability of a type 1 error, which is the incorrect rejection of a true null hypothesis. On the other hand, this raises the possibility of Type II error, or failing to reject a false null hypothesis.

Conversely, raising the alpha level increases the probability of Type I error but decreases the probability of Type II error. It's a delicate balance, and researchers often need to consider the relative costs and consequences of each type of error when determining the appropriate alpha level for their study. [4]

Table1. Type I error, Type II error and the power of test

Observation	Reality	
	H_0 is true	$H_0$is false
H0 is rejected	Type I error false positive (α)	Correct conclusion $(1-\beta)$, Power
Ho is accepted	Correct conclusion (1- α)	Type II error false negative (β)

2.4 Variability

The variance or the standard deviation, represents variability. These measures indicate how spread out the data points are within a certain population. Larger variances or standard deviations indicate greater variability, meaning the data points are more dispersed from the mean. Generally, larger sample sizes are needed to increase the reliability of the results when there is high variability. This is because more data points are needed to capture the variability of the population. Conversely, when variability is low, smaller sample sizes may be sufficient to achieve reliable results [(3)].

When the outcome is dichotomous (i.e., it falls into one of two categories), such as success/failure, yes/no, or presence/absence, the calculation of sample size may not directly require the standard deviation. Instead, it often relies on parameters such as the expected proportion of successes or the anticipated effect size. Information from pilot studies or previous research can be valuable in estimating parameters such as variance or success proportions for sample size calculations in future studies.

2.5 Effect size

Effect size represents the magnitude of the difference or relationship between two groups or variables being studied. The effect size plays a crucial role in determining the necessary sample size for analytical studies, when the effect size is large, it means that there is a substantial difference or relationship between the groups or variables. In such cases, a smaller sample size may be sufficient to detect this effect with statistical significance. Conversely, when the effect size is small, a larger sample size is typically required to detect the effect accurately [(1)].

Estimating the effect size can indeed be subjective and often relies on clinical judgment or prior research findings. Pilot studies or a thorough review of the existing literature can provide valuable insights into what effect sizes are reasonable or expected in a given context. [(6)].

Table2. Factors affecting on the sample size of a study and the direction of their impact

Factor	Magnitude	Required sample size	Direction
Type I error	Small	Large	↑↓
	Large	Small	
Power	Low	Small	↑↑
	High	Large	
Variability	Small	Small	↑↑
	Large	Large	
Effect size	Small	Large	↑↓
	Large	Small	

3. Sample size calculations

3.1 Sample size for descriptive studies

Descriptive studies mean that no comparison made between groups. The most popular types of descriptive study designs are case reports, case series, and cross-sectional surveys. Case reports provide a comprehensive account of a single patient's symptoms, signs, diagnosis, course of therapy, response to that course of care, and post-treatment follow-up. On the other hand, case series present a group of patients who share a same or related medical condition. [(3)]. Descriptive cross-sectional studies describe the prevalence or distribution of one or more health outcomes in a given population[(7)]. When conducting descriptive study, an investigator estimates the mean for continuous variables or the proportion for dichotomous variables, then presents the confidence intervals surrounding the estimates to evaluate the accuracy of the sample estimate [(5)].

3.1.1Sample size for continuous variable

When dealing with continuous variables like blood pressure, BMI, and LDL cholesterol level, the sample size can be determined by applying equation (1)

$$n = \frac{4s^2 Z_\alpha{}^2}{d^2} \qquad (1)$$

Where

S =Standard deviation derived from pilot surveys or prior studies.

Z_α= Critical point from Z-table

d= Width of the confidence interval

Keep in mind that since statistical power is only relevant for statistical comparisons, Equation (1) is independent of it.

Example: Assume a researcher wishes to determine the average BMI in diabetes Type 1 Community who are between the ages of 18 and 65, with a 95% confidence interval of no more than ±0.3. Given that the standard deviation from previous research was 1.25. Therefore, the necessary sample size would be,

$$n = \frac{4 \times 1.25^2 \times 1.96\ ^2}{0.6^2} = 67\ patients\ required.$$

3.1.2 Sample size for dichotomous variable

Spouse that the researcher wishes to determine the disease's prevalence in a specific population. He must decide on the width of the confidence interval and desired level of confidence; equation (2) can be used to determine the necessary sample size.

$$n = \frac{4P(1-P){Z_\alpha}^2}{W^2} \qquad (2)$$

P =Expected proportion in the population, derived from prior studies or pilot investigations.

Z_α= Critical point from Z-table
W= Confidence interval's width.

For example, suppose that investigators carry out a cross-sectional study to determine the obesity prevalence in a city. A previous study indicated that the proportion of obese people was 12%. His goal is to determine the sample size using a 95% confidence interval that is no greater than ±0.04. Thus, in accordance with (2), the sample size will be:

$$n = \frac{4 \times 0.12 \times 0.88 \times 1.96^2}{0.08^2} = 254\ subjects\ required$$

3.2 Sample size for analytical studies

Studies with an analytical approach are employed to test hypotheses or to examine the relationship between exposure and outcome. Their roles are understanding the "how" and "why" of disease occurrence. Prospective and retrospective cohort studies follow participants from exposure to outcome over time. Case-control studies, on the other hand, go backward, connecting exposure to outcome [(3)].

While analytical cross-sectional studies enable the researcher to examine the association between exposure and result, they do not provide the drawing of conclusions about the time sequences between exposure and disease [(8)].

3.2.1Sample size for analytical cross-sectional studies

In this type of studies, the investigators need to determine the estimated prevalence of the outcome, the minimum clinically meaningful difference (effect size) between exposed and unexposed groups that they aim to detect, the required power and the significance level.

Let m stand for the desired sample size from the unexposed population and cm ($0<c<\infty$) for the exposed population, where c is predetermined. N=(c+1) n is the total number of samples. The two samples are of similar size when c=1.

- Zα= the critical value of the standard normal distribution corresponding to a specific significance level α.
- $p_1: proportion\ of\ desise\ in\ unexposed\ group$

- p_2: $proportion\ of\ desise\ in\ exposed\ group$
- c: $number\ of\ subjects\ not\ exposed\ for\ every\ subject\ exposed$
- Z_β = the critical value of the standard normal distribution corresponding to a specific level of statistical power β.

The investigators can use Fleiss's formula

$$n = \frac{n'}{4}\left[1 + \sqrt{1 + \frac{2(c+1)}{n'c|p_2 - p_1|}}\right]^2 \qquad (3)$$

for the required sample size from the unexposed population, and cn for that from the exposed population.

Where,

$$n' = \frac{\left[Z\alpha/2\sqrt{(c+1)p(1-p)} + Z_\beta\sqrt{cp_1(1-p_1) + p_2(1-p_2)}\right]^2}{c(p_2 - p_1)^2}, \quad p = \frac{p_1 + cp_2}{c+1} \qquad (4)$$

The values of $p_1 and\ p_2$may be obtained through pilot study or previous studies.[9]

Example: Assuming that an investigator wishes to examine the effect of smoking on lung cancer and according to previous studies, proportion of lung cancer in unexposed may be around 8% and in exposed it can be around 15%. The investigator plans to study double as many nonsmokers as smokers. With significance level at 5% and the power at $1 - \beta = 0.80$
so that $c = 2$. The value of n' in (4) is

$$n' = \frac{\left[2.576\sqrt{3 \times 0.13 \times 0.87} + 0.84\sqrt{2 \times 0.08 \times 0.92 + 0.15 \times 0.85}\right]^2}{2(0.15 - 0.08)^2}$$
$$= 384.34$$

so that, by (3), the required sample size from the population of nonsmokers is

$$n = \frac{384.34}{4}\left[1 + \sqrt{1 + \frac{2 \times 3}{384.34 \times 2 \times 0.07}}\right]^2 = 406$$

The required sample sizes for nonsmokers and smokers are therefore $406\ and\ 406 \times 2 = 812$ respectively. The total number of individuals required is1218.

3.2.2 Sample size for case-control studies

In this type of studies, investigators compare a group of individuals who have the disease or condition under consideration (cases) with a group of individuals who do not have the disease or condition (controls). Then, they examine the past to determine whether there are any differences between the cases and controls regarding their exposure to a specific risk factor or multiple factors[10]. The formula (5) is used to calculate the sample for this design

$$n = \frac{(c+1)(1+(\vartheta-1)P_0)^2}{cP_0^2(P_0-1)^2(\vartheta-1)^2}\left[Z_{1-(\frac{\alpha}{2})}\sqrt{(c+1)P_t{}^*(1-P_t{}^*)} + Z_{1-\beta}\sqrt{\frac{\vartheta P_0(1-P_0)}{[1+(\vartheta-1)P_0]^2}+cP_0(1-P_0)}\right]^2 \quad (5)$$

$$P_t{}^* = \frac{P_0}{c+1}\left(\frac{c\vartheta}{1+(\vartheta-1)P_0}+1\right) \quad (6)$$

Where,
n = the number of participants in the sample.
P_0 = the expected exposure prevalence in the control group.
c = is the ratio of individuals in the case group to those in the control group.
ϑ= the expected odds ratio[11].
Example: Suppose that an investigator is considering designing a case–control study in a country to assess the association between esophageal cancer and alcohol consumption dichotomized at less than or more than or equal to 80 g/day[12].Suppose also that previous surveys have shown that around20% of men without esophageal cancer in that country consume more than or equal 80g/day, and suppose that an odds ratio of ϑ = 2.5 is hypothesized. He also decides to preset α = 0.05 and wants to design a study with equal sample sizes so that its power to detect the hypothesized odds ratio of ϑ = 2.5 is 80% (or β = 0.20). To estimate the sample size for this study first we will calculate $P_t{}^*$ from equation (6)

$$P_t{}^* = \frac{0.20}{1+1}\left(\frac{2.5}{1+(2.5-1)0.20}+1\right) = 0.29$$

Then from equation (5) we find that

$$n = \frac{(1+1)(1+(2.5-1)\times 0.2)^2}{0.2^2\,(0.2-1)^2(2.5-1)^2}\left[1.96\sqrt{(1+1)\times 0.29\times 0.71} + 0.84\sqrt{\frac{2.5\times 0.2\times 0.8}{[1+(2.5-1)\times 0.2]^2}+0.2\times 0.8}\right]^2$$

= 576
In total, 576 men will be enrolled by in the study as 288 cases and 288 controls.

3.2.3 Sample size for cohort studies

In cohort studies, exposed and non-exposed subjects are followed over time period for the development of an outcome event. One can have used Fleiss's formula to calculate sample size for this design

$$n = \frac{c+1}{c(\theta-1)^2\pi^2}\left[Z_{1-\alpha/2}\sqrt{(c+1)P_p(1-P_p)} + Z_{1-\beta}\sqrt{\theta\pi(1-\theta\pi)+c\pi(1-\pi)}\right]^2 (7)$$

$$P_p = \frac{\pi(c+\theta)}{c+1} \quad (8)$$

Since,
n = the number of participants in the sample.
c = is the ratio of individuals in the exposed group to those in the unexposed group.
θ = the predicted incidence risk ratio.
π = the predicted outcome prevalence in the non-exposed group.
P_p = the paired prevalence over the two groups [9,11].

Example: suppose that an investigator wants to study rates of pancreas cancer in smokers. The three-year study period will be evaluated. According to a prior study, 15% of non-smokers develop pancreas cancer. If the researcher wants to run the study with identical sample sizes such that, at a significance level of 5%, its power to identify the relative risk of θ = 3 that has been proposed is 90% (or β = 0.10).The sample size according to (7) will be:

$$n = \frac{1+1}{1(3-1)^2 \times 0.15^2}\left[1.96 \times \sqrt{2 \times 0.3 \times 0.7} + 1.28\sqrt{3 \times 0.15 \times (1-3 \times 0.15) + 0.15 \times (1-0.15)}\right]^2 = 94$$

The study demands 94 people in total—47 smokers and 47 non-smokers—to be sampled in order to achieve its requirements.

3.3 Sample size for randomized clinical trials

Randomized controlled trials (RCTs) are a common kind of experimental studies. Experimental studies are those, in which the researchers make an intervention to the participants and then compare the outcomes between experimental and control groups i.e., a researcher wants to demonstrate whether the new treatment outperforms the standard treatment. In RCTS, the investigators assign participant randomly to either receive the intervention or a control (placebo) condition. The three primary types of designs for RTCS are superiority, equivalency, and non-inferiority trials.

A superiority trial aims to demonstrate that a new treatment is superior to a standard treatment (placebo).The hypothesis in a this type is that the effect of the new treatment is significantly greater than that of the control i.e.,

$H_1: \mu_T - \mu_C > \varphi\ , for\ mean$ or $H_1: P_T - P_C > \varphi\ , for\ proprtion.$

Where φ is a predefined margin.

In non-inferiority trial, the goal is to verify that a new treatment is not significantly worse than an existing standard treatment

A non-inferiority trial seeks to demonstrate that, by a modest pre-specified amount (-δ,δ>0), a new treatment is not inferior to an active control (placebo) treatment that is currently in use. This quantity is known as the non-inferiority margin. The purpose of an equivalency trial is to demonstrate that a clinically acceptable difference's lower and higher equivalency margins correspond to the

true treatment difference. An RCT is referred to as a superiority trial when its goal is to demonstrate the superiority of one treatment over another [3]. Table 3 illustrates these trials regarding to the outcome type.

Table3. Hypothesis tests in clinical studies depending on the type of study and the outcome of interest

Design Type	Outcome Type	Testing Statistics	H_0	H_1
Non inferiority	Continuous	Mean	$\mu_T - \mu_C \leq -\delta$	$\mu_T - \mu_C > -\delta$
	Dichotomous	Proportion	$P_T - P_C \leq -\delta$	$P_T - P_C > -\delta$
Equivalence	Continuous	Mean	$\lvert\mu_T - \mu_C\rvert \geq \delta$	$\lvert\mu_T - \mu_C\rvert < \delta$
	Dichotomous	Proportion	$\lvert P_T - P_C\rvert \geq \delta$	$\lvert P_T - P_C\rvert < \delta$
Superiority	Continuous	Mean	$\mu_T - \mu_C \leq \delta$	$\mu_T - \mu_C > \delta$
	Dichotomous	Proportion	$P_T - P_C \leq \delta$	$P_T - P_C > \delta$

In this paper, we will focus on the sample size calculation for equality test. For non-inferiority and superiority tests, the interested reader's interested readers can refer to the article [11].

Formula for sample size calculation for comparison between two means:

The sample size for a comparison between two groups can be determined using the formula (9) when the outcome is continuous, such as height, BMI, blood pressure, etc.

$$n_C = \left(1 + \frac{1}{r}\right)\left(S \times \frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\frac{\beta}{2}}}{|\mu_T - \mu_C| - \delta}\right)^2 \qquad (9)$$

$$n_T = r.\, n_C \qquad (10)$$

Where,

μ_T= the expected mean of the outcome in the treatment group.

μ_C = the expected mean of the outcome in the control group.

S = the expected standard deviation of the outcome for both the treatment and control groups.

r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group.

$Z_{1-\frac{\alpha}{2}}$= value from the standard normal curve corresponding to the desired level of confidence.

$Z_{1-\frac{\beta}{2}}$= value from the standard normal curve corresponding to the desired study power.

δ = the effect size.[9,11]

Formula for sample size calculation for comparison between two proportions:

The formula below can be used to determine the sample size needed to compare two groups when the outcome is dichotomous, like diseased-non, diseased alive-dead, etc.

$$n_C = \left[\frac{P_T(1-P_T)}{r} + P_C(1-P_C)\right]\left(\frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\frac{\beta}{2}}}{|P_T - P_C| - \delta}\right) \qquad (11)$$

$$n_T = r.n_C \qquad (12)$$

P_T = the expected probability of success in the treatment group.
P_C = the expected probability of success in the control group.
r = expected number of subjects in the treatment group divided by the expected number of subjects in the control group.
$Z_{1-\frac{\alpha}{2}}$= value from the standard normal curve corresponding to the specified level of confidence.
$Z_{1-\frac{\beta}{2}}$= value from the standard normal curve corresponding to the specified study power.
δ = the effect size. [13]

The formula for estimating sample size will yield the number of participants needed to attain the desired level of statistical significance for a particular hypothesis. Participants may be dropped from studies for a variety of reasons i.e., loss to follow up, lack of compliance, etc. Consequently, in order to guarantee power even in the event that some participants withdraw from a study, an adjusted sample size that accounts for dropout rates is necessary (Tips for problematic sample-size computation). Similar previous studies can be used to estimate the dropout rate. Adding a percentage of the total to the anticipated sample size as an adjusted sample size for dropout is the most common mistake made when adjusting sample size. For example, the miscalculated adjusted sample size is equal to $n + n\frac{d}{100}$, for instance, if N is the adjusted sample size, n is the calculated sample size, and d% is the expected dropout rate. To get the adjusted sample size (N), the proper formula is to divide the estimated number of sample size (n) by 1 minus the dropout rate. $N = \frac{n}{(1-\frac{d}{100})}$

Example: Assume that 400 participants are required to achieve the desired power, and based on prior research, the investigator anticipates a 10% dropout rate during the study. Therefore, $N = 400 + 400 \times \frac{10}{100} = 440$ participants is the miscalculated adjusted sample size, while $N = \frac{400}{(1-\frac{10}{100})} = 445$ participants is the correctly calculated adjusted sample size.

4. Strategies to reduce the sample size

When designing a clinical study, determining the right sample size frequently leads to an unfeasible sample size. Ethical considerations can prohibit recruiting this many participants, or the researchers might lack the funding to carry out a study of this magnitude. Numerous techniques exist for reducing the sample size.

Minimize statistical power: The sample size will decrease if the power is, for instance, decreased from 90% to 80%. The quality of the data that will be gathered is not enhanced by this.

Instead of using dichotomous measurements, use continuous ones: Continuous variables typically allow for smaller sample sizes than dichotomous variables in research.

Using paired measurements: In certain research, it is possible to take measurements of each participant in pairs. For the same individual, an investigator can measure the variable once at baseline and once more at the end of the study. The difference between these two measures is the outcome variable. The SD for changes among individuals is smaller than the SD for overall differences between groups of individuals, which accounts for the smaller sample size.

Expand the Minimum Expected Difference: If the suggested study is preliminary, it is advisable to expand the minimal detectable difference. The findings of an initial investigation may serve as justification for a more comprehensive follow-up study with a lower minimum difference and larger sample size.

Using Different Group Sizes: It is well established that, for a given total number of subjects, an equal number of subjects in each group typically yields the maximum power; but, studying more individuals still has benefits, even if those extra participants are all members of one of the groups. But in a lot of real-world scenarios, recruiting subjects for one group is easier or less expensive than for the other. [3, 6].

5. Conclusion

Estimating sample size is a common concern for inexperienced researchers. We hope that this paper has contributed to the demystification of this procedure by identifying some of the key concepts and factors involved in determining the minimum sample size required for epidemiological research. An adequate sample size contributes to ensure that the time and resources dedicated to a study will provide a significant research conclusion. Sample size estimation must be reported in the published articles to enable the reader to evaluate the study. It very important to know that sample size estimation varies from one type of study to another and one blanket formula for sample size calculation cannot be used for all study designs.

References

1. *Le, C. T., & Eberly, L. E. (2016). Introductory biostatistics.*
2. *Gupta, K. K., Attri, J. P., Singh, A., Kaur, H., & Kaur, G. (2016). Basic concepts for sample size calculation: Critical step for any clinical trials!. Saudi journal of anaesthesia, 10(3), 328–331.*
3. *Wang, X., & Ji, X. (2020). Sample Size Estimation in Clinical Research: From Randomized Controlled Trials to Observational Studies. Chest, 158(1S), S12–S20.*
4. *Rothman, K. J., Greenland, S., & Lash, T. L. (2008). Modern epidemiology (Vol. 3). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.*
5. *Patra, P.K. (2012). Sample size in clinical research, the number we need - . International Journal of Medical Science and Public Health, 1, 5-9.*
6. *Eng J. (2003). Sample size estimation: how many individuals should be studied?. Radiology, 227(2), 309–313.*
7. *Wang, X., & Cheng, Z. (2020). Cross-Sectional Studies: Strengths, Weaknesses and Recommendations. Chest Journal, 158, S65-S71.*
8. *Chattopadhyay, A. (2011). Oral health epidemiology: Principles and practice. Sudbury, Mass: Jones and Bartlett Publishers.*
9. *Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). Statistical methods for rates and proportions. Hoboken, N.J: Wiley-Interscience.*
10. *Eng J. (2003). Sample size estimation: how many individuals should be studied?. Radiology, 227(2), 309–313.*
11. *Stevenson M. A. (2021). Sample Size Estimation in Veterinary Epidemiologic Research. Frontiers in veterinary science, 7, 539573.*
12. *Gerstman, B. Burt. Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology. 3rd ed.*
13. *Chow, S.-C., Shao, J., & Wang, H. (2008). Sample size calculations in clinical research. Boca Raton: Chapman & Hall/CRC.*