



Important considerations in calculating and reporting of sample size in randomized controlled trials

Ameneh Ebrahim Valojerdi¹, Kiarash Tanha¹, Leila Janani*¹

Received: 1 June 2017

Published: 25 Dec 2017

Abstract

Background: The calculation of the sample size is one of the most important steps in designing a randomized controlled trial. The purpose of this study is drawing the attention of researchers to the importance of calculating and reporting the sample size in randomized controlled trials.

Methods: We reviewed related literature and guidelines and discussed some important issues in sample size calculation and reporting in randomized controlled trials.

Conclusion: The calculation of the sample size is one of the most important steps in designing a randomized controlled trial. According to the CONSORT (Consolidated Standards of Reporting Trials) guideline and other standard guidelines for designing and reporting of RCTs, sample size calculations should be reported and justified in all published RCTs. Because sample size calculations are prone to bias and because of the high ethical and financial costs related to conducting an RCT, we recommend involving a biostatistician at the designing stage of the study and to ask for statistical advice for sample size calculations.

Keywords: Randomized controlled trials (RCTs), Sample size, Post hoc power

Copyright© Iran University of Medical Sciences

Cite this article as: 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 (25 Dec);31:127. <https://doi.org/10.14196/mjiri.31.127>

Introduction

Randomized controlled trials (RCTs) are considered as the “gold standard” in evidence-based medicine to exploring the effects of different interventions (1). The quality of RCTs as a concept is not easy to define and it depends on the internal and external validity of the study (2). However, the strength of this type of study is unverified where the sample size was determined incorrectly. An investigation of all reports of randomized trials indexed in PubMed in years 2000 and 2006 shows that the components of the sample size calculation were reported only in 27% of 519 trials and 45% of 616 trials in years 2000 and 2006 respectively (3). In another review in surgery literature, Maggard et al showed that the sample size calculations were not provided in about 60% of surgical RCTs and only 50% of studies had adequate sample sizes to detect differences between treatment groups (4).

However, sample size calculation for a controlled trial is almost always a matter of compromise between the resources and other objectives, such as safety with small effects and efficacy with large effects (5, 6). A study with low sample size has low power and a study with low statistical power has a less chance of detecting a true effect (7). On the other side, an oversized trial is a waste of resources and also lead to unethical trials because of

confronting the patients to unnecessary risks which are not fair (8).

There are numerous articles about sample size calculation methods and formulas, but the aim of this study is reviewing the practical and important issues about this topic in RCTs and drawing the attention of researchers to the importance of sample size calculation and reporting in RCTs.

Guidelines and sample size determination

The majority of studies fail to report sample size calculation in their reports yet (9). Altman reported that “sample size calculations and important aspects of statistical analysis methods were often incompletely described in protocols and publications” (3). In 1993, the Consolidated Standards of Reporting Trials (CONSORT) was published with the aim of developing a scale to assess the quality of randomized controlled trial (RCT) reports (10). The authors should discuss how proper sample size was determined in the method section. Also, any interim analyses and stopping rules should be explained when applicable (11). Nowadays, submitting an RCT requires reporting the trial according to the CONSORT statement. In a study that was published in 2011, the Cochrane Collaboration, states some aspects of trials in their tool

Corresponding author: Dr Leila Janani, leila_janani@yahoo.com

¹ Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran.

which has been established for assessing the risk of bias in randomized trials. It was described that calculating sample size, is not directly related to the risk of bias but it is important to be reported (12). In another article, Delphi Consensus developed a criteria list for quality assessment of randomized controlled trials with one item about details of sample size calculation (13). In three ICH GCP (Good Clinical Report) guidelines: E3 (structure and content of clinical study reports), E8 (general considerations for controlled trials) and E9 (statistical principles for controlled trials) calculating and reporting sample size with details were emphasized (14-16). According to the above-mentioned description, most of the RCT guidelines have emphasized that sample size calculation and reporting are one important item in the valid report of RCTs.

Sample size determination based on primary outcome

Primary endpoints are the response variables that are chosen to assess drug (or another intervention) effects which it may consist of efficacy or safety of the intervention. Secondary endpoints assess other effects that may relate or not related to the primary endpoint. The primary and secondary outcomes should be defined in protocols before starting trials (14). The number of subjects in RCTs is usually determined by the primary objective of the trial. If the sample size is determined on other bases, such as important secondary outcome, this should be made clear and justified (15).

Statistical analysis and sample size

According to guidelines, the RCTs should be analyzed in accordance with the analyzing plan which has been stated in the protocol (14). One important item in protocols is a description of sample size calculation. So any statistical analysis such as interim analysis, subgroup analysis and etc. should be considered in sample size calculation in the plan of RCTs. For instance, for planned interim analysis, the sample size should be adjusted based on the interim analysis (17). Moreover, different designs of RCTs require different sample size approaches. Because of correlations between the repeated measurements of the outcome in a crossover design, we need fewer sample size than the parallel design. As the correlation increases, fewer subjects are needed for a crossover trial (18). Furthermore, type of hypothesis test that is related to the aim of the study, could have an effect on sample size. For example, in the one-sided test, a fewer sample size is needed. Selecting one type of non-inferiority, superiority or equivalence study can have an effect on the method of sample size calculation. In a superiority trial which needs less sample size than non-inferiority and equivalency, the aim is demonstrating the superiority of a new therapy compared to an established therapy. Non-inferiority trials determine whether a new treatment is at least as good as an established treatment. The sample size of an equivalence trial or a non-inferiority trial should be based on a confidence interval for the treatment difference. In the non-inferiority trials,

the sample size would be less than equivalence designs because the level of the upper confidence limit is not of primary interest (19). According to all above points, the researcher should consider the consistency between the type of hypothesis, design, statistical analysis and methods of sample size calculation.

Standard Reporting of Sample size calculation

A review in 2005-2006 showed that 50% of articles didn't report details of the sample size calculations (20). It's important that details of the calculation are reported such that others could recalculate and perform our study's sample size. The sample size calculation in different design of RCTs may depend on different items, but generally, for sample size calculation, primary variable, the endpoint or outcome measure, statistical analysis methods, the probability of erroneously rejecting the null hypothesis (the type I error) and the probability of erroneously failing to reject the null hypothesis (the type II error) should be considered (14, 21). Sometimes, the margin of error or clinical important is one of the critical and challenging parameters. The challenge here is to define a difference between test and reference which can be considered clinically meaningful. This is not easily available and should be decided based on clinical judgment and other literature. Furthermore, in practice, there is a tendency to 'adjust' on other factors or conducting the subgroup analysis which it requires larger sample size (22); Another consideration is the rate of attrition; patients who didn't finish their therapy for reasons which not related to the disease under treatment or the therapy. Suppose R is attrition rate which we expect, an adequate adjustment is provided:

$$N_d = N / (1 - R)$$

where N is the sample size calculated assuming no attrition and N_d that required with attrition (5). According to all the points mentioned, guidelines recommend that details of sample size calculation should be reported in RCTs.

Post hoc vs. Observed power

The power of a study reflects the probability of detecting a difference when this difference exists. More power means less risk for Type II errors and more chances to detect a difference when it exists.

Commonly, the power calculations have not been performed before conducting the trial (23, 24) and when facing non-significant results, investigators use the observed difference and variability and the sample size of the trial to compute the power of study which is known as observed power (25, 26), while true post hoc power is determined using effect size and variability of the literature for primary outcome. Observed power analyses have little statistical meaning for two reasons (27, 28). First, because there is a one-to-one relationship between p-values and observed power and it does not give us more information than the p-value. When the p-value close to zero, the observed power is large and when the p-value is large, the observed power is small. If the p-value is greater than 0.05, the observed power has to be smaller

Table 1. Important Steps of sample size calculation in RCT

1)	Considering the object of RCT	
2)	Considering the type of the outcome	
3)	Calculating important clinical difference	
4)	Calculating the estimation of variation	Effect size
5)	Considering type I statistical error and type II statistical error of study	
6)	Considering the design of RCT	
7)	Considering statistical methods which will be used in the final analysis (Such as: adjustments, multiple comparisons, confounders and...)	
8)	Calculating sample size using software or formula	
9)	Considering the attrition rate	
10)	Reporting all items which were used in sample size calculation with detail	

than 50% (28). Second, when investigators computing observed power, they implicitly make this assumption that the observed difference is clinically meaningful and more important than null hypothesis in statistics. In the theory of hypothesis testing, in posterior, the use of confidence intervals is preferable to judge the relevance of a finding. The confidence interval is related directly to sample size and conveys more information than p-values or observed power (25).

Summary

In Table 1, we summarized the main steps in calculating and reporting the sample size in RCTs.

In addition, there are different software programs that can be applied to calculate the sample size for several types of designs. Examples of validated programs are nQuery Advisor, PASS and 'Power and Precision'.

Conclusion

In conclusion, the calculation of the sample size is one of the most important steps in designing a randomized controlled trial. According to the CONSORT (consolidated standards of reporting trials) guideline and other standard guidelines for designing and reporting of RCTs, sample size calculations should be reported and justified in all published RCTs. Readers of a published trial should be able to find all assumptions considered in the calculation of the sample size. Because sample size calculations are prone to bias and because of the high ethical and financial costs related to conducting an RCT, we recommend involving a biostatistician at the designing stage of the study and to ask for statistical advice for sample size calculations.

Conflict of Interests

The authors declare that they have no competing interests.

References

- LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337(8):536-42.
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized controlled trials: is blinding necessary? *Contemp. Clin. Trials*. 1996;17(1):1-12.
- 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 (Clinical research ed)*. 2010;340:c723.
- Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: are they done correctly? *Surgery*. 2003;134(2):275-9.
- Lachin JM, Marks JW, Schoenfeld LJ, Tyor MP, Bennett PH, Grundy SM, et al. Design and methodological considerations in the National Cooperative Gallstone Study: A multicenter controlled trial. *Contemp. Clin. Trials*. 1981;2(3):177-229.
- Lachin J. Sample size considerations for controlled trials of potentially hepatotoxic drugs. *Guidelines for Detection of Hepatotoxicity Due to Drugs and Chemicals* Washington, DC: US Department of HEW, National Institutes of Health, NIH Publication. 1979(79-13):119-30.
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Rev Neurosci*. 2013;14(5):365-76.
- Kumar GS. Importance of sample size in controlled trials. *Int J Clin Exp Physiol*. 2014;1(1):10.
- Fayers P, Cuschieri A, Fielding J, Craven J, Uscinska B, Freedman L. Sample size calculation for controlled trials: the impact of clinician beliefs. *Br J Cancer*. 2000;82(1):213.
- Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC med*. 2010;8(1):18.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol*. 2001;1(1):2.
- Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj*. 2011;343:d5928.
- Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized controlled trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998;51(12):1235-41.
- Guideline IHT. General considerations for controlled trials E8. *Current Step*. 1997;4.
- Phillips A, Haudiquet V. ICH E9 guideline 'Statistical principles for controlled trials': a case study. *Stat Med*. 2003;22(1):1-11.
- Harmonisation ICo. ICH harmonised tripartite guideline for statistical principles for controlled trials: Brookwood Medical Publ.; 1998.
- Montori VM, Devereaux P, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203.
- Pocock SJ. *Controlled trials: a practical approach*: John Wiley & Sons; 2013.
- Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46(5):947-54.
- Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ*. 2009;338:b1732.
- Walters SJ, Campbell MJ, Paisley S. Methods for determining sample sizes for studies involving health-related quality of life measures: a tutorial. *Health Serv Outcomes Res Methodol*. 2001;2(2):83-99.
- Senn SS. *Statistical issues in drug development*: John Wiley & Sons; 2008.
- Bailey CS, Fisher CG, Dvorak MF. Type II error in the spine surgical literature. *Spine*. 2004;29(10):1146-9.
- Freedman K, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. *Bone Joint J*. 2001;83(3):397-402.
- Biau DJ, Kernéis S, Porcher R. *Statistics in brief: the importance of*

- sample size in the planning and interpretation of medical research. *Clin Orthop Relat Res.* 2008;466(9):2282-8.
26. Walters SJ. Consultants' forum: should post hoc sample size calculations be done? *Pharm Stat.* 2009;8(2):163-9.
27. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med.* 1994;121(3):200-6.
28. Hoening JM, Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am Stat.* 2001;55(1):19-24.