

Adequate reporting of the sample size calculation in surgical randomized controlled trials

Benjamin Speich, PhD^{1,2}

¹Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and
Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom

²Basel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University
Hospital Basel, University Basel, Basel, Switzerland

Benjamin Speich, PhD
Centre for Statistics in Medicine
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
University of Oxford
Oxford OX3 7LD
United Kingdom
Phone: +44 1865 737904
Email: benjamin.speich@ndorms.ox.ac.uk

Conflicts of interest: None

Key words: surgery; randomized controlled trials as topic; sample size; reporting; CONSORT

Word count: 1566

Abstract

Background The sample size calculation for a randomized controlled trial (RCT) should make sure that a statistically significant and clinically relevant difference can be detected with a high probability. To make sure that readers of an RCT can critically judge whether the assumptions for the sample size calculation were realistic, a number of specific sub-items should be reported.

Methods In this mini-review the frequency of adequate reporting of sample size specific sub-items was assessed from surgical RCTs that were published in leading general surgical journals in 2017. Sub-items from the CONSORT (CONsolidated Standards for Reporting Trials) and also from the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) Statement were considered.

Results A total of 85 relevant RCTs were identified over PubMed. The overall adequate reporting of the sample size according to CONSORT was 53% (45 of 85). When considering additional sub-items from the SPIRIT Statement, which are not explicitly requested by the CONSORT Statement, the adherence dropped to 12% (10 of 85).

Conclusion The adherence in reporting the sample size sub-items from the CONSORT Statement was moderate among surgical RCTs. It is crucial that surgeons who plan and conduct RCTs are aware of these sub-items.

Introduction

Randomized controlled trials (RCT) provide the most reliable evidence when a medical intervention is assessed [1]. It is crucial that the most important items are adequately reported, so that readers (e.g. clinicians, researchers, patients, policy makers) can adequately judge the quality of a published RCT.. Most medical journals endorse the CONSolidated Standards for Reporting Trials (CONSORT) Statement which was designed to improve the quality of reporting in RCTs [2]. The implementation of the CONSORT Statement did lead to a certain improvement in adequate reporting. However, room for improvement remained [3]. RCTs assessing non-pharmacological interventions are associated with lower reporting quality when considering the CONSORT items [4]. Nagendran and colleagues published in 2013 that the adequate reporting in surgical studies is in general poor [5]. Reasons that adequate reporting is unsatisfying might be that some authors perceive the checklist containing 25 items as a barrier or might not be aware which specific sub-items need to be reported [6]. For example the sample size is crucial to make sure that the sample size is large enough to have a high probability of finding a statistically significant difference [2]. Stating for example that an intervention shows no effect is problematic when the study was not adequately powered to reveal such an effect [7]. To adequately report this item a sample size description should include, according to the CONSORT explanation and elaboration paper, “(1) the estimated outcomes in each group (which implies the clinically important target difference between the intervention groups); (2) the α (type I) error level; (3) the statistical power (or the β (type II) error level); and (4), for continuous outcomes, the standard deviation of the measurements”.[2] Additionally, the primary outcome should be mentioned and, obviously, the sample size (total and per treatment arm). The SPIRIT Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) Statement [8], which is the reporting guideline for study protocols for clinical trials, was published after the last CONSORT update. The SPIRIT Statement lists the following additionally 3 crucial sub-items for the adequate reporting of the sample size: (1) the rationale (e.g. reference) for each of the assumed outcome values; (2) the statistical test used; and (3) adjustment for missing data (Table 1). This short report has the intention to raise awareness for all the aforementioned sub-items. Therefore, the proportion of adequately reported sub-items (according to CONSORT as well as SPIRIT) for the “sample size item” was assessed in surgical randomized controlled trials.

Methods

PubMed was searched from the 1.1.2017 until the 31.12.2017 for any surgical RCTs in the five general surgical journals with the highest impact factor (i.e. Annals of surgery, JAMA surgery, The British journal of surgery, Journal of the American College of Surgeons, and Surgery; according to the Thomson Reuters InCites Journal Citation Reports; search strategy within the Appendix). Included were RCTs presenting results on the primary outcome. Excluded were studies which were clearly labeled as pilot or feasibility studies and studies that performed additional, secondary analyses on an RCT.

From the included studies the following general characteristics were extracted: The study hypothesis (i.e. superiority, non-inferiority), if the planned sample size was mentioned at all, and whether the sample size derived statistically or was justified with another rational. The adequate reporting was checked for the following sub-items: reporting of the primary outcome in the sample size calculation, the estimated outcome in each group, the type I error, the statistical power (type II error), the standard deviation for the measurement (only for continuous outcomes), the total (and per group) planned sample size, the rational for the assumed outcome values in each arm, if the statistical test for the sample size calculation was provided, and adjustment for missing data. If an RCT did not mention at all a rational for the sample size calculation, it was assessed if the study refers to a study protocol. When a clear reference to a study protocol was found, this document was used to assess the aforementioned sub-items. The proportion of adequate reporting for each sub-item was calculated. Additionally, it was assessed if the item "sample size" was adequately reported (i.e. all applicable sub-items adequately reported) according to CONSORT [2] and SPIRIT [8] Statements.

Results

A total of 217 potentially relevant articles was found on PubMed. Based on titles, abstracts and full text 132 articles were excluded (Figure 1a). The 85 included RCTs were published in The British journal of surgery (n=30; 35%), Annals of surgery (n=27; 32%), Surgery (n=11; 13%), Journal of the American College of Surgeons (n=10; 12%), and JAMA surgery (n=7; 8%). A superiority-hypothesis was tested in 66 RCTs, while 14 tested for non-inferiority, and 2 had a superiority and non-inferiority hypothesis. In 3 RCTs it was unclear what the hypothesis was. Out of the 85 included RCTs 9 (11%) did not mention the planned sample size in the published paper at all. Two of these 9 RCTs mentioned explicitly a study protocol which contained a sample size calculations. The assessment of the reporting of all items and

sub-items was conducted under the considerations of these 2 study protocols. All sample sizes which were mentioned (n=78) derived statistically. The primary outcome in the sample size was reported in 73 out of the 85 RCTs (86%) and the assumed values in each group in 54 (64%) RCTs (Figure 1b). The type I error and the type II error (power) were reported in 75 (88%) and 77 (91%) RCTs, respectively. The standard deviation was reported in 18 (46%) out of 39 RCTs (46 RCTs did not report a continuous outcome, Figure 1b). The total and per group planned sample size was reported in 76 (89%) of the included RCTs. The overall adequate reporting of the item "sample size" according to the descriptions in the CONSORT explanation and elaboration paper [2] was 54% (45 of 85 RCTs). The adherence to the additional 3 sub-items from the SPIRIT explanation and elaboration paper [8] were the following: 44 (54%) reported the rational for the assumed values, 31 (36%) described the statistical test used for the sample size, and 37 out of 73 (51%) reported, if they adjusted the sample size for missing data. 12 RCTs assessed primary endpoints for which the loss of follow up was judged as highly unlikely (e.g. length of hospital stay, intraoperative blood loss). The overall adequate reporting of the item "sample size" according to the descriptions in the SPIRIT explanation and elaboration paper [8] was 12% (10 of 85 RCTs).

Discussion

This short report indicates that 53% (45 of 85) of surgical RCTs reported the sample size adequately. This result is between the proportion of adherence found by Nagendran et al. [5], (6%; 3 of 54) and a recently published study by Limb and colleagues (76.3%; 119 of 156).[9] Both of these studies did not present an assessments of sub-items. The more recently published SPIRIT Statement is for study protocols for clinical trials. When considering the SPIRIT items, the adequate reporting of the item "sample size" dropped to 12%. Even though the SPIRIT Statement is not intended for the reporting of published RCTs, the sub-items listed for the sample size are also relevant and applicable for published RCTs. It might be possible that these sub-items will also be included in an updated version of the CONSORT Statement. In the assessed surgical trials, these SPIRIT sub-items seemed to be less often reported. For example, the adjustment for missing data is crucial to make sure that the RCT is not underpowered due to a high number of loss of follow-up. Furthermore, also a rational for the assumed values (e.g. data from observational studies, internal unpublished data, data from pilot studies) should be given. This should avoid that unrealistic large effects are assumed to conduct an RCT with a relatively

small sample size. In reality “small but clinically meaningful true differences are much more likely than large differences to exist, but large trials are required to detect them” [2]. An assessment by Ahmed Ali and colleagues showed that approximately one third of surgical trials are underpowered to detect a clinically significant difference and that there was no improvement over time [10]. Conducting a trial and randomizing patients to different intervention with a low chance of detecting a difference (in case such a difference exists) is unethical, a waste of resources and should be avoided. Instead, investigators could try to include multiple study sites, so that an adequate sample size could be achieved.

This mini-review has the following limitations which are worth mentioning. First, data extraction was not conducted in duplicate. It is therefore likely that some items were judged incorrectly. However, it seems unlikely that the overall message of this article would have changed if data was extracted in duplicate. Second, the CONSORT extension for non-pharmacologic trials requests that, when relevant, “details of whether and how the clustering by care providers or centers was addressed” [11]. To assess if this sub-item would have been applicable would have required expert knowledge about the specific surgical interventions tested in the different RCTs. It is likely that the overall adherence would have further decreased when this item would have been assessed.

In conclusion, the adherence in reporting to the sample size sub-items from the CONSORT Statement was moderate among surgical RCTs. Room for improvement exists especially in reporting the assumed values in each treatment arm. Furthermore, other important aspects of the sample size which are only requested by the SPIRIT Statement were less frequently reported. It is crucial that surgeons who plan and conduct RCTs are aware of these sub-items. Journals can play a vital role in this process by requesting those sub-items in published RCTs as well as in published protocols.

Conflict of Interests:

Benjamin Speich is supported by an Advanced Postdoc.Mobility grant from the Swiss National Science Foundation (P300PB_177933). No conflicts of interests occur.

Acknowledgements:

Benjamin Speich is supported by an Advanced Postdoc.Mobility grant from the Swiss National Science Foundation (P300PB_177933).

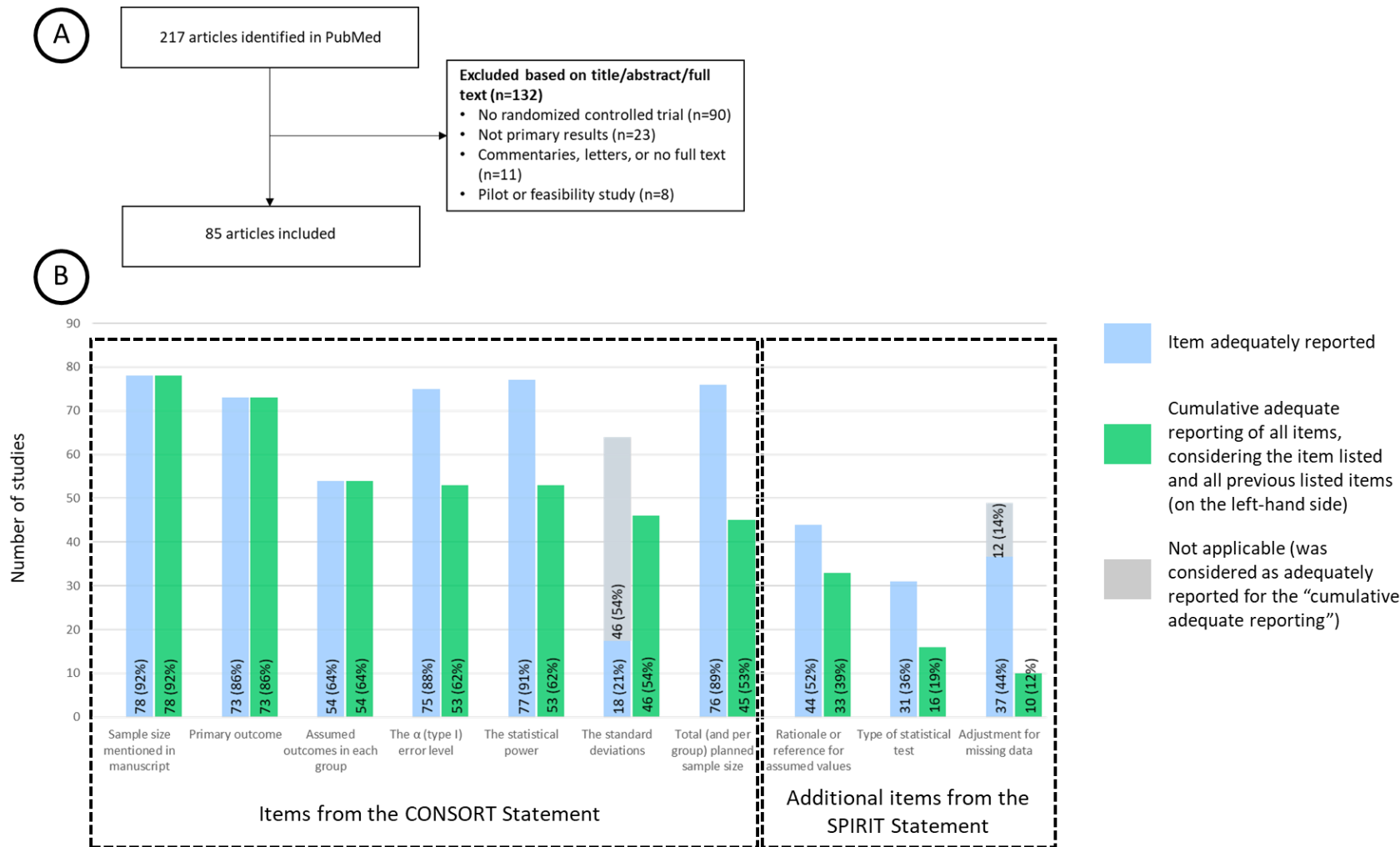
158 **References**

- 159 1. Duley L, Antman K, Arena J, et al (2008) Specific barriers to the conduct of randomized trials.
160 Clin Trials 5:40-48
- 161 2. Moher D, Hopewell S, Schulz KF, et al (2010) CONSORT 2010 explanation and elaboration:
162 updated guidelines for reporting parallel group randomised trials. BMJ 340:c869
- 163 3. Turner L, Shamseer L, Altman DG, et al (2012) Does use of the CONSORT Statement impact
164 the completeness of reporting of randomised controlled trials published in medical journals? A
165 Cochrane review. Syst Rev 1:60
- 166 4. Jin Y, Sanger N, Shams I, et al (2018) Does the medical literature remain inadequately
167 described despite having reporting guidelines for 21 years? - A systematic review of reviews:
168 an update. Journal of multidisciplinary healthcare 11:495-510
- 169 5. Nagendran M, Harding D, Teo W, et al (2013) Poor adherence of randomised trials in surgery
170 to CONSORT guidelines for non-pharmacological treatments (NPT): a cross-sectional study.
171 BMJ Open 3:e003898
- 172 6. Clavien PA, Lillemoe KD (2014) A new policy to implement CONSORT guidelines for surgical
173 randomized controlled trials. Ann Surg 260:947-948
- 174 7. Altman DG, Bland JM (1996) Absence of evidence is not evidence of absence. Australian
175 veterinary journal 74:311
- 176 8. Chan AW, Tetzlaff JM, Gotzsche PC, et al (2013) SPIRIT 2013 explanation and elaboration:
177 guidance for protocols of clinical trials. BMJ 346:e7586
- 178 9. Limb C, White A, Fielding A, et al (2017) Compliance of Randomized Controlled Trials
179 Published in General Surgical Journals With the CONSORT 2010 Statement. Ann Surg
- 180 10. Ahmed Ali U, Ten Hove JR, Reiber BM, et al (2018) Sample size of surgical randomized
181 controlled trials: a lack of improvement over time. J Surg Res 228:1-7
- 182 11. Boutron I, Altman DG, Moher D, et al (2017) CONSORT Statement for Randomized Trials of
183 Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for
184 Nonpharmacologic Trial Abstracts. Ann Intern Med 167:40-47

185

186 **Figure Legend:**

187 **Figure 1:** Study selection of randomized controlled trials from general surgical journals (A), and the assessment of adequate reporting of the
188 sample size and its relevant sub-items (B).



190

191 **Table 1:** List of all the sub-items that should be included in the description of the sample size
 192 according to the CONSORT (Consolidated Standards of Reporting Trials) and SPIRIT
 193 (Standard Protocol Items: Recommendations for Interventional Trials) explanation and
 194 elaboration papers [2, 8].

Reporting guideline	Listed sub-items that should be considered
CONSORT (item 7a)	"The primary outcome on which the calculation was based" [2]
	"The estimated outcomes in each group (which implies the clinically important target difference between the intervention groups)" [2]
	The α (type I) error level
	The statistical power (also called β or type II error level)
	"For continuous outcomes, the standard deviation of the measurements" [2]
SPIRIT* (item 14)	The sample size (total and per treatment arm)
	"A rationale or reference for the outcome values assumed for each study group" [8]
	The statistical test
	Adjustment for missing data (if relevant)

202 *The SPIRIT items does also consist all listed CONSORT sub-items for the sample size. The here listed SPIRIT-
 203 sub-items are in addition to the ones mentioned in the CONSORT Statement.

204