Author/s:
Bunzli, S;Choong, E;Shadbolt, C;Wall, L;Nelson, E;Schilling, C;Wilding, H;Lohmander, LS;Balogh, ZJ;Paolucci, F;Clarke, P;Choong, PFM;Dowsey, MM

Title:
Placebo Surgery Controlled Trials: Do They Achieve What They Set Out to Do? A Systematic Review

Date:
2021-06-01

Citation:
Bunzli, S., Choong, E., Shadbolt, C., Wall, L., Nelson, E., Schilling, C., Wilding, H., Lohmander, L. S., Balogh, Z. J., Paolucci, F., Clarke, P., Choong, P. F. M. & Dowsey, M. M. (2021). Placebo Surgery Controlled Trials: Do They Achieve What They Set Out to Do? A Systematic Review. Annals of Surgery, 273 (6), pp.1102-1107. https://doi.org/10.1097/SLA.0000000000004719.

Annals of Surgery

DOI: 10.1097/SLA.0000000000004719

"Placebo surgery controlled trials: do they achieve what they set out to do? A systematic review"

*Dr Samantha Bunzli, PhD, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

Ms Emma Choong, B-BMED, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

Mr Cade Shadbolt, MA, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

Ms Laura Wall, BPsyc(Hons), The University of Newcastle Faculty of Business and Law, Newcastle Business School

Dr Elizabeth Nelson, PhD, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

Dr Chris Schilling, PhD, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

Ms Helen Wilding, GradDipInfoMgt, St Vincent's Hospital Melbourne, Library Service

Prof Stefan Lohmander, MD, University of Lund, Sweden, Department of Orthopaedics

Prof Zsolt Balogh, MD, The University of Newcastle, School of Medicine and Public Health

Prof Francesco Paolucci, PhD, The University of Newcastle Faculty of Business and Law, Newcastle Business School

Prof Philip Clarke, PhD, Oxford University, Health Economics Research Centre

Prof Peter Choong, MD, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

Assoc/Prof Michelle Dowsey, PhD, The University of Melbourne, Department of Surgery, St. Vincent's Hospital Melbourne

*Corresponding author

Dr. Samantha Bunzli, The University of Melbourne, Department of Surgery, St. Vincent's Hospital, Melbourne. Level 2, Clinical Sciences Building, 29 Regent street, Fitzroy 3065, Victoria, Australia. Email: samantha.bunzli@unimelb.edu.au. Ph. +61 3 9231 2365

Funding: This work was funded by an Australian National Health and Medical Research Council grant (APP1163613). The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication

Word count: 2893

Mini Abstract

We conducted a systematic review to determine whether placebo surgery controlled trials achieve what they set out to. Based on the analysis of 24 trials, we suggest that trialists factor 88% additional time and 87% more recruitment sites into the design of future trials to increase the likelihood of reaching completion.

Abstract

Objective: To explore whether placebo surgery controlled trials achieve what they set out to do by investigating discrepancy between projected and actual design aspects of trials identified through systematic review methods.

Summary background: Interest in placebo surgery controlled trials is growing in response to concerns regarding unnecessary surgery and the societal cost of low-value healthcare. As questions about the justifiability of using placebo controls in surgery have been addressed, attention is now being paid to more practical concerns.

Methods: Six databases were searched from inception - May 2020 (MEDLINE, Embase, Emcare, APA PsycInfo, CINAHL, Cochrane Library). Placebo surgery controlled trials with a published protocol were included. Three authors extracted 'projected' design aspects from protocols and 'actual' design aspects from main findings papers. Absolute and relative difference between projected and actual design aspects were presented for each trial. Trials were grouped according to whether they met their target sample size ('completed') and were concluded in a timely fashion. Pairs of authors assessed risk of bias.

Results: Of 24 trials with data available to analyse; three were completed and concluded within target timeframe; ten were completed and concluded outside the target timeline; four were completed without clear target timeframes; two were incomplete and concluded within the target framework; five were incomplete and concluded outside the target timeline. Trials which reached the recruitment target underestimated trial duration by 88% and number of recruitment sites by 87%.

Conclusions: Trialists need to factor additional time and sites into future placebo surgery controlled trials. A robust reporting framework of projected and actual trial design is imperative for trialists to learn from their predecessors.

Review Registration: PROSPERO (CRD42019133296)

Introduction

Economic pressures, variable patient satisfaction and arguable clinical benefits have placed the spotlight not only on the indication for certain surgical procedures in many conditions, but also the efficacy of surgery itself. In response, there has been a growing interest in the field of placebo surgery controlled trials[1-4].

Over the last decade, concerns about the safety and justifiability of placebo surgery controlled trials have received careful attention. Following a systematic review of placebo controlled trials of invasive procedures, Wartolowska et al[3] concluded that the risks associated with placebo procedures were generally small, and that such trials were a feasible way of testing efficacy. This was supported by a subsequent review by Probst et al[1], who concluded that placebo surgery controlled trials can be a valid way of testing surgical efficacy, provided that the ethical conditions of equipoise, consent, and risk minimisation are met.

As fundamental questions about the justifiability of placebo surgery controlled trials have been exhaustively addressed, greater attention is now being paid to more practical concerns. Challenges faced by those undertaking placebo controlled trials of invasive procedures have previously been explored in an extension to Wartolowska et al's[4] earlier review. None of the presumed challenges examined (i.e. funding source, anaesthesia, blinding procedures, and recruitment rate) were reported as obstacles in the main findings papers of the trials included in their analysis. This led the authors to conclude that placebo controlled trials are feasible, at least for minor procedures with lower levels of invasiveness. However, encountering challenges and reporting them in a main findings paper are two separate things, which may have resulted in Wartolowska et al[4] underestimating the barriers faced by those conducting these trials. Here, we take an alternative approach to highlighting potential barriers to the timely completion of placebo surgery controlled trials, by comparing features of trial planning and design reported in clinical trial registries or published protocols with those reported in manuscripts describing trial outcomes. We aim to: (i) document discrepancies between projected and actual aspects of trial design, and (ii) explore potential reasons for these discrepancies.

Methods

Protocol Design

We conducted a systematic review of published placebo surgery controlled trials, in accordance with the PRISMA Statement[5] and Synthesis Without Meta-analysis (SWiM) reporting guidelines[6]. This review was registered with PROSPERO (CRD42019133296).

Search Strategy

We systematically searched MEDLINE, Embase, Emcare, APA PsycInfo, CINAHL, and the Cochrane Library from inception to 21 May 2020 (see full search strategy in online supplement, http://links.lww.com/SLA/C871). Database searches used keywords and database specific vocabulary to cover two key conceptual groups: randomised controlled trials, and placebo surgery. Search strategies were not limited by language. In Endnote, duplicate references, records in languages other than English, French, Spanish or Italian, and records for irrelevant types of publications (e.g. case reports, editorials, comments) were

identified and removed. Reference management and database searches were conducted by a senior medical librarian (HW).

Selection criteria

Randomised controlled trials were eligible for inclusion only if: (i) the participants were adults aged >18 years; ii) they compared the efficacy of a surgical procedure to a placebo procedure, and (iii) a published protocol paper or clinical trial registry entry was available for the trial (see online supplement, http://links.lww.com/SLA/C871)

Adapting the definition of Probst et al[1], a surgical procedure was defined as a medical operation involving an anaesthesia and skin incision with instruments, to change the structural anatomy with an expectable physiological response in the target body compartment. We excluded trials of procedures which did not involve a skin incision to access the target tissue (e.g. bronchoscopy, endoscopy) and trials of procedures which did not seek to change the structural anatomy of the target body part (e.g. radiofrequency, neural ablation procedures).

A placebo procedure was initially defined restrictively, to only include participant-blinded surgical procedures involving anaesthesia and skin incision only, with no change to the structural anatomy and without an expectable physiological response in the target body compartment. While there is no consensus on the definition of a surgical placebo[7], our intent was to mimic as closely as possible the surgical version of a pharmaceutical placebo. We also excluded trials in which the placebo intervention was performed alongside a concomitant intervention that in and of itself may have a therapeutic effect. A post hoc change to this definition was made as only one study met our initial eligibility criteria. We broadened our definition to include any surgical procedure meant to mimic the index procedure, while excluding parts of the index procedure thought to be therapeutically necessary (see online supplement, http://links.lww.com/SLA/C871).

Study selection

Study selection was conducted in Covidence. Title and abstract screening (EN, EC, SB) and full-text screening (SB, EC, PCh, MD) were both conducted by a team of reviewers. At both stages of this process, each study was independently screened for inclusion by two members of the relevant team. Disagreements were resolved through discussion among the multi-disciplinary team. Studies excluded because of our initial definition of surgical placebos were rescreened following the post hoc change to the eligibility criteria. For each trial included following full-text screening, we searched for an entry in a clinical trial registry or published protocol paper. In the first instance, we looked for a trial registration number and/or published protocol paper cited in the main findings paper. Where this was not provided, we manually searched clinical trial registries (see online supplement, http://links.lww.com/SLA/C871) and for a published protocol paper using the authors names. Only trials with a registry entry or published protocol paper were included in our analysis.

Data extraction and Risk of Bias Assessment

Data extraction was conducted by a team of four reviewers (SB, EC, LW, CSh). Data from each study was extracted independently by two members of this team. Clinical and

methodological features of the studies (e.g. surgical specialty, description of the index and placebo procedures) were extracted from the main findings paper (the first publication reporting the primary outcomes of the trial). Data on planned trial design (e.g. sample size, time to conduct trial, number of sites) were extracted from the earliest entry in a clinical trial registry or, if this was not available, a published protocol. Data on actual trial design were extracted from the main findings paper or, when this was not possible, from the final entry in clinical trial registry. To account for differing means of defining surgical placebos, we also characterised eligible trials by drawing on the ASPIRE guideline's notion of placebos having differing levels of 'fidelity' to the index procedure[8]. Two reviewers (SB, CSh) individually characterised trials as high, medium or low fidelity, with any disagreement settled through discussion with a third reviewer (MD) (see online supplement, http://links.lww.com/SLA/C871).

Risk of bias assessment was conducted by a team of four reviewers (SB, EC, LW, MD) using the Cochrane Collaboration tool for randomised controlled trials[9]. Each study was assessed independently by two members of this team. The Cochrane Collaboration tool comprises of six domains, with provisions for assessing other sources of bias defined by the assessors. We included three additional domains: (i) sufficient sample size (we considered this to be different from incomplete outcome data as follow-up rates for enrolled participants were generally high), (ii) imbalance in the proportion of participants who underwent their allocated intervention (we considered this to be different from selection bias or incomplete outcome data) , and (iii) incidence of crossover (i.e. crossover between the study arms during the follow-up period where the threshold for cross over was lower in the placebo than the intervention arm. While ineffective blinding procedures may explain crossover, other factors such as treatment failure may also play a role and we therefore chose to include 'incidence of crossover' as an assessment criterion) . Discrepancies were resolved through discussion with the senior author (MD).

Data Synthesis

The main outcomes were trial completion and timely conclusion. Trials were considered complete only if their actual sample size was within 10% of the target sample size presented in their protocol. Trials were considered to have concluded in a timely fashion if the actual trial length was within 10% of the target trial length presented in their protocol. Importantly, at the conclusion of a trial, not all trials were deemed to have been completed (i.e. some failed to meet their target sample before ending the trial).

These outcomes were used to group trials for the purpose of narrative and tabular synthesis. To explore how trial characteristics differed between outcome groups, we reported the proportion of trials within each group. For each trial, we also presented the absolute and relative difference between projected and actual: time to completion; sample size; and number of recruitment sites. Due to a lack of reporting in trial protocols on projected time to recruit, data were presented for actual time to recruit only. Quantitative exploration of the association between trial characteristics and trial outcomes was not viable due to the limited number of eligible trials.

Results

The database search returned 12889 records. Following removal of duplicates and limiting to language and publication type, 5518 records underwent title and abstract screening and 170 underwent full text screening. The only trial which was originally eligible for inclusion in our analysis was a trial conducted by Roos et al.[10] (see online supplement, http://links.lww.com/SLA/C871).

Following adjustment to the eligibility criteria, we assessed the 73 articles which had not met our earlier definition of a placebo procedure against our broadened eligibility criteria. Twenty-nine trials met our broadened criteria following full text review, and 24 had a published protocol available (see Figure 1. PRISMA flow-chart).

< Insert Figure 1. PRISMA flow-chart about here >

The 24 trials could be grouped into three fields of research: orthopaedic surgery (n=9)[10-18]; cardiovascular surgery (n=8)[19-26] and neurosurgery (n=7)[27-33]. The trials were conducted in Europe (n= 7)[10,13,14,17,18,26,27]; the United States (n=7)[19,24,28,29,31-33]; the United Kingdom (n=2)[12,20]; Australia (n=2)[11,15]; Canada (n=1)[21]; New Zealand (n=1)[30] and internationally (n=4)[16,22,23,25]. The average number of study sites was 9 (range 1 - 32). The trials took on average 5 years (range 1.5 -10.3 years) to complete. The length of the recruitment period was on average 3.3 years (range 0.75 - 8.5 years) and the average sample size was 109 (range 18 - 313). The placebo procedure was classified as 'high fidelity' in 16 trials[11-15,17-27]; 'moderate fidelity' in 6 trials[28-33]; and 'low fidelity' in 2 trials[10,16] (see online supplement, http://links.lww.com/SLA/C871). A summary of each included trial is presented in Table 1.

< Insert Supplementary Table 1, http://links.lww.com/SLA/C872. Summary of placebo surgery controlled trials >

Outcomes from the risk of bias assessment are presented in Figure 2.

< Insert Figure 2. Risk of Bias Assessment about here >

Based on the mean percentage difference between target and actual outcomes among the trials which reached the recruitment target and had available data (Table 2; Completed trials, concluded within target timeframe and Completed trials, not concluded within target timeframe, http://links.lww.com/SLA/C873), trial duration was underestimated by 88% and the number of recruitment sites was underestimated by 87%.

< Insert Supplementary Table 2. Completion outcomes with reference to target timeframe >

Completed trials, concluded within target timeframe

Three trials (n=3/24, 13%) were completed in a timely fashion[19,21,28] (see Supplementary Table 2; Completed trials, concluded within target timeframe, http://links.lww.com/SLA/C873) When qualitatively compared to incomplete trials and trials completed outside of their target timeframe, these trials appeared to have a smaller target sample size on average (63 vs 146). Only one of these trials (n=1/3, 33%) required more trial sites than anticipated[28]. Two trials scored unclear risk of bias on criterion six: 'selective

reporting'[19,21]. One trial scored 'unclear' risk of bias on criterion seven: 'sufficient sample size' as the effect size used in the power calculation was not justified[28].

Completed trials, not concluded within target timeframe

Ten completed trials were not concluded within their target timeframe (n=10/24, 42%)[11-14,17,18,20,25,30,33] (see Supplementary Table 2; Completed trials, not concluded within target timeframe, http://links.lww.com/SLA/C873). On average, they took 2.6 years longer to complete than anticipated, and in six trials (n=6/10, 60%) the time required to recruit participants exceeded the target duration of the entire trial[11-14,20,33]. Five trials (n=5/10, 50%) required additional trial sites to meet their target sample[12-14,20,25], with two trials adding 18 sites over the duration of the trial[12,25]. Two trials scored low risk of bias on all criteria[14,20]. Risk of bias was introduced in one criterion for one trial[17] and in two or more criteria for seven trials[11-13,18,25,30,33].

Completed trials, without clear timeframe

One completed trial (n= 1/24, 4%) did not have available data on the actual length of the trial[27], and three (n=3/24, 12.5%) did not provide details on expected trial length[22,29,34] (see Supplementary Table 2; Completed trials, without clear timeframe, http://links.lww.com/SLA/C873). One single site trial took 2.3 years to recruit the sample[27]. An international trial which involved 22 recruitment sites, took 0.75 years to recruit the sample[22]. Two trials scored low risk of bias[22,34] and two trials were unclear risk of bias on one criterion[27,29].

Incomplete trials, concluded within target timeframe

Two trials (n= 2/24, 8%) failed to reach 50% of the target sample size and concluded within the target timeframe[10,15] (see Supplementary Table 2; Incomplete trials, concluded within target timeframe, http://links.lww.com/SLA/C873). One trial scored high risk of bias on one criterion[15] and the other scored high risk of bias on five criteria[10].

Incomplete trials, not concluded within target timeframe

Five trials (n= 5/24, 21%) were not completed and concluded outside of the target timeframe[23,24,26,32,35] (see Supplementary Table 2; Incomplete trials, not concluded within target timeframe, http://links.lww.com/SLA/C873). These trials fell short of the target sample size by an average of 44% despite recruiting for an average of 3.9 years and extending the trial duration by an average of 53%. In addition, three trials (n=3/24, 12.5%) added recruitment sites over the duration of the trial[23,24,32], with one trial adding 28 sites[24]. Only one trial reported a planned interim analysis to review evidence of treatment efficacy or adverse effects[35]. Following interim analysis of the Kallmes et al. trial [35] a decision was made to revise down the power calculation to reduce the target sample size due to early difficulties with recruitment. For all five trials, high risk of bias was introduced on criterion seven: 'sufficient sample size'. In four trials, high risk of bias was introduced in two or more criteria[24,26,32,35].

Discussion

Based on our analysis of completed trials with published findings, we suggest that trialists are currently underestimating trial duration by 88% and the number of recruitment sites by 87%. Eighty-seven percent of trials were not completed within the projected timeframe and 29% did not meet their target sample size by the conclusion of the trial. These are likely to be conservative estimates given that this review did not take into consideration trials which were abandoned prior to collecting sufficient data to warrant publication. Future studies may be able to avoid this issue by initially searching for placebo surgery controlled trials in Clinical Trial Registries. This approach may identify additional failed trials, which never published any results and could also provide insight into the experiences of trials which are open and in the process of recruiting. However, as our review was able to identify six published trials that met all inclusion criteria other than having a registry entry[36-41], this approach is unlikely to identify all relevant trials.

In addition to highlighting difficulties in reaching completion, the findings of this review underscore the need for improved standards in the reporting of projected and actual aspect of trial design.

That only two trials met our initial definitions of 'surgical procedure' and 'placebo procedure' is an important finding. In our systematic review, we differentiated between 'surgical procedures' involving a skin incision and 'invasive procedures' involving access to the body via natural orifice (e.g. endoscopic procedures) or injection. As such, minimally invasive procedures included in previous reviews[2,4] were excluded from our review. The majority of studies we screened failed to meet our initial definition of a surgical placebo as an anaesthetic and skin incision only, with no change to the structural anatomy and without an expectable physiological response in the target body compartment. So why do placebo surgery controlled trials overwhelmingly favour invasive placebos? It may be that more invasive placebo procedures are necessary to ensure participant and assessor blinding. Another reason why trialists may favour more invasive placebo procedures is the perception that they 'offer' something to participants randomised to the placebo arm e.g. a diagnostic image of the body part[42-44]. The perception of 'benefit' may make these trials more acceptable in the eyes of institutional review boards, surgeons and patients, which in turn may facilitate timely completion. However, as Beard et al.[8] have also discussed, 'inducing' people to participate in a trial by the perception that a benefit exists, is ethically problematic given that equipoise is a prerequisite condition. Unfortunately, the number of trials available to make comparisons between studies involving high, moderate and low fidelity placebo procedures was small. Updating this review when more trials have been completed may allow for quantitative exploration of the association between placebo invasiveness and trial completion.

Almost half of the included trials were at high risk of being underpowered. Failing to reach adequate power means that a trial is unable to test the *a priori* trial hypothesis, thereby risking that the trial has been done for no potential knowledge gain. Our analysis highlights that reaching sufficient power before funds run out presents an optimisation problem for trialists. In addition to lengthening trial timelines, strategies trials employed to reach completion included adding recruitment sites and revising down the sample size calculation.

While lengthening the trial timeline and adding sites is associated with additional financial costs, factoring sufficient time and trial sites into future funding applications is important. While we did not locate funding information for each of the included trials, it is important to consider the role that funding plays in trial outcomes. Retrieving detail about the source and amount of funding a trial received presents a challenge. While information about publicly funded trials can be readily accessed, it is common for trials to seek additional funding from other sources over the duration of the trial. Asking trialists to share information about the projected and actual costs is one potential strategy that would provide useful insight into the economic impact of extending trial timelines and adding recruitment sites.

Slow recruitment is unlikely to fully account for discrepancy between projected and actual trial duration. Additional barriers to timely completion faced by the trials in this review may have included, for example, delays in approval by institutional review boards, difficulties recruiting surgeons, and time taken to analyse data or write manuscripts. In this review, we took authors' reports of 'trial length' on face value (e.g. "*This trial took place between January 2016 and March 2018*"). However it was not clear what milestones these dates correspond to; it may be that in some cases these dates corresponded with the start of recruitment (first patient in) to last follow-up (last patient out); while in other cases they may have corresponded to the date of ethical approval and completion of data analysis. Given that the latter can add significant time to the trial length, this is likely to have been a source of heterogeneity in our data.

While all of the included trials had a protocol registered in a clinical trial database, the quality of information provided in the protocols varied widely. Further, protocol papers were often published after the trial had commenced when recruitment was already underway and thus were likely to have reported a 'revised' timeline. Routinely including detail in trial protocols about anticipated sample size (accompanied by justified sample size calculation), number of recruitment sites, length of the recruitment period and trial duration as well as any changes to these anticipated aspects of trial design, would enable future trialists to learn from the experiences of their predecessors and build more realistic timeframes into the design of placebo surgery controlled trials.

The lack of explicit requirement to report the nature of any challenges incurred over the course of the trial makes it difficult for trialists to learn from the experiences of others. We advocate for embedding a robust reporting framework into existing checklists such as the CONSORT checklist[45]. Explicitly providing information on projected versus actual sample size (including justification for adjustments to the sample size calculation), number of recruitment sites, length of the recruitment period and trial duration will not only assist trialists to design trials which are more likely to be completed, but will also enable readers to make more informed judgements about trial quality. Based on the findings of this review we suggest that trialists are currently underestimating trial duration by at least 88% and the number of recruitment sites by at least 87%. We recommend that trialists, grant reviewers and funders factor additional time and recruitment sites into the design and budgets of proposed placebo surgery controlled trials to increase the likelihood of reaching the target sample size before running out of time and/or money.

References:

1. Probst P, Grummich K, Harnoss JC, et al. Placebo-Controlled Trials in Surgery: A Systematic Review and Meta-Analysis. *Medicine (Baltimore).* 2016;95(17):e3516.
2. Cousins S, Blencowe NS, Tsang C, et al. Reporting of key methodological issues in placebo-controlled trials of surgery needs improvement: a systematic review. *J Clin Epidemiol.* 2020;119:109-116.
3. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. *BMJ.* 2014;348:g3253.
4. Wartolowska K, Collins GS, Hopewell S, et al. Feasibility of surgical randomised controlled trials with a placebo arm: a systematic review. *BMJ Open.* 2016;6(3):e010194.
5. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Ann Intern Med.* 2009;151:W-65 - W-94.
6. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *Bmj.* 2020.
7. Nunn R. It's time to put the placebo out of our misery. *BMJ* 2009;338:b1568.
8. Beard DJ, Campbell MK, Blazeby JM, et al. Considerations and methods for placebo controls in surgical trials (ASPIRE guidelines). *The Lancet.* 2020;395(10226):828-838.
9. Higgins J, Altman D, Gotzsche P, al. e. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343(d5928).
10. Roos EM, Hare KB, Nielsen SM, Christensen R, Lohmander LS. Better outcome from arthroscopic partial meniscectomy than skin incisions only? A sham-controlled randomised trial in patients aged 35–55 years with knee pain and an MRI-verified meniscal tear. *BMJ Open.* 2018;8(2).
11. Clark W, Bird P, Gonski P, et al. Safety and efficacy of vertebroplasty for acute painful osteoporotic fractures (VAPOUR): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet.* 2016;388(10052):1408-1416.
12. Beard DJ, Rees JL, Cook JA, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. *Lancet.* 2018;391(10118):329-338.
13. Paavola M, Malmivaara A, Taimela S, et al. Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial. *BMJ.* 2018;362:k2860.

14. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med.* 2013;369(26):2515-2524.

15. Buchbinder R, Osborne RH, Ebeling PR, et al. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med.* 2009;361(6):557-568.

16. Kallmes DF, Comstock BA, Heagerty PJ, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med.* 2009;361(6):569-579.

17. Firanescu CE, de Vries J, Lodder P, et al. Vertebroplasty versus sham procedure for painful acute osteoporotic vertebral compression fractures (VERTOS IV): randomised sham controlled clinical trial. *BMJ.* 2018;361:k1551.

18. Schroder CP, Skare O, Reikeras O, Mowinckel P, Brox JI. Sham surgery versus labral repair or biceps tenodesis for type II SLAP lesions of the shoulder: a three-armed randomised clinical trial. *BJSM online.* 2017;51(24):1759-1766.

19. Siddiqui AH, Zivadinov R, Benedict RH, et al. Prospective randomized trial of venous angioplasty in MS (PREMiSe). *Neurology.* 2014;83(5):441-449.

20. Al-Lamee R, Thompson D, Dehbi HM, et al. Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial. *Lancet.* 2018;391(10115):31-40.

21. Traboulsee AL, Machan L, Girard JM, et al. Safety and efficacy of venoplasty in MS. *Neurology.* 2018;91(18):e1660-e1668.

22. Feldman T, Mauri L, Kahwash R, et al. Transcatheter Interatrial Shunt Device for the Treatment of Heart Failure With Preserved Ejection Fraction (REDUCE LAP-HF I [Reduce Elevated Left Atrial Pressure in Patients With Heart Failure]): A Phase 2, Randomized, Sham-Controlled Trial. *Circulation.* 2018;137(4):364-375.

23. Verheye S, Jolicoeur EM, Behan MW, et al. Efficacy of a device to narrow the coronary sinus in refractory angina. *N Engl J Med.* 2015;372(6):519-527.

24. Tobis JM, Charles A, Silberstein SD, et al. Percutaneous Closure of Patent Foramen Ovale in Patients With Migraine: The PREMIUM Trial. *J Am Coll Cardiol.* 2017;70(22):2766-2774.

25. Witte KK, Lipiecki J, Siminiak T, et al. The REDUCE FMR Trial: A Randomized Sham-Controlled Study of Percutaneous Mitral Annuloplasty in Functional Mitral Regurgitation. *JACC: Heart Failure.* 2019;7(11):945-955.

26. Zamboni P, Tesio L, Galimberti S, et al. Efficacy and safety of extracranial vein angioplasty in multiple sclerosis: A randomized clinical trial. *JAMA Neurol.* 2018;75(1):35-43.

27. Boelens OB, van Assen T, Houterman S, Scheltinga MR, Roumen RM. A double-blind, randomized, controlled trial on surgery for chronic abdominal pain due to anterior cutaneous nerve entrapment syndrome. *Annals of Surgery.* 2013;257(5):845-849.

28. Gross RE, Watts RL, Hauser RA, et al. Intrastriatal transplantation of microcarrier-bound human retinal pigment epithelial cells versus sham surgery in patients with advanced Parkinson's disease: a double-blind, randomised, controlled trial. *The Lancet Neurology.* 2011;10(6):509-519.

29. LeWitt PA, Rezai AR, Leehey MA, et al. AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial. *The Lancet Neurology.* 2011;10(4):309-319.

30. Snow B, Mulroy E, Bok A, et al. A phase IIb, randomised, double-blind, placebo-controlled, dose-ranging investigation of the safety and efficacy of NTCELLÂ® [immunoprotected (alginate-encapsulated) porcine choroid plexus cells for xenotransplantation] in patients with Parkinson's disease. *Parkinsonism & related disorders.* 2019;61:88-93.

31. Marks WJ, Bartus RT, Siffert J, et al. Gene delivery of AAV2-neurturin for Parkinson's disease: a double-blind, randomised, controlled trial. *The Lancet Neurology.* 2010;9(12):1164-1172.

32. Olanow CW, Bartus RT, Baumann TL, et al. Gene delivery of neurturin to putamen and substantia nigra in Parkinson disease: A double-blind, randomized, controlled trial. *Ann Neurol.* 2015;78(2):248-257.

33. Rafii MS, Tuszynski MH, Thomas RG, et al. Adeno-Associated Viral Vector (Serotype 2)-Nerve Growth Factor for Patients With Alzheimer Disease: A Randomized Clinical Trial. *JAMA Neurol.* 2018;75(7):834-841.

34. Marks WJ, Jr., Bartus RT, Siffert J, et al. Gene delivery of AAV2-neurturin for Parkinson's disease: a double-blind, randomised, controlled trial. *Lancet Neurology.* 2010;9(12):1164-1172.

35. Kallmes D, Comstock B, Heagerty P, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *The New England Journal of Medicine.* 2009;361:569-579.

36. Freed C, Greene P, Breeze R, et al. Transplantation of embryonic dopamine neurons for severe parknsons disease. *The New England Journal of Medicine.* 2001;344(10):710-719.

37. Jarrell J, Mohindra R, Ross S, Taenzer P, Brant R. Laparoscopy and Reported Pain among Patients with Endometriosis. *Journal of Obstetrics and Gynaecology Canada.* 2005;27(5):477-485.

38. Molegraaf MJ, Torensma B, Lange CP, Lange JF, Jeekel J, Swank DJ. Twelve-year outcomes of laparoscopic adhesiolysis in patients with chronic abdominal pain: A randomized clinical trial. *Surgery.* 2017;161(2):415-421.

39. Abbott J, Hawe J, Hunter D, Holmes M, Finn P, Garry R. Laparoscopic excision of endometriosis: a randomized, placebo-controlled trial. *Fertil Steril.* 2004;82(4):878-884.

40. Moseley B, O'Malley K, Peterson N, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *The New England Journal of Medicine.* 2002;347(2):81-88.

41. Kroslak M, Murrell GAC. Surgical Treatment of Lateral Epicondylitis: A Prospective, Randomized, Double-Blinded, Placebo-Controlled Clinical Trial. *Am J Sports Med.* 2018;46(5):1106-1113.

42. Campbell MK, Entwistle VA, Cuthbertson BH, et al. Developing a placebo-controlled trial in surgery: issues of design, acceptability and feasibility. *Trials.* 2011;12:50.

43. Gooberman-Hill R, Jinks C, Boucas SB, et al. Designing a placebo device: involving service users in clinical trial design. *Health Expect.* 2013;16(4):e100-110.

44. Baldwin MJ, Wartolowska K, Carr AJ. A survey on beliefs and attitudes of trainee surgeons towards placebo. *BMC Surgery.* 2016;16(1).

45. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;267(8):637-639.

Figure legends:

Figure 1. PRISMA flow diagram of study selection process

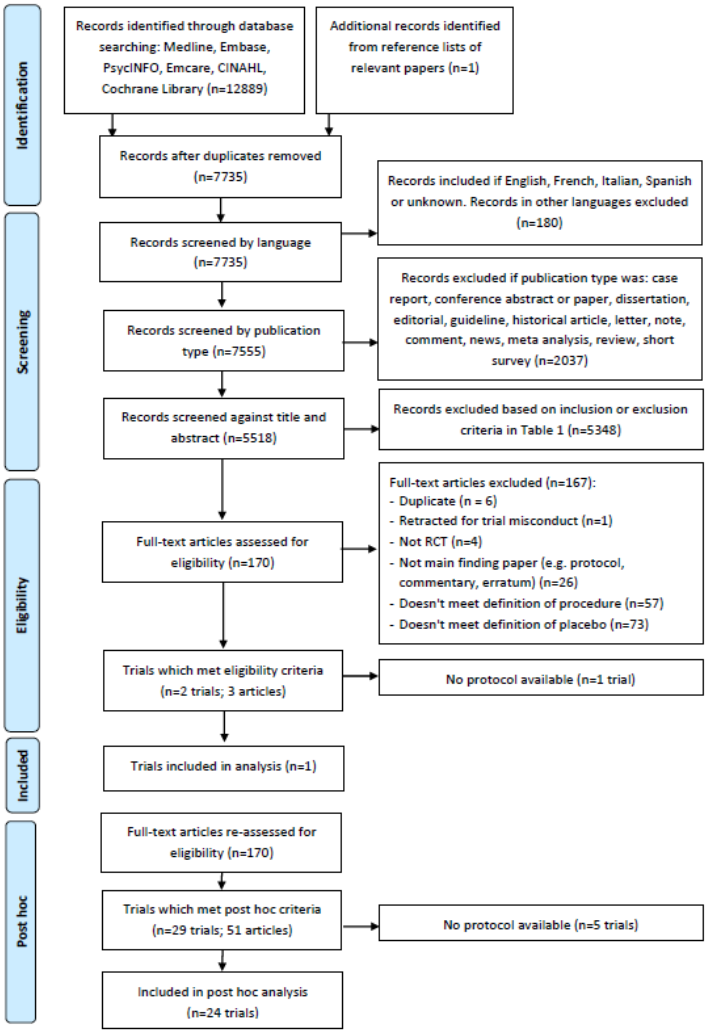

Figure 2. Risk of Bias Assessment.

+ Low Risk of bias; − High risk of bias; ? Unclear risk of bias