Sample Size Estimation in Clinical Research

From Randomized Controlled Trials to Observational Studies

Xiaofeng Wang, PhD; and Xinge Ji, MS

Sample size determination is an essential step in planning a clinical study. It is critical to understand that different study designs need different methods of sample size estimation. Although there is a vast literature discussing sample size estimation, incorrect or improper formulas continue to be applied. This article reviews basic statistical concepts in sample size estimation, discusses statistical considerations in the choice of a sample size for randomized controlled trials and observational studies, and provides strategies for reducing sample size when planning a study. To assist clinical researchers in performing sample size calculations, we have developed an online calculator for common clinical study designs. The calculator is available at http://riskcalc.org:3838/samplesize/. Finally, we offer our recommendations on reporting sample size determination in clinical studies. CHEST 2020; 158(1S):S12-S20

KEY WORDS: dropout; hypothesis testing; power; sample size; study designs

General Overview

Estimating sample size is a critical step in conducting a clinical study. Sample size must be planned carefully to ensure that the research time, personnel effort, and costs are not wasted. It is not uncommon that a study fails to detect even large treatment effects because of insufficient sample size. Consulting with a biostatistician helps to address this key part of the planning stage of clinical studies. When a biostatistician is not available, clinical investigators can determine the appropriate sample size for standard study designs using relatively straightforward calculations.

There is a vast clinical literature discussing sample size estimation, including Wittes,[1] Eng,[2] Kasiulevičius et al,[3] Sakpal,[4] and Noordzij et al.[5] Technical calculation details are described in many conventional statistical textbooks.[6-10] There are many formulas available, which can make it tricky for clinical investigators to decide which method to use. All sample size formulas depend on the type of study design and the type of outcome.

The current paper addresses basic concepts in sample size estimation, provides an overview of the commonly used clinical study designs and their corresponding hypothesis tests, and displays a checklist for determining sample size in a study. We then discuss several strategies for reducing sample size when planning a study. An online calculator has been developed to assist clinical researchers in performing sample size calculations based on the study design. The online calculator is available at http://riskcalc.org:3838/samplesize/. A few examples of how to perform the sample size

ABBREVIATIONS: LDL = low-density lipoprotein; RCT = randomized controlled trial

AFFILIATIONS: From the Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH.

CORRESPONDENCE TO: Xiaofeng Wang, PhD, Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave/JJN3-01, Cleveland, OH 44195; e-mail: wangx6@ccf.org

DOI: https://doi.org/10.1016/j.chest.2020.03.010

estimation using the calculator are provided. Finally, some advice is offered on reporting sample size determination in clinical studies.

Basic Statistical Concepts in Sample Size Estimation

An appropriate sample size generally depends on the specified statistical hypotheses and a few study design parameters. These include the minimal meaningful detectable difference (effect size), estimated measurement variability, desired statistical power, and significance level. Some basic statistical concepts in sample size estimation are given here.

Null and Alternative Hypotheses

The null and alternative hypotheses are two mutually exclusive statements about a population. The null hypothesis is generally denoted as H_0, which is set up to be rejected. It states the opposite of what an investigator expects (eg, the opposite being there is no difference between two groups). The alternative hypothesis is generally denoted as H_1 or H_a. It makes a statement that suggests a potential result expected by the investigator (eg, there is a difference between two groups). A hypothesis test uses sample data to determine whether to reject the null hypothesis. Notice that not being able to reject the null hypothesis does not mean that it is true; it means that we do not have enough evidence to reject it.

One-Sided and Two-Sided Tests

In a one-sided test, the alternative hypothesis is directional. The one-sided test is to determine whether the population parameter is either greater than or less than the hypothesized value, or the parameter of group one is either greater than or less than the parameter of group two. In a two-sided test, the alternative hypothesis is nondirectional. The two-sided test is to determine whether the population parameter differs from the hypothesized value, or the parameter of group one differs from the parameter of group two regardless of which one is larger.

Type I Error and Significance Level

A type I error is the rejection of a true null hypothesis, which is referred to as a false positive. The type I error rate is known as the significance level, which is the probability of rejecting the null hypothesis given that it is true. It is commonly denoted by $\alpha$and is also called the alpha level. Conventionally, the significance level is set to 0.05 (5%), implying that it is acceptable to have a 5% probability of incorrectly rejecting the null hypothesis.

Type II Error and Power

A type II error is the nonrejection of a false null hypothesis, which is referred to as a false negative. The type II error rate is the probability that the null hypothesis fails to be rejected when it is false. The type II error is often denoted by β. The power of a test equals $1 - \beta$. Conventionally, the power is set to 80% or 90% when calculating the sample size.

Minimal Detectable Difference

In a clinical trial, the minimal detectable difference refers to the smallest difference between treatments that is considered as clinically significant.

Variance or SD

The variance or SD tells us how spread out the data points are in a specific population. Variance is defined as the average squared deviation from the mean, and the SD is the square root of the variance. They can be obtained either from previous studies or a pilot study. When the outcome is binary, SD is not required for sample size calculation.

Study Designs and Hypothesis Tests in Clinical Research

The sample size estimation formulas can be very different, depending on the type of study design, the type of outcome, and the hypothesis test an investigator specifies. Clinical research studies can be classified into two general categories: experimental and observational.[11,12] Experimental studies, also called interventional studies, are those in which the researcher intervenes at some point during the study. Experimental studies can also be subdivided into two: randomized and nonrandomized. Randomized controlled trials (RCTs) are the most reliable way to establish causal inference and study the efficacy of new treatments in clinical studies. The major categories of RCT study designs are parallel, crossover, cluster, and factorial designs. In a parallel design, subjects will be randomly assigned either to receive or not to receive an intervention. In a crossover design, each participant receives or does not receive intervention in a random sequence over time. In a cluster design, preexisting groups of subjects are randomly selected to receive (or not receive) an intervention. In a factorial design, each participant is randomly assigned to a group that receives a particular combination of interventions or nonintervention.

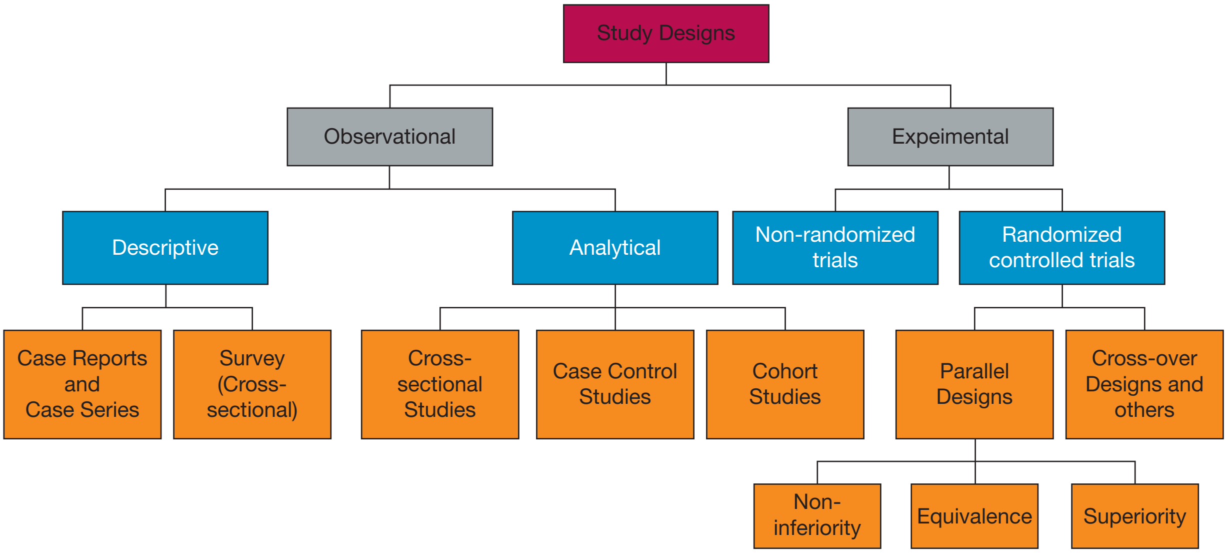


Figure 1 – *Types and subtypes of research study designs.*

Unfortunately, certain interventions, such as surgical interventions, often do not readily lend themselves to RCTs. Nonrandomized trials are used in such situations, where the patients are allocated to treatment groups according to the research protocol.[13]

Observational studies can be either descriptive or analytical. Case reports, case series, and cross-sectional surveys are the most common descriptive study designs. A case report is a detailed description of the symptoms, signs, diagnosis, treatment, response to treatment, and follow-up after treatment of an individual patient, whereas a case series describes a group of patients with an identical or similar condition.[14] A descriptive cross-sectional survey collects data to provide an assessment of a population of interest at one point in time. In general, descriptive studies do not have a comparison group, whereas analytical studies feature a group comparison. Within analytical observational studies, cohort studies (both prospective and retrospective) track subjects forward in time from exposure to outcome. By contrast, case-control studies work backward, tracing back from outcome to exposure. Analytical cross-sectional studies are useful to determine the prevalence, which measures both exposure and outcome at one time point.[12]

The types and subtypes of research study designs are illustrated in Figure 1. The results from RCTs often rank high with respect to the levels of clinical evidence, as RCTs are designed to have less risk of systematic errors.

TABLE 1] Commonly Used Hypothesis Tests in Parallel RCTs

Design Type	Outcome Type	Testing Statistics	Null Hypothesis	Alternative Hypothesis
Noninferiority	Continuous	Mean	$\mu_T - \mu_C \leq -\delta$	$\mu_T - \mu_C > -\delta$
	Dichotomous	Proportion	$p_T - p_C \leq -\delta$	$p_T - p_C > -\delta$
		OR	$OR \leq \exp(-\delta)$	$OR > \exp(-\delta)$
	Time-to-event	HR	$HR \leq \exp(-\delta)$	$HR > \exp(-\delta)$
Equivalence	Continuous	Mean	$\lvert \mu_T - \mu_C \rvert \geq \delta$	$\lvert \mu_T - \mu_C \rvert < \delta$
	Dichotomous	Proportion	$\lvert p_T - p_C \rvert \geq \delta$	$\lvert p_T - p_C \rvert < \delta$
		OR	$\lvert \log(OR) \rvert \geq \delta$	$\lvert \log(OR) \rvert < \delta$
	Time-to-event	HR	$\lvert \log(HR) \rvert \geq \delta$	$\lvert \log(HR) \rvert < \delta$
Superiority	Continuous	Mean	$\mu_T - \mu_C \leq \delta$	$\mu_T - \mu_C > \delta$
	Dichotomous	Proportion	$p_T - p_C \leq \delta$	$p_T - p_C > \delta$
		OR	$OR \leq \exp(\delta)$	$OR > \exp(\delta)$
	Time-to-event	HR	$HR \leq \exp(\delta)$	$HR > \exp(\delta)$

HR = hazard ratio; RCTs = randomized controlled trials.

TABLE 2] Commonly Used Hypothesis Tests in Observation Studies

Design Type	Outcome Type	Testing Statistics	Null Hypothesis	Alternative Hypothesis
Cohort	Dichotomous	Proportion	$p_0 = p_1$	$p_0 \neq p_1$
		RR	$RR = 1$	$RR \neq 1$
	Time-to-event	HR	$HR = 1$	$HR \neq 1$
Case-control	Dichotomous	Proportion	$p_0 = p_1$	$p_0 \neq p_1$
		OR	$OR = 1$	$OR \neq 1$
Cross-sectional	Continuous	Mean	$\mu_0 = \mu_1$	$\mu_0 \neq \mu_1$
	Dichotomous	Proportion	$p_0 = p_1$	$p_0 \neq p_1$

RR = relative risk. See Table 1 legend for expansion of other abbreviation.

However, RCTs can have limited applicability to patients in clinical settings. Observational studies are viewed as having less internal validity but can complement findings from RCTs by assessing treatment effectiveness in day-to-day clinical practice, extending validity and diversity. In terms of the strength of evidence in observational studies, prospective cohort studies are generally more reliable than retrospective studies, and cohort and case-control studies are more reliable than cross-sectional studies and other descriptive studies.

RCTs can be parallel, crossover, or other advanced study designs.[8] The majority of RCTs in clinical research are parallel-group trials. An analysis of the RCTs indexed in PubMed between 2000 and 2006 found that 78% of RCTs were parallel designs, and 16% were crossover.[15] For brevity, we only discuss the sample size estimation of the parallel design. The book by Chow et al[8] discusses other subtypes of RCT designs, and Lipsitz and Parzen[16] and Bernardo et al[17] discuss nonrandomized interventional studies.

In a parallel RCT, there are three commonly used trial designs: noninferiority, equivalence, and superiority trials.[18] A noninferiority trial aims to show that a new treatment is not worse than an active control treatment already in use by a small prespecified amount. This amount is known as the noninferiority margin. An equivalence trial is designed to show that the true treatment difference lies between a lower and an upper equivalence margin of clinically acceptable differences. When an RCT aims to show that one treatment is superior to another, the trial (test) is called a superiority trial (test).

In observational studies, investigators often compare the difference of outcomes for two groups (case vs control in case-control studies, exposed vs unexposed group in cohort studies). If the outcome of interest is continuous, two means are compared. If the outcome is dichotomous, two proportions are compared. If the outcome is a time-to-event variable, such as time to death or time to discharge, then the hazard ratio for the two groups is assessed.

Tables 1 and 2 summarize commonly used hypothesis tests in clinical studies depending on the type of study and the outcome of interest. For RCTs in Table 1, the μ_T and μ_C denote the means of the treatment group and the control group, respectively, when the outcome is continuous. The $p_T$and $p_C$denote the proportions of the treatment group and the control group, respectively, when the outcome is dichotomous. OR is defined as $OR = p_T(1-p_C)/p_C(1-p_T)$. The HR is the hazard ratio between the two groups. For observational studies in Table 2, the p_0 and p_1 denote the proportions of two different groups when the outcome is dichotomous. The μ_0 and μ_1 denote the means of two different groups when the outcome is continuous. The $RR = p_1/p_0$ is the relative risk.

General Considerations for Sample Size Estimation

With knowledge of the design information detailed in the previous sections, the calculation of an appropriate sample size involves using a suitable formula. For example, in a cross-sectional study, the investigator may want to compare the means of two groups. If we assume that the sample sizes in both groups are equal, the equation of the sample size is given by

$$n = \frac{2\sigma^2\left(z_{crit} + z_{pow}\right)^2}{D^2},$$

where n is the sample size for each group, σ^2 is the variance of either group (assumed to be equal for both groups), and D is the minimal detectable difference between the two means. The z_{crit} and z_{pow} are the standard normal deviates at a level of significance and at

$1 - \beta$power, respectively. A standard normal deviate is a realization of a standard normal random variable. The z_{crit} is 1.96 at 5% level of significance for two-sided tests. The z_{pow} is 0.84 at 80% power and 1.28 at 90% power.

The sample size formula can be very complicated, depending on the type of study design and the type of outcome. In e-Appendix 1, we present the appropriate sample size estimation formulas for various RCT and observational study designs. We have also developed an online sample size calculator (http://riskcalc.org:3838/samplesize/) that closely parallels the equations presented in e-Appendix 1.

In brief, we outline the basic steps for calculating sample size at the design stage of a clinical study: (1) define the population of the study; (2) select the type of study design; (3) specify the null and alternative hypotheses, along with the significance level and power; (4) gather information relevant to the parameters of interest (means or proportions, minimal detectable difference, and variance are the main expected parameters); (5) calculate sample size over a range of reasonable parameters and select an appropriate one for the study; and (6) evaluate the expected rate of dropouts and adjust the sample size as needed to finalize the estimate.

Loss to follow-up is important to consider when designing a clinical study. Any sample size calculated should be inflated to account for the expected dropouts. A common method is to multiply a factor $1/(1 - r_d)$, where r_d is the dropout rate.

It is noteworthy that not all clinical studies involve the comparison of groups. In cross-sectional surveys, the purpose is often to describe one or more characteristics of a specified group by using means or proportions. These studies are not involved in hypothesis testing. We need to know the desired margin of error to compute the sample size. The margin of error is defined as half the width (or "radius") of a CI for a statistic from a survey. It reflects how precise the statistic, such as mean or proportion, is expected to be. For instance, in a study designed to estimate the prevalence of a disease, the sample size equation[9] is

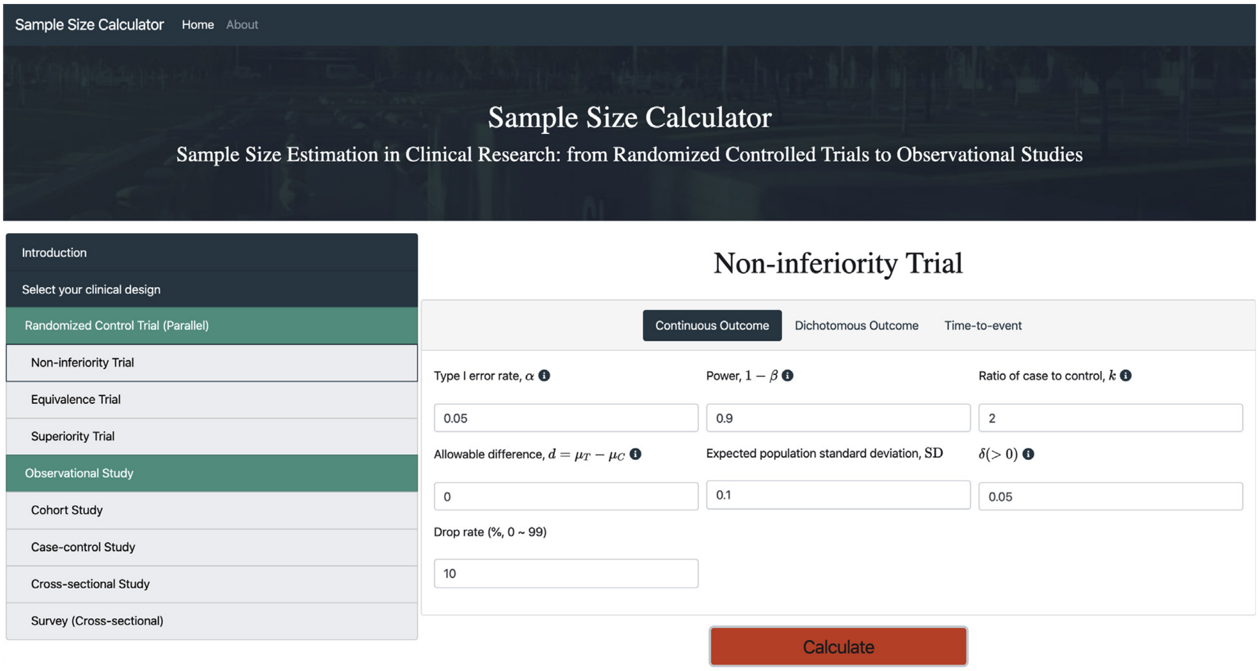
$$n = \frac{z_{crit}^2 p(1 - p)}{e^2},$$

where $p$is the estimate of the prevalence rate, and e is its margin of error. Here the sample size is driven by the number of cases of a disease rather than the total number of subjects in a particular population.

Cases of Sample Size Estimation

Example 1

Consider an RCT for evaluation of the effect of a test drug on cholesterol in patients with coronary heart



Sample Size Calculator

Sample Size Estimation in Clinical Research: from Randomized Controlled Trials to Observational Studies

Non-inferiority Trial

Type I error rate, α: 0.05

Power, $1 - \beta$: 0.9

Ratio of case to control, k: 2

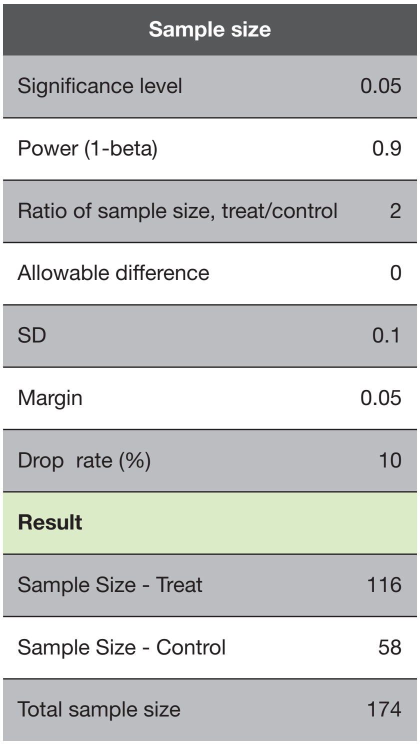
Allowable difference, $d = \mu_T - \mu_C$: 0

Expected population standard deviation, SD: 0.1

$\delta(> 0)$: 0.05

Drop rate (%, 0 ~ 99): 10

Figure 2 – *Interface of the online calculator.*



Sample size	
Significance level	0.05
Power (1-beta)	0.9
Ratio of sample size, treat/control	2
Allowable difference	0
SD	0.1
Margin	0.05
Drop rate (%)	10
Result	
Sample Size - Treat	116
Sample Size - Control	58
Total sample size	174

Figure 3 – *Sample size estimation results using the online calculator for the noninferiority trial example.*

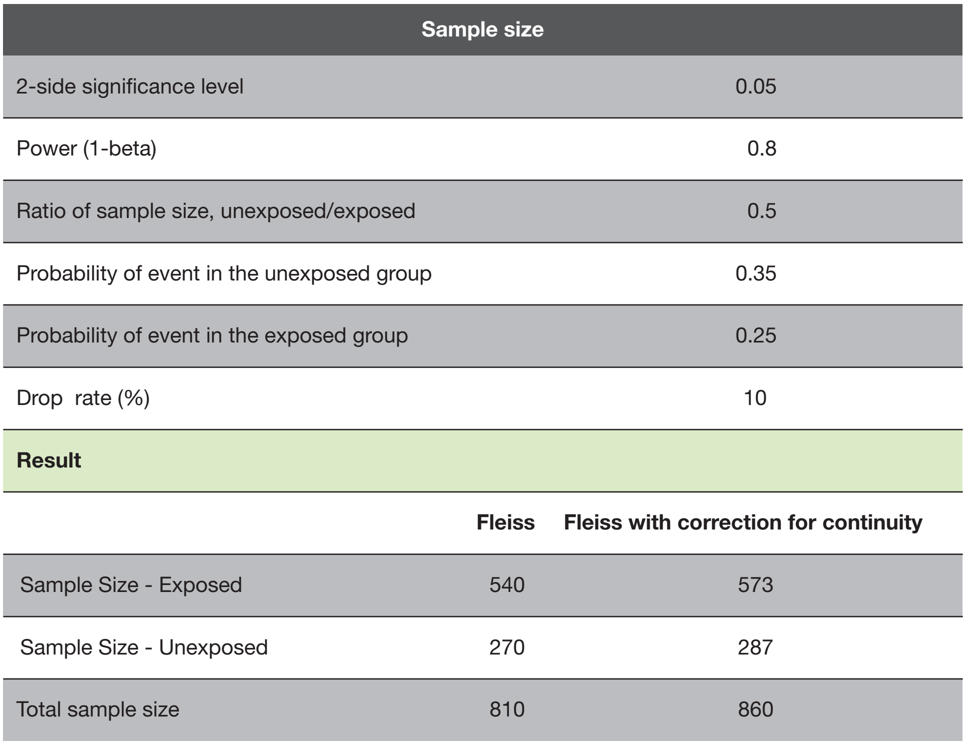
disease.[8] Investigators are interested in conducting a clinical trial to compare two cholesterol-lowering agents for the treatment of patients with coronary heart disease through a parallel design. The primary outcome is the percent change in low-density lipoprotein (LDL), a continuous variable. Suppose that the investigators want to establish the noninferiority of the test drug compared vs the active control agent. The testing hypotheses are:

$$H_0\colon\ \mu_T - \mu_C \leq -\delta \text{ vs. } H_1\colon\ \mu_T - \mu_C > -\delta$$

where μ_T and μ_C are the mean parameters for the treatment group and control group, respectively, and $\delta > 0$ denotes the noninferiority margin, a (clinically meaningful) minimal detectable difference.

To compute the sample size, we want to gather the following parameters: (1) a significance level (or type I error rate) [we set to 5% here]; (2) power [we set to 90% here]; (3) the sample size ratio of treatment to control [we set to 2:1]; (4) the true (allowable) difference in mean LDL between groups [we set to 0%]; (5) SD [from previous studies, it is assumed to be 10% in this study]; (6) clinically meaningful difference [a difference of 5% change of LDL is considered as the clinically meaningful difference]; and (7) dropout rate [we assume a 10% dropout rate in this study].

Using the online calculator that we developed, we can easily perform the calculation based on this information. Figure 2 shows the interface of the online



Sample size		
2-side significance level	0.05	
Power (1-beta)	0.8	
Ratio of sample size, unexposed/exposed	0.5	
Probability of event in the unexposed group	0.35	
Probability of event in the exposed group	0.25	
Drop rate (%)	10	
Result		
	Fleiss	**Fleiss with correction for continuity**
Sample Size - Exposed	540	573
Sample Size - Unexposed	270	287
Total sample size	810	860

Figure 4 – *Sample size estimation results using the online calculator for the cohort study example.*

calculator. We navigate to the page of "Noninferiority Trial," input the values into the corresponding entries, and then click the "Calculate" button. The results for the sample size estimation in this case study are displayed in Figure 3. We can see that the sample size for the treatment group is 116, and the sample size for the control group is 58.

Example 2

Suppose that we want to conduct a prospective cohort study to examine the effect of beta-blockers in the management of COPD. Our primary goal is to assess the effect of beta-blockers on hospital mortality of patients with COPD. We assume that the mortality rate for the treatment (exposed) group is 25% and the mortality rate for the control (unexposed) group is 35%. We further assume that the sample size ratio of the unexposed to exposed groups is 1:2. In the online calculator, we can navigate to the page of independent proportional outcome for cohort study. The results for the sample size estimation with a significance level of 5% and a power of 80% are displayed in Figure 4. The sample size for the exposed group is 540, and the sample size for the unexposed group is 270.

In the sample size estimation for dichotomous outcomes, the statistical method assumes that a normal distribution approximates a binomial one when computing the sample size. More accurate approximations can be obtained when a continuity correction is used.[10] If we consider the correction for continuity in the analysis, the sample size will increase to 573 and 287, respectively. In general, we prefer to use the results with a continuity correction in a study.

Strategies for Reducing Sample Size

When planning a clinical study, finding an appropriate sample size often results in a sample size that is too large to be feasible. Investigators may not have the resources to conduct such a large study, or ethical reasons may prevent enrolling this many subjects. Reducing the sample size usually involves some compromise, such as accepting a small loss in power. We highlight here several strategies for reducing the sample size. Some strategies involve modifications of the research hypothesis. Investigators should carefully consider whether the new hypothesis still answers an interesting research question. Consulting with a biostatistician is important at this stage. More discussions regarding reducing sample size can be found in Browner et al[19] and Eng.[2]

1. Reduce statistical power. Reducing the power, for example, from 90% to 80% will reduce the sample size. This does not improve the quality of the data that will be collected.
2. Use continuous outcomes. When continuous variables are an option in a study, they usually permit smaller sample sizes than dichotomous variables.
3. Enrich the subject population. Variability in outcomes can be reduced by making the subject population more homogeneous. The tradeoff is that the generalizability of the study suffers. Investigators could identify an enriched subgroup within the larger enrolled group, prespecify the primary outcome for the enriched subgroup, and use all subjects to study a secondary outcome.
4. Use paired measurements. In some studies, paired measurements can be made in each subject. For example, we measure the variable one time at baseline and another time at the end of the study, for the same subject. The outcome variable is the change between these two measurements. A paired test can be used in this situation, which often permits a smaller sample size.
5. Reduce the dropout rate. Henry[20] suggested allocating funds earmarked for data collection into intensive follow-up instead. Investigators could put resources into follow-up, which can reduce the dropout rate and, in turn, reduce the sample size.
6. Use unequal group sizes. It is known that an equal number of subjects in each group usually provides the greatest power for a given total number of subjects. However, in many real situations, it is less expensive or easier to recruit subjects for one group than the other. Benefit is gained by studying more subjects even if the additional subjects all come from one group. From the perspective of feasibility and cost, the gain in power is considerable when the size of one group is increased to twice the size of the other. Tripling and quadrupling one of the groups provide much smaller gains.[19]
7. Expand the minimal detectable difference. Determination of the minimal detectable difference is often based on clinical experience or literature review. If the planned study is preliminary, a larger expected difference could be justified. The results of the preliminary study could be used to plan a more rigorous large study with a smaller minimum difference.[2]
8. Use surrogate outcomes. Surrogate outcomes are measurements that are highly correlated with the

primary outcome and, hopefully, with treatment effect. Typically, they are easier to measure, have less variability, and/or occur sooner than the primary outcome.

9. Increase the event rate. There are several ways to increase the event rate, including expanding the follow-up period, using a surrogate outcome, and using a composite outcome. A composite outcome combines multiple outcomes in one. It tends to be useful for diseases that have multiple effects. An example of a composite outcome would be an outcome of "death or complications." Designing a composite outcome can be complex, because investigators may need to weight each component.

Reporting Considerations and Available Standards

The sample size should be determined early in the planning of a study. An appropriate sample size helps ensure that the research time and support costs invested in a study lead to a meaningful research conclusion.

The Consolidated Standards of Reporting Trials (CONSORT) statement [21] suggests standard elements for authors to include in reports of trial findings. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement[22] provides similar guidance for reporting observational studies. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement[23] aims to improve the transparency of the reporting of a prediction model study. These three statements all include items related to sample size reporting. The method for determining sample size in a study should be described with enough detail to allow its use in other protocols. The authors should provide enough details so that a knowledgeable reader can reproduce the sample size calculation. The power, significance level, mean or rate for the control group, minimal detectable difference, variance, and dropout rate should be clearly documented. Any other factors that formed the basis of the sample size calculation should be included.

Let us go back to Example 1 in the last section. As an example, we could report the sample size determination for Example 1 as follows: "Calculation of the sample size was based on the ability to detect a clinically relevant difference in the percent change of LDL (primary outcome) of 5% between the two trial arms. Smith et al found an SD of 10% for the percent change of LDL in their study; we used this value as our reference for the sample calculation. The sample size ratio of treatment to control was set to 2:1. By assuming a 10% dropout rate in our study, the sample size calculation indicated that 116 participants were required in the treatment group and 58 participants were required in the control group. These sample sizes gave us approximately 90% power (with a 5% level of significance) to reject the null hypothesis that the new treatment was not less effective than the active control treatment."

Short List of Questions to Guide the Reviewer

We provide some questions that the reviewer should ask regarding sample size when he or she reviews a manuscript. It is hoped that these will inform the review process.

1. **The elements used to calculate the sample size.** Is the sample size and its calculation clearly reported? Is the target population clearly defined? Are the power, significance level, mean or rate parameters, minimal detectable difference, variance, and dropout rate clearly documented?

Acknowledgments

Financial/nonfinancial disclosures: None declared.

Additional information: The e-Appendix can be found in the Supplemental Materials section of the online article.

References

1. Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev*. 2002;24(1):39-53.
2. Eng J. Sample size estimation: how many individuals should be studied? *Radiology*. 2003;227(2):309-313.
3. Kasiulevičius V, Šapoka V, Filipavičiūtė R. Sample size calculation in epidemiological studies. *Gerontologija*. 2006;7(4):225-231.
4. Sakpal TV. Sample size estimation in clinical trial. *Perspect Clin Res*. 2010;1(2):67-69.
5. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. *Nephrol Dial Transplant*. 2010;25(5):1388-1393.
6. Altman DG. *Practical Statistics for Medical Research*. Boca Raton, FL: CRC Press; 1990.
7. Breslow NE, Day NE, Heseltine E, Breslow NE. *Statistical Methods in Cancer Research: The Design and Analysis of Cohort Studies*. Lyon, France: International Agency for Research on Cancer; 1987.
8. Chow SC, Shao J, Wang H, Lokhnygina Y. *Sample Size Calculations in Clinical Research*. 3rd ed. Boca Raton, FL: Chapman and Hall/CRC; 2017.
9. Cochran WG. *Sampling Techniques*. New York, NY: John Wiley & Sons; 1977.
10. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. New York, NY: John Wiley & Sons; 2013.

11. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25):1887-1892.

12. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet*. 2002;359(9300):57-61.

13. Axelrod DA, Hayward R. Nonrandomized interventional study designs (quasi-experimental designs). In: *Clinical Research Methods for Surgeons*. New York, NY: Springer; 2006:63-76.

14. Gagnier JJ, Kienle G, Altman DG, et al. The CARE guidelines: consensus-based clinical case report guideline development. *J Clin Epidemiol*. 2014;67(1):46-51.

15. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.

16. Lipsitz SR, Parzen M. Sample size calculations for non-randomized studies. *J R Stat Soc Ser D Stat*. 1995;44(1):81-90.

17. Bernardo MVP, Lipsitz SR, Harrington DP, Catalano PJ. Sample size calculations for failure time random variables in non-randomized studies. *J R Stat Soc Ser D Stat*. 2000;49(1):31-40.

18. Lesaffre E. Superiority, equivalence, and non-inferiority trials. *Bull NYU Hosp Jt Dis*. 2008;66(2):150-154.

19. Browner WS, Newman TB, Hulley SB. Estimating sample size and power: applications and examples. *Designing Clin Res*. 2007;3:367.

20. Henry GT. *Practical Sampling*. vol 21. Newbury Park, CA: SAGE Publications Ltd.; 1990.

21. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8(1):18.

22. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495-1499.

23. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.