International Journal of Surgery Protocols 21 (2020) 1–4

Contents lists available at ScienceDirect

International Journal of Surgery Protocols

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

Characteristics, results, and reporting of contemporary surgical trials: A systematic review and analysis

N. Bryce Robinson [a], Ajita Naik [a], Irbaz Hameed [a], Yongle Ruan [a], Mohamed Rahouma [a], Viola Weidenmann [a], Marco A. Zenati [b], Deepak L. Bhatt [c], Leonard N. Girardi [a], Paul Kurlansky [d], Shahzad G. Raja [e], David Moher [f], Stephen Fremes [g], Joanna Chikwe [h], Mario Gaudino [a,*]

[a] Department of Cardiothoracic Surgery, Weill Cornell Medicine, New York, NY, USA
[b] BHS Department of Cardiothoracic Surgery, West Roxbury, MA, USA
[c] Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA, USA
[d] Department of Surgery, Columbia University Medical Center, New York, NY, USA
[e] Department of Cardiac Surgery, Harefield Hospital, London, UK
[f] Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[g] Schulich Heart Centre, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, Canada
[h] Department of Cardiac Surgery, Smidt Heart Institute, Cedars-Siani Medical Center, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 5 February 2020
Received in revised form 9 March 2020
Accepted 10 March 2020
Available online 19 March 2020

Keywords:
Surgery
Randomized control trial
Spin
Outcomes
Methodology

ABSTRACT

Introduction: The complexities and risks inherent to the field of surgery and surgical interventions present unique challenges to the design and analysis of surgical randomized controlled trials (RCT). Prior studies have investigated the practical and methodologic challenges posed by surgical RCTs. To date, however, a comprehensive analysis of the contemporary literature across multiple surgical subspecialties does not exist. In this descriptive analysis, we set out to characterize surgical RCTs over the past 10 years across six major surgical specialties.

Methods and analysis: A literature search by a medical librarian will be performed to identify all surgical randomized clinical trials published between January 2009 and December 2019 in the two journals with the highest impact factor for six surgical specialties as well as two large general medicine journals. Two reviewers will independently screen the citations retrieved from the literature search and extract data according to a previously described protocol via a pre-defined data collection form. Categorical variables will be reported as counts and percentages. Following assessment of normality, continuous variables will be reported as mean (standard deviation) or median (inter-quartile range). Based on normality of data, independent t-test or the Mann-Whitney U test will be used to compare continuous variables and chi-square and Fisher's exact tests to compare categorical variables. Comparisons across multiple sets will be performed using ANOVA or Kruskak-Wallis tests. Two-sided significance testing will be used and a p-value <0.05 will be considered significant without adjustment for multiple testing. All analyses will be performed using SPSS version 24 and R within RStudio. PROSPERO (ID number: 162797).

Ethics and dissemination: There are no ethical concerns directly pertinent to this systematic review. The retrieved data will be made available upon request. The study will be written in English and submitted for publication in a peer-reviewed journal.

1. Background

According to the World Health Organization, there are more than 200 million surgical procedures performed each year [1]. Many of the interventions applied across multiple surgical specialties have little, if any, randomized data demonstrating their safety and utility. Unlike the fields of medicine and pharmacology, where the RCT is the gold standard of research, surgery relies heavily on observational and retrospective data to drive innovation and introduce novel techniques and devices [2]. In an analysis of surgical RCTs, McCulloch found that randomized trials represented less than 10% of the published literature investigating operative techniques and outcomes [3].

* Corresponding author at: Department of Cardiothoracic Surgery, Weill Cornell Medicine, 525 E 68th St, New York, NY 10065, USA.
E-mail address: mfg9004@med.cornell.edu (M. Gaudino).

https://doi.org/10.1016/j.isjp.2020.03.002
2468-3574/

The lack of randomized evidence is likely secondary to the unique challenges that exist in the design, conduct, and analysis of surgical RTCs. Blinding and allocation concealment, for example, can be difficult when compared to medical trials and introduce an important element of bias for surgeons and assessors. Sham surgery has been proposed to eliminate this risk, however this remains controversial in that it introduces risk to the patient without benefit [4]. Additional issues include surgeon expertise, variability in technique, surgeon and patient preference, use of intention-to-treat or as-treated analysis, follow-up, and crossover between treatment arms [5].

In this study, we set out to characterize the state of surgical RCTs published over the previous 10 year period in the two highest impact factor journals from six separate surgical specialties (general surgery, cardiothoracic surgery, neurosurgery, orthopedic surgery, transplantation surgery, and vascular surgery) as well as two large medical journals (*Lancet* and *New England Journal of Medicine*).

2. Methods and analysis

2.1. Search strategy and definition of surgical trial

A literature search by a medical librarian will be performed to identify all surgical randomized clinical trials (RCT) published between January 2009 and December 2019 in the two journals with the highest impact factor for general medicine (*The New England Journal of Medicine* and *The Lancet*) and each of the following surgical specialties: cardiothoracic surgery (*The Journal of Thoracic and Cardiovascular Surgery* and *The Annals of Thoracic Surgery*), general surgery (*JAMA Surgery* and *Annals of Surgery*), neurosurgery (*Neurosurgery* and *Journal of Neurosurgery*), orthopedic surgery (*Journal of Bone and Joint Surgery American Volume* and *Arthroscopy-The Journal of Arthroscopic and Related Surgery*), transplant surgery (*Journal of Heart and Lung Transplantation* and *American Journal of Transplantation*), and vascular surgery (*European Journal of Vascular and Endovascular Surgery* and *Journal of Vascular Surgery*).

A clinical trial will be defined as surgical if it involves evaluation of a surgical intervention in both the experimental and control arms. Surgery will be defined as any procedure performed by a trained specialist with the goal of correcting deformities or defects, repairing injuries, or for the cure of certain diseases, as specified by the National Centre for Biotechnology Information (NCBI). Non-surgical interventional trials and medical trials will be excluded.

2.2. Extraction of trial data

The following data will be recorded for each trial: journal of publication and impact factor (according to Thomson Reuters-Clarivate Analytics), year of publication, type of intervention, single- or multi-center study, geographical locations of the participating centers, details of the primary outcome (definition of the outcome, composite or non-composite endpoints), number of screened patients and percentage of screened patients enrolled in the trial, sample size, statistical power, treatment effect (relative risk reduction) size estimation used for sample size calculation, length of the follow-up, number of events, number of patients lost to follow-up, number of crossovers, number of citations on Scopus/Web of Science, blinded or unblinded assessment of outcomes, details of the primary analysis (intention-to-treat, as treated or per protocol, superiority, equivalence or non-inferiority), and adjustment for multiple testing in case of multiple primary outcomes, trial sponsor, declared conflict of interest of first and last authors. Willingness to share data, involvement of a clinical trials unit in trial design and/or conduct, date of trial registration, trial start date, and the number of revisions on registry will also be recorded. A detailed assessment of blinding will also be done.

The methods used to deal with the possible learning curve effect and assure deliverability of the intervention (experience cut-off, pre-trial training, expertise-based design) and to monitor the quality of the intervention (statistical monitoring of crossover or outliers, video-recording, etc) will also be recorded. Data will also be collected on the level of details of the experimental procedure described in the trial protocol (used a semiquantitative scale: none, limited, detailed).

To determine the trials' primary outcome(s), the following will be examined sequentially: the methods, trial design, the primary aim of the study, and the outcome used in the sample size calculation. If no primary outcome is clearly identified (i.e. explicitly specified in the article, in a sample size calculation, or in the primary study objectives), the trial will be ineligible for the primary analysis, but will still be included in all other analyses. Primary trial outcomes will be classified as major or minor clinical events based on a pre-defined classification scheme that will be reported in the manuscript.

The conflicts of interest of the first and last authors will be identified from the disclosure statements published in the trials or supplementary material. For trials listing co-first authors, disclosure of both authors in the list will be considered. Authors' conflicts of interest will be defined as any report of consulting, advisory, or speaking fees or honoraria, stock ownership, affiliation, or employment by the study sponsor.

Two reviewers will independently screen the citations retrieved from the literature search and extract all data following previously described methodology and using a pre-defined data collection form [6–8]. A third reviewer will resolve any discrepancy.

2.3. Trials analysis

2.3.1. Classification of trial results

Consistent with previous reports [9–11], trials will be classified as "favorable" or "unfavorable" for the experimental therapy based on the results: a trial will be classified as "favorable" if, for at least one primary outcome among those defined in the protocol, the experimental therapy is significantly better than the control therapy ($p < 0.05$ or a 95% confidence interval (CI) which excludes the null value in superiority trials), the experimental therapy is not substantially worse than the control therapy (in non-inferiority trials), or the effects of the treatments differ by no more than the equivalence margin (in equivalence trials).

2.3.2. Appraisal of spin

In studies reporting a non-significant difference in the primary outcome, the presence and amount of distortion or misrepresentation of benefit, or "spin", will be evaluated as previously described [12,13]. Spin will be defined as the use of specific reporting strategies to suggest that the experimental treatment is beneficial or non-inferior despite a statistically non-significant difference for the primary outcome, or to distract the reader from statistically non-significant results [8].

For each selected article, two readers will independently read the full manuscript and the online appendices. The reviewers will independently assess article contents using a pretested and standardized data abstraction form as previously described [12]. Discrepancies will be resolved by a third reviewer. The presence of spin will be assessed in the following sections of the manuscript: title, abstract results; abstract conclusion; main-text results, discussion, and conclusions. Following a described method, the strategies of spin considered will be (1) a focus on secondary statistically significant results (within-group comparison, secondary outcomes,

subgroup analyses, modified population of analyses); (2) interpreting statistically non-significant results for the primary outcomes as showing treatment equivalence or comparable effectiveness; (3) claiming or emphasizing the beneficial effect of the experimental treatment despite statistically non-significant results; and 4) claiming or emphasizing non-inferiority despite not establishing non-inferiority boundaries or when data are inconclusive. Other spin strategies that are not classified according to this scheme will be systematically recorded and classified as "others" [12,13]. The extent of spin across a study will be defined as the number of sections with spin in the entire article.

2.3.3. Assessment of discrepancy between the registered and published primary outcomes

For each trial, we will identify the registration number in the published articles or clinical trial registries (ClinicalTrials.gov, ISRCTN register, or country-specific registries). Only trials prospectively registered that clearly describe the primary outcome in the registry will be considered in this analysis. Consistent with previous definitions [14,15], a major discrepancy between the registered and published primary outcomes will be identified if the outcomes are different or assessed at different time points. Major discrepancies will be defined as: (1) a pre-specified primary outcome in the trial registration protocol reported as a secondary outcome in the final published article; (2) the published primary outcome described as a secondary outcome in the registry; (3) the pre-specified primary outcomes in the trial registration not reported in the published article; (4) a new primary outcome introduced in the published article; and (5) the timing of assessment of the primary outcome in the registered protocol and published article differing [15].

The trials will be analyzed by two reviewers independently. All discrepancies will discussed to obtain consensus, and if needed, the article will be discussed with a third reviewer.

2.3.4. Calculation of the Fragility Index

For superiority-design trials reporting at least one statistically significant dichotomous primary outcome ($p < 0.05$ or a 95% CI excluding the null value), we will quantify how robust the results are by using the Fragility Index described by Walsh et al. [6]. The Fragility Index is defined as the number of patients whose status would need to switch from non-event to event to render a statistically significant difference non-significant. The results for each outcome will be entered in a 2×2 contingency table following which the p-value for each outcome will be calculated using the two-sided Fisher's exact test. Single participants will be iteratively shifted one at a time in the lower-incidence treatment group from "non-event" to "event" and the p-value for the 2×2 table will be re-calculated. The Fragility Index for an outcome shall equal the smallest number of patients required to turn the re-calculated p-value non-significant (≥ 0.05). Lower values will indicate less robust results.

2.3.5. Appraisal of trial pragmatism

For each surgical trial, the PRECIS-2 tool will be used to investigate how pragmatic or explanatory the trial is, and the overall level or pragmatism of the surgical trials over the past decade. Using previously described methodology [16], the PRECIS-2 tool will be used to evaluate nine domains of trial design: eligibility criteria, recruitment, setting, organization, the flexibility of intervention delivery, the flexibility of adherence to the intervention, follow-up, primary outcome, and primary analysis. A 5-point Likert scale will be used to rate the level of pragmatism in each trial design domain as follows: (1) very explanatory, (2) rather explanatory, (3) equally pragmatic/explanatory, (4) rather pragmatic, and (5) very pragmatic. The trials will be analyzed by two reviewers independently. All discrepancies will discussed to obtain consensus, and if needed, the article will be discussed with a third reviewer.

2.3.6. Sponsor

Trials will be classified as commercially-sponsored if they are industry-initiated and sponsored, or investigator-initiated studies that receive commercial support. Trials will be classified as non-commercially-sponsored if they are investigator-initiated and report local government or federal or hospital or university sponsorship, or no sponsors. For commercially-sponsored trials, the body of the articles, supplementary materials and original trial designs will be additionally analyzed for report of commercial or sponsor involvement in the trial design, conduct, analysis, or reporting.

2.4. Statistical analysis

Categorical variables will be reported as counts and percentages. Following assessment of normality, continuous variables will be reported as mean (standard deviation) or median (inter-quartile range). Based on normality of data, independent t-test or the Mann-Whitney U test will be used to compare continuous variables and X^2 and Fisher's exact tests to compare categorical variables. Two-sided significance testing will be used and a p-value <0.05 will be considered significant without adjustment for multiple testing. Comparisons across multiple sets will be performed using ANOVA or Kruskak-Wallis tests. All analyses will be performed using SPSS version 24 (IBM, Chicago, IL, USA) and R (version 3.4.2 R Project for Statistical Computing) within RStudio.

This study protocol has been prospectively registered on the International Prospective Register of Systematic Reviews (PROSPERO ID Number: 162797).

3. Ethics and dissemination

Ethical approval is not required for this study as no patient records or direct contact with patients or animals will occur. This study will identify current issues that remain in surgical RCTs over the past 10 years. This will inform the surgical community about areas where improvement is necessary. This study will be published in English, with plans to present at national meetings.

Funding

Stephen Fremes is supported in part by the Bernard S. Goldman Chair in Cardiovascular Surgery

Authors contributions

NBR, AN, IH, YR, VW, MAZ, DLB, LNG, PK, SGR, DM, SF, JC, and MG all contributed to the design of the study, and the writing and editing of this manuscript. MR, SF, DM, PK, and MG contributed to the design of the statistical analysis and variables to be extracted. All authors contributed equally and have given final approval of this manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T.G. Weiser, S.E. Regenbogen, K.D. Thompson, A.B. Haynes, S.R. Lipsitz, W.R. Berry, et al., An estimation of the global volume of surgery: a modelling strategy based on available data, Lancet 372 (9633) (2008) 139–144.

[2] J. Ellis, I. Mulligan, J. Rowe, D.L. Sackett, Inpatient general medicine is evidence based. A-Team, Nuffield Department of Clinical Medicine, Lancet Lond Engl. 346 (8972) (1995) 407–410.

[3] P. McCulloch, Randomised trials in surgery: problems and possible solutions, BMJ 324 (7351) (2002) 1448–1451.

[4] B. Speich, Blinding in surgical randomized clinical trials in 2015, Ann Surg 266 (1) (2017) 21–22.

[5] F. Farrokhyar, P.J. Karanicolas, A. Thoma, M. Simunovic, M. Bhandari, P.J. Devereaux, et al., Randomized controlled trials of surgical interventions, Ann Surg 251 (3) (2010) 409–416.

[6] B. Djulbegovic, M. Lacevic, A. Cantor, K.K. Fields, C.L. Bennett, J.R. Adams, et al., The uncertainty principle and industry-sponsored research, Lancet Lond. Engl. 356 (9230) (2000) 635–638.

[7] L.L. Kjaergard, B. Als-Nielsen, Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ, BMJ 325 (7358) (2002) 249.

[8] B. Als-Nielsen, W. Chen, C. Gluud, L.L. Kjaergard, Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events?, JAMA 290 (7) (2003) 921–928.

[9] M.E. Flacco, L. Manzoli, S. Boccia, L. Capasso, K. Aleksovska, A. Rosso, et al., Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor, J. Clin. Epidemiol. 68 (7) (2015) 811–820.

[10] S. Hopewell, K. Loudon, M.J. Clarke, A.D. Oxman, K. Dickersin, Publication bias in clinical trials due to statistical significance or direction of trial results, Cochrane Database Syst. Rev. 1 (2009), MR000006.

[11] J.P. Ioannidis, Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials, JAMA 279 (4) (1998) 281–286.

[12] I. Boutron, S. Dutton, P. Ravaud, D.G. Altman, Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes, JAMA 303 (20) (2010) 2058–2064.

[13] A.V. Hernandez, V. Pasupuleti, A. Deshpande, P. Thota, J.A. Collins, J.E. Vidal, Deficient reporting and interpretation of non-inferiority randomized clinical trials in HIV patients: a systematic review, PLoS One 8 (5) (2013) e63272.

[14] A.-W. Chan, A. Hróbjartsson, M.T. Haahr, P.C. Gøtzsche, D.G. Altman, Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles, JAMA 291 (20) (2004) 2457–2465.

[15] T. Chen, C. Li, R. Qin, Y. Wang, D. Yu, J. Dodd, et al., Comparison of clinical trial changes in primary outcome and reported intervention effect size between trial registration and publication, JAMA Netw. Open 2 (7) (2019) e197242.

[16] N. Sepehrvand, W. Alemayehu, D. Das, A.K. Gupta, P. Gouda, A. Ghimire, et al., Trends in the explanatory or pragmatic nature of cardiovascular clinical trials over 2 decades, JAMA Cardiol. (2019).