The 7th IEEE International Conference on E-Health and Bioengineering - EHB 2019
Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania, November 21-23, 2019

Guidelines for Small Size Samples Biostatistics in Current Medical Practice

George Calin Seritan [1], Felix-Constantin Adochiei [1], Bogdan-Adrian Enache [1], Gladiola Petroiu-Andruseac [2], Cristina-Daniela Enache [3], Dana Adriana Dascultu [3], Cristina Adriana Becheanu [3], Alexandru Ulmeanu [3]

[1] DMAECS, University Politehnica of Bucharest, Romania, bogdan.enache2207@upb.ro

[2] Faculty of Medical Bioengineering, Grigore T. Popa University of Medicine and Pharmacy, Iaşi, Romania, agladi@yahoo.com

[3] Toxicology Department, Emergency Children's Hospital "Grigore Alexandrescu" Bucharest, Romania, cristina.tipluica@yahoo.com

***Abstract*—The present study aims at bringing useful guidelines to a physician for determining the sample size when designing any research project. The sample size is an estimate of the number of patients determined in concordance with the objectives and the budget of the project. This paper presents several ways for sample size determination i.e. classic methods, Bayesian approach, dedicated software. All of them are analyzed considering that the end-user is not a statistician, but a doctor and pointers for choosing the best way when dealing with a small SS are highlighted.**

Keywords—small sample size, sample size determination, Bayesian methods, clinical trials

I. INTRODUCTION

In any clinical research in order to be able to confirm the study hypotheses, there must be a valid relationship between the number of observations, their possible repetitions and the quality of the evidence [1]. This number of observations is called the sample size (SS) and is usually referred to as the letter *n*.

SS determination is an essential part in designing any clinical trial due to practical, ethical and financial reasons. This is usually done by applying a series of mathematical formulas that ensure the precision is sufficient in estimating the population parameters or the obtained results are significant [1]. It is important to determine the SS before the study is conducted because in this way the optimal number of patients are selected and there is no waste of resources [2].

A huge SS can lead to increased costs, delays in dissemination, ineffective use of treatment, etc. [3]. On the other hand, a tiny SS can lead to underpowered trials that will produce, in most case, inconclusive results [1], [3].

Previous studies have shown that a lot of trials, from different fields, are frequently underpowered. Statistical power represents the probability that a study will produce a true result if there is one that can be detected. Underpowered studies have a low chance of determining this and have been correlated with false-positive results and low reproducibility [4],[5].

In [6] 90 surgery trials from 1988 and 1989 were analyzed and only 24% of them had a statistical power greater than 80% to detect a 50% difference in therapeutic effect. In [7] 117 orthopaedic randomized controlled trials (RCTs) were studied and found that the type-II error rate for primary outcomes was 91%, which for nonsurgical specialities represents a problem.

Using a different approach, in [4] the percent of sufficient powered studies were determined over a period of 40 years from 1975 to 2014. It was determined that percent of sufficiently powered studies rose from 5.1% in 1975 to 8.5% in 2014. The conclusion was that only 7% of individual trials had enough power to detect an observable effect, but the trend is ascendant and in the next years, this percent will rise constantly.

In order to avoid a low power statistical study, there are several methods described in the literature, that can be used to calculate the appropriate SS. For single-arm (non-randomized) trials, which are common in Phase 1 and 2 testings and in which every participant receives the experimental treatment, log-rank test and its weighted versions are the most common [8]. This method is based on asymptotic normality of the test statistic and is recommended for moderate-to-large SSs trials. Another researcher [9] proved that the correctness of the one-sample log-rank test when dealing with small samples is dependent on the correct specification of the underlying distribution of the standard population. The number of patients needed for a standard double-arm trial (also known as an RCT), which is common in Phase 3 testing, can be easily calculated using online resources [10] or other methods described in Section 3 of this paper. These approaches can lead to problems when dealing with small samples, because the study will be underpowered and the result has a high chance of being incorrect. The alternative is applying a Bayesian approach which uses external and subjective data expressed as a prior probability distribution, which combined with the trial evidence generates a posterior probability distribution for trail effect [11].

978-1-7281-2603-6/19/$31.00 ©2019 IEEE

The main contributions of this paper are highlighting the factors that influence the determination of the SS (Section 3), creating an extended comparative study between the classic methods and the Bayesian approach for small size calculations (Section 4 and 5) and providing a simple to use algorithm in choosing the appropriate method for sample calculation (Section 6).

II. How small is a small sample?

There is no definitive answer to this question. It depends on several criteria, that will be discussed later, and is closely related to the advancements in statistics and software development.

Over the years the rule of thumb was that regardless of the type of the study the lowest sample should be 25 items [12]. In practice, several studies were conducted with samples lower than 25, due to the type of the study, the rarity of the disease, or many interim analyses [13], and their results offered valuable information.

In a very inspiring paper [14], the trend of the SS was analyzed for a period of 15 years from 2000 to 2014. It was concluded that the median value of the SS is increasing for parallel studies from 32 items in 2000 to 36 in 2006 and for crossover studies from 15 to 20, in the same period. When the research ended in 2014 median SSs for parallel studies and crossover studies were 43 respectively 24. This increase is caused by the researchers' awareness of the importance of the SS over time, and the number of studies with a similar topic which acts as a starting point in SS determination.

III. Factors influencing the sample size

In determining the SS, a series of factors have to be taken into consideration regardless of the method used. Among them the most important are:

The type of study: descriptive, observational or experimental. A descriptive study is one used to determine population parameters, i.e. means, standard deviation (STD), proportions etc. Observational or experimental studies involve comparisons of the parameters. Although the type of study does not influence directly the size of the sample, it is an important factor when determining the type of error required.

The type of the error: type I (also known as α), or type II (a.k.a. β). α error is the rejection of a true null hypothesis (rejection of a false-positive result), while β error is a non-rejections of a false null hypothesis (a false negative finding). α errors are required in descriptive studies, while observational and experimental studies need both types of errors [1],[15]. These values are correlated with the significance criterion P, which is the difference between the compared parameters that are considered statistically significant – Table I.

TABLE I. The type of error

Truth	***Decision***	
	Accept H_0	Reject H_0
Null Hypothesis True (H_0 True)	Correct decision within P (1-α)	Type I error within P (α – significance)
Null Hypothesis False (H_0 False)	Type II error within P (β)	Correct decision within P (1-β – power)

The default values for these errors are $\alpha = 0.05$ and $\beta = 0.1$ or 0.2.

The statistical power of the study. There is always a trade-off between the desired statistical power and the feasible number of individuals [16], [17].

The type of the response variables: if they are quantitative (means, proportions, etc), or qualitative (true or false).

The minimum difference is the smallest measurable difference between the compared parameters. It is used preponderant in observational or experimental studies and should be chosen on clinical bases. In descriptive studies is reflected by the value of the confidence interval. The rule of thumb is lower the difference the greater the SS will be.

One or two-tailed tests. If the difference between the compared parameters is possible in only one direction one-tailed test is the best choice because it requires a smaller sample. Otherwise, the two-tailed test is mandatory [16], [18].

The presented factors act as design parameters for the determination of the appropriate SS. Once they are established all it remains is applying the appropriate methodology classical or Bayesian.

IV. Classic methodology for SS determination

In the classic approach, there are several methods that are used to determine the SS: imitating a SS of a similar study, using statistical tables, applying formulas or using appropriate software.

The simplest approach seems to be imitating the SS of a similar study, but this comes with a cost. Without verifying the procedures used in the study, the risk of repeating the errors that were made in determining the SS is great. This can be overcome by doing a literature review and comparing the SSs for several similar studies.

Another approach is to use published statistical tables for a given set of criteria. This approach is less popular these days because it is very restrictive regarding the conditions of the study and is usually applicable for large samples [19].

Although tables can provide a rapid determination of the SS, for a particular case is better to apply the formula, that is actually implied by the table. In the literature, there are several formulas that are used, but the most famous ones are the formula for estimation of a proportion and estimation of a mean for finite population and the formula for comparison of two means. The first two are mostly used in descriptive studies

while the latter is mostly for experimental or observational ones.

The formula for estimation of a proportion in a finite population is given by equation 1 [1].

$$n = \frac{t_\alpha^2 \cdot p \cdot q \cdot N}{(N-1)\cdot e^2 + t_\alpha^2 \cdot p \cdot q} \quad (1)$$

where n is the SS to be determined, t_α – is the value of the normal curve for type I error, p – the expected percentage of the response variable, $q = 1- p$, N – the size of the entire population, e – accepted margin of error.

The formula for estimation a mean – eq. 2 is similar to the one above:

$$n = \frac{t_\alpha^2 \cdot s^2 N}{(N-1)\cdot e^2 + t_\alpha^2 \cdot s^2} \quad (2)$$

where s^2 – is the variance of the variable for which the mean is calculated.

Determination of the SS when comparing two means is done using eq. 3.

$$n = \frac{2 \cdot s^2 \cdot (t_\alpha + t_\beta)^2}{(\bar{x}_1 - \bar{x}_2)^2} \quad (3)$$

where t_β – is the value of the normal curve for type II error, x_1 and x_2 – are the means of the two groups that are compared and s^2 – is the mean estimated variance of the compared groups.

Even with the help of the formulas, the determination of a minimum SS can be difficult for a non-statistician, because it involves detailed knowledge of choosing the proper coefficients. Sometimes different values for the effect size have to be tested in order for the best solution to be determined. In this case, the use of specialized software is necessary. We recommend PASS software [20] which is a software solution specialized for sample determination, IBM SPSS Statistics which is a complete package of statistical tools and the toolset for Statistical computation from MATLAB [21], [22].

The classic methodology faces some problems caused by the medical relevance of the specification of null and alternative hypotheses. This can lead to hard choices for a physician and can cause high error rates and low statistical power of the study. This is not the case for the Bayesian approach because it does not require a particular value of the alternative hypothesis [23].

V. Bayesian approach for SS determination

The Bayesian approach is based on external and subjective information expressed as a prior probability distribution. This will be combined with the trial data in order to produce a posterior probability distribution for the size of the effect. This approach ensures that the research is reducing the uncertainty about the effect from a level that already exists [11].

If for a study exist strong prior evidence, then even a small trial that is in concordance with the data can be conclusive. However, the situation changes if the research data and the prior information are contradictory.

The Bayesian approach usually starts with a broad literature review for gathering evidence regarding previous trials, even including studies that are only tangential related to the search topic. This information can then be used to construct a prior distribution for the effect with weights allocated in relation to pertinence, validity and precision. This first step works well when is not applied to extreme data sets. If the quantity of data is large, it can conduct to a time-consuming process that can delay the research. Otherwise, the lack of information could make probability distribution unreliable. In these extreme cases, the Bayesian approach can still be applied by using a non-informative prior distribution. This is, in essence, a uniform probability distribution based on the fact that every size of the effect is equally likely because there is no evidence to sustain otherwise. The reason that this method is still useful is that it enables the results to be expressed in terms of direct probabilities within a certain range based purely on the results from the research [11].

If applied correct the Bayesian method produces a smaller sample than the classic methods [24], [25] which is a strong argument in helping clinicians to make treatment decisions in small sample trials.

VI. Results

All the presented methods have advantages and disadvantages and sometimes can be confusing for a non-statistician. In order to make this process easier, we summarized all this information in Table II and propose a simple algorithm.

TABLE II. Classical and Bayesian methods comparison

Method	***Advantages***	***Disadvantages***
Imitating SS from other studies	• Fast • Easy to use	• Requires verifications
Statistical Tables	• Fast • Easy to use	• Is for large samples • Is for predefined conditions
Known formulas	• Any SS • Tailored for the study	• Requires some statistical knowledge
Specialized software	• Fast • Run multiple scenarios	• Requires some statistical knowledge • High cost
Bayesian approach	• Produces the smaller sample	• Requires prior data

The algorithm was designed to provide useful guidelines for a doctor facing a small sample trial. We recommend that this should be applied from the early stages of the research because it can provide a common ground between the medical team and the statisticians and also can help steer the project in the right direction.

The proposed algorithm consists of 7 steps:

Step 1 – Establish the type of the study and based on it the type of the error.

Step 2 – Determine the minimum difference and if the test is one or two-tailed.

Step 3 – Search the literature for other similar studies to have an idea about the SS.

Step 4 – Apply the known formulas for the specific conditions of the study.

Step 5 – If the obtained result is feasible and within the budget verify it by calculating the statistical power of the study.

Step 6 – If the study requires multiple scenarios the use of specialized software is appropriate.

Step 7 – If Steps 5 and 6 do not produce the desired result, apply the Bayesian method.

In the end, we would like to point out that the accuracy of the SS depends on the accuracy of the estimated parameters. Therefore, it is prudent to consider more than the minimum number of individuals calculated to compensate for the loss of precision due to estimation or other causes of errors.

VII. Conclusions

This paper presents useful guidelines to a physician for determining the SS when facing a small sample trial. A detailed comparison between the classical methods and the Bayesian approach is presented and based on it a simple algorithm for determining the optimal SS is developed. Based on the presented evidence it can be concluded that the Bayesian approach will produce a small sample than the classical method, but the elaborate process that is required makes it a valid candidate only when the classic approach is not feasible.

Acknowledgment

This paper was developed within the program PubArt from Politehnica University of Bucharest.

References

[1] S. Espa, M. M. Rodríguez, and A. R. González-ramírez, “Allergologia et immunopathologia Sample size calculation,” Allergol. Immunopathol. (Madr)., vol. 42, no. 5, pp. 485–492, 2016.

[2] U. Ahmed Ali, J. R. ten Hove, B. M. Reiber, P. C. van der Sluis, and M. G. Besselink, “Sample size of surgical randomized controlled trials: a lack of improvement over time,” J. Surg. Res., vol. 228, pp. 1–7, 2018.

[3] B. Copsey et al., “Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review,” J. Clin. Epidemiol., vol. 104, pp. 52–61, 2018.

[4] H. J. Lamberink et al., “Statistical power of clinical trials increased while effect size remained stable: an empirical analysis of 136,212 clinical trials between 1975 and 2014,” J. Clin. Epidemiol., vol. 102, pp. 123–128, 2018.

[5] F.-C. Adochiei, R.-I. Ciucu, I.-R. Adochiei, S. D. Grigorescu, G. C. Seri?an, and M. Casian, “A WEB Platform for Rendring and Viewing MRI Volumes using Real-Time Raytracing Principles,” in 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), 2019, pp. 1–4.

[6] J. B. Dimick, M. Diener-West, and P. A. Lipsett, “Negative results of randomized clinical trials published in the surgical literature: equivalency or error?,” Arch. Surg., vol. 136, no. 7, pp. 796–800, 2001.

[7] H. V Lochner, M. Bhandari, and P. Tornetta III, “Type-II error rates (beta errors) of randomized trials in orthopaedic trauma,” JBJS, vol. 83, no. 11, pp. 1650–1655, 2001.

[8] M. A. Phadnis, “Sample size calculation for small sample single-arm trials for time-to-event data: Logrank test with normal approximation or test statistic based on exact chi-square distribution?,” Contemp. Clin. Trials Commun., vol. 15, no. April, p. 100360, 2019.

[9] J. Wu, “Sample size calculation for the one-sample log-rank test,” Pharm. Stat., vol. 14, no. 1, pp. 26–33, 2015.

[10] M. A. Kohn, M. S. Jarrett, and J. Senyak, “Sample size calculators,” Clin. Transl. Sci. Institute.[atualizado em 2017]. Dispon{\'\i}vel em http//www. samplesize. net/correlation-sample-size, 2016.

[11] L. Billingham, K. Malottki, and N. Steven, “Small sample sizes in clinical trials: a statistician’s perspective,” Clin. Investig. (Lond)., vol. 2, no. 7, pp. 655–657, 2012.

[12] M. A. Bujang, P. A. Ghani, S. A. Soelar, and N. A. Zulkifli, “Sample size guideline for exploratory factor analysis when using small sample: Taking into considerations of different measurement scales,” ICSSBE 2012 - Proceedings, 2012 Int. Conf. Stat. Sci. Bus. Eng. "Empowering Decis. Mak. with Stat. Sci., pp. 447–451, 2012.

[13] J. Shao and H. Feng, “Group sequential t-test for clinical trials with small sample sizes across stages,” Contemp. Clin. Trials, vol. 28, no. 5, pp. 563–571, 2007.

[14] P. H. Lee and A. C. Y. Tse, “The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed,” Eur. J. Intern. Med., vol. 40, pp. 16–21, 2017.

[15] S. D. Grigorescu et al., “Robotic Platform with Medical Applications in the Smart City Environment,” in 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), 2019, pp. 1–6.

[16] J. Eng, “Sample size estimation: How many individuals should be studied?,” Radiology, vol. 227, no. 2, pp. 309–313, 2003.

[17] A. I. Drosu and O. M. Drosu, “A Bio-potential Measurement Solution for the Acquisition of Electromyogram,” in 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), 2019, pp. 1–6.

[18] C. Cepisca, F. C. Adochiei, S. Potlog, C. K. Banica, and G. C. Seritan, “Platform for bio-monitoring of vital parameters in critical infrastructures operation,” in 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2015, p. E--7.

[19] Kish and leslie, “Survey sampling. John Willey and Sons,” Univ. florida, Inst. Food Agric. Sci., pp. 1–5, 1965.

[20] M. A. Bujang and N. Baharum, “Guidelines of the minimum sample size requirements for Cohen’s Kappa,” Epidemiol. Biostat. Public Heal., vol. 14, no. 2, pp. e12267-1-e12267-10, 2017.

[21] M. Stanculescu, S. Marinescu, and G. Cheregi, “Matlab in electrical engineering,” J. Electr. Electron. Eng., vol. 2, no. 2, 2009.

[22] T. Cazangiu, F.-C. Argatu, B.-A. Enache, V. Vita, and G. Stavros, “Device for monitoring people with Alzheimer’s disease,” in 2018 International Symposium on Fundamentals of Electrical Engineering (ISFEE), 2018, pp. 1–5.

[23] S. K. Sahu and T. M. F. Smith, “A Bayesian method of sample size determination,” pp. 235–253, 2006.

[24] F. Sadia and S. S. Hossain, “Contrast of bayesian and classical sample size determination,” J. Mod. Appl. Stat. Methods, vol. 13, no. 2, pp. 420–431, 2014.

[25] H. Costin and C. Rotariu, "Vital Signs Telemonitoring by Using Smart Body Area Networks, Mobile Devices and Advanced Signal Processing," Advances in Biomedical Informatics, DOI: 10.1007/978-3-319-67513-8_10, 2018