Allergol Immunopathol (Madr). 2014;42(5):485–492

Allergologia et immunopathologia

Sociedad Española de Inmunología Clínica, Alergología y Asma Pediátrica

www.elsevier.es/ai

SERIES: BASIC STATISTICS FOR BUSY CLINICIANS (XI)

Sample size calculation

MM Rodríguez del Águila[a,*], **AR González-Ramírez**[b]

[a] *UGC Medicina Preventiva, Vigilancia y Promoción de la Salud. Hospital Virgen de las Nieves, Granada, Spain*
[b] *Fundación Pública Andaluza para la Investigación Biosanitaria de Andalucía Oriental. Hospital Clínico San Cecilio, Granada, Spain*

Received 29 January 2013; accepted 25 March 2013
Available online 23 November 2013
Series' Editor: V. Pérez-Fernández.

Abstract When designing any research project, definition is required of the sample size needed in order to carry out the study. This sample size is an estimate of the number of patients required in accordance with the pursued study objective. In this context, it is more efficient in terms of both cost and time to use samples than to work with the entire population.

The present article describes the way to establish sample size in the kinds of studies most frequently found in health research, and how to calculate it using the *epicalc* package included in the shareware R program. A description is provided of the formulae used to calculate sample sizes for the estimation of a mean and percentage (referring to both finite and infinite populations) and for the comparison of two proportions and two means. Likewise, examples of the application of the mentioned statistical package are provided.

Introduction

The design phase of any epidemiological study includes the determination of the sample size needed to carry out the study.[1–4] In order to confirm the established study hypotheses, there must be a coherent relationship among the amount or number of observations made and their possible repetitions, their representativeness and the quality of the evidence – in addition to a solid and rigorous experimental design.[4] This number of observations or samples is called the *sample size* (SS),[5] referred to as the letter *n*.

The SS calculation involves the application of a series of mathematical formulae that have been designed to secure precision in estimating the population parameters or to obtain significant results in those studies that compare several treatment regimens or groups. It is important to establish the SS before the study is carried out, since in this way we can be sure of recruiting an adequate number of patients. If this is not done, we run the risk of conducting an unnecessary number of tests, with the associated waste of time and money, or of collecting an insufficient body of data – thereby generating imprecision and very probably leading to failure to detect significant differences, when in fact such differences might indeed exist.[6,7] It is common for the number of observations to be defined by the investigator, according to the existing economic and human resources, and on the time available for carrying out the study.[8]

* Corresponding author.
E-mail address: mmar.rodriguez.sspa@juntadeandalucia.es (M. Rodríguez del Águila).

0301-0546/$ – see front matter © 2013 SEICAP. Published by Elsevier España, S.L.U. All rights reserved.
http://dx.doi.org/10.1016/j.aller.2013.03.008

As has been mentioned, the SS of a study is determined using mathematical formulae designed to the effect. Accordingly, we will need prior information that can be obtained from historical studies, the literature, or from a pilot study. In this context, pilot studies are small studies carried out under the same conditions as the global or larger study, but involving a limited sample size of 10 or 20 subjects – thereby allowing us to correct possible errors in implementing the project. The preliminary results afforded by such pilot studies produce information for establishing the definitive sample size.

There are studies in which it proves difficult to recruit the necessary number of patients; in most such cases the study involves a rare disease for which the number of cases is limited, such as for example idiopathic solar urticaria. Even in these cases, however, it is advisable to determine the SS, attempting to carry out the study on a multicentre basis in which each participating centre contributes a certain number of cases. When a study of this type is not possible, the considerations referring to SS are made according to the maximum number of patients that can be recruited in the course of the study – but this implies the important inconvenience of a decrease in precision.

It is not unusual to find studies that establish two primary objectives and/or several secondary objectives. In theory, each primary objective should be associated to its own SS. Choosing a smaller SS results in diminished statistical power. Theoretically, we should choose the largest SS of the primary objectives, since failure to do so would cause the primary objectives without a sufficient SS to automatically become secondary objectives or simply exploratory objectives.[9]

The SS is partially dependent on the size of the population of origin. In order to establish the necessary number of patients in a study, we generally start by assuming that populations of unknown or infinite size must be sampled. In some studies we will need to sample populations of finite size (or *N*), particularly in descriptive surveys where this size must be incorporated into the calculations. In fact, the SS in the formulae that include *N* in the calculation tends to converge with the size in which this parameter is not included. Most authors consider a population to be finite if *N* is less than 100,000 subjects.

Factors influencing the calculation of sample size

In calculating the sample size, we must first take a number of factors into account, since they condition the different formulae used to establish SS.[10] These factors are the following:

(1) The type of study involved: descriptive, observational or experimental. In descriptive studies with finite populations, we also need to know the population size, *N*.
(2) The α (type I error) and β (type II error) errors we are willing to accept. In case of doubt, we adopt $\alpha = 0.05$ and $\beta = 0.10$ or 0.20 as standard values, with the following exceptions:
    (a) In descriptive studies we only require the α error or confidence level in the estimation, together with the precision (magnitude or width of the confidence interval).
    (b) In experimental or observational studies we require both α and β error.
(3) The response variables to be observed and their level of measurement (i.e., whether they are quantitative or qualitative: means or proportions (%)).
(4) The minimum difference to be detected between the treatment groups or between the null hypothesis and the alternative hypothesis. This will depend on the study involved. The smaller the difference we wish to detect, the larger the number of subjects we must include in the study. This difference should be not only clinically significant but also realistic. In descriptive studies, the difference is reflected by the amplitude of the confidence interval calculated in the estimation.
(5) When the variables analysed in the study are of a quantitative nature, their variability must be considered, measured in terms of variance or standard deviation (SD). If there is little variability, the required number of subjects is much smaller than when the variability referred to the analysed characteristic is large. The variability can be obtained from the literature sources or from pilot studies.
(6) Skewness (laterality) of the hypothesis test: i.e., whether it is a one- or two-tailed test. Studies involving one-tailed testing generally require a smaller sample size than those with two-tailed testing, though the former should only be contemplated when the direction of the test is evident.
(7) Losses referred to patient localisation or follow-up. These losses should be added to the sample calculation made.
(8) The different groups to be compared and the comparisons to be made between them. When several groups are contrasted, the formulae used to determine SS must document information on the number of groups considered in the study. Failure to do so can result in the propagation of α error, which would exceed the initially defined level of 5%.[11]

Example: Suppose we wish to examine the effectiveness of four different treatments (A, B, C, and D) in patients with atopic dermatitis, evaluating the number of successes or failures with each of them. If we perform two-by-two comparisons, a total of six comparisons would have to be made (A–B, A–C, A–D, B–C, B–D and C–D). The probability of obtaining a correct decision referred to H_0 in one such test would be $(1 - \alpha) = 95\%$, and the probability of obtaining a correct decision in all the above tests therefore would be $0.95^6 = 0.74$. The probability of a wrong decision in any of them would be $1 - 0.74 = 0.26$, which is far higher than the generally established value of $\alpha = 0.05$.

The formulae referred to SS are little affected by the magnitude of *N* referred to the population, since the larger the latter, the more stable the SS value tends to become.

The result of applying a formula for calculating SS generally yields a non-whole number. In this sense, SS is taken to be the rounded next higher whole number or integer (e.g., for $n = 120.34$ we take 121).