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.

© 2016 European Federation of Internal Medicine. Published by Elsevier B.V. All rights reserved.

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/© 2016 European Federation of Internal Medicine. Published by Elsevier B.V. All rights reserved.

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.

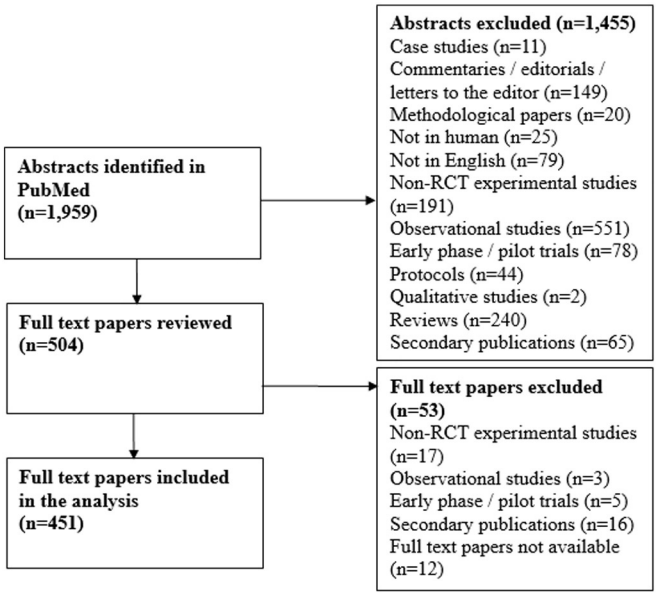

Fig. 1. Paper screening procedure.

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

endorsement of CONSORT guidelines, and impact factor categories on the provision of an *a priori* sample size calculation and information for sample size calculation (level of significance, power, and effect size) were examined using a chi-square test. The association of the above factors with the percentage difference and absolute percentage difference was assessed using multiple linear regression. Due to the small number of studies providing target sample size in trial registries (n = 98), Mann–Whitney *U* test was used (instead of regression) to examine the association between the above factors with the percentage difference between trial registries-reported sample size, paper-reported sample size, and the calculated sample size. To evaluate the difference between the quality of sample size calculation reported in trial registries and in papers, the within-paper difference was examined using scatter plots, Bland-Altman plots, and Wilcoxon signed-ranked test. R 3.2.0 was used to conduct the statistical analysis.

3. Results

Table 1 shows the characteristics of the 451 papers analyzed. The median achieved sample size of these studies was 96 (inter-quartile range 48;254). The median sample sizes for parallel studies and crossover studies were 43 (IQR 21.5;117) per group and 24 (IQR 15;48), respectively. Most of the studies were parallel group design (391, 86.7%) and 31 of the studies (6.9%) were crossover trials. Less than half of the trials (191, 42.4%) were drug interventions, 124 (27.5%) received industrial funding, and 255 (56.5%) received institutional funding. A total of 48 papers (10.6%) were published in journals not indexed in the Journal Citation Report, and 150 (33.3%), 99 (22.0%), 91 (20.2%), and 63 (14.0%) papers were published in journals with impact factors of 0.001-3, 3.001-5, 5.001-10, and >10, respectively. A total of 46 papers (10.2%) published in general medical journals, and 250 (55.4%) published in a journal that endorsed the CONSORT guidelines. The study type was not associated with the reporting of the sample size calculation ($p > 0.05$), but institutional supported studies ($p < 0.01$), papers published in journals that endorsed the CONSORT guidelines ($p < 0.05$), and journals with higher impact factors were more likely to report the sample size calculation ($p < 0.001$).

Table 2 shows the characteristics of the 262 papers that reported an *a priori* sample size calculation. Most of the studies were superiority trials (247, 94.3%). The level of significance was reported in 256 papers (97.7%), and 232 of the studies (90.6%) used a 0.05 level. The desired power was reported in 253 papers (96.6%), and 163 (64.4%) and 55 (21.7%) of the papers used 80% and 90%, respectively. Papers published in journals with higher impact factor were more likely to report the expected effect size ($p = 0.01$) and more likely to provide all necessary information to calculate the sample size required ($p = 0.03$).

Supplemental Fig. 1 shows the scatter plot of the reported *versus* calculated square roots of the sample sizes in the 184 papers that provided all necessary information to calculate the sample size required.

Table 1
Characteristics of the included studies (n = 451).

	A priori sample size calculation	
	Not reported	Reported
Achieved sample size*** (median, IQR)	73 (37;154)	120 (60;301)
Study type (frequency, %)		
Parallel group	163 (41.7%)	228 (58.3%)
Crossover	16 (51.6%)	15 (48.4%)
Others#	10 (34.5%)	19 (65.5%)
Drug trial (frequency, %)		
Yes	80 (41.9%)	111 (58.1%)
No	109 (41.9%)	151 (58.1%)
Industrial funding (frequency, %)		
Yes	45 (36.3%)	79 (63.7%)
No	144 (44.0%)	183 (56.0%)
Institutional funding* (frequency, %)		
Yes	94 (36.9%)	161 (63.1%)
No	95 (48.5%)	101 (51.5%)
Journal type** (frequency, %)		
General medical	11 (23.9%)	35 (76.1%)
Specialty	178 (44.0%)	227 (56.0%)
CONSORT guidelines*** (frequency, %)		
Endorsed	80 (32.0%)	170 (68.0%)
Not endorsed	109 (54.2%)	92 (45.8%)
Impact factor 2014*** (frequency, %)		
Not indexed	30 (62.5%)	18 (37.5%)
0.001-3	75 (50.0%)	75 (49.3%)
3.001-5	42 (42.4%)	57 (57.0%)
5.001-10	28 (30.8%)	63 (68.5%)
>10	14 (22.2%)	49 (77.8%)
Total	189 (41.9%)	262 (58.1%)

Cluster (n = 7), factorial (n = 8), intra-individual (n = 12), cluster crossover (n = 1), cluster factorial (n = 1).
* Significant at 5% level.
** Significant at 1% level.
*** Significant at 0.1% level.

Table 2
Reporting of *a priori* sample size calculation by study type and journal impact factor (n = 262).

	Sample size (median, IQR)	Information for sample size calculation (frequency, %)			All essential information provided (frequency, %)
		Level of significance	Power	Effect size	
Study type					
Parallel	126 (70–307)***	223 (93.3%)	220 (96.5%)	171 (75.0%)	162 (71.1%)
Crossover	24 (15–155)	14 (93.3%)	14 (93.3%)	7 (46.7%)	7 (46.7%)
Others#	364 (31–3947)	19 (100.0%)	19 (100.0%)	14 (73.7%)	14 (73.7%)
Drug trial					
Yes	141 (67–338)	109 (98.2%)	75 (94.9%)	83 (74.8%)	78 (70.3%)
No	101 (50–280)	148 (97.4%)	179 (97.3%)	110 (71.9%)	106 (69.7%)
Industrial funding					
Yes	163 (75–507)**	76 (96.2%)	75 (94.9%)	7 (46.7%)	54 (68.4%)
No	100 (50–270)	181 (98.4%)	179 (97.3%)	171 (75.0%)	130 (70.7%)
Institutional funding					
Yes	134 (61–353)*	161 (99.4%)*	160 (98.8%)*	128 (78.5%)*	125 (77.2%)**
No	92 (55–238)	96 (95.0%)	94 (93.1%)	65 (64.4%)	59 (58.4%)
Journal type					
General medical	364 (237–1195)***	35 (100.0%)	34 (97.1%)	28 (80.0%)	27 (77.1%)
Specialty	94 (51–254)	222 (97.4%)	220 (96.5%)	165 (72.1%)	157 (68.9%)
CONSORT guidelines					
Endorsed	142 (67–339)**	168 (98.2%)	166 (97.1%)	134 (78.4%)**	128 (74.9%)*
Not endorsed	86 (48–184)	88 (96.7%)	87 (95.6%)	58 (63.0%)	55 (60.4%)
Impact factor 2014					
Not indexed	82 (39–121)***	17 (94.4%)	16 (88.9%)	9 (47.4%)**	8 (44.4%)*
0.001-3	73 (44–-139)	74 (98.7%)	74 (98.7%)	53 (70.7%)	51 (68.0%)
3.001-5	97 (61–239)	55 (96.5%)	54 (94.7%)	37 (64.9%)	36 (63.2%)
5.001-10	120 (62–389)	62 (96.9%)	62 (96.9%)	54 (85.7%)	51 (79.7%)
>10	343 (197–891)	49 (100.0%)	48 (98.0%)	39 (79.6%)	38 (77.6%)
Total	120 (60–302)	257 (97.7%)	254 (96.6%)	192 (73.3%)	184 (70.0%)

Cluster (n = 6), factorial (n = 5), intra-individual (n = 6), cluster crossover (n = 1), cluster factorial (n = 1).
* Significant at 5% level.
** Significant at 1%level.
*** Significant at 0.1% level.

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

ARTICLE IN PRESS

Supplemental Fig. 2 shows the Bland-Altman plot of the reported *versus* calculated square roots of the sample sizes in these 184 papers. Both plots show that the reported and calculated sample sizes were very similar; the median values (inter-quartile range) of the percentage difference and absolute percentage difference were 0.0% (−4.6%;3.0%) and 3.5% (0.5%;20.9%), respectively. The multiple linear regressions (Table 3) showed that papers published in journals that endorsed the CONSORT guidelines had a marginally-significant lower percentage difference of −12% (95% CI: −25%, 1%, $p = 0.06$), and papers published in journals indexed in the Journal Citation Report 2014 had a lower percentage difference (ranging from −25% to −22%) and absolute percentage difference (ranging from −32% to −22%) compared with papers not published in these journals.

Of the 451 papers included in the analysis, 17 (3.8%) of them were non-inferiority trials and the remaining were superiority trials. The non-inferiority trials on average had a larger sample size than the superiority trials, with a median (inter-quartile range) sample size of 291 (92;810). Most of the non-inferiority trials (n = 15, 88.2%) reported an *a priori* sample size calculation, and 12 (80.0%) provided all the necessary information to calculate the sample size required. The quality of the reported and calculated sample sizes were high; the median values (inter-quartile range) of the percentage difference and absolute percentage difference were 0.0% (−0.1%;9.0%) and 1.4% (0.0%;39.1%), respectively.

A total of 193 (42.8%) papers had registered on trial registries and 98 (50.8%) of them provided targeted sample size. Table 4 shows that only two–third of these papers (n = 62) reported sample size calculation. Studies supported by industrial funding were more likely to report sample size calculation ($p = 0.04$). Among papers that reported sample size calculation, only 25 (40.3%) had no discrepancy with the reported number in the trial registries, and 14 (22.6%) had a discrepancy of >30%.

Supplemental Figs. 3 and 4 show the scatter and Bland-Altman plots of the trial registry-reported *versus* paper-reported target sample size, and Table 5 (first two columns) shows that the percentage difference and absolute percentage difference of the target sample size reported in trial registry and that reported in paper. Out of the 63 studies that provided target sample size in both trial registry and paper, 25 (39.7%) of them were identical with the target sample size report in the paper. The percentage difference and absolute percentage difference were not associated with any of the study characteristics.

Supplemental Figs. 5 and 6 show the scatter and Bland-Altman plots of the trial registry-reported *versus* calculated square roots of the sample sizes, and Table 5 shows that the percentage difference and absolute percentage difference of the target sample size reported in trial registries and the calculated sample size. The median values (inter-quartile range) of the percentage difference and absolute percentage difference of the target sample size reported in trial registries and the calculated sample size were 0.3% (−8.1%;15.1%) and 9.3% (2.8%;37.8%), respectively, which were larger than the difference found between the target sample size reported in paper and the calculated sample size ($p = 0.005$ and 0.016, respectively). The percentage difference and absolute percentage difference were not associated with any of the study characteristics.

Table 3
Multiple linear regression on percentage difference and absolute percentage difference of the reported and calculated sample sizes (n = 184).

	Percentage difference (95% CI)	Absolute percentage difference (95% CI)
Achieved sample size	2.09×10^{-6} (-1.33×10^{-5}, 1.75×10^{-5})	-8.07×10^{-7} (-1.48×10^{-5}, 1.32×10^{-5})
Study type		
Parallel group	Ref	Ref
Crossover	0.10 (−0.19, 0.38)	0.08 (−0.18, 0.33)
Others#	−0.13 (−0.34, 0.09)	0.02 (−0.17, 0.22)
Drug trial		
Yes	−0.10 (−0.21, 0.02)	−0.08 (−0.18, 0.03)
No	Ref	Ref
Industrial funding		
Yes	−0.07 (−0.20, 0.07)	0.002 (−0.12, 0.12)
No	Ref	Ref
Institutional funding		
Yes	0.03 (−0.09, 0.16)	0.08 (−0.04, 0.19)
No	Ref	Ref
Journal type		
General medical	−0.01 (−0.20, 0.18)	0.03 (−0.14, 0.20)
Specialty	Ref	Ref
CONSORT guidelines		
Endorsed	−0.12 (−0.25, 0.01)	−0.05 (−0.17, 0.07)
Not endorsed	Ref	Ref
Impact factor 2014		
Not indexed	Ref	Ref
0.001-3	−0.23 (−0.50, 0.05)	−0.22 (−0.47, 0.03)
3.001-5	−0.24 (−0.52, 0.05)	−0.24 (−0.50, 0.02)
5.001-10	−0.25 (−0.53, 0.04)	−0.26 (−0.51, 0.004)
>10	−0.22 (−0.52, 0.08)	−0.32 (−0.59, −0.05)*

Cluster (n = 7), factorial (n = 8), intra-individual (n = 12), cluster crossover (n = 1), cluster factorial (n = 1).
* Significant at 5% level.
** Significant at 1% level.
*** Significant at 0.1% level.

Table 4
Reporting of *a priori* sample size calculation within studies with target sample size reported in trial registries (n = 98).

	A priori sample size calculation	
	Not reported in paper	Reported in paper
Achieved sample size*** (median, IQR)	80 (36;203)	121 (74;288)
Study type (frequency, %)		
Parallel group	35 (39.3%)	54 (60.7%)
Crossover	1 (20.0%)	4 (80.0%)
Others#	0 (0.0%)	4 (100.0%)
Drug trial (frequency, %)		
Yes	17 (32.7%)	35 (67.3%)
No	19 (41.3%)	27 (58.7%)
Industrial funding* (frequency, %)		
Yes	23 (31.1%)	51 (68.9%)
No	13 (54.2%)	11 (45.8%)
Institutional funding (frequency, %)		
Yes	11 (40.7%)	16 (59.3%)
No	25 (35.2%)	46 (64.8%)
Registered at ClinicalTrials.gov (frequency, %)		
Yes	11 (40.7%)	16 (59.3%)
No	25 (35.2%)	46 (64.8%)
Total	36 (36.7%)	62 (63.3%)

cluster (n = 2), factorial (n = 2).
* Significant at 5% level.
** Significant at 1% level.
*** Significant at 0.1% level.

4. Discussion

In this review and analysis of 451 RCT papers published in December 2014, more than half of the studies (58.1%) reported the sample size calculation, and only 40% of the studies reported all the information essential for replication. Institutionally-supported studies and papers published in journals that endorsed the CONSORT guidelines had better reporting, but the reporting of sample size calculation was still unacceptable as many essential information was missing. Compared with trials with ethics approval received in 1994–1995 that 34% reported all information [14], no improvement was observed. The study type was not associated with the quality of reporting. One possible explanation is that too few non-parallel studies were included to detect statistical significance. In the 31 crossover studies included in the analysis, only 15 (48.4%) reported the sample size calculation, 7 (22.6%) provided all the essential information for the replication of sample size calculation, and only 3 reported a sample size within 20% of the calculated one. These findings were in accordance with the previous reviews of RCTs published in PubMed-indexed journals, which found that only 14% of the crossover studies published in 2000 and 33% published in 2006

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

Table 5
Quality of target sample size reported in trial registries (n = 98).

	Report (registry) *vs* Report (paper) (n = 98)		Report (registry) *vs* Calculated (n = 45)		Report (paper) *vs* Calculated (n = 63)	
	Percentage difference (IQR)	Absolute percentage difference (IQR)	Percentage difference (IQR)	Absolute percentage difference (IQR)	Percentage difference (IQR)	Absolute percentage difference (IQR)
Drug trial						
Yes	0.00 (0.00;0.53)	0.06 (0.00;0.56)	0.01 (−0.05;0.45)	0.07 (0.03;0.35)	0.00 (−0.02;0.01)	0.02 (0.00;0.10)
No	0.03 (0.00;0.15)	0.06 (0.00;0.17)	0.001 (−0.12;0.15)	0.12 (0.02;0.35)	0.00†† (−0.12;0.01)	0.03† (0.00;0.15)
Industrial funding						
Yes	0.00 (0.00;0.01)	0.00 (0.00;0.18)	−0.03 (−0.09;0.002)	0.06 (0.01;0.40)	−0.01 (−0.09;0.00)	0.03 (0.003;0.13)
No	0.03 (0.00;0.21)	0.07 (0.00;0.25)	0.03 (−0.08;0.17)	0.10 (0.03;0.38)	0.03†† (−0.08;0.17)	0.02† (0.00;0.14)
Institutional funding						
Yes	0.03 (0.00;0.22)	0.07 (0.00;0.30)	0.002 (−0.10;0.15)	0.11 (0.04;0.41)	0.00†† (−0.11;0.01)	0.02† (0.001;0.14)
No	0.00 (0.00;0.5)	0.01 (0.00;0.09)	0.005 (−0.01;0.34)	0.05 (0.002;0.34)	0.00 (−0.01;0.03)	0.005 (0.00;0.05)
Registered at ClinicalTrials.gov						
Yes	0.03 (0.00;0.63)	0.07 (0.00;0.63)	0.04 (0.002;0.54)	0.05 (0.02;0.54)	0.001† (−0.00;0.07)	0.01 (0.001;0.11)
No	0.00 (0.00;0.15)	0.05 (0.00;0.22)	−0.01* (−0.13;0.07)	0.11 (0.05;0.30)	0.00 (−0.13;0.003)	0.02 (0.00;0.14)
Total	0.00 (0.00;0.18)	0.06 (0.00;0.22)	0.00 (−0.08;0.15)	0.09 (0.03;0.38)	0.00†† (−0.07;0.01)	0.02† (0.00;0.14)

Cluster (n = 2), factorial (n = 2).
* Significant between-group difference at 5% level.
** Significant between-group difference at 1% level.
*** Significant between-group difference at 0.1% level.
† Significant within-group difference (Report (registry) *vs* Calculated - Report (report) *vs* Calculated) at 5% level.
†† Significant within-group difference (Report (registry) *vs* Calculated - Report (report) *vs* Calculated) at 1% level.
††† Significant within-group difference (Report (registry) vs Calculated - Report (report) vs Calculated) at 0.1% level.

reported sample size calculations [10,11]. Although there was a clear trend towards improvement, the quality of reporting a sample size calculation for crossover studies needs amelioration. In addition, it is noted that non-inferiority trials had better reporting for sample size estimation, which may due to its frequent use in industrial studies where the methodology is rigorously reviewed.

We observed a trend of increasing sample sizes in the published RCTs. Reviews of the RCTs published in PubMed-indexed journals in 2000 and 2006 [10,11] found that the median sample sizes per group of the parallel design studies were 32 in 2000 and 36 in 2006 and that the median values of the crossover studies were 15 and 20, in the respective years. Here, we found that the median sample sizes for parallel studies and crossover studies in 2014 were 43 (IQR 21.5;117) per group and 24 (IQR 15;48), respectively. Such an increasing trend in the sample sizes may be due to researchers' awareness of the importance of sample size over time, and the number of samples recruited increased over time. However, it is trivial that just having a larger sample size only will not affect the overall quality of the report, which is necessary for publishing in high quality journals. This is supported by our results, showing that papers published in higher impact journals were also better at reporting their sample size calculation, both in terms of the information provided and the accuracy of the estimated sample size.

A review of sample size calculation reporting in two parallel groups superiority RCTs published in six leading general medical journals showed that the quality of reporting was high. Only 5% of the papers failed to report sample size calculation, and 78% included all the required information for the sample size calculation [5]. Furthermore, the reported sample sizes were accurate, with 65% were within 5% of that calculated by the assessors. Two similar reviews in leading medical journals showed similar results, with 70%–80% of their published RCTs reporting the sample size calculation [15,16]. These findings agreed with our data, suggesting that the RCTs published in higher impact journals were more likely to provide the necessary information for sample size calculation (especially for the estimated effect size of the treatment). However, recent reviews of RCTs published in top journals in ophthalmic surgery, plastic surgery, herbal intervention, endocrinology, and urology discovered that the prevalence of the reporting of the sample size calculation was only 10% to 40% [17–21], closed to that in research protocols submitted for ethical approval (48%) [22]. Therefore, the guidelines in these medical journals should be modified to reinforce the importance of reporting the sample size calculation in RCTs. Furthermore, the editors of these medical journals should seek statistical advice from peer reviewers to evaluate the accuracy of the sample size calculation for any submitted RCT manuscripts. Indeed, it is strongly recommended to stick to CONSORT principles even when not reporting an RCT. For example, the study objectives (2b) and the eligibility criteria for participants (4a) should be clearly stated.

While the comparison of reported and calculated sample size was not completely novel, this is an important study evaluating the quality of target sample size reported in trial registries among general medical papers. We found that nearly 60% of the reviewed studies changed the *a priori* sample size in the papers, and this is comparable with findings in discrepancies across proposals submitted for ethical approval (53% [14] to 66% [23]). The results were not surprising as it was also found that in 32% of the trials in registries had changed the stated primary outcome [24]. This disagreement between the targeted sample size reported in the trial registry and the *a priori* sample size reported in the paper evidenced an unreported change in plan or enrollment problem that should be specified in the report. However, none of these changes were reported in our reviewed papers.

Since the essential information for sample size calculation was not provided in all trial registries, we used an indirect approach to compare the quality between trial registry-reported and paper-reported sample size calculation, which is the percentage difference with the calculated sample size. We found that the quality of paper-reported sample size calculation was higher, which suggested that the reliability of target sample size reported in trial registries remains questionable. Person-in-charge governing the trial registries should enforce explicit reporting guidelines regarding target sample size.

Inappropriate calculation of the sample size calculation will downgrade the quality of an RCT. Some researchers argued that underpowered trials are in fact useful if their methodological quality is high and if the study design is rigorous [25–27]. Of course, the inappropriate reporting of the sample size estimation indirectly reflects that there may be other major flaws in the design, methodology, or data analysis in the trial, which is a very serious problem. A follow-up study of one thousand trials found that 10% were discontinued due to poor recruitment, as only 40% of the targeted sample size could be recruited [28]. This finding led us to postulate that whether the most commonly (>85% of RCTs included in this study) adopted desired power levels,

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

ARTICLE IN PRESS

80% or 90%, should always be targeted. In some scenarios, it may not be feasible to recruit this number of samples, and researchers should take this into account during the planning stages.

This study had several limitations. Of the 504 papers reviewed, 12 were excluded as the full text papers were not accessible. Hence, the quality of these papers could not be assessed. We believe this limitation is minor, as these papers were only equal to 2.6% of the papers analyzed. We used only one database (PubMed) and those RCTs published in journals not indexed in PubMed were excluded. In the current analysis, study types were classified as parallel group, crossover, and others, which included cluster, factorial, and intra-individual designs. The quality of reporting the sample size calculation for all these designs were not assessed separately, as there were too few studies of these designs. A review of the clustered RCTs found that approximately half of the studies reported sample size calculation and two-thirds accounted for the cluster effect [29]. However, very few studies examined the quality of reporting in other study designs, and these therefore warrant further investigation.

In conclusion, the reporting of the sample size calculation in RCTs published in PubMed-indexed journals was poor, and journals should endorse and enforce explicit reporting guidelines regarding sample size calculations.

Supplementary data to this article can be found online at doi:10.1016/j.ejim.2016.10.008.

Author contributions

PHL designed the study, conducted the systematic review, analyzed the data, and drafted the manuscript. ACYT conducted the systematic review and critically reviewed the manuscript.

Conflict of interest

No conflicts of interest to declare.

Funding/support

None.

Acknowledgements

We thank Dr. An-Wen Chan (University of Toronto) for his helpful comments.

References

[1] Wittes J. Sample size calculations for randomized controlled trials. Epidemiol Rev 2002;24(1):39–53.
[2] Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Lancet 2010;375(9721):1136.
[3] Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006;295:1152–60.
[4] Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ 2004;328:702–8.
[5] Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ 2009;338:b1732.
[6] Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 2012;1:60.
[7] Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev 2012;11, MR000030.
[8] Dechartres A, Charles P, Hopewell S, Ravaud P, Altman DG. Reviews assessing the quality or the reporting of randomized controlled trials are increasing over time but raised questions about how quality is assessed. J Clin Epidemiol 2011;64:136–44.
[9] Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. Int J Epidemiol 2002;31:150–3.
[10] Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 2005;365:1159–62.
[11] Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ 2010;340:c273.
[12] Garfield E. Journal impact factor: a brief review. CMAJ 1999;161:979–80.
[13] Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. JAMA 2002;287:2805–8.
[14] Chan AW, Hrobjartsson A, Jorgensen K, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ 2008;337:a2299.
[15] Mills E, Wu P, Gagnier J, Devereaux PJ. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. Contemp Clin Trials 2005;26:480–7.
[16] Mills E, Wu P, Gagnier J, Heels-Ansdell D, Montori VM. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. J Clin Epidemiol 2005;58:662–7.
[17] Agha RA, Camm CF, Doganay E, Edison E, Siddiqui MRS, Orgill DP. Randomised controlled trials in plastic surgery: a systematic review of reporting quality. Eur J Plast Surg 2014;37:55–62.
[18] Pratoomsoot C, Sruamsiri R, Dilokthornsakul P, Chaiyakunapruk N. Quality of reporting of randomised controlled trials of herbal interventions in ASEAN plus six countries: a systematic review. PLoS One 2015;10(1), e108681.
[19] Rios LP, Odueyungbo A, Moitri MO, Rahman MO, Thabane L. Quality of reporting of randomized controlled trials in general endocrinology literature. J Clin Endocrinol Metab 2008;93:3810–6.
[20] Scales Jr CD, Norris RD, Preminger GM, Vieweg J, Peterson BL, Dahm P. Evaluating the evidence: statistical methods in randomized controlled trials in the urological literature. J Urol 2008;180(4):1463–7.
[21] Yao AC, Khajuria A, Camm CF, Edison E, Agha R. The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. Eye 2014;28:1341–9.
[22] Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. BMJ 2013;346:f3115.
[23] Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. J Eval Clin Pract 2002;8(3):353–9.
[24] Ramagopalan S, Skingsley AP, Handunnetthi L, Klingel M, Magnus D, Pakpoor J, et al. Prevalence of primary outcome changes in clinical trials registered on ClinicalTrials.gov: a cross-sectional study. F1000 Res 2014;3:77.
[25] Edwards SJL, Lilford RJ, Braunholtz D, Jackson J. Why "underpowered" trials are not necessarily unethical. Lancet 1997;350:804–7.
[26] Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet 2005;365:1348–53.
[27] Bacchetti P, McCulloch C, Segal MR. Being 'underpowered' does not make a study unethical. Stat Med 2012;31(29):4138–9.
[28] Kasenda B, Von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA 2014;311(10):1045–51.
[29] Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, et al. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000–8. BMJ 2011;343:d5886.

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