

Title: CURRENT PRACTICE IN METHODOLOGY AND REPORTING OF THE SAMPLE SIZE CALCULATION IN RANDOMISED TRIALS OF HIP AND KNEE OSTEOARTHRITIS: A SYSTEMATIC REVIEW

Authors: Copsey B, Thompson JY, Vadher K, Ali U, Dutton SJ, Fitzpatrick R, Lamb SE, Cook JA.

Abstract: (3946 / 4000 characters including spaces)

Purpose (the aim of the study)

This study aims to review the methodology and reporting of sample size calculation in a contemporary sample of hip and knee osteoarthritis trials. The choice of sample size is a key aspect of designing a clinical trial, ensuring that the trial is powered to detect clinically important treatment effects, is ethically sound and does not overuse resources. Accurate and complete reporting of the sample size calculation allows the reader to interpret the trial results alongside the aims and assumptions made when the trial was designed.

Methods

This study is a systematic review of randomised controlled trials in hip and/or knee osteoarthritis published in 2016. Studies were identified by searching MEDLINE, Cochrane library, CINAHL, EMBASE, PsycINFO, PEDro and AMED. Data was extracted on study characteristics, methods used to calculate the sample size, and reporting and justification of individual components used in the sample size calculation. The reported information was used to replicate the sample size calculation. The standard deviation assumed in the sample size calculation was compared to the corresponding value in the study results. The results were synthesised using the number and proportion of studies for categorical outcomes and the median and interquartile range for continuous outcomes. Subgroup analysis was conducted to test for differences in reporting quality based on funding source, type of intervention and control treatment, and number of study centres.

Results

We identified 116 eligible trials for inclusion in this review. The majority were parallel-group, superiority, single-centre trials of knee osteoarthritis funded by non-industry sources. The number of randomised participants ranged from 20 to 633 (median 73). Of these, 78/116 (67%) reported a power calculation. Less than a quarter of studies reported all core components of the sample size calculation (21%, 16/78). The power and level of statistical significance were reported in almost all trials (96%, 75/78). The level of attrition assumed was not reported or unclear in a quarter of trials (27% n=21/78). Of trials powered on a single continuous outcome, two-thirds of studies reported the standard deviation (67%, 51/76) and less than half provided a justification for the value used (46%, 35/76). In trials with continuous outcomes, the justification for the mean difference was provided in 40 trials (53%, 40/76). The mean difference was most commonly based on the results of a previously published trial (29%, 22/76) or a published minimum clinically important difference (17%, 13/76).

Of the 78 trials which reported a power calculation, the sample size was only reproducible in half of the trials (53%, n=41/78) and assumptions were often needed when information was ambiguous or not reported. A quarter of trials did not provide sufficient information for the sample size calculation to be replicated (28%, n=22/78). The replicated calculation produced a sample size over 10% larger than the reported value in 12% of studies (n=9/78). There were 4 studies where the difference between the reproduced and reported sample size was greater than 50 participants.

When comparing the standard deviation of the primary outcome, the estimate used in the sample size calculation was accurate (within 10%) in only a third of studies where they could be compared (24%, n=9/29). In 6 studies, the follow-up standard deviation was over 30% larger than the value assumed in the sample size calculation leading to a reduction in power.

Conclusions

Sample size calculations in trials of hip and knee osteoarthritis are not reported adequately. Even where there is sufficient information, the calculation cannot be accurately reproduced. Furthermore,

the assumptions made can also be insufficient. This raises questions about the robustness of the findings, and whether these trials are able to address their stated principal aim.