Sample size in clinical and experimental trials

Hélio Amante Miot*

Investigative clinical-epidemiological or experimental studies - have the objectives of describing phenomena or comparing the behavior of variables in subgroups of a population. To accomplish those objectives, the population universe is not studied in its entirety, usually because it is not accessible or viable, but mainly because it is not necessary, when a representative sample is available for correlations with the target population[1,2].

The sample planning of a study determines the numerical dimension and also the sampling technique (collection/selection) of the elements of the study. It is essential in the elaboration of the project, and problems with such planning may compromise the final data analysis and interpretation of its results. A proper sample planning depends on basic knowledge of the study statistics and deep knowledge of the problem under investigation, in order to combine the statistical significance of the tests with the clinical meaning of the results[1,3,4].

Most biostatistical tests assume that the study sample is probabilistically representative of the population. Some samples of convenience, e.g. like choosing consecutive patients of a specific outpatient clinic, may not properly represent all the study population. The investigator should be alert to possible selection biases resulting from the availability of patients in consecutive sampling, because increasing sample size would not correct the effect of biased samples. In addition, strategies of non-probability stratified sampling, by using sample quotas, complex sampling (conglomerates, multi-levels), voluntary response, saturation of variables,

"snowball" type or by using non-randomized methods of data collection should be designed, sized and analyzed with expert statistical support. This paper discusses the principles of simple random sample size calculations[4].

The selection of a population fraction that makes up the study sample implies the investigator will assume a certain degree of error for the estimated values of population parameters to each variable; such sample error is quantifiable, and inversely proportional to the sample size[4,5].

In order to describe the population estimate represented by a quantitative variable (discrete or continuous), one should obtain the population standard deviation of that variable and determine the significance level of the estimate and the maximum tolerable sample error (in units of mean value) (Chart 1)[2].

In order to describe the population estimate represented by a qualitative variable (nominal or ordinal), one should obtain the population frequency of the variable results and determine the significance level of the estimate and the maximum tolerable sample error (in percentage) (Chart 1). When a qualitative variable is not dichotomic, sample sizing should be considered for the proportion of each category that constitutes the variable[4].

When the population standard deviation or frequencies of the variable are unknown, and the literature does not present any similar data, a pre-test should be conducted with 30-40 subjects and the behavior of this subgroup should be considered as the population estimate[2].

Chart 1. Formulas for sample sizing to describe quantitative and qualitative variables in a population.

	Quantitative variable	Qualitative variable
Infinite population	$n = \left(\dfrac{Z\alpha/2 \cdot \delta}{E} \right)^2$	$n = \left(\dfrac{Z\alpha/2 \cdot \sqrt{p \cdot q}}{E} \right)^2$
Finite population (<10000)	$n = \dfrac{N \cdot \delta^2 \cdot (Z\alpha/2)^2}{(N-1) \cdot (E)^2 + \delta^2 \cdot (Z\alpha/2)^2}$	$n = \dfrac{N \cdot p \cdot q \cdot (Z\alpha/2)^2}{(N-1) \cdot (E)^2 + p \cdot q \cdot (Z\alpha/2)^2}$

n – sample size; $Z_{\alpha/2}$ – critical value for the desired confidence degree, usually: 1.96 (95%); δ – population standard deviation of the variable; E – standard error, usually: ±5% of the proportion of cases (absolute precision), or ±5% of the mean value (1.05×mean); N – (finite) population size; p – proportion of favorable results of the variable in the population; q – proportion of unfavorable results in the population (q=1-p).

*Assistant Professor of theDermatology and Radiotherapy Department, Faculdade de Medicina de Botucatu of Universidade Estadual Paulista (UNESP) - Botucatu (SP), Brazil.

In addition, sample sizing formulas assume populations of unlimited size. A special situation occurs when limited populations (<10000 subjects) are studied, as in these cases, each sampled unit represents a significant fraction of the finite sample universe. In such cases, these formulas may be adjusted using a correction factor for finite populations, thereby minimizing the required sample size (Chart 1)[2].

Example 1: To describe the measurements of mean arterial pressure from a specific population of patients that has never been described before, with tolerable error of ±5 mmHg, the sample size would have to be based on the standard deviation considering the values from this group. If a pre-test with 30 patients showed the standard deviation of 15 mmHg, the sample size, using the formula presented in Chart 1, would be:

$$n=(1.96 \times 15/5)^2 = 34.6 \text{ patients}$$

Example 2: To describe the prevalence of venous insufficiency of the lower limbs, with tolerable error of ±5%, in the population of morbidly obese patients from a specific obesity outpatient clinic with 315 patients (630 limbs), the sample size calculation could be based on the results obtained by Seidel et al.[6], who estimated the proportion of 69.3% of affected limbs. The sample size calculation for a finite population uses the formula presented in Chart 1:

$$n=[630 \times 0.693 \times 0.307 \times (1.96)^2]/\{[(630-1)$$
$$\times (0.05)^2]+[0.693 \times 0.307 \times (1.96)^2]\}=215.5 \text{ limbs}$$

The sample size calculation for subgroup comparison (hypothesis testing) within a sample depends on the selected statistical test, differences between the groups and the investigator's tolerance to detecting differences when they do not exist (type I error) or failure to detect differences between the subgroups when they really exist (type II error). The probabilities associated with errors of types I and II are standardized as α and β, and values of 5% (bilateral) and 20% are usually adopted, but other values may be used judiciously (Chart 2)[1,2].

One strategy that enables to reduce the variability of measurements, increasing the comparability between the individuals in a sample, and, consequently, reducing the sample size required to detect a phenomenon, is pairing (or matching) of observations (Chart 2). It can be used when the same subject is observed at different moments (longitudinal study) or submitted to measurements in different areas of the body, e.g., the comparison of a treatment in the right lower limb versus the left lower limb, provided that the ethical limits of this comparison are respected. Another type of pairing - more elaborated - is the selection of subjects presenting the same characteristics: age, gender, ethnical group, social class, among other variables that can control the individual variability. In these cases, the measurement is made between pairs, rather than using a direct comparison of subgroups[1].

Example 3: To compare the flow measurements of two limbs of dogs submitted to two different procedures of arterial revascularization, with the minimum tolerable difference of ±50 mL/min to consider efficient one of the procedures, a pilot study would have to indicate the standard deviation of the differences between flows (e.g.: 60 mL/min). The sample size, considering the formula presented in Table 2, would be:

$$n=[(1.96+0.84) \times 60/50]^2 = 11.3 \text{ animals}$$

Example 4: To compare the healing rates of two surgical procedures, the traditional method resulting in a 70% healing rate and the study procedure at least 10% better than the conventional system, the minimum sample size calculation of a clinical trial should consider the formula presented in Chart 2:

$$n=\{[(0.7 \times 0.3)+(0.8 \times 0.2)] \times (1.96+0.84)^2\}/(0.7-0.8)^2$$
$$= 290.4 \text{ patients (each group)}$$

In studies where several variables are important for the analysis of the studied outcome, i.e., are not only control or

Chart 2. Formulas for sample sizing to compare two groups according to quantitative and qualitative variables and according to pairing of cases.

	Quantitative variable	Qualitative variable
Non-paired sample	$n = (Sa^2 + Sb^2) \cdot \left(\dfrac{Z\alpha/2 + Z\beta}{d}\right)^2$	$n = \dfrac{(p1 \cdot q1 + p2 \cdot q2) \cdot (Z\alpha/2 + Z\beta)^2}{(p1-p2)^2}$
Paired sample	$nP = \left(\dfrac{(Z\alpha/2 + Z\beta) \cdot Sd}{\overline{D}}\right)^2$	$nP = \dfrac{\left(Z\alpha/2 + 2 \cdot Z\beta \cdot \sqrt{pa \cdot qa}\right)^2}{4 \cdot pd \cdot (pa - 0{,}5)^2}$

n – sample size (for each subgroup); nP – number of pairs; $Z_{\alpha/2}$ – value of error α, usually: 1.96 (5%); Z_β – value of error β, usually: 0.84 (20%); d – minimum difference between the mean values; Sa and Sb – standard deviation of the variable in each group; Sd – standard deviation of the difference between the pairs; $\overline{D}$ – mean value of the difference between the pairs; p1 and p2 – proportion of favorable results in subgroup 1 or 2; q1 and q2 – proportion of unfavorable results in subgroup 1 or 2; pa – proportion of unmatched pairs for group 1; qa – proportion of matched pairs for group 1; pd – sum of the proportion of unmatched pairs for the two groups.

correction variables, it is necessary to calculate the sample size to each important variable of the study.

Tests for equivalence, non-inferiority and agreement require specific sample sizing methods, different from the tests of differences between mean values and proportions commonly used. In addition, multivariate analyses, comparison of subgroups to different numerical proportions, or multiple longitudinal comparisons, also involve higher complexity in sample sizing calculation. All these items exceed the scope of this paper[1,5,7-10].

Sample size calculation for trials that involve the estimate of linear correlation between two quantitative variables is dependent solely on the linear correlation coefficient (Chart 3).
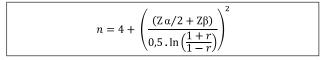
Example 5: To establish the correlation between the measurement of muscle force of quadriceps and the maximum distance covered by patients with history of intermittent claudication, the sample size calculation could be based on the study conducted by Pereira et al.[11], which described a linear correlation coefficient of 0.87. According to the formula presented in Chart 3:

$$n = 4 + \{(1.96 + 0.84)/[0.5 \times \ln(1+0.87)/(1-0.87)]\}^2$$
$$= 8.4 \text{ patients}$$

Longitudinal studies (prospective cohorts and clinical trials), as they require the patients' follow-up over long periods, can be affected by subjects who leave, quit, drop out, die or are excluded from the study. The initial sample calculation correction is recommended, increasing it at least 30%, in order to overcome such sample losses. Dropout subjects should be studied judiciously regarding their reasons for leaving and whether they present difference in the study variables in relation to the other study subjects, to identify factors specifically linked with the dropouts. When more than 30% of the subjects are lost to follow-up, the results of the whole sample may be compromised, regardless of the number of cases.

Provided that the conclusions of a study can be generalized only to the population under study, it is possible that repeating the study in other centers may yield different results, reflecting the reality of the other populations. Such results may indeed exceed the confidence interval limits for the primarily estimated parameter, not necessarily meaning lack of

internal validation of either study. This is one of the risks of using results from other investigators when sizing the sample of a different population. A preliminary analysis of the first fraction of cases (pre-test) is strongly recommended, making it easier to estimate the sample required to each reality, and prevent analytical constraints at the end of the study[12].

Whenever the sample size is very small (<30 measurements), the analysis of subgroups is more difficult and the performance of statistical trials is compromised. One should be, however, careful to prevent sample supersizing, which usually occurs when the access to large computer databases are available. Increasing the sample reduces the confidence intervals of estimates and allows the detection of differences between subgroups which, even if statistically significant, do not present clinical relevance[3,12-14].

At last, there are different formulas for the sample size calculation to specific statistical trials besides those presented here, depending on the mathematical model considered, which can be easily found in the literature or on the Internet[1,15,16]. There are some free software applications in Portuguese, such as intuitive BioEstat, that offer sample sizing modules[17]. However, sample sufficiency should be regarded as an important part of a study methodological planning, which has be integrated into the elaboration of hypothesis, study design, sampling techniques and data analysis and interpretation, for a successful investigation.

References

1. Norman GR, Streiner DL. Biostatistics. The bare essentials. 3rd ed. Shelton, Connecticut: People's Medical Publishing House; 2008.

2. Fontelles MJ, Simões MG, Almeida JC, Fontelles RGS. Metodologia da pesquisa: diretrizes para o cálculo do tamanho da amostra. Rev Paran Med. 2010;24:57-64.

3. Paes AT. Itens essenciais em bioestatística. Arq Bras Cardiol. 1998;71:575-80.

4. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Co.; 1987.

5. Azevedo RS. Qual o tamanho da amostra ideal para se realizar um ensaio clínico? Rev Assoc Med Bras. 2008;54:289.

6. Seidel AC, Mangolim AS, Rossetti LP, Gomes JR, Jr FM. Prevalência de insuficiência venosa superficial dos membros inferiores em pacientes obesos e não obesos. J Vasc Bras. 2011;10:124-30.

7. Katz MH. Multivariable analysis. A practical guide for clinicians. 2nd ed. Cambridge, UK: Cambridge University Press; 2006.

8. Ortega Calvo M, Cayuela Dominguez A. Unconditioned logistic regression and sample size: a bibliographic review. Rev Esp Salud Publica. 2002;76:85-93.

9. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical therapy. 2005;85:257-68.

Chart 3.   Formula for sample sizing for linear correlation between quantitative variables.

$$n = 4 + \left(\frac{(Z\alpha/2 + Z\beta)}{0{,}5 \cdot \ln\left(\frac{1+r}{1-r}\right)} \right)^2$$

n – sample size; $Z_{\alpha/2}$ – value of error α, usually: 1.96 (5%); Z_β – value of error β, usually: 0.84 (20%); r – linear correlation coefficient (Pearson or Spearman).

10. Pinto VF. Estudos clínicos de não-inferioridade: fundamentos e controvérsias. J Vasc Bras. 2010;9:141-4.

11. Pereira DAG, Faria BMA, Gonçalves RAM, Carvalho VBF, Prata KO, Saraiva PS, et al. Relação entre força muscular e capacidade funcional em pacientes com doença arterial obstrutiva periférica: um estudo piloto. J Vasc Bras. 2011;10:26-30.

12. Mourão Jr CA. Questões em bioestatística: o tamanho da amostra. Rev Interdisc Est Experim. 2009;1:26-8.

13. Coutinho ESF, da Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. Rev Bras Psiquiatr. 2005;27:146-51.

14. Weyne GRS. Determinação do tamanho da amostra em pesquisas experimentais na área de saúde. Arq Med ABC. 2004;29:87-90.

15. Laboratório de Epidemiologia e Estatística - LEE - Pesquisa. 2000 [cited 2011 Sep 16]. Available from: http://www.lee.dante.br/pesquisa.html.

16. UCSF Biostatistics - Power and Sample Size Programs. 2006. [cited 2011 Sep 16]. Available from: http://www.epibiostat.ucsf.edu/biostat/sampsize.html.

17. BioEstat 5.3 - Instituto de desenvolvimento sustentável Mamirauá. 2011. [cited 2011 Sep 16]. Available from: http://www.mamiraua.org.br/download/.