

A myriad of methods: Calculated sample size for two proportions was dependent on the choice of sample size formula and software

Melanie L. Bell^{a,b}, Armando Teixeira-Pinto^{c,*}, Joanne E. McKenzie^d, Jake Olivier^e

^aPsycho-Oncology Co-Operative Research Group, School of Psychology, University of Sydney, Australia

^bMel and Enid Zuckerman College of Public Health, University of Arizona, 295 N Martin Ave, Tucson, AZ 85724, USA

^cSchool of Public Health, Edward Ford Building (A27), University of Sydney, Sydney NSW 2206, Australia

^dSchool of Public Health and Preventive Medicine, Alfred Centre, Monash University, Melbourne VIC 3004, Australia

^eSchool of Mathematics and Statistics, The Red Centre, The University of New South Wales, Sydney 2052, Australia

Accepted 21 October 2013; Published online 16 January 2014

Abstract

Objectives: Several methods exist to calculate sample size for the difference of proportions (risk difference). Researchers are often unaware that there are different formulae, different underlying assumptions, and what the impact of choice of formula is on the calculated sample size. The aim of this study was to discuss and compare different sample size formulae for the risk difference.

Study Design and Setting: Four sample size formulae were used to calculate sample size for nine scenarios. Software documentation for SAS, Stata, G*Power, PASS, StatXact, and several R libraries were searched for default assumptions. Each package was used to calculate sample size for two scenarios.

Results: We demonstrate that for a set of parameters, sample size can vary as much as 60% depending on the formula used. Varying software and assumptions yielded discrepancies of 78% and 7% between the smallest and largest calculated sizes, respectively. Discrepancies were most pronounced when powering for large risk differences. The default assumptions varied considerably between software packages, and defaults were not clearly documented.

Conclusion: Researchers should be aware of the assumptions in power calculations made by different statistical software packages. Assumptions should be explicitly stated in grant proposals and manuscripts and should match proposed analyses. Crown Copyright © 2014 Published by Elsevier Inc. All rights reserved.

Keywords: Sample size; Power; Difference in proportions; Risk difference; Statistical software; Continuity correction; Binary

1. Introduction

Research studies need to be designed with an adequate sample size that reliably addresses the study's hypotheses. Adequate sample size ensures that resources are used efficiently and that patients are not exposed to the potential harms of research unnecessarily [1,2]. The sample size formula should be appropriate for the type of outcome and should align with planned analyses. Furthermore, the formula and software used to calculate sample size and the parameters used in the calculation should be reported [3]. Sample size calculations are often checked by grant review

committees, ethics boards, and journal reviewers to ensure that sample size is appropriate.

There are several statistical software packages, as well as many online calculators available to calculate sample size. For certain tests such as the two-sample *t*-test, there are very few formulae for sample size determination. However, for testing the equality of two proportions, or the risk difference, there are multiple formulae, which can yield very different results. More than 60 asymptotic tests and many exact tests have been considered for 2×2 tables [4]. Many researchers may not be aware that there are multiple ways to calculate sample size for 2×2 table designs, and even with this awareness, they may not know which approach and assumptions their software uses.

The aims of this report were to discuss for the risk difference:

1. assumptions that can be made in sample size calculations, including exact vs. asymptotic tests, standard deviation (SD) pooling, and continuity correction;

Conflict of interest: The authors declare no conflict of interests.

Funding: A.T.-P. was supported by the Australian National Health and Medical Research Council Grant 402764 to the Screening and Test Evaluation Program.

* Corresponding author. Tel.: + 61 2 9351 5424; fax: +61 2 9036 6247.

E-mail address: armando.teixeira-pinto@sydney.edu.au (A. Teixeira-Pinto).

What is new?

Key findings

- Many sample size formulae, with different underlying assumptions, for the risk difference exist.
- The choice of the formula can impact importantly on the required sample size, particularly when powering for large risk differences.
- There is variation in the sample size formulae and default settings implemented in statistical softwares.
- Calculation of sample size using the default settings across statistical packages can result in important differences in sample size.

What is the implication and what should change now?

- Better documentation of the sample size formulae and default settings is required in many statistical software packages.
- Improved reporting of the parameters used in sample size calculations, assumptions of the selected formula, and statistical software used to calculate the sample size is required.

2. various sample size formulae;
3. default assumptions used by statistical software packages SAS, version 9.2 (SAS Institute Inc, Cary, NC, USA) [5]; Stata, version 11 (StataCorp LP, College Station, TX, USA) [6]; PASS, version 11 (NCSS, LLC, Kaysville, UT, USA) [7]; G*Power, version 3.1.3 (<http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>) [8]; StatXact, version 10 (Cambridge, MA, USA) [9]; and various R libraries (R Foundation for Statistical Computing, Vienna, Austria) [10].

We show that the calculated sample size for the risk difference can differ substantially depending on what assumptions are made.

2. Methods

2.1. Assumptions

Consider a two-arm randomized trial, with equal allocation (1:1), where the statistic of primary interest is the difference in proportions using a two-sided test. (Note that the effect of unequal allocation is, in general, to increase the overall sample size, albeit only slightly in some cases [11].) Let p_0 be the proportion in the control population and p_1 be the proportion in the intervention population.

Although we focus on randomized trials, our results generalize to any design that uses the risk difference.

2.1.1. Asymptotic vs. exact tests

We consider two common approaches that can be used to calculate sample size and power: the asymptotic (or z -test) approach (formulae given later) and enumeration used for exact tests such as Fisher exact conditional test (conditional because the treatment margins are fixed for equal randomization) and Boschloo exact test. The idea behind enumeration is to list all possible 2×2 tables with given row and column margins. The probability of each possible table is then calculated using the hypergeometric distribution. We should note, however, that rather than the difference of proportions, Fisher constructed his test based on the null hypothesis that the odds ratio = 1 [12]. But in this article, we will consider this equivalent to testing the equality of two proportions.

Within the asymptotic approach, there are two assumptions underlying the differences in sample size formulae for the test of the difference in proportions. The first is the method used to estimate the SD (whether it is pooled or not) and the second is whether the continuity correction is used.

2.1.2. SD pooling

The SD of a proportion is a function of the population proportion. Therefore, in the sample size calculation, there arises the choice of whether to pool p_0 and p_1 . A pooled estimate averages the two proportions, whereas the unpooled does not. An argument for pooling in sample size calculations is that the test is performed under the assumption of the null hypothesis, that is, $p_0 = p_1$. Under this assumption, the best estimate of the SD is the pooled estimate. The commonly used conditional formula is a hybrid; see Ref. [13] for a derivation. See the book by Agresti [14] for more information on derivations and rationale.

2.1.3. Continuity correction

The Yates continuity correction was developed to compensate for the approximation of a discrete distribution (binomial) with a continuous one (normal). The continuity correction decreases the value of the chi-square statistic, increases the P -value, and decreases power for a set sample size. The continuity corrected sample size gives a good approximation of the required sample size for Fisher exact test [15], which is sometimes advocated when sample sizes are small.

2.1.4. The arcsine transform

The arcsine transform has been proposed to stabilize the variance and normalize proportions and is sometimes used in sample size calculations and analyses for proportions [16]. We mention this transform because some software packages have implemented it as an option; however, we do not investigate this transform further.

Download English Version:

<https://daneshyari.com/en/article/10513860>

Download Persian Version:

<https://daneshyari.com/article/10513860>

[Daneshyari.com](https://daneshyari.com)