

STUDY PROTOCOL

Open Access



Current practice in methodology and reporting of the sample size calculation in randomised trials of hip and knee osteoarthritis: a protocol for a systematic review

Bethan Copsey^{1*} , Susan Dutton¹, Ray Fitzpatrick², Sarah E. Lamb¹ and Jonathan A. Cook¹

Abstract

Background: A key aspect of the design of randomised controlled trials (RCTs) is determining the sample size. It is important that the trial sample size is appropriately calculated. The required sample size will differ by clinical area, for instance, due to the prevalence of the condition and the choice of primary outcome. Additionally, it will depend upon the choice of target difference assumed in the calculation. Focussing upon the hip and knee osteoarthritis population, this study aims to systematically review how the trial size was determined for trials of osteoarthritis, on what basis, and how well these aspects are reported.

Methods: Several electronic databases (Medline, Cochrane library, CINAHL, EMBASE, PsycINFO, PEDro and AMED) will be searched to identify articles on RCTs of hip and knee osteoarthritis published in 2016. Articles will be screened for eligibility and data extracted independently by two reviewers. Data will be extracted on study characteristics (design, population, intervention and control treatments), primary outcome, chosen sample size and justification, parameters used to calculate the sample size (including treatment effect in control arm, level of variability in primary outcome, loss to follow-up rates). Data will be summarised across the studies using appropriate summary statistics (e.g. *n* and %, median and interquartile range). The proportion of studies which report each key component of the sample size calculation will be presented. The reproducibility of the sample size calculation will be tested.

Discussion: The findings of this systematic review will summarise the current practice for sample size calculation in trials of hip and knee osteoarthritis. It will also provide evidence on the completeness of the reporting of the sample size calculation, reproducibility of the chosen sample size and the basis for the values used in the calculation.

Trial registration: As this review was not eligible to be registered on PROSPERO, the summary information was uploaded to Figshare to make it publicly accessible in order to avoid unnecessary duplication amongst other benefits (<https://doi.org/10.6084/m9.figshare.5009027.v1>); Registered January 17, 2017.

Keywords: Sample size, Osteoarthritis, Reporting, Systematic review

* Correspondence: bethan.copsey@csm.ox.ac.uk

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

Full list of author information is available at the end of the article



Background

The sample size and target difference of a clinical trial is a key feature that impacts on how the trial is designed and conducted [1]. There are multiple factors which contribute to the determination of the required sample size with choice of target difference arguably the most important [2, 3]. It is recognised that an overly large sample size is undesirable as it increases the costs of the trial and likely delays dissemination of the findings [2]. Too large a sample size is also unethical as it may result in additional participants receiving a treatment when there is already sufficient evidence to show that it is inferior to an alternative [4]. Conversely, too small a sample size also poses ethical issues as it will result in a study lacking sufficient power to detect a clinically important treatment effect [5, 6].

Many studies have found that sample sizes are often inadequately reported and based on inaccurate assumptions [7, 8]. Discrepancies in the assumptions for parameters in sample size calculations can impact on power [9, 10]. In particular, for continuous outcomes, underestimation of the standard deviation can lead to underpowered studies. Tavernier and Giraudeau hypothesised that this mis-specification may be due to the imprecision and more homogeneous populations of pilot studies often used to estimate the parameter [9].

OARSI (Osteoarthritis Research Society International) recently published recommendations to use realistic and clinically important effect sizes when calculating the sample size for an osteoarthritis trial, suggesting that previously some trials have not done so [11]. Keen et al. found that rheumatology trials published in 2001–2002 were often underpowered and sample size calculations were poorly reported [12]. Since then, there have been considerable efforts to improve reporting of randomised trials; for example, the Consolidated Standards of Reporting Trials (CONSORT) Statement includes an item to report how the ‘sample size was determined’ [13]. However, it is unclear how investigators currently determine and report the sample size for trials of hip and knee osteoarthritis.

Aim and objectives

The primary objective is to summarise the methodology used (including the assumptions made and justifications provided) to determine the sample size calculation for randomised trials of hip and knee osteoarthritis.

The secondary objectives are to assess the reporting and reproducibility of the sample size calculation.

Methods

The PRISMA-P Checklist for this review protocol is available as Additional file 1.

Inclusion criteria

Studies will be eligible for inclusion if they are randomised controlled trials of hip and/or knee osteoarthritis with two treatment arms (one intervention and one comparator) published in 2016. Inclusion will not be restricted by study outcomes, intervention or control treatments.

Exclusion criteria

Articles on non-randomised studies will be excluded, including case-control or cross-sectional studies. Quasi-randomised studies or studies which do not state that the allocation was randomised will be excluded. Factorial design and cross-over trials will be excluded. Trials with more than two arms will be excluded.

Pilot studies will be excluded, as will studies which refer to themselves as ‘feasibility’, ‘proof of concept’ or ‘exploratory’ studies. Studies which intend to use results to inform future definitive phase III trials will not be included. Studies which do not consider treatment evaluation (e.g. compare different screening methods) will be excluded.

Studies on the prevention of osteoarthritis or trials with mixed populations will be excluded; for example, a combination of rheumatoid arthritis and osteoarthritis. Trials of, for example, total knee arthroplasty will only be eligible if it is clearly stated that all participants had knee osteoarthritis.

Non-English language articles will be excluded as this review focusses on study reporting.

Articles will be excluded if they are conference abstracts or study protocols. Only the primary report of a trial will be eligible; separate publications for secondary analyses will be excluded, including long-term follow-up or subgroup analyses.

Identification of studies

Articles reporting the results of clinical trials will be identified using electronic searches of databases including Medline, Cochrane Central Register of Controlled Trials (CENTRAL), CINAHL, EMBASE, PsycINFO, PEDro and AMED. An example search strategy is given in Additional file 2.

A preliminary search to inform the search strategy indicated that this would lead to around 100 included studies, which was considered to be sufficiently large for a methodological review of this kind [14–16].

Selection of studies

All search results will be combined and duplicates will be removed. Titles and abstracts will be screened independently by two reviewers. Following this, full texts for the records considered to be potentially included will be obtained and also screened independently by two researchers. Any disagreements will be resolved through

discussion between the two reviewers assessing the papers, with involvement of a further reviewer where necessary.

Data extraction and management

Data will be extracted using a standardised form. The data-extraction form will be piloted to ensure that all relevant information is recorded and to allow refinement prior to formal use. The form will then be enhanced by adding and clarifying items to extract. Data extraction will be performed by a second reviewer on a sample of included studies (at least 20%) in order to check accuracy. Relevant data will be extracted from the study protocol where this is cited in the main trial results publication; where there are conflicts between the information in the protocol and main publication, information from the main publication will be used.

The following information will be extracted from each article when reported:

- Article: Country
- Design: Study design (e.g. superiority, non-inferiority)
- Population: Condition (including how osteoarthritis was defined), setting, eligibility criteria (in particular age, gender, and disease severity).
- Treatment: Intervention, comparator.
- Outcome: Primary outcome measure(s)
- Sample size details: Statistical approach (conventional, other), chosen sample size, method for calculation, values used and justification (e.g. effect size, target difference, standard deviation, adjustment for loss to follow-up, sidedness, significance level, power), whether sample size could be replicated, whether sample size re-estimation was planned (e.g. using interim analysis), whether sensitivity analysis was conducted to examine impact of assumptions on sample size. Note: Post-hoc sample size calculations will not be considered.
- Follow-up: Number of participants randomised, number lost to follow-up, whether compliance was measured.

Data synthesis

Data will be summarised across the studies, including the general characteristics of the included studies using appropriate summary statistics (e.g. *n* and %, median and interquartile range (IQR)).

The proportion of studies that justify the sample size and target difference, and which report each component of the sample size calculation will be calculated.

The target difference expressed as a standardised effect size (e.g. Cohen's *d* for continuous outcome) will be presented graphically to compare across the studies and, where there are a sufficient number of studies, within

conditions (i.e. considering hip osteoarthritis and knee osteoarthritis separately) [17].

Sample size replication

Using the reported values, an attempt will be made to replicate the sample size calculation. It will be assumed that 80% power and 5% two-sided significance level were used unless otherwise stated and that a conventional (Neyman-Pearson) approach to the sample size calculation has been adopted [1]. The sample size calculation will be replicated using statistical software such as the 'power twomeans' command in STATA IC 14 [18].

When comparing the replicated to stated sample size, the ratio of replicated/reported values will be calculated. Ratios will be presented in a box plot [19]. The proportion of studies where the ratio is ≥ 1.1 or ≤ 0.9 , and ≥ 1.3 or ≤ 0.7 (i.e. out by at least 10% or 30%) will be calculated.

Subgroup analysis

Subgroup analysis will be used to explore the associations between study-level characteristics and key aspects of the sample size calculation: (1) observed sample size (number of participants randomised), (2) whether the sample size calculation was fully specified and (3) replicability. Data will be summarised within subgroups and presented using box plots (median, interquartile range (IQR) and range).

For subgroup comparisons, the following aspects will be compared:

- (1) Type of intervention: surgical vs non-surgical trials
- (2) Centres: single vs multi-centre
- (3) Funding: industry-funded (all or in part) vs no industry funding (or not reported)
- (4) Comparator: placebo/waitlist vs active control

If sufficient studies are reflected across the respective subgroups, we will formally compare groups: (1) sample size will be compared between subgroups using the median difference and 95% confidence interval. Absolute risk differences with 95% confidence intervals will be used to compare between subgroups for (2) reporting of a sample size calculation, (3) reporting of core sample size components and (4) the replicated sample size being > 10% larger than the reported sample size.

Formal subgroup comparisons will only be conducted where a sufficient number of studies are present within each group. A two-sided significance level of 0.05 will be used. As the subgroup analysis is exploratory, no adjustment will be made for multiple testing.

Discussion

This review will examine the current practice for sample size calculation in randomised trials of hip and knee

osteoarthritis, which will include the target difference that studies are designed to detect, the chosen sample size and justification of key inputs. It will also provide evidence on the completeness of the reporting of the sample size calculation and the accuracy of the sample size calculation (i.e. whether the calculation was reproducible). This systematic review will also provide insight into the number of randomised trials being conducted on hip and knee osteoarthritis and the variety of interventions being evaluated. Focussing on a specific clinical area will permit a more detailed assessment of the methodology within a more homogeneous sample of trials.

Subgroup analysis will explore whether the sample size used and reporting of the sample size calculation differ based on type of intervention, number of centres, funding source and comparator. Surgical and non-surgical studies will be compared since several studies have highlighted the complexities of surgical trials and have highlighted their poor methodological quality and reporting [20–22]. Studies have also suggested that multi-centre trials may have higher methodological quality than single-centre trials [20, 23, 24]. The effect of funding source will be examined since previous reviews have shown that industry-funded studies may differ in terms of transparency and outcome reporting [25–27]. Finally, trials with an active comparator will be compared to those with a placebo or ‘no treatment’ control since trials with an active control arm may have methodological differences, e.g. using a smaller target difference and thus requiring a larger sample size [28, 29].

This systematic review will be limited in that it relies primarily on information from the trials’ results publication(s), which may not be transparent about modification once the study had begun and thus may not accurately reflect the a-priori sample size calculation when the study was planned. There is some evidence to suggest that practice is more complex than trial reports suggest [30]. Nevertheless, the reported sample size should reflect the final design and is the natural one to assess, at least in the first instance.

The findings of this systematic review will provide evidence on whether sufficient information is being reporting in sample size calculations and explore variability in the chosen sample size and reporting based on study design features (including the justification of key inputs). This may highlight areas for improvement in the reporting and conduct of sample size calculations for hip and knee osteoarthritis trials, and to an extent, trials of other conditions.

Additional files

Additional file 1: This includes the PRISMA Checklist. The line numbers relate to the original manuscript submission. (DOCX 29 kb)

Additional file 2: This details the search terms used for the MEDLine database. (DOCX 13 kb)

Abbreviations

RCT: Randomised controlled trial; IQR: Interquartile range; OARS: Osteoarthritis Research Society International

Acknowledgements

Not applicable

Funding

This project is funded by a doctoral studentship from the EPSRC (Engineering and Physical Sciences Research Council) and Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences via the Medical Sciences Division of the University of Oxford. The funders had no input into the development of the protocol.

Availability of data and materials

Not applicable

Authors’ contributions

All authors read and approved the final manuscript. BC designed the study and drafted the manuscript. SD, RF and SL contributed to the design of the study, critical revision of manuscript and project supervision. JC contributed to the design and conception of the study, drafting of manuscript, critical revision of manuscript and project supervision.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. ²Nuffield Department of Population Health, University of Oxford, Oxford, UK.

Received: 12 September 2017 Accepted: 25 September 2017

Published online: 10 October 2017

References

- Cook JA, Hislop J, Adewuyi TE, Harrild K, Altman DG, Ramsay CR, Fraser C, Buckley B, Fayers P, Harvey I et al. Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference Elicitation in TriAls) review. *Health Technol Assess*. 2014, 18(28):v-vi, 1–175.
- Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*. Philadelphia: Lippincott Williams & Wilkins; 2013.
- Cook JA, Hislop J, Altman DG, Fayers P, Briggs AH, Ramsay CR, Norrie JD, Harvey IM, Buckley B, Fergusson D. Specifying the target difference in the primary outcome for a randomised controlled trial: guidance for researchers. *Trials*. 2015;16(1):12.
- Lenth RV. Some practical guidelines for effective sample size determination. *Am Statistician*. 2001;55(3):187–93.
- Altman DG. Statistics and ethics in medical research: III How large a sample? *BMJ*. 1980;281(6251):1336.
- Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA*. 2002;288(3):358–62.
- Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ*. 2009;338:b1732.
- 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:f1135.

9. Tavernier E, Giraudeau B. Sample size calculation: inaccurate a priori assumptions for nuisance parameters can greatly affect the power of a randomized controlled trial. *PLoS One*. 2015;10(7):e0132578.
10. Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol*. 2003;56(8):717–20.
11. Losina E, Ranstam J, Collins J, Schnitzer T, Katz J. OARSI clinical trials recommendations: key analytic considerations in design, analysis, and reporting of randomized controlled trials in osteoarthritis. *Osteoarthritis Cartilage*. 2015;23(5):677–85.
12. Keen HI, Pile K, Hill CL. The prevalence of underpowered randomized clinical trials in rheumatology. *J Rheumatol*. 2005;32(11):2083–8.
13. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8(1):18.
14. Hopewell S, Collins GS, Boutron I, Yu L-M, Cook J, Shanyinde M, Wharton R, Shamseer L, Altman DG. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ*. 2014;349:g4145.
15. Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. The quality of reporting in cluster randomised crossover trials: proposal for reporting items and an assessment of reporting quality. *Trials*. 2016;17(1):575.
16. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *BMJ*. 2009;339:b3496.
17. Cohen J. *Statistical power analysis for the behavioural sciences*. Hillside: Lawrence Erlbaum Associates; 1988.
18. StataCorp L. *Stata Statistical Software: Release 14*. College Station: StataCorp LP; 2015.
19. Williamson DF, Parker RA, Kendrick JS. The box plot: a simple visual method to interpret data. *Ann Intern Med*. 1989;110(11):916–21.
20. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA*. 2002;287(22):2973–82.
21. Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ*. 2013;346:f2820.
22. Wenner DM, Brody BA, Jarman AF, Kolman JM, Wray NP, Ashton CM. Do surgical trials meet the scientific standards for clinical trials? *J Am Coll Surg*. 2012;215(5):722.
23. Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. *BMJ*. 2012;344:e813.
24. Pengel LH, Barcena L, Morris PJ. The quality of reporting of randomized controlled trials in solid organ transplantation. *Transpl Int*. 2009;22(4):377–84.
25. Djulbegovic B, Lacey M, Cantor A, Fields KK, Bennett CL, Adams JR, Kuderer NM, Lyman GH. The uncertainty principle and industry-sponsored research. *Lancet*. 2000;356(9230):635–8.
26. Lundh A, Sismondo S, Lexchin J, Busuioic OA, Bero L. Industry sponsorship and research outcome. *Cochrane Libr*. 2012;12:MR000033.
27. Schott G, Pacht H, Limbach U, Gundert-Remy U, Lieb K, Ludwig W-D. The financing of drug trials by pharmaceutical companies and its consequences: part 2: a qualitative, systematic review of the literature on possible influences on authorship, access to trial data, and trial registration and publication. *Dtsch Arztebl Int*. 2010;107(17):295.
28. Guideline IHT. Choice of control group and related issues in clinical trials E10. Choice. 2000;E10:CPMP/ICH/364/96.
29. Mhaskar R, Djulbegovic B, Magazín A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol*. 2012;65(6):602–9.
30. Cook JA, Hislop JM, Altman DG, Briggs AH, Fayers PM, Norrie JD, Ramsay CR, Harvey IM, Vale LD. Use of methods for specifying the target difference in randomised controlled trial sample size calculations: two surveys of trialists' practice. *Clin Trials*. 2014;11(3):300–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

