ELSEVIER

Journal of Clinical Epidemiology 58 (2005) 859–862

Journal of Clinical Epidemiology

Sample size calculation should be performed for design accuracy in diagnostic test studies

Antoine Flahault[a,b,*], Michel Cadilhac[a,b], Guy Thomas[b]

[a]*Unité de biostatistique et d'Informatique médicale, Hôpital Tenon, Paris, France*
[b]*Institut National de la Santé et de la Recherche Médicale (INSERM), unité 707, Université Pierre & Marie Curie, 27, rue Chaligny, F-75571 Paris cedex 12, France*

Accepted 8 December 2004

Abstract

Background and Objectives: Guidelines for conducting studies and reading medical literature on diagnostic tests have been published: Requirements for the selection of cases and controls, and for ensuring a correct reference standard are now clarified. Our objective was to provide tables for sample size determination in this context.

Study Design and Setting: In the usual situation, where the prevalence Prev of the disease of interest is <0.50, one first determines the minimal number N_{cases} of cases required to ensure a given precision of the sensitivity estimate. Computations are based on the binomial distribution, for user-specified type I and type II error levels. The minimal number $N_{controls}$ of controls is then derived so as to allow for representativeness of the study population, according to $N_{controls} = N_{cases}$ [(1 − Prev)/Prev].

Results: Tables give the values of N_{cases} corresponding to expected sensitivities from 0.60 to 0.99, acceptable lower 95% confidence limits from 0.50 to 0.98, and 5% probability of the estimated lower confidence limit being lower than the acceptable level.

Conclusion: When designing diagnostic test studies, sample size calculations should be performed in order to guarantee the design accuracy. © 2005 Elsevier Inc. All rights reserved.

Keywords: Sensitivity; Specificity; Sample size; Binomial distribution; Diagnostic test

1. Introduction

The importance of sample size calculation in medical research is emphasized in all sets of good clinical practice guidelines. Previous articles have dealt with this issue under various circumstances, and in particular for two-group comparisons within clinical trials [1]. Simel et al. [2] deal with sample sizes based on desired likelihood ratios confidence intervals. Knottnerus and Muris [3] deal with the whole strategy needed for development of diagnostic tests, but do not provide practical tables for calculating sample sizes in the very situation that clinician epidemiologists are in when dealing with sensitivity or specificity confidence intervals.

From a statistical point of view, sample size issues for diagnostic test assessment studies have formal counterparts within the field of clinical trials, so that answers could be derived from published equations and tables [4], at least in principle; how to perform such derivations may not be clear to clinicians.

Moreover, in the case of binary (yes/no) outcome tests, a normal approximation to the binomial distribution is often used [5]. Although the accuracy of the approximation is usually good, modern software allows for exact calculations to be carried out at virtually no extra cost. For instance, a SAS macro is available to compute exact binomial confidence limits [6].

Our objective was to describe the determination of sample size for binary diagnostic test assessment studies, and to provide exact tables based on the binomial distribution.

2. Methods

2.1. Definitions

Assessing a diagnostic test procedure with binary (yes/no) outcome entails determining the operating characteristics of the test with respect to some disease of interest. The intrinsic characteristics of the test are sensitivity and specificity. Sensitivity (Se) is the probability that the test outcome is positive in a patient who has the disease, and is estimated by the proportion of positive test results among a sample of patients with the disease (cases). Specificity (Sp) is the probability that the test

* Corresponding author. Tel.: +33-1-44738441; fax: +33-1-44738454.
E-mail address: flahault.a@wanadoo.fr (A. Flahault).

0895-4356/05/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi: 10.1016/j.jclinepi.2004.12.009

outcome is negative in a subject who is free from the disease of interest, and is estimated by the proportion of negative results in a sample of disease-free subjects. The positive (or negative) predictive value of the test in a given population is the probability that a test positive (or negative) subject has (or does not have) the disease. Although predictive values are of obvious clinical and epidemiological relevance, they are not intrinsic to the test, insofar as they also depend on the prevalence of the disease in the population under study. These issues are discussed by Altman and Bland [7,8].

In addition, to evaluate the accuracy of the sensitivity or specificity estimate, the experimenter must further estimate some confidence limit. The $1 - \alpha$ lower confidence limit for Se (or Sp) can be thought of as the lowest value of Se (or Sp) that is not rejected by a one-sided test of level α of the null hypothesis $Se = Se_L$ (or $Sp = Sp_L$) against the alternative hypothesis $Se > Se_L$ (or $Sp > Sp_L$). Upper confidence limits are defined in an analogous manner, but are irrelevant here, because the concern is that the test actually performs worse, not better, than indicated by the observed proportion of positive (or negative) outcomes in the trial sample.

2.2. Number of cases

Assume first that we wish to determine the number of cases to estimate the sensitivity of a new diagnostic test.

The process of determining the sample size in this context is formally identical to that when comparing an observed proportion to a known proportion. Prior to sample size computation, the experimenter must thus specify (i) the expected sensitivity Se of the test and (ii) the maximal distance δ from Se within which the $1 - \alpha$ lower confidence limit is required to fall, with probability $1 - \beta$. Although α and $1 - \beta$ retain their usual meanings in terms of type I error and power, respectively, δ is analogous to the effect size, and $Se - \delta$ plays the role of the known proportion. Relying on the normal approximation to the binomial distribution, sample sizes could thus be determined according to equation (A1) in the Appendix. For small values of δ, sample sizes may be further approximated by halving the numbers given in Table 3.1 of Machin et al. [4], but this table is ill-suited to the context of diagnostic test assessment studies.

Moreover, because in cases of interest the expected sensitivity will typically be close to one, the normal approximation may be somewhat inaccurate, and one should therefore fall back on exact equations based on the binomial distribution (see Appendix).

2.3. Number of controls

To determine the number of controls needed to estimate the specificity of a diagnostic test, the procedure is identical with that described in the preceding section, substituting specificity for sensitivity.

In practice, the clinician will want to estimate both sensitivity and specificity within a study population containing cases and controls. In this case, to ensure that the study population is representative of the population to which the test will be applied, the proportions of cases and controls should take account of the prevalence Prev of the disease, according to

$$N_{\text{controls}} = N_{\text{cases}}\,[(1-\text{Prev})/\text{Prev}] \tag{1}$$

For the vast majority of diseases, $\text{Prev} < 0.50$, and so equation (1) yields $N_{\text{controls}} > N_{\text{cases}}$. In such instances, provided the accuracy requirements are similar for Sp and Se, the experimenter should first determine the minimal number of cases from the tables, and then compute the number of controls from equation (1). If $\text{Prev} > 0.50$, first read N_{controls} in the tables, then compute N_{cases} from equation (1).

Sample sizes were computed according to equation (A3) in the Appendix, using *Mathematica* software [9]. The code is available from the corresponding author.

So far, we have considered situations where cases and controls are sampled separately. Often, however, the investigator must sample from a population without prior knowledge of the individual case–control status. In such instances, the sample size n must be determined such that, with high probability (e.g. 95%), the sample contains sufficient numbers of cases and controls. To meet this requirement, a possible strategy is to choose n as the smallest integer such that

$$\sum_{x=N_{\text{cases}}}^{n} \binom{n}{x} \text{Prev}^x (1-\text{Prev})^{n-x} \geqslant 0.95, \tag{2}$$

where Prev is the population disease prevalence, and N_{cases} is determined from the tables. If $\text{Prev} > 0.50$, use the same equation with N_{controls} in place of N_{cases}.

3. Results

Sample sizes corresponding to lower 95% confidence limits, to be violated with probability $< 5\%$, are presented in Tables 1 and 2.

Whenever disease prevalence is <0.50, the following guidelines should be followed. The first step requires an assumption on the expected value of the new diagnostic test sensitivity. The second step is to specify the minimum acceptable lower confidence limit, together with the required probability (which was set here at 0.95) that this limit is not violated. The minimal sample size for the group of cases is then read from the tables. The corresponding number of controls is obtained from equation (1).

For example, suppose we wish to investigate a new diagnostic procedure with expected sensitivity and specificity of 0.90, in a population where the disease prevalence is 0.10, and we require the lower 95% confidence limit to be > 0.75 with 0.95 probability. From Table 1, $N_{\text{cases}} = 70$; from equation (1), $N_{\text{controls}} = 630$.

Download English Version:

https://daneshyari.com/en/article/10514552

Download Persian Version:

https://daneshyari.com/article/10514552

Daneshyari.com