Journal of Clinical Epidemiology 104 (2018) 52–61

Journal of Clinical Epidemiology

REVIEW ARTICLE

Sample size calculations are poorly conducted and reported in many randomized trials of hip and knee osteoarthritis: results of a systematic review

Bethan Copsey[a,*], Jacqueline Y. Thompson[a], Karan Vadher[a], Usama Ali[a], Susan J. Dutton[a], Raymond Fitzpatrick[b], Sarah E. Lamb[a], Jonathan A. Cook[a]

[a]*Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Botnar Research Centre, Windmill Road, Headington, Oxford OX3 7LD, UK*

[b]*Nuffield Department of Population Health, University of Oxford, Richard Doll Building, Old Road Campus, Oxford OX3 7LF, UK*

Accepted 17 August 2018; Published online 23 August 2018

Abstract

Objectives: To review the methodology and reporting of sample size calculations in a contemporary sample of trials in osteoarthritis.

Study Design and Setting: Randomized trials in hip and/or knee osteoarthritis published in 2016 were identified by searching MEDLINE, Cochrane library, CINAHL, EMBASE, PsycINFO, PEDro, and AMED until March 31, 2017. Data were extracted on study characteristics, methods used to calculate the sample size, and the reporting and justification of components used in the sample size calculation. We attempted to replicate the sample size calculation using the reported information.

Results: This review included 116 trials. Seventy-eight (67%, $n = 78/116$) reported a power calculation. Less than a quarter reported all core components of the sample size calculation (21%, $n = 16/78$). The sample size calculation was only reproducible in 53% of the trials that reported a power calculation ($n = 41/78$). The replicated calculation produced a sample size over 10% larger than the reported value in 12% of trials ($n = 9/78$). Insufficient information was reported to allow the sample size calculation to be replicated in a quarter of trials (27%, $n = 21/78$).

Conclusion: Sample size calculations in trials of hip and knee osteoarthritis are not adequately reported, and the calculation frequently cannot be reproduced.

Keywords: Sample size; Osteoarthritis; Reporting; Systematic review; Clinical trial; Research methods

1. Introduction

Sample size calculation is a key part of designing a clinical trial and is important for ethical, practical, and financial reasons. An overly large sample size can increase trial costs, delay dissemination of study findings, and result in more participants receiving a treatment when there is already sufficient evidence to show it is inferior to an alternative [1]. An overly small sample size can lead to underpowered trials that are more likely to ''miss'' a clinically important treatment effect, should it exist [2,3].

Altman et al. emphasized the importance of reporting the justification for the target sample size, especially when the trial does not recruit as many participants as planned [4]. When the sample size calculation is adequately reported, the reader can understand what the study was designed to achieve. The difference between the treatments that the trial was designed to statistically detect (the target difference), with associated assumptions, should be

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 study design, the collection, analysis and interpretation of data, the writing of the report, or in the decision to submit the article for publication. Sarah E Lamb, Ray Fitzpatrick, Usama Ali, and Bethan Copsey receive funding from the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust. Jonathan Cook is funded by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Conflict of interest: None.

* Corresponding author. Tel.: +44 0 1865 737923.

E-mail address: bethan.copsey@csm.ox.ac.uk (B. Copsey).

https://doi.org/10.1016/j.jclinepi.2018.08.013

What is new?

Key findings

- Sample size calculations in hip and knee osteoarthritis trials are often poorly reported, and where reported, the calculation often cannot be replicated.
- The standard deviation assumed in the sample size calculation was a poor estimate of the observed standard deviation in a substantial proportion of trials.

What this adds to what was known?

- This is the first review of sample size calculations in trials of hip and knee osteoarthritis and shows that the problems of poor reporting and lack of reproducibility of sample size calculations exist in this clinical area.

What is the implication and what should change now?

- Trialists and reviewers should ensure that sample size calculations are reported clearly and completely to facilitate the interpretation of trial results and prevent the conduct of underpowered trials.
- Trialists should perform a sensitivity analysis at the design stage to explore how a difference in the estimate of the standard deviation could affect the power of the study.

specified [5]. If well justified, the target difference can inform the interpretation of the trial findings, clarifying the presence (or absence) of a meaningful difference. Appropriate calculation of the sample size and reporting of the calculation help to avoid research waste, preventing the conduct of trials that are likely to produce inconclusive and potentially misleading results.

Previous systematic reviews have found that power calculations are often not performed, inadequately reported, or based on inaccurate assumptions [5−7]. A study may be underpowered if the parameters used to calculate its sample size are based on inaccurate assumptions [8−10]. Reviews of trials in a handful of specific conditions, such as back pain and rheumatology, have highlighted poor reporting of sample size calculations [11−13]. Focusing on a specific clinical area reduces the heterogeneity in the assumptions made in the sample size calculation. For instance, oncology trials are more likely to be powered on survival, which is not usually applicable to low-mortality conditions such as osteoarthritis [14].

We explored whether recently published osteoarthritis trials also poorly reported their sample size calculations. To our knowledge, the sample size calculations of hip and knee osteoarthritis trials have not previously been reviewed. Few reviews of any clinical area have attempted to replicate the sample size calculation of published trials [5,7,15,16]. Even fewer have compared the standard deviation assumed in the sample size calculation with the observed values in the trial results [7,9].

1.1. Objectives

Primary objective of this study was to summarize current practice in calculating the sample size for trials of hip and knee osteoarthritis, including the sample size, target difference, and justification for the chosen inputs.

Secondary objectives were to assess the reporting and reproducibility of these sample size calculations.

2. Materials and methods

The study methods were described in a published protocol and are summarized below [17].

2.1. Identification of studies

Seven databases were searched to identify relevant articles published in 2016: MEDLINE, Cochrane library (CENTRAL), CINAHL, EMBASE, PsycINFO, PEDro, and AMED (MEDLINE search strategy in Appendix B). The final search was performed on March 31, 2017, to allow for a 3-month lag between publication and database indexing.

2.2. Selection of studies

Abstracts and full texts were each screened independently by two of four reviewers (B.C., U.A., K.V., and J.Y.T.). Disagreements were resolved by discussion with a third reviewer (J.A.C.).

2.2.1. Inclusion criteria

Studies were eligible for inclusion if the article was the primary report of a randomized controlled trial of two treatment arms in a hip and/or knee osteoarthritis population. Included articles were published online or in a journal issue in 2016.

2.2.2. Exclusion criteria

The following article and study types were excluded:

- Conference abstracts
- Study protocols
- Non-English language articles
- Quasirandomized and nonrandomized studies
- Pilot and feasibility studies
- Factorial designs
- Cross-over trials
- Trials with three or more arms