European Journal of Internal Medicine xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Original Article

The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed

Paul H. Lee [a,*], Andy C.Y. Tse [b]

[a] School of Nursing, Hong Kong Polytechnic University, Hong kong
[b] Department of Health and Physical Education, The Education University of Hong Kong, Hong kong

ARTICLE INFO

Article history:
Received 17 February 2016
Received in revised form 29 September 2016
Accepted 10 October 2016
Available online xxxx

Keywords:
Crossover
Factorial
Protocol
Sample size
Trials

ABSTRACT

Background: There are limited data on the quality of reporting of information essential for replication of the calculation as well as the accuracy of the sample size calculation. We examine the current quality of reporting of the sample size calculation in randomized controlled trials (RCTs) published in PubMed and to examine the variation in reporting across study design, study characteristics, and journal impact factor. We also reviewed the targeted sample size reported in trial registries.

Methods: We reviewed and analyzed all RCTs published in December 2014 with journals indexed in PubMed. The 2014 Impact Factors for the journals were used as proxies for their quality.

Results: Of the 451 analyzed papers, 58.1% reported an *a priori* sample size calculation. Nearly all papers provided the level of significance (97.7%) and desired power (96.6%), and most of the papers reported the minimum clinically important effect size (73.3%). The median (inter-quartile range) of the percentage difference of the reported and calculated sample size calculation was 0.0% (IQR −4.6%;3.0%). The accuracy of the reported sample size was better for studies published in journals that endorsed the CONSORT statement and journals with an impact factor. A total of 98 papers had provided targeted sample size on trial registries and about two–third of these papers (n = 62) reported sample size calculation, but only 25 (40.3%) had no discrepancy with the reported number in the trial registries.

Conclusions: The reporting of the sample size calculation in RCTs published in PubMed-indexed journals and trial registries were poor. The CONSORT statement should be more widely endorsed.

1. Introduction

In presenting the results of randomized controlled trials (RCTs), an *a priori* sample size calculation should be reported because an RCT with too small sample size lacks statistical power and will lead to inconclusive results. In contrast, a sample size that is too large may lead to ethical issues, such as unnecessary exposures to potential harm [1]. The Consolidated Standards Of Reporting Trials (CONSORT) statement recommends reporting of how the sample size was determined [2]. In addition, the extensions of the CONSORT statement that were available for reporting the sample size calculation in non-inferiority trials [3] and cluster trials [4] suggested that the margin of non-inferiority and design effect should be reported, respectively.

Previous reviews showed that RCTs published in CONSORT-endorsed journals were more likely to report their sample size calculation [5–7]. This finding was not surprising because the authors were obliged to report this information while submitting their papers to these journals. Many researchers have conducted reviews regarding the compliance with the CONSORT statement in different specialty fields, but <10% of these reviews investigated the compliance in the reporting sample size estimation [8]. Most importantly, few reviews assessed the quality of reporting of information that are essential for replication of the calculation (such as desired level of power and expected effect size of the treatment) and examined the accuracy of the sample size calculation. We know that most two-arm parallel group RCTs published in six leading general medical journals reported sample size calculations, and that the provision of information and accuracy of the calculated sample sizes were of acceptable level [5]. However, the quality of RCTs in other types of study designs and of those published in other journals, and the associations between quality of reporting and study characteristics, remain unknown.

Here, we reviewed and analyzed all RCTs published in December 2014 in journals indexed in PubMed. Because previous results showed that the quality of the journals was associated with the quality of reporting, we also examined whether the study type, study characteristics (drug trial *versus* non-drug trial, funding source), journal type (general medical *versus* specialty), endorsement of CONSORT guidelines, and impact factor of the journals were associated with quality of

* Corresponding author at: School of Nursing, GH527, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
E-mail address: paul.h.lee@polyu.edu.hk (P.H. Lee).

http://dx.doi.org/10.1016/j.ejim.2016.10.008
0953-6205/

Please cite this article as: Lee PH, Tse ACY, The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed, Eur J Intern Med (2016), http://dx.doi.org/10.1016/j.ejim.2016.10.008

reporting of sample size calculation. We hypothesized that studies which received funding, were published in journals that endorse the CONSORT guidelines, and were published in high quality journals would have better quality in reporting sample size calculation.

In addition to the sample size calculation provided in the published papers, we also reviewed the targeted sample size reported in trial registries, for example ClinicalTrials.gov or International Standard Randomized Controlled Trial Number (ISRCTN). The discrepancies between the reported sample size in trial registries and in the papers, as well as the quality of the reported sample size in trial registries, would also be evaluated.

2. Methods

2.1. Search strategy

We used the Cochrane highly sensitive search strategy (phase 1) [9] to search PubMed for papers reporting randomized controlled trials published in December 2014 and indexed by 30 April 2015. Two reviewers (PHL and ACYT) independently screened the abstracts and full texts to determine their eligibility. All online-only materials relevant to the sample size calculation were downloaded and analyzed.

2.2. Inclusion and exclusion criteria

We adopted inclusion criteria similar to those reviewing PubMed-indexed papers published in 2000 and 2006 [10,11]. An article had to satisfy the following criteria to be included in the analysis: 1) the study subjects were humans, 2) the trial had to involve health-care interventions, 3) the participants had to be randomly allocated into at least two study groups with different interventions, and 4) the article had to be published in English. We included all trial types, including parallel group, crossover, clustered, and factorial designs. The exclusion criteria were as follows: 1) a cost-effectiveness, diagnostic, or methodological study; 2) a secondary publication; and 3) an early phase or pilot trial in which the sample size was not calculated based on hypothesis testing.

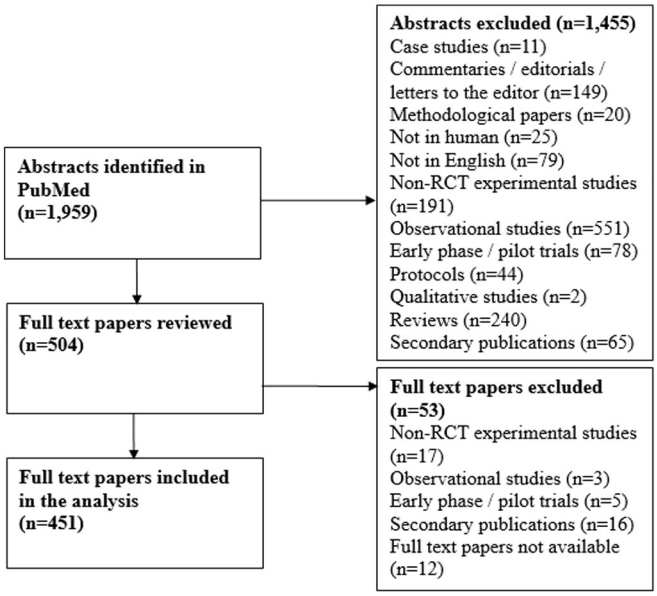

Fig. 1. Paper screening procedure.

2.3. Search results

Fig. 1 shows the search results. A total of 1959 abstracts were identified by the search strategy, 504 full text papers were reviewed, and 451 papers were included in the analysis.

2.4. Data extraction

Information regarding the sample size calculation were extracted, including 1) the type of study (parallel group: each participant was randomly assigned to one of the study groups; crossover: each participant was required to undergo all study groups in a random sequence; and others), 2) the 2014 Impact Factor of the journal (as a proxy for the quality of the journal [12,13]), 3) the endorsement of the CONSORT guidelines from the author guidelines of the journal, 4) the specialty of the journal, 5) whether the trial was drug-related, 6) the source of the funding (institutional, industrial, both, or none), 7) the sample size of the analyzed dataset, 8) the *a priori* calculated sample size of the study (if any), and if yes, then 9) the level of significance adopted, 10) the desired power, and 11) the expected effect size of the treatment (the mean and SD for continuous outcomes, the proportions of all groups for binary outcome, or the non-inferiority margin for non-inferiority trials).

The trial registration numbers for all papers were obtained from the main text. Only those with *a priori* calculated sample size were collected whilst we excluded those with actual sample size reported. All target sample sizes reported in trial registries were then multiplied by the estimated attrition rates reported in the corresponding papers.

2.5. Sample size calculation

For papers that reported adequate information for the sample size calculation (the level of significance adopted, the desired power, and the expected effect size of the treatment for superiority trials or the non-inferiority margin for non-inferiority trials), we calculated the sample size required to achieve the reported desired level of power and level of significance. If the information regarding the tail type of the statistical test used was missing, then all superiority trials were assumed to use two-tailed tests and all non-inferiority trials were assumed to use one-tailed tests. The formulas used for sample size calculation can be found in Supplemental Material 1: eMethods.

Three comparisons of sample sizes were made. First, we compared the differences between the reported and calculated sample sizes in the analyzed papers. For papers that provided a targeted sample size in a trial registry, we additionally made two more comparisons as follows. We compared the differences between the targeted sample sizes reported in the trial registries and that reported in the analyzed papers, as well as the differences between the targeted sample sizes reported in the trial registries and the calculated sample sizes in the papers. Instead of using percentage difference of these two sample sizes [5], the percentage difference of the square root of these two sample sizes (that is, $\frac{\sqrt{n_1}-\sqrt{n_2}}{\sqrt{n_2}}$, where n_1 and n_2 are the reported and calculated sample sizes in the analyzed paper respectively) was used, as this equals the percentage difference of the true effect size which could be detected (compared with the reported effect size assumption) under the reported levels of significance and power. To examine the absolute error, the absolute percentage difference of the square root of these two sample sizes (that is, $\frac{|\sqrt{n_1}-\sqrt{n_2}|}{\sqrt{n_2}}$) was also used.

2.6. Statistical analysis

The Impact Factor was grouped into five categories (not indexed in Journal Citation Report, 0.001-3, 3.001-5, 5.001-10, and >10). The journal type was grouped into general medical or specialty. The associations between the study type, drug trial, funding source, journal type,

Please cite this article as: Lee PH, Tse ACY, The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed, Eur J Intern Med (2016), http://dx.doi.org/10.1016/j.ejim.2016.10.008