

BMJ Open Feasibility of surgical randomised controlled trials with a placebo arm: a systematic review

Karolina Wartolowska,^{1,2} Gary S Collins,^{2,3} Sally Hopewell,^{2,3} Andrew Judge,^{1,2,4} Benjamin J F Dean,^{1,2} Ines Rombach,^{1,2} David J Beard,^{1,2,5} Andrew J Carr^{1,2,5}

To cite: 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**:e010194. doi:10.1136/bmjopen-2015-010194

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-010194>).

Received 6 October 2015
Revised 19 January 2016
Accepted 15 February 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Karolina Wartolowska;
karolina.wartolowska@ndorms.ox.ac.uk

ABSTRACT

Objectives: To find evidence, either corroborating or refuting, for many persisting beliefs regarding the feasibility of carrying out surgical randomised controlled trials with a placebo arm, with emphasis on the challenges related to recruitment, funding, anaesthesia or blinding.

Design: Systematic review.

Data sources and study selection: The analysis involved studies published between 1959 and 2014 that were identified during an earlier systematic review of benefits and harms of placebo-controlled surgical trials published in 2014.

Results: 63 trials were included in the review. The main problem reported in many trials was a very slow recruitment rate, mainly due to the difficulty in finding eligible patients. Existing placebo trials were funded equally often from commercial and non-commercial sources. General anaesthesia or sedation was used in 41% of studies. Among the reviewed trials, 81% were double-blinded, and 19% were single-blinded. Across the reviewed trials, 96% (range 50–100%) of randomised patients completed the study. The withdrawal rate during the study was similar in the surgical and in the placebo groups.

Conclusions: This review demonstrated that placebo-controlled surgical trials are feasible, at least for procedures with a lower level of invasiveness, but also that recruitment is difficult. Many of the presumed challenges to undertaking such trials, for example, funding, anaesthesia or blinding of patients and assessors, were not reported as obstacles to completion in any of the reviewed trials.

INTRODUCTION

Progress in surgery is based on practical experience.¹ Surgical randomised controlled trials (RCTs) are uncommon;² only about 15% of published RCTs are related to surgical interventions.³ Novel procedures tend to be developed through an iterative process of trial and error,⁴ and only 24% of the currently used surgical therapies are supported by results of RCTs.¹

Strengths and limitations of this study

- Review of all published surgical randomised controlled trials (RCTs) with a placebo arm, spanning the years 1959 to 2014.
- Owing to the nature of this review, we could not investigate the obstacles that prevented initiation or completion of trials and, subsequently, our observations are limited to the successfully published trials. However, this review of all published trials provides different evidence than a report from a single discontinued trial.
- Many of the problems reported in reviewed trials are not unique to placebo-controlled surgical trials, but are also relevant to other surgical trials and randomised controlled trials in general.

Apart from not being necessary for approval of new treatment,⁵ several reasons have been mentioned in the literature that may explain why surgical RCTs are scarce. Such studies are perceived as expensive^{2 6} and unlikely to attract funding.^{3 5 7} They are considered to be difficult to design and conduct because of challenges posed by randomisation, blinding, differences in skills and experience of surgeons, variability of patients, as well as lack of consensus on surgical outcomes.^{1 2 6–8} Moreover, patient recruitment is also believed to be a problem.⁶ The inclusion of a placebo control adds another level of complexity to a RCT.^{6 9} For example, some authors suggest that many patients may be unwilling to undergo an invasive procedure if there is no clear direct benefit to them, which may result in slow recruitment.⁶ Others believe blinding of patients and outcome assessors is not feasible, and that the surgeon can never be blinded.¹⁰ As a result of that, very few interventional procedures have been validated using a placebo-controlled RCT.^{1 2 5 9 11 12} It is important to note that some of these opinions come from personal experience

from a single trial, while others are just perceptions and assumptions. There have also been many publications discussing placebo in surgery that concentrate on ethical concerns, such as general equipoise and minimising the risks,^{13 14} and on conceptual problems, for example, whether surgeons will be willing to test efficacy of an already established procedure.^{10 15} However, very little has been written on the methodological challenges of such studies^{6 16} and, to the best of our knowledge, no one has attempted to summarise the evidence from all the published placebo-controlled surgical trials.

When we previously performed a systematic review examining the harms and benefits of placebo-controlled surgical RCTs, we found that there clearly are obstacles to completing such trials, as less than a hundred have been published between 1959 and 2013.¹² Therefore, we conducted a secondary review of these studies to find evidence corroborating or refuting persisting beliefs regarding the feasibility of carrying out placebo-controlled surgical trials.

METHODS

Selection criteria

The criteria used to select placebo-controlled surgical RCTs have been described previously.¹² In brief, studies were eligible if they were randomised trials in which the efficacy of surgery was compared with placebo. Surgery was defined as any interventional procedure that changes the anatomy and requires a skin incision or the use of endoscopic techniques; dental studies were excluded. We used the term 'placebo' to refer to a surgical placebo, a sham surgery, or a procedure intended to mimic the active intervention. A quasi-placebo, that is, a diagnostic procedure that could imitate the surgery, was also included. The important criterion was that patients were under general anaesthesia or blinded in some other way, and could not distinguish whether they underwent the actual surgery or placebo. We did not limit the inclusion criteria to any particular condition, patient group, intervention, or type of outcome. We excluded studies investigating anaesthesia or other pharmacological substances used perioperatively.

In this review, we used the term 'surgical placebo'. The word 'sham' is preferred by some authors because surgical placebo has to involve an imitation of the investigated intervention in order to resemble it closely; therefore, it is different from an inactive 'sugar pill' placebo used in pharmacological trials.¹⁷ The word 'sham' has negative associations, and it suggests that a procedure is fake and deceitful; however, in many trials, the placebo involved an accepted surgical procedure such as endoscopy or arthroscopy, which was used also for diagnostic purposes with real benefits to the patients.

Search strategy

We searched the MEDLINE, EMBASE and the Cochrane Central Register of Controlled Trials

databases from the date of their inception to 14 November 2013, with no restriction on language. We did not systematically search for studies reported only as conference abstracts. Search terms were published previously.¹²

Three reviewers (KW, IR and BJFD) independently screened the initial set of records identified from the search, and then screened the full text of any potentially relevant articles. Each reviewer assessed the eligibility of each study, and the final list of included studies was agreed by consensus. Moreover, we searched ClinicalTrials.gov (on 14 November 2013), a database of registered randomised clinical trials, to identify any recently completed or ongoing studies. On 15 June 2014, we checked whether results of any of the trials identified in the ClinicalTrials.gov database have been published since the original search.

Dealing with duplicate publications

When there were several articles reporting outcomes from a single trial, that is, with the same authors, location, patient population and recruitment dates, we only included the paper reporting the primary outcome for the trial, and excluded pilot and follow-up reports.

Data extraction

We used a standardised data extraction form to collect information about the characteristics of each study including: year of publication, country in which the trial was conducted, funding source, details of the active and placebo intervention as well as the type of anaesthesia, blinding, number of patients who were assessed, eligible, randomised and who declined participation, as well as those who completed the trial. To reduce errors, the three review authors (KW, IR and BJFD) extracted data separately and checked the entries for consistency; a single set of data was agreed by all three reviewers.

Data synthesis

We have performed a descriptive analysis of the characteristics of each individual study and presented data in a table.

RESULTS

Study selection

We analysed the studies identified as a part of the systematic review on harms and benefits, including seven trials that were excluded from the systematic review due to lack of a direct comparison between the surgical and the placebo group. We also checked whether the trials identified in the ClinicalTrials.gov database had their results published between November 2013 and June 2014, and found three additional trials.^{18–20} This resulted in 63 full-text articles, which were included in this review (figure 1).

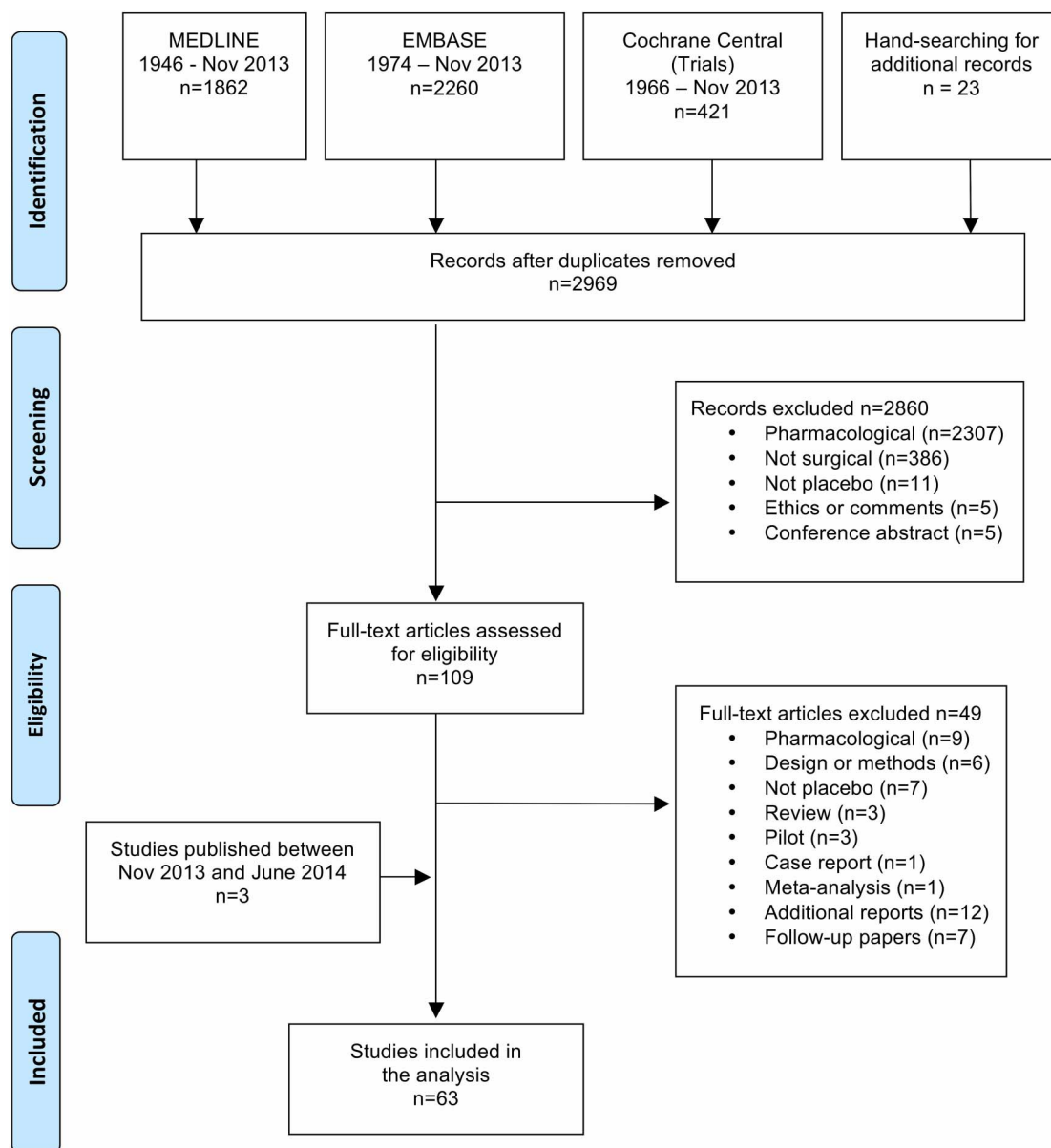


Figure 1 PRISMA flow diagram.

Placebo-controlled surgical RCT characteristics

The number of published placebo-controlled surgical trials was small; however, 73% (n=46/63) of included RCTs were published after the year 2000, suggesting an increasing interest in performing such studies. Half the trials (n=35/63, 55%) used a key-hole surgery, including endoscopy (n=28/63), laparoscopy (n=4/63), arthroscopy (n=2/63) and bronchoscopy (n=1/63). The remaining trials involved other types of minimally invasive interventions, for example, using catheters for vascular access, or needles for injection of fat or exogenous materials to remodel tissue. Very few studies investigated open techniques, such as exposure of the internal mammary artery (n=2/63), or exposure of scalp muscles (n=1/63). Fifteen trials used implants and an additional seven used gastric balloons or bubbles (characteristics of

the reviewed trials are presented in online supplementary appendix 1).

Funding sources

One-third of the studies (n=21/63, 33%) were non-commercially funded, and almost as many were funded by a commercial company (n=18/63, 29%), often the manufacturer of the implant or the endoscope. The source of funding in the remaining studies (n=24/63, 38%) was not reported.

Over half the trials were undertaken in the USA (n=35/63, 56%), the others were in Canada, the UK, Germany, The Netherlands, Belgium, Italy, Finland, Sweden, Portugal, Norway, Greece, Denmark, Australia and Brazil.

Sample size

The majority (n=47/63, 75%) of the identified studies were small, with fewer than 100 participants. The median number of patients randomised in a trial was 61 (IQR 66, range 10–298).

About half the RCTs (n=33/63; 52%) reported a formal sample size calculation, but only a quarter (n=16/63) allowed for dropouts and attrition. Most of the trials that included a sample size calculation (n=23/33, 70%) attained their prespecified sample size (without accounting for attrition). Ten trials under-recruited, such that the number of randomised patients was lower than the calculated sample size. All 10 trials were terminated early: three due to slow recruitment,^{21–23} one because at the interim analysis the surgery was highly effective,²⁴ and two because at the interim analysis the active procedure lacked efficacy.^{18 25} Two studies were stopped because of serious adverse events either in the trial²⁶ or at another centre using a similar procedure.²⁷ One trial was terminated when the sponsoring company was sold.²⁸ Finally, one study was stopped because the investigated procedure was approved as a standard care and the equipoise ceased to exist, despite the fact that the study did not show its superiority over placebo.²⁹ Finally, one trial recruited the intended sample size, but due to a high drop-out rate the number of patients who completed the trial was lower than the required sample size.³⁰

Recruitment and screening

Recruitment, sometimes as slow as 1–2 patients per month,^{21 31} was a common problem,^{18 21 27–35} and was the reason for an early termination of three trials.^{21–23}

Many of the analysed studies did not provide any details about screening and recruitment; they either stated that they recruited consecutive patients fulfilling the criteria,²¹ or that they randomised patients who were willing to participate and were eligible.³⁶ About one-third of the trials specified the number of screened (n=24/63, 38%) and eligible patients (n=27/63, 43%), and stated how many patients declined participation or withdrew before the treatment (n=22/63, 34%); only one-fifth of the trials (n=13/63, 21%) reported all three numbers (see online supplementary appendix 1). The available data suggest that the initial assessment of

eligibility was the main obstacle in recruitment as patients did not meet the inclusion criteria or were not eligible due to exclusion criteria.

On average, it was necessary to screen more than five patients in order to randomise one, but three in four eligible patients started the trial (table 1). The number of patients who had to be screened before the necessary group was recruited varied greatly. This variance was, at least partly, related to the method of identifying potential participants. The trial with the largest number of screened patients recruited participants by using TV and newspaper advertising: out of 4523 screened patients only 260 were eligible and were willing to participate; however, 196 had negative discography, and only 64 patients were randomised.²³ More targeted recruitment from specialist centres had much higher success rates, but often required a multisite effort.³⁵

Many trials had additional inclusion and exclusion criteria that could only be verified after the patient entered the trial, for example, a verification of diagnosis by positive findings during the endoscopy, or on diagnostic imaging. As a consequence of this, many patients were excluded because either they did not have the investigated condition or they had some concomitant condition that precluded their participation in the trial and, sometimes, required appropriate treatment. Moreover, any technical complications during the assessment or study procedures potentially resulted in patients' drop-out. For example, in the trial on laparoscopic adhesiolysis for abdominal pain by Swank and colleagues,³² nine patients did not have adhesions, one of them had a hernia and was treated laparoscopically, three patients had stricturing adhesions that required therapeutic adhesiolysis, and in one instance a pneumoperitoneum could not be achieved; therefore, out of 121 assessed patients, 13 were excluded during laparoscopy.

Fluctuating symptoms were a problem in a few studies, for example, patients became asymptomatic while waiting for the procedure and had to be excluded from the trial,¹⁹ or did not report symptoms during the study visit, and did not undergo the treatment, but were included in the intention-to-treat analysis.³⁷ This problem also complicated the post-treatment assessment,³⁸ especially that only one trial included an observational control group.³⁹

Table 1 Participants flow through the reviewed trials as a percentage of the number of patients who were randomised into each trial

| | Number of studies | Median per cent of sample randomised | First and third quartile (%) | Minimum and maximum (%) |
|------------------|-------------------|--------------------------------------|------------------------------|-------------------------|
| Screened | 24 | 530 | 243, 773 | 100, 7067 |
| Eligible | 27 | 132 | 108, 172 | 100, 448 |
| Declined | 22 | 18 | 8, 144 | 3, 4942 |
| Sample size | 33 | 96 | 90, 110 | 49, 323 |
| Outcome assessed | 61 | 96 | 90, 100 | 52, 100 |

'Number of studies' refers to the number of trials that provided relevant data. Trials terminated early are included in these analyses. 'Sample size' refers to the sample size required to reach statistical power, not inflated to account for drop-outs. 'Outcome assessed' refers to total number of patients, in both arms. The denominator is the number of patients actually randomised into each trial.

Refusal to participate

Some of the approached patients declined participation in the trial, withdrew their initial consent, refused to be randomised or to comply with the requirements of the protocol, and had a strong preference for one of the treatment options. Most of the trials did not report the reasons for patients' refusal to participate, and the available data did not allow us to quantify the percentage of patients who refused to enter the study. Only 22 reviewed trials stated the number of patients who declined to participate, but it was not always clear whether these numbers referred to patients at the screening stage or to patients already identified as eligible. The median percentage of patients who declined participation as a percentage of randomised patients was 18%, and varied from 3% to 48.42%. It is important to note, that the two trials with high numbers of patients refusing to participate investigated vertebroplasty, which, at the time, was an established procedure; therefore, patients could easily receive the treatment from a different medical centre, without participating in a trial.^{34 35}

Patient retention

In general, recruitment was more problematic than retention and, once recruited, patients usually remained in the trial. Across the reviewed trials, 96% of randomised patients completed the study (table 1). A lower completion rate in five trials was caused by an early termination^{25 29 40} as well as withdrawals or change of patients' health status.^{41 42} In general, the predicted attrition, by which the required sample size was inflated to account for drop-outs, was 10% (median) with the range from 5% to 24%, whereas the actual patients' attrition between randomisation and outcome assessment was 4% (range 0–50%).

The completion rate was similar in the active and in the placebo arm, except for two trials: one¹⁸ where five times as many patients were lost to follow-up in the active group than in the placebo group, and one³⁰ where the drop-out rate was three times higher in the placebo group. Neither of these studies could explain this difference.

Most of the drop-outs occurred before randomisation. The reported reasons for drop-out during the trial were withdrawals, loss to follow-up, or discontinuation without known cause,^{20 39 41 43 44} patients' request to be unblinded,²⁴ adverse events,^{25 26} change of medical status such as pregnancy or concurrent illness.^{23 45 46} A long wait between the screening and procedure did not necessarily result in patient withdrawal.⁴³ A variable reporting did not allow us to evaluate quantitatively the reasons for drop-outs.

Blinding was possible, and some studies attempted to blind surgeons

In 12 trials (19%), only patients were blinded, but in the majority of RCTs (n=51/63, 81%) both patients and outcome assessors were blinded; including three trials, in which there was also an attempt to blind the operator.

For example, in two trials, the implant delivery system was preloaded by the manufacturer—the devices looked identical but only one contained an implant.^{28 47} In another trial, the surgeon placed the catheter but then handed the procedure over to a technician who delivered the treatment according to the randomisation.²¹

Authors of the reviewed trials went to great lengths to imitate the visual, verbal and physical cues, and to make the placebo as similar as possible to the active procedure. For example, patients wore goggles, or had the view obscured, so that they could not see the device.⁴⁸ The preparation for the placebo intervention was done in the same way as for the active procedure.^{35 49} Similar verbal instructions were given as during the surgery,^{42 50} and there were attempts to imitate the noises made by the devices.⁵¹ In trials that used exogenous substances, the container was opened so that the distinct smell was also present during the placebo condition.³⁵ Some researchers attempted to keep the duration of the procedure the same in both arms,^{39 43 52} whereas others thought that it was more ethical to shorten the placebo intervention.³¹

Very few studies assessed the success of blinding. Often authors thought that it was reasonable to assume that patients in the study were not able to distinguish between placebo and surgery due to minimally invasive characteristics of the procedure, and minimally post-operative treatment-related symptoms.^{53 54} In one trial, the post-treatment symptoms were believed to be a sign of correctly placed effective gastropliation, as patients with these symptoms had better outcomes.³⁹ Blinding was reported as successful in n=13/63 (21%) studies. In four trials,^{35 42 55 56} a larger proportion of patients in the active group guessed correctly; however, the placebo group did not guess the treatment allocation. In one study, two patients were definitely unblinded early due to implant extrusion.⁵⁷

Anaesthesia

In the reviewed trials, patients in both groups received some type of anaesthesia. General anaesthesia or sedation was used in n=26/63 trials (41%), including one trial in which general anaesthesia was used in the surgical group, but patients in the placebo group were sedated without intubation.²⁹ Local analgesia was used in n=16/63 (25%) RCTs, four studies used a mixture of methods, and n=17/63 (27%) trials did not describe the type of anaesthesia used. None of the trials reported that anaesthesia was a barrier in conducting their study.

DISCUSSION

This review has demonstrated that surgical RCTs with a placebo arm are feasible, at least for procedures with a lower level of invasiveness. Many of the presumed challenges, such as funding, anaesthesia or blinding of patients and assessors, were not reported as obstacles in any of the reviewed trials. The main hurdle in

completing a trial was finding a sufficient number of eligible patients.

We found that, although, there were very few surgical RCTs with a placebo arm published between 1959 and 2014, there was a rising trend. This may be related to an increasing interest in placebo and placebo-controlled trials in general,⁵⁸ or to the increasing popularity of minimally invasive procedures since 1980s. The latter explanation is supported by the fact that most of the reviewed trials used some type of key-hole surgery.

The analysed placebo-controlled trials were funded equally often by industry as by non-commercial funding bodies. The number of commercially funded older trials may be underestimated in our review because surgical RCTs funded by industry have lower odds of being published.⁵⁹ However, the recent trials are registered in the ClinicalTrials.gov database and would have been identified. The distribution of the source of funding was similar to that described by other authors.¹¹ This is encouraging, as it shows that there is an interest within the industry to validate the efficacy of their products, and also that the non-commercial bodies are willing to investigate the efficacy of surgical procedures. The costs of running surgical RCTs are high,² but in the long run, preferential funding of treatment with proven efficacy may help to improve the allocation of resources and to lower the costs of healthcare.⁶⁰ For example, the trial by Moseley and colleagues⁵² demonstrated that arthroscopic surgery had no benefits because the outcomes in the arthroscopic debridement arm and the lavage arm were not better than in the placebo group and, consequently, there was a decline in the use of this procedure for knee osteoarthritis.⁶¹

Recruitment into placebo-controlled surgical trials was possible but was often very slow, and resulted in an early termination of several trials. Slow recruitment is the most frequent reason for discontinuation of RCTs, including surgical RCTs. For example, 21% of reviewed surgical RCTs were discontinued early, and 44% of these were due to problems with recruitment.⁵⁹ Authors often underappreciate the fact that the target population in surgical trials is small; therefore, it may be challenging to recruit a required number of patients in a reasonable period of time.² The right timing of a trial may also affect its completion,⁷ for example, initiating a trial too early in the intervention's development may result in more procedure-related adverse events,²⁶ whereas, when a procedure has been already established, like vertebroplasty, it may be difficult to recruit participants.^{34 35}

In the reviewed trials, the number of patients who had to be screened in order to recruit a necessary participant group was larger than in other RCTs, but the proportion of eligible patients who started the study was comparable with other types of RCTs.⁶² This is another argument suggesting that the main challenge in those trials was finding suitable patients rather than persuading potential participants to enter the trial, and this is a bigger problem than in other types of RCTs.⁶²

Reporting of the recruitment process and eligibility was generally poor and often difficult to interpret, as the reviewed studies usually did not describe in detail why eligible patients did not enter the trial, which is in line with observations from other reviews.⁶² The quality of reporting in analysed RCTs was poor, but this is a known problem in surgical trials.^{7 62}

There is an assumption that patients are unwilling to take part in surgical RCTs, especially patients in severe pain.⁶³ Interestingly, in the trial by Moseley and colleagues,⁵² patients in more pain were more likely to agree to participate. Also patients tend to choose the new treatment even if it was not proven to be superior over placebo. For example, in the trials on Parkinson's disease, patients actually opted for the transplantation when they were given a choice after the end of the trial, despite the fact that it was not demonstrated to be more effective than placebo.⁴⁹ In a recent orthopaedic placebo-controlled RCT, patients were willing to participate, and screening failures were a larger problem than refusals or withdrawals.¹⁶ The clinical characteristics of patients who entered into a placebo-controlled RCT were comparable to the non-enrolment group, as well as to patients in other trials.¹⁶

Only about half the published trials reported a sample size calculation, which is in line with another review of surgical trials, which found that sample size calculations were reported only in 63% of RCTs.¹¹ However, it is important to note that some of the reviewed trials were published before the CONSORT (Consolidated Standards of Reporting Trials) were introduced, and before the sample size calculation became required by the board review. Some trials were small because of the author's assumption that surgical studies have a large effect size; therefore, inferring that a smaller sample size is required in surgical trials than in drug trials.^{47 64} However, surgical RCTs may require larger numbers of patients to reach the required sample size.⁶⁵ Recent systematic reviews demonstrated that the effect size of the surgical procedure in comparison with placebo in the existing trials was often small.^{12 66} It is likely that the apparent lack of difference between the active treatment and placebo might have been related to the small sample size and the effect not reaching the statistical significances.^{47 67 68} It might be also caused by a large placebo effect; however, the magnitude of the placebo effect in surgical procedures is unknown. The magnitude of response in the placebo arm is related not only to the placebo effect, that is, response directly related to the placebo intervention, but also to non-specific changes, such as regression to the mean, natural history of disease, or effect of participation in the trial.⁶⁹ Only one reviewed trial included a non-interventional group to control for these non-specific effects.³⁹

A placebo procedure can successfully imitate a minimally invasive surgery. Blinding in interventional trials is more challenging than in pharmaceutical ones;^{11 70} however, there are many strategies to blind the patients

and outcome assessors,⁷⁰ and the reviewed trials often used ingenious methods to achieve blinding. The success of blinding was rarely assessed, but it is not necessary, according to the current reporting standards. The requirement to assess blinding was removed from the CONSORT checklist because of evidence that testing for blindness is not valid because it cannot distinguish the success of blinding from 'hunches' about the treatment's efficacy.⁷¹

Blinding of patients and outcome assessors is especially important if the outcomes are subjective or difficult to quantify.⁷² Softer outcomes are difficult to evaluate in unblinded trials due to patient-related or assessor-related bias, which may distort the treatment effect.^{73 74} In this analysis, we have demonstrated that the withdrawal rate was generally low and was similar in the active and the placebo groups. This provides supporting evidence that blinding reduces the attrition bias, as patients do not know to which treatment they had been allocated.⁷⁵

Future implications for clinicians and unanswered questions

What remains to be understood is why eligible patients decline participation or withdraw their consent before randomisation.⁷⁶ Addressing these issues may improve the recruitment procedure in future trials.

There is also a need to estimate the magnitude of placebo effect in interventional trials. Several authors have highlighted the fact^{12 65 66} that for softer outcome measures, the magnitude of placebo effect in surgical trials is underestimated while the effect size of the surgical intervention is overestimated and, as a result of that, many trials do not recruit sufficient numbers of patients to detect differences between the effects of surgery and placebo.

Journals should encourage authors to report the details of patient recruitment and allocation, including the reasons for withdrawals and screening failures. Data like this are very useful when planning future trials. There has been an improvement in the reporting quality of recent trials,²⁰ and these guidelines were included in the CONSORT extension for non-pharmacological interventions.⁷⁷

In conclusion, not every surgical procedure has a viable placebo control; however, surgical RCTs with a placebo arm are feasible for many less invasive procedures. Although placebo-controlled surgical RCTs are challenging, they should not be dismissed as a potential trial design in surgical research. There is a need to better understand the factors that make those trials challenging so that future trials are not terminated early, and contribute good quality evidence to surgical practice.

Author affiliations

¹Oxford NIHR Musculoskeletal Biomedical Research Unit, Oxford, UK

²Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Institute of Musculoskeletal Sciences, University of Oxford, Oxford, UK

³Centre for Statistics in Medicine, Oxford, UK

⁴MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, UK

⁵Royal College of Surgeons of England Clinical Trials Unit, Botnar Institute of Musculoskeletal Sciences, Oxford, UK

Contributors KW contributed to study concept and design. KW, AJ, SH and GSC contributed to design of the search strategy and statistical analysis. KW, IR and BJFD contributed to data collection and management. KW contributed to drafting of the manuscript and the guarantor of the study. KW, AJC, DJB, IR, BJFD, SH and GSC contributed to critical revision of the manuscript for important intellectual content. All authors had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. All authors revised and approved the final version of the article.

Funding The study was funded by the National Institute of Health Research Oxford Musculoskeletal Biomedical Research Unit.

Competing interests The authors are involved in a placebo-controlled surgical trial on shoulder pain (NCT01623011).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/4.0/>

REFERENCES

- Howes N, Chagla L, Thorpe M, *et al*. Surgical practice is evidence based. *Br J Surg* 1997;84:1220–3.
- Gelijns AC, Ascheim DD, Parides MK, *et al*. Randomized trials in surgery. *Surgery* 2009;145:581–7.
- Wente MN, Seiler CM, Uhl W, *et al*. Perspectives of evidence-based surgery. *Dig Surg* 2003;20:263–9.
- Carr AJ, Robertsson O, Graves S, *et al*. Knee replacement. *Lancet* 2012;379:1331–40.
- McCulloch P, Altman DG, Campbell WB, *et al*. No surgical innovation without evaluation: the ideal recommendations. *Lancet* 2009;374:1105–12.
- 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.
- Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. *Trials* 2009;10:9.
- Ergina PL, Cook JA, Blazeby JM, *et al*. Challenges in evaluating surgical innovation. *Lancet* 2009;374:1097–104.
- Miller FG, Kaptchuk TJ. Sham procedures and the ethics of clinical trials. *JRSM* 2004;97:576–8.
- Dowrick AS, Bhandari M. Ethical issues in the design of randomized trials: to sham or not to sham. *J Bone Joint Surg Am* 2012;94(Suppl 1):7–10.
- Wenner DM, Brody BA, Jarman AF, *et al*. Do surgical trials meet the scientific standards for clinical trials? *J Am Coll Surg* 2012;215:722–30.
- Wartolowska K, Judge A, Hopewell S, *et al*. Use of placebo controls in the evaluation of surgery: systematic review. *BMJ* 2014;348:g3253.
- Miller FG. Sham surgery: an ethical analysis. *Am J Bioeth* 2003;3:41–8.
- London AJ, Kadane JB. Placebos that harm: sham surgery controls in clinical trials. *Stat Methods Med Res* 2002;11:413–27.
- Wright JG, Katz JN, Losina E. Clinical trials in orthopaedics research. Part i. Cultural and practical barriers to randomized trials in orthopaedics. *J Bone Joint Surg Am* 2011;93:e15.
- Hare KB, Lohmander LS, Roos EM. The challenge of recruiting patients into a placebo-controlled surgical trial. *Trials* 2014;15:167.
- Wright JG. Placebo surgery research: a blinding imperative. [Comment]. *J Clin Epidemiol* 2007;60:536.
- Eid GM, McCloskey CA, Eagleton JK, *et al*. Stomach vs a sham procedure for revisional surgery to reduce regained weight in roux-en-y gastric bypass patients: a randomized clinical trial. *JAMA Surg* 2014;149:372–9.

19. Sihvonen R, Paavola M, Malmivaara A, *et al*. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med* 2013;369:2515–24.
20. Cotton PB, Durkalski V, Romagnuolo J, *et al*. Effect of endoscopic sphincterotomy for suspected sphincter of oddi dysfunction on pain-related disability following cholecystectomy: the episode randomized clinical trial. *JAMA* 2014;311:2101–9.
21. Freeman BJ, Fraser RD, Cain CM, *et al*. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy versus placebo for the treatment of chronic discogenic low back pain. *Spine* 2005;30:2369–77.
22. Thompson CC, Chand B, Chen YK, *et al*. Endoscopic suturing for transoral outlet reduction increases weight loss after roux-en-y gastric bypass surgery. *Gastroenterology* 2013;145:129–137.e3.
23. Pauza KJ, Howell S, Dreyfuss P, *et al*. A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J* 2004;4:27–35.
24. Bajbouj M, Becker V, Eckel F, *et al*. Argon plasma coagulation of cervical heterotopic gastric mucosa as an alternative treatment for globus sensations. *Gastroenterology* 2009;137:440–4.
25. Silverberg GD, Mayo M, Saul T, *et al*. Continuous csf drainage in ad: results of a double-blind, randomized, placebo-controlled study. *Neurology* 2008;71:202–9.
26. Lee PE, Kung RC, Drutz HP. Periurethral autologous fat injection as treatment for female stress urinary incontinence: a randomized double-blind controlled trial. *J Urol* 2001;165:153–8.
27. Lindor KD, Hughes RW Jr, Ilstrup DM, *et al*. Intra-gastric balloons in comparison with standard therapy for obesity—a randomized, double-blind trial. *Mayo Clin Proc* 1987;62:992–6.
28. Gillespie MB, Wylie PE, Lee-Chiong T, *et al*. Effect of palatal implants on continuous positive airway pressure and compliance. *Otolaryngol Head Neck Surg* 2011;144:230–6.
29. Jarell J, Mohindra R, Ross S, *et al*. Laparoscopy and reported pain among patients with endometriosis. *J Obstet Gynaecol Can* 2005;27:477–85.
30. Corley DA, Katz P, Wo JM, *et al*. Improvement of gastroesophageal reflux symptoms after radiofrequency energy: a randomized, sham-controlled trial. *Gastroenterology* 2003;125:668–76.
31. Fleischer D. Endoscopic Nd:YAG laser therapy for active esophageal variceal bleeding. A randomized controlled study. *Gastrointest Endosc* 1985;31:4–9.
32. Swank DJ, Swank-Bordewijk SC, Hop WC, *et al*. Laparoscopic adhesiolysis in patients with chronic abdominal pain: a blinded randomised controlled multi-centre trial. *Lancet* 2003;361:1247–51.
33. Bradley JD, Heilman DK, Katz BP, *et al*. Tidal irrigation as treatment for knee osteoarthritis: a sham-controlled, randomized, double-blinded evaluation. *Arthritis Rheum* 2002;46:100–8.
34. Buchbinder R, Osborne RH, Ebeling PR, *et al*. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med* 2009;361:557–68.
35. Kallmes DF, Comstock BA, Heagerty PJ, *et al*. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med* 2009;361:569–79.
36. Genco A, Cipriano M, Bacci V, *et al*. BioEnterics Intra-gastric Balloon (BIB): a short-term, double-blind, randomised, controlled, crossover study on weight reduction in morbidly obese patients. *Int J Obes (Lond)* 2006;30:129–33.
37. Scolapio JS, Gostout CJ, Schroeder KW, *et al*. Dysphagia without endoscopically evident disease: to dilate or not? *Am J Gastroenterol* 2001;96:327–30.
38. Toouli J, Roberts-Thomson I, Kellow J, *et al*. Manometry based randomised trial of endoscopic sphincterotomy for sphincter of oddi dysfunction. *Gut* 2000;46:98–102.
39. Schwartz MP, Wellink H, Gooszen HG, *et al*. Endoscopic gastroplication for the treatment of gastro-oesophageal reflux disease: a randomised, sham-controlled trial. *Gut* 2007;56:20–8.
40. Fockens P, Cohen L, Edmundowicz SA, *et al*. Prospective randomized controlled trial of an injectable esophageal prosthesis versus a sham procedure for endoscopic treatment of gastroesophageal reflux disease. *Surg Endosc* 2010;24:1387–97.
41. Abbott J, Hawe J, Hunter D, *et al*. Laparoscopic excision of endometriosis: a randomized, placebo-controlled trial. *Fertil Steril* 2004;82:878–84.
42. Benjamin SB, Maher KA, Cattau EL, *et al*. Double-blind controlled trial of the garren-edwards gastric bubble: an adjunctive treatment for exogenous obesity. *Gastroenterology* 1988;95:581–8.
43. 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. *Lancet Neurology* 2011;10:509–19.
44. Shaheen NJ, Sharma P, Overholt BF, *et al*. Radiofrequency ablation in Barrett's esophagus with dysplasia. *N Engl J Med* 2009;360:2277–88.
45. Montgomery M, Håkanson B, Ljungqvist O, *et al*. Twelve months' follow-up after treatment with the EndoCinch endoscopic technique for gastro-oesophageal reflux disease: a randomized, placebo-controlled study. *Scand J Gastroenterol* 2006;41:1382–9.
46. Sutton CJ, Ewen SP, Whitelaw N, *et al*. Prospective, randomized, double-blind, controlled trial of laser laparoscopy in the treatment of pelvic pain associated with minimal, mild, and moderate endometriosis. *Fertil Steril* 1994;62:696–700.
47. Maurer JT, Sommer JU, Hein G, *et al*. Palatal implants in the treatment of obstructive sleep apnea: a randomised, placebo-controlled single-centre trial. *Eur Arch Otorhinolaryngol* 2012;269:1851–6.
48. Stone GW, Teirstein PS, Rubenstein R, *et al*. A prospective, multicenter, randomized trial of percutaneous transmural laser revascularization in patients with nonrevascularizable chronic total occlusions. *J Am Coll Cardiol* 2002;39:1581–7.
49. Freed CR, Greene PE, Breeze RE, *et al*. Transplantation of embryonic dopamine neurons for severe Parkinson's disease. *N Engl J Med* 2001;344:710–19.
50. Rothstein R, Filipi C, Caca K, *et al*. Endoscopic full-thickness plication for the treatment of gastroesophageal reflux disease: a randomized, sham-controlled trial. *Gastroenterology* 2006;131:704–12.
51. Mathus-Vliegen EM, Tytgat GN, Veldhuyzen-Offermans EA. Intra-gastric balloon in the treatment of super-morbid obesity. Double-blind, sham-controlled, crossover evaluation of 500-milliliter balloon. *Gastroenterology* 1990;99:362–9.
52. Moseley JB, O'Malley K, Petersen NJ, *et al*. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81–8.
53. Baeck LJ, Liukko T, Rantanen I, *et al*. Radiofrequency surgery of the soft palate in the treatment of mild obstructive sleep apnea is not effective as a single-stage procedure: a randomized single-blinded placebo-controlled trial. *Laryngoscope* 2009;119:1621–7.
54. Stuck BA, Sauter A, Hormann K, *et al*. Radiofrequency surgery of the soft palate in the treatment of snoring. A placebo-controlled trial. *Sleep* 2005;28:847–50.
55. Castro M, Rubin AS, Laviolette M, *et al*. Effectiveness and safety of bronchial thermoplasty in the treatment of severe asthma. *Am J Respir Crit Care Med* 2010;181:116–24.
56. Hogan RB, Johnston JH, Long BW, *et al*. A double-blind, randomized, sham-controlled trial of the gastric bubble for obesity. *Gastrointest Endosc* 1989;35:381–5.
57. Friedman M, Schalch P, Lin HC, *et al*. Palatal implants for the treatment of snoring and obstructive sleep apnea/hypopnea syndrome. *Otolaryngol Head Neck Surg* 2008;138:209–16.
58. Enck P, Klosterhalfen S, Weimer K, *et al*. The placebo response in clinical trials: more questions than answers. *Philos Trans R Soc Lond B Biol Sci* 2011;366:1889–95.
59. Chapman SJ, Shelton B, Mahmood H, *et al*. Discontinuation and non-publication of surgical randomised controlled trials: observational study. *BMJ* 2014;349:g6870.
60. Chandra A, Jena AB, Skinner JS. The pragmatist's guide to comparative effectiveness research. *J Econ Perspect* 2011;25:27–46.
61. Hawker G, Guan J, Judge A, *et al*. Knee arthroscopy in england and ontario: patterns of use, changes over time, and relationship to total knee replacement. *J Bone Joint Surg* 2008;90:2337–45.
62. Toerien M, Brookes S, Metcalfe C, *et al*. A review of reporting of participant recruitment and retention in RCTs in six major journals. *Trials* 2009;10:52.
63. Bono CM, Heggeness M, Mick C, *et al*. North American spine society: newly released vertebroplasty randomized controlled trials: a tale of two trials. *Spine J* 2010;10:238–40.
64. Olanow C. Double-blind, placebo-controlled trials for surgical interventions in Parkinson disease. *Arch Neurol* 2005;62:1343–4.
65. Fairbank J. Randomised controlled trials in surgery. *Lancet* 1999;354:257.
66. Holte Dahl R, Brox JI, Tjomsland O. Placebo effects in trials evaluating 12 selected minimally invasive interventions: a systematic review and meta-analysis. *BMJ Open* 2015;5:e007331.
67. MacLeod IA, Mills PR, MacKenzie JF, *et al*. Neodymium yttrium aluminium garnet laser photocoagulation for major haemorrhage from peptic ulcers and single vessels: a single blind controlled study. *Br Med J (Clin Res Ed)* 1983;286:345–8.
68. Fullarton GM, Birnie GG, Macdonald A, *et al*. Controlled trial of heater probe treatment in bleeding peptic ulcers. *Br J Surg* 1989;76:541–4.

69. Ernst E, Resch KL. Concept of true and perceived placebo effects. *BMJ* 1995;311:551–3.
70. Boutron I, Guittet L, Estellat C, *et al.* Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007;4:e61.
71. Sackett DL. Commentary: measuring the success of blinding in rcts: don't, must, can't or needn't? *Int J Epidemiol* 2007;36:664–5.
72. Hróbjartsson A, Thomsen ASS, Emanuelsson F, *et al.* Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *Can Med Assoc J* 2013;185:E201–11.
73. Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *J Intern Med* 2004;256:91–100.
74. Wood L, Egger M, Gluud LL, *et al.* Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601–5.
75. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002;136:254–9.
76. Treweek S, Lockhart P, Pitkethly M, *et al.* Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. *BMJ Open* 2013;3:e002360.
77. Boutron I, Moher D, Altman DG, *et al.* Extending the consort statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008;148:295–309.