ELSEVIER

Journal of Clinical Epidemiology 108 (2019) 140–146

Journal of Clinical Epidemiology

LETTERS TO THE EDITOR

Sample size calculations need to be adequate and parsimonious

Although the article of Flahault et al. [1] and the letter of Chu and Cole [2] have been published long ago, we would like to present our perspective on the topics they discussed at the time. Flahault et al. [1] have the merit of having introduced the exact sample size (s-size) calculation for binomial tests in trials evaluating diagnostic tests; Chu and Cole [2] proposed choosing the s-size (n_2) for which the power is always more than the prefixed threshold rather than the smaller s-size ($n1$) calculated when the power first exceeds the prefixed threshold. We would like to point out that Cesana et al. [3] also showed this s-size calculation for a statistical approach that combines the power of the statistical test with the power of Clopper-Pearson's confidence interval (CI), and that it has long been implemented in both s-size calculation statistical software such as PASS [4] and general statistical software such as SAS [5]. The saw-toothed behavior of the power (1-β) and significance (α) functions of discrete distributions [3,6] is widely known (see Table 1), and SAS [5] includes graphics that are similar to those of Chu and Cole's Figure 1 [2], which requires a more sophisticated SAS code than the one reported [2]. Chu and Cole's proposal [2] has been supported by Julious and Campbell [7], who believed that PASS [4] was wrong as the calculated s-size corresponds to n_1, but we have to disagree as "the number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed" [8] and increasing an s-size increase by "as much as 34% higher than n_1" [2] is not justified by reasonable statistical requirements. Khan et al. [8] suggested accepting an "α-level or power, not exactly equal to conventionally accepted levels" to obtain a material s-size "reduction, particularly for studies with say <50 patients." It is also necessary to avoid an overlarge increase in s-size due to an excessive need for precision (as has been shown by Cesana and Antonelli) [9], and the very desirable attempt to combine statistical and clinical significance must be considered critically because of such an increase [10] A moderate increase in s-size may be justified in a "superiority trial" or to achieve a satisfactory probability of obtaining both a statistically significant result and the required precision such as the width of the CI [5,9]. Although it is widely acknowledged that exact statistical methods should generally be preferred (particularly when exact *P*-values can be relatively easily

Table 1
π_0 π_1 N (c) Power (1-β)

π_0	π_1	N (c)	Power (1-β)	Alpha (α)
0.87	0.98	[a]40 (39)	0.8095	0.0266
0.87	0.98	41 (40)	0.8023	0.0236
0.87	0.98	42 (41)	0.7950	0.0210
0.87	0.98	43 (42)	0.7876	0.0186
0.87	0.98	44 (43)	0.7803	0.0165
0.87	0.98	45 (44)	0.7729	0.0147
0.87	0.98	46 (45)	0.7655	0.0130
0.87	0.98	47 (45)	0.9323	0.0462
0.87	0.98	48 (46)	0.9288	0.0417
0.87	0.98	49 (47)	0.9252	0.0376
0.87	0.98	50 (48)	0.9216	0.0339
0.87	0.98	51 (49)	0.9179	0.0305
0.87	0.98	52 (50)	0.9141	0.0275
0.87	0.98	53 (51)	0.9102	0.0247
0.87	0.98	54 (52)	0.9063	0.0222
0.87	0.98	55 (53)	0.9023	0.0200
0.87	0.98	56 (54)	0.8982	0.0180
0.87	0.98	57 (55)	0.8941	0.0161
0.87	0.98	58 (55)	0.9711	0.0464
0.87	0.98	59 (56)	0.9695	0.0423
0.87	0.98	60 (57)	0.9678	0.0385
0.87	0.98	61 (58)	0.9661	0.0350
0.87	0.98	62 (59)	0.9643	0.0318
0.87	0.98	63 (60)	0.9625	0.0289
0.87	0.98	64 (61)	0.9606	0.0262
0.87	0.98	65 (62)	0.9586	0.0238
0.87	0.98	66 (63)	0.9566	0.0215
0.87	0.98	67 (64)	0.9546	0.0195
0.87	0.98	68 (64)	0.9882	0.0490
0.87	0.98	69 (65)	0.9875	0.0449
0.87	0.98	70 (66)	0.9868	0.0412

π_0 and π_1 are the probabilities to be compared; N is the sample size calculated for a nominal significance value of 0.05, (c) is the cut-off value, power and alpha are the power and statistical significance functions of the exact binomial test.

[a] Indicates n_1 and $ indicates n_2. It should be noted that, from $n_2 = 47$, n_2 is not only greater than the required expected threshold of 0.80 but also that of 0.85. The power function would also be greater than 0.90, if there were not two values (56 and 57) whose corresponding power values are slightly lower than 0.90 (0.8982 and 0.8941, which would actually be rounded up to 0.90).

Conflict of interest: The authors have no conflict of interest to declare.

computed), in the case of biomedical research involving human beings, it is crucial to satisfy the ethical requirements of ensuring that the s-size is strictly related to the aims of the study and therefore appropriately calculated in the most parsimonious way. Finally, absolutely justified approximated procedures generally requiring a smaller s-size may be preferred, given the well-known conservative nature of the binomial test and the unsatisfactory coverage of Clopper-Pearson's CI.

Bruno Mario Cesana*
Department of Clinical Sciences and Community Health
Unit of Medical Statistics
Biometry and Bioinformatics "Giulio A. Maccacaro"
Faculty of Medicine and Surgery
University of Milan
Milan, Italy

Paolo Antonelli
Retired professor of Medical Statistics
University of Brescia
Italy
E-mail address: apaolo52@gmail.com (P. Antonelli)
*Corresponding author. Department of Clinical Sciences and Community Health, Unit of Medical Statistics, Biometry and Bioinformatics "Giulio A. Maccacaro", Faculty of Medicine and Surgery, University of Milan, Milan, Italy. Tel.: +39 0250320854.
E-mail address: bruno.cesana@guest.unimi.it (B.M. Cesana)

References

[1] Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J Clin Epidemiol 2005;58:859−62.
[2] Chu H, Cole SR. Sample size calculation using exact methods in diagnostic test studies issued on the. J Clin Epidemiol 2007;60:1201−2.
[3] Cesana BM, Reina G, Marubini E. Sample size for testing a proportion in clinical trials: a "two-step" procedure combining power and confidence interval expected width. Am Stat 2001;55:28892.
[4] Hintze J. PASS 13. Chapter 120 single-stage phase II clinical trials. Kaysville, Utah, USA: NCSS, LLC; 2014: Available at www.ncss.com.
[5] SAS Institute Inc. SAS/STAT9.2 user's guide. 2nd ed. Cary, NC: SAS Institute Inc.; 2009: Example 67.2 The Sawtooth Power Function in Proportion Analyses.
[6] Chernick MR, Liu CY. The saw-toothed behavior of power versus sample size and software solutions: single binomial proportion using exact methods. Am Stat 2002;56:149−55.
[7] Julious SA, Campbell MJ. Tutorial in biostatistics: sample sizes for parallel group clinical trials with binary data. Stat Med 2012;31:2904−36.
[8] Khan I, Sarker S-J, Hackshaw A. Smaller sample sizes for phase II trials based on exact tests with actual error rates by trading-off their nominal levels of significance and power. Br J Cancer 2012;107:1801−9.
[9] Cesana BM, Antonelli P. A new approach to sample size calculations for the power of testing and estimating population means of Gaussian distributed variables. Biomed Stat Clin Epidemiol 2010;4:67−78.
[10] Cesana BM, Antonelli P. Sample size calculations in clinical research should also be based on ethical principles. Trials 2016;17:149.

https://doi.org/10.1016/j.jclinepi.2018.11.018

Conflict of interest is not associated with positive conclusions in toothpaste trials: a systematic survey

Letter to the Editor

Conflict of interest (COI) is defined as divergence between individual's private interests and professional obligations—whether professional actions or decisions are motivated by personal gain, such as direct financial relationships, academic advancement, clinical revenue streams, or community standing [1,2]. In a recent review, 70% of randomized controlled trial (RCT) protocols reported a contract on publication rights between industry and academic investigators; and in 86% of these, industry had the authority to disapprove or review articles before journal publication [3]. Many television commercials abound related to dentin hypersensitivity (DH) toothpastes, prompting our interest.

As part of an ongoing systematic review and network meta-analysis of desensitizing toothpaste trials to treat DH (Prospero #CRD42018086815), we considered this potential relationship by examining study characteristics, directionality of conclusions, and industry funding associated with COI. We developed search strategies and conducted electronic searches up to February 2018 in Medline, Embase, Cochrane Reviews, CENTRAL, ProQuest, Clinical Trials, and the WHO International Clinical Trials Registry Platform. Pairs of independent reviewers screened titles/abstracts, selected full texts, and extracted data. We constructed multivariable logistic regression models for COI and positive conclusions. We included the following variables for the COI model: sample size, percentage of females enrolled in the study, year of publication, loss to follow-up, number or intervention arms, time of follow, region, and conclusion. In the model of positive conclusion, we included the same variables, additionally including reported funding.

We included 121 RCTs; the majority (70, 58%) was judged as having COI and reported a positive conclusion (82, 68%). Although the large majority of trials were funded through dental industries (70, 58%) and government (8, 7%), many (43, 36%) were unfunded or did not report funding.

In our multivariable model, studies with COI had larger sample sizes and more females enrolled (Table 1). In addition, similar to other reviews, RCTs with larger samples were less

[1] Present address: Department of Pediatric Dentristry and Orthodontics, Federal University of Minas Gerais−Avenida Presidente Antônio Carlos 6627, Pampulha, Belo Horizonte, Brazil Phone: + 55 31 3049 2398 E-mail: carolcm10@hotmail.com

Funding: C.C.M. received a post doc fellowship by CAPES (Coordination for the Improvement of Higher Educational Personnnel, Ministry of Education, Brazil, process #88881.119166/2016-01), during her post doctoral studies at McMaster University. J.J.R. is supported by a PhD training award from the NCMIC Foundation. FAPEMIG (Research Foundation of the State of Minas Gerais, Brazil). Funding sources had no involvement in the research (preparation of the article, data collection, analysis, or interpretation of data).

Conflict of interest: The authors have no conflict of interest to report.