Title: Risk of bias and the reporting of surgeons' experience in randomised controlled trials of total hip and total knee arthroplasty: A systematic review.

Authors:

Christopher Carroll PhD[1]

Faizan Mahmood, BA[1]

[1]School of Health and Related Research (ScHARR), University of Sheffield, S1 4DA, UK

***Corresponding author:**

Christopher Carroll, Reader in Systematic Review and Evidence Synthesis, Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, Regent Street, Sheffield, S1 4DA, UK

c.carroll@shef.ac.uk

Tel: +44(0)114 22 20864

Fax: +44(0)114 22 20749

Running title:

Risk of bias and reporting of surgeons' experience

Keywords:

Systematic review; bias; reporting; randomised controlled trial; CONSORT; arthroplasty

Abstract:

Rationale, aims and objectives: The potential bias introduced by surgeons' lack of comparable, relevant experience when performing the procedures in different arms of randomised controlled trials (RCTs) is arguably not well-managed or reported. The aim of this work was to review the frequency and nature with which surgeons' relevant experience is reported in RCTs of total hip (THA) and total knee arthroplasty (TKA), and to relate this to other risk of bias domains for this study design.

Methods: A systematic review of RCTs comparing different minimally invasive procedures for TKA and comparisons of THA and hemiarthroplasty (HA). We searched MEDLINE, EMBASE, Science Citation Index, The Cochrane Library, Conference Proceedings Citation Index- Science (CPCI-S), Current Controlled Trials and Clinical Trials.gov.

Results: 75 relevant RCTs were identified, 65 RCTs comparing minimally invasive with standard or other minimally invasive approaches to TKA, and 10 for THA compared with HA. Risk of bias based on the reported details of surgeons' relevant experience was categorised as low, high or unclear. There was a clear distinction before and after 2009, with a substantial decrease in trials at high or unclear risk of bias after this date. There were no strong associations between this domain and other, standard risk of bias domains for RCTs.

Conclusion: The surgeons' relevant experience in an evaluated procedure is often poorly reported but has improved since 2009. The variable is not adequately captured by any other risk of bias domain. Future work should concentrate on conducting research on a much larger sample of studies and in procedures other than knee and hip arthroplasty.

1. INTRODUCTION

'Bias' is a systematic error affecting the internal validity and results of a study; it can operate in either direction (leading to under- or overestimation of an intervention's effect) and can vary in magnitude (the potential confounding effect might be small or large)[1]. Bias can also be difficult to quantify, so assessments are usually made in terms of the 'risk of bias'. Such assessments are undertaken to determine how close the reported results of a trial are to the 'truth' of the relative efficacy or safety of any interventions being tested[1]. The relative efficacy of different interventional procedures is commonly assessed in randomised controlled trials (RCTs). However, while other potential confounders and sources of bias that might compromise the internal validity of a trial can be controlled for by randomisation and allocation concealment (e.g. such as differences in patients' age, gender, condition severity, comorbidities etc.) or the blinding of patients or clinical outcome assessors, the potential effect of trial surgeons' ability to perform the two procedures or techniques to exactly the same standard (even though one technique might be quite novel) is arguably not managed to the same degree[2].

In order for a comparison to be valid (all other confounders being controlled for), the surgeons performing the procedure in each arm of an explanatory randomized controlled trial must be equally or adequately adept at each of the respective techniques[2]. Otherwise, for example, one procedure might appear relatively much less efficacious or safe than another simply because of a relative lack of relevant experience between the surgeons performing the procedures in the two arms of a trial. Such considerations must also take into account that one procedure might be more technically challenging than the other[2]. The surgeons' experience might therefore be considered an additional source of bias in trials of interventional procedures;[2, 3] this has been called 'differential expertise bias'[2].

Indeed, there is a sizeable body of literature on the learning curve associated with surgical techniques (i.e. the more relevant experience the surgeon has in a technique, the more accurate the outcomes for a technique) and, consequently, surgeons' experience in certain interventional procedures is an acknowledged potential confounder of outcomes[4-9]. Currently, this is not explicitly taken into

account or assessed by any standard critical appraisal tool for RCTs, even in tools specifically for critically appraising trials of surgical or interventional procedures[3, 10]. A previous systematic review has appraised the reporting of RCTs of surgical interventions, including descriptions of the participating surgeons, but did not consider the relative difference in surgeons' relevant experience across arms or how this might introduce bias into the trial[10]. A more recent review has considered the possibility of 'expertise bias' within spinal surgery RCTs: it found very limited reporting of this variable and did not seek to assess its impact[11]. The aim of this current review, therefore, is to assess the reporting and potential impact of surgeons' relevant experience in RCTs of total hip arthroplasty (THA) and total knee arthroplasty (TKA). These procedures have been chosen because it has been demonstrated that there is a particular learning curve associated with arthroplasty techniques, which can require the performance of a minimum number of procedures (50) if a surgeon is to be considered to have sufficient 'relevant experience'[6, 12]. However, there is no definitively accepted threshold for prior cases for all of the procedures being assessed, so such a number could not be applied universally across trials (although numbers were sometimes reported). Nevertheless, an assessment of perceived necessary experience in respective techniques could still be made, and this is the focus of this review.

Two recent reviews form the basis of this work and identified sufficient numbers of RCTs for the foundation of this exploratory study: one review compared total hip arthroplasty with hemiarthroplasty[13] and one compared subvastus (SV), midvastus (MV) or quadriceps-sparing (QS) approaches with medial parapatellar (MP) approaches to total knee arthroplasty (TKA)[14]. These reviews included eight and 32 RCTs respectively. The current systematic review updates these two reviews by identifying more recently-published, relevant RCTs, in order to compile the sample for this analysis. These procedures have been chosen because the reviews indicated that there was an adequate evidence base of RCTs for assessment, and because the surgeon's experience variable has been considered briefly in one of the reviews but was not the subject of analysis[13]. The protocol for this systematic review is registered with PROSPERO (CRD42017056755). This systematic review seeks to answer the following question: what is the frequency and nature of the reporting of surgeons' relevant

experience in RCTs of total hip and total knee arthroplasty and is there an association between risk of bias based on surgeons' relevant experience and risk of bias across other domains (e.g. selection, performance or detection bias), as well as between this variable and outcomes?

2. METHODS

A systematic review of the evidence was undertaken and reported following the general principles recommended in the PRISMA statement (http://www.prisma-statement.org/). Inclusion criteria are outlined in Table 1. The protocol for the review was registered in the PROSPERO database (CRD42017056755).

<insert Table 1: Inclusion criteria>

2.1 Search strategy

This review was to include all 40 of the RCTs included in the two 'foundation' reviews[13, 14], as well as any additional relevant RCTs published since the conduct of those reviews. Two sets of searches of electronic databases were undertaken. Comprehensive searches were undertaken to identify RCTs, systematic reviews and meta-analyses comparing total hip arthroplasty and hemiarthroplasty in patients with fractures of the femoral neck. This involved combining terms for total hip arthroplasty and hemiarthroplasty with terms for RCTs and systematic review or meta-analysis. An example MEDLINE search strategy is reported in Appendix 1. The aim of the strategy was to identify all trials and reviews comparing total hip arthroplasty with hemiarthroplasty published since 2010 (the date of the searches performed for the first review[13]). The same process was also followed to identify all trials or reviews comparing TKA approaches published since 2013 (the date of the searches performed for the second review[14]). The following electronic databases were searched from 2010 or 2013 to February 2017, depending on the intervention, for published and unpublished research evidence:

- MEDLINE (via Ovid)
- EMBASE (via Ovid)
- Science Citation Index (via ISI Web of Science)
- The Cochrane Library including the Cochrane Systematic Reviews Database, Cochrane Controlled Trials Register, DARE, HTA and NHS EED databases
- Conference Proceedings Citation Index- Science (CPCI-S) (via ISI Web of Science)
- Current Controlled Trials: https://www.isrctn.com/
- Clinical Trials.gov up: https://clinicaltrials.gov/

All citations were imported into Endnote® and duplicates deleted.

2.2 Study selection

All titles and abstracts of unique citations were screened independently by two reviewers (CC, FM) using the inclusion criteria outlined in Table 1. Full papers were retrieved of any citation identified by at least one reviewer as being potentially relevant. These full papers were then screened for inclusion by both reviewers and any disagreement resolved by consensus. The reference lists of all relevant, identified systematic reviews or meta-analyses were also checked for additional trials; cross-referencing with RCTs identified as relevant by the database searches was performed. Only full publications were included in the review because details of surgeons' experience were only likely to be covered in full publications. However, abstracts and records of 'unpublished' trials (from registers) were also checked to trace any potentially relevant trials that were not identified by the conventional search.

2.3 Data extraction and critical appraisal

After piloting the form on three trials by two reviewers (CC, FM), data were extracted from all included studies by one reviewer (FM) and checked by a second (CC) and any discrepancies were resolved by discussion. Data were extracted from the full papers of all included RCTs (data previously extracted in the 'foundation' reviews were not used). The following data were extracted: brief characteristics of the included RCTs, including location, population, intervention and comparator details; the number of surgeons; their reported experience; any efforts made to control for the variable of surgeons' relevant experience; and any available data on the outcomes listed in Table 1. Quality assessment of included RCTs was undertaken by two reviewers using the Cochrane risk of bias tool[1]; any disagreements were resolved by consensus.

2.4 Surgeon's experience

In order to simplify the data and to render them comparable with other (Cochrane) risk of bias domains (low, high and unclear), the description of trial surgeons' experience was categorised into fairly crude, but distinct and differing levels of risk of bias (see Table 2). This ranged from publications that reported details of the included surgeons' experience in the relevant procedures and/or whether an explicit or implicit effort was made to control for this variable between arms (and therefore designated for the purposes of this study as being at 'low risk of bias'), to those that report who did the surgical procedures but did not report on the operating surgeons' relevant experience ('high risk of bias') and to those where the publication made no mention at all of the surgeons' experience in either procedure ('unclear risk of bias'). Illustrative examples of each category are provided in Table 2. The authors considered this to be a reasonable categorisation for distinguishing between the level of relevant detail reported by the trial publications.

<insert Table 2: Categories of risk of bias based on reported experience of surgeon(s) performing the procedures>

2.5 Data analysis

Key data were tabulated and discussed in a narrative synthesis. These data enabled an assessment of the prevalence and nature of reporting of surgeons' relevant experience in these trials. Analyses were also undertaken using a simple chi-squared test to investigate whether there was an association between the suggested risk of bias based on surgeons' relevant experience and the risk of bias across other domains (e.g. selection, performance or detection bias): i.e. what was the likelihood of trials categorised as being at low risk of bias for surgeons' relevant experience also being categorised as being at low risk of selection bias? This was performed to assess whether other domains (such as selection bias), which are currently included in standard checklists, actually capture the potential risk of bias arising from uncertainty surrounding the adequately comparable delivery of the two 'interventions' by surgeons. Where the trial evidence was sufficiently homogeneous, and appropriate outcome data were reported (e.g. means and standard deviations [SDs] for continuous data), included studies were combined in a meta-analysis using a random effects model (RevMan® version 5.1) for determine possible impact of the surgeon variable on outcomes. For TKA comparisons, only outcomes with 10 or more relevant studies were meta-analysed. The standardised mean difference was reported for continuous data and risk ratios (RRs) for dichotomous data (e.g. revision or dislocation event rates). Statistical heterogeneity between trials was assessed using the I^2 statistic. Separate meta-analyses were performed for relevant outcomes, where the data permitted, based on the surgical approach being undertaken, e.g. minimally invasive (MV, SV or QS) compared with standard medial parapatellar (MP) approaches, or for total hip arthroplasty compared with hemiarthoplasty, with sub-groups based on the different levels of risk of bias (based on surgeons' reported experience). This enabled an exploratory assessment of the potential impact of surgeons' experience on outcomes. Results of these analyses were tabulated.

3. RESULTS

3.1 Quantity of evidence

The searches of the electronic databases retrieved 354 unique citations, of which 108 were relevant to this review. After checking relevant systematic reviews and meta-analyses, the final number of relevant RCTs was 75 (a full list of these included trials is available in Appendix 2). The processes of inclusion and exclusion are reported in the PRISMA flowchart in the Figure. A full list of excluded studies, with reasons, is provided in Appendix 3.

<insert Figure: PRISMA flowchart>

There were 65 RCTs comparing minimally invasive with standard or other minimally invasive approaches to TKA. Details of these trials are summarised in Table 3, including the risk of bias category assigned to each trial based on the reported description of the surgeons' relevant experience. This represented 34 new TKA RCTs that were not included in the 2014 review (one trial from that review was excluded here because it was published in Chinese[15]). With a single exception[16], all of the 'new' TKA trials were identified by the search of electronic databases and verified by cross-referencing with published systematic reviews and meta-analyses. In terms of the trial evidence for TKA, the principal comparisons were different or novel minimally invasive approaches to knee arthroplasty, especially mini-vastus (MV), sub-vastus (SV) or quadriceps-sparing (QS) approaches, compared with the standard medial parapatellar (MP) approach. 32 RCTs compared MV with MP approaches, 21 compared SV with MP approaches, six compared QS and MP approaches, and 10 conducted other comparisons (e.g. MS vs SV, QS vs SV, MP with and without patellar eversion).

<insert Table 3: Basic characteristics of included RCTs and risk of bias categorisation according to the reported expertise of surgeons>

The number of surgeons in any trial ranged from one (32 studies) to seven[12] in the TKA trials, and from two[17, 18] to as many as 14[19] in the THA trials. 51% (32/65) of TKA trials had only one surgeon. There were 10 RCTs of total hip arthroplasty compared with hemiarthroplasty (this represented

the publication of only two new relevant RCTs[17, 18] since the 2011 systematic review[13]), details of which are also summarised in Table 2. Both of the 'new' trials were identified by the search of electronic databases and verified by cross-referencing with published systematic reviews and meta-analysis.

3.2 Surgeons' experience

Overall, the majority of RCTs included in this sample were assessed as being at high or unclear risk of bias on this variable (see Table 3): 39/65 TKA trials (60%) and 5/10 hip arthroplasty trials (50%) (11 of the 13 trials at 'Unclear risk of bias' did not even report the number of surgeons conducting the procedures). Therefore, only 26 TKA RCTs (40%) and 5 hip arthroplasty RCTs (50%) were deemed to be at low risk of bias. It should be noted that the reporting of the 'surgeons' relevant experience' domain, especially in the TKA trials, appears to improve from 2009 onwards, the date when the CONSORT statement extension for nonpharmacological trials was published, i.e. CONSORT-NPT[3]). Compared with 2008 or before, the proportion of TKA RCTs categorised as being at 'low' risk of bias increases from 17% to 46%; the proportion of trials categorised as being at a 'high' risk of bias decreases from 58% to 41%, and those at 'unclear' risk of bias, from 25% to 12% (see Table 4).

<insert Table 4: Number of RCTs and risk of bias categorisation by date (median)>

The data suggests that there were changes in reporting standards over time within this sample, with a clear trend in improved reporting of surgeons' relevant experience as part of the trial. There are small changes in reporting standards within the hip arthroplasty RCT sample over time also, with a small trend in improved reporting and control of surgeons' relevant experience as part of the trial. Finally, there was no specific pattern by country for either TKA or total hip arthroplasty (without taking into account date).

In 63% of the included trials (47/75), no details were reported at all concerning the operating surgeons' relevant experience in performing the respective procedures, i.e. therefore deemed to be at high or

unclear risk of bias for this variable. There were some differences between trials depending on the procedures being evaluated (trials of SV TKA tended to be at lower risk of bias for this domain), but most noticeably there is a clear trend by date, with more recent trials tending to be better reported and therefore at lower risk of bias for this domain (see Table 4). Some of the trials in the 'low risk of bias' category might even be considered equivalent in some ways to the 'expertise-based RCTs' described elsewhere[2]. However, not all recently-published trials adequately report the key details (53% of TKA trials published in 2009 or later are all still at a high or unclear risk of bias for this domain: see Table 4).

Where the data were appropriate for pooling, the results of a series of analyses are reported in Table 5. The choice of analyses was determined by the availability and appropriateness of the data (e.g. the provision of means and standard deviations for continuous outcomes) so, despite the potentially substantial number of trials for inclusion (up to 65 for the TKA sample), only data from between 13 and 20 trials could be pooled in any meta-analysis. There was a high degree of statistical heterogeneity in the sample for each analysis of TKA trials, but low or moderate statistical heterogeneity in the sample of hip trials. For the TKA trials, there were no significant differences in any outcome between trials at high or unclear risk of bias based on the surgeons' reported expertise and those at low risk of bias on this variable (all confidence intervals overlapped).

<insert Table 5>

However, with the exception of blood loss, the findings from the pooled data of the studies at low risk of bias were much more uncertain, i.e. had much wider confidence intervals, than the findings for those studies considered to be at high or unclear risk of bias on this variable. Forest plots for these analyses are available in Appendix 4.

3.3 Other risk of bias domains

Only the following domains were assessed: selection bias (randomisation and allocation concealment), performance bias (patient blinding) and detection bias (blinding of outcome assessors). Unlike drug trials, sealed (and opaque) envelopes were often used as both a means of randomisation and allocation concealment in these trials of surgical procedures. In this sample, 35% of trials (26/75) reported using this method and, despite known issues with sealed envelopes, it can be a robust method of both randomisation and allocation concealment[1]. The results of the critical appraisal of all included trials are presented in Table 6 for both the TKA and hip arthroplasty RCTs.

<insert Table 6: Risk of bias for other domains>

Blinding of patients can also be a problem for comparative studies of certain surgical procedures[20]. However, in this sample (across the principal comparisons) it appears to have been possible because a number of trials report making efforts to ensure blinding of patients and outcome assessors by indicating that the incision made for the comparative procedures was in the same location and of the same length[16, 21-27]. However, such inconsistency in the conduct and reporting of blinding has been demonstrated in surgical trials previously[20].

Assessments of 'reporting bias' are not presented here for two reasons. First, this domain has been recognised as problematic in its assessment and the application of its findings to synthesis[28]. Second, unlike drug trials, this sample of trials of surgical interventions usually did not register any protocol, so no reliable or valid assessment could be made of whether there was complete consistency between intended and published outcomes[1]. In only three studies (4%)[29-31] were any outcomes reported in the Results that were not specified in the Methods. As a result, almost all trials would have been assessed as being at low risk of bias for this domain. The generally low risk of bias across the domains relating to detection and attrition bias (findings for the latter not reported here) might be a reflection of the small number of patients in trials, their generally short or very short follow-ups (few TKA trials had follow-ups longer than 12 months) and the relative ease of blinding assessors to interventions compared with the practical problems inherent in minimising or nullifying learning curve effects.

It is apparent that the evidence base for the comparison of THA and HA was generally of low quality, with a high risk of bias across most domains (except attrition bias, data not reported), similar to the risk of bias inherent in the domain of surgeons' relevant experience (see Tables 3 and 6). Based on this sample, the standards of reporting of surgical RCTs, across most of these domains, appear to have improved in recent years[32]. It is noteworthy, however, that the trend for improved reporting is also apparent across this sample of TKA RCTs for other risk of bias domains, for example randomisation (see Table 6).

Chi-squared tests did find significant associations, but only at the $p<0.05$ level, between risk of bias due to the reported details of surgeons' relevant experience and risk of bias due to both allocation concealment (X^2 statistic 5.07, p=0.024) and attrition bias (X^2 statistic 7.82, p=0.05). This suggests that, when the risk of bias was categorised as low for the 'surgeons' domain, then it was also categorised as low for these other two domains. However, there was no association at all for randomisation (X^2 statistic 2.43, p=0.119) or detection bias (X^2 statistic 1.40, p=0.237). Therefore, there does not appear to a particularly strong association in this sample between the risk of bias assessed as being due to the surgeons' relevant experience, and the risk of bias as adjudged for other domains. Performance bias, as determined by patient blinding, was not assessed as this was often unclear.

DISCUSSION

The standards of reporting regarding surgeons' relevant experience in performing the procedures being assessed in total knee and total hip arthroplasty trials is not very good, but does appear to be better than that reported elsewhere. For example, in this sample, 77% (58/75) of the trials reported the number of surgeons involved, which compares favourably with only 32% (51/158) in a previous review of RCTs of various surgical interventions from 2004[10]. In this sample the trend is certainly towards improved reporting of details regarding surgeons' relevant experience, as well as other risk of bias domains, such as randomisation procedures, a trend which had not always been found with reporting of surgical RCTs more generally[10, 33]. This trend is further highlighted by comparison with previous reviews. In a

similar review, but of open spinal surgery RCTs published between 2005 and 2010, only 10% (10/99) reported any details about the operating surgeons' experience or expertise[11], while another review, published in 2006[10], reported that the surgeons were described only as 'experienced' in 19% of the trials, with only 11% reporting a surgeon as having experience of the experimental intervention. This compares with the 37% (28/75) of the current sample, categorised as being at low risk of bias on account of the acknowledgment of 'experience' (see Table 4). In this previous review, in terms of reported efforts to 'standardise' procedures, only 6% of trials reported supervision by a senior surgeon, and 1% the use of protocol guidelines and video assessment[10]. However, it is unclear if these procedures were performed with the intention of standardising practice across centres or specifically to facilitate comparability of surgeons' relevant experience across all of the procedures being evaluated. The former is suggested by other applications of the CONSORT-NPT checklist[34]. Even if it did indicate the reporting of attempts to 'standardise' levels of surgeons' relevant experience across arms, the reported rates were certainly much lower (no more than 6%) than in this sample of TKA and THA RCTs (37%).

The reason for the improved reporting of this domain from 2009 onwards is not entirely clear, but might be due in some part to the publication in 2008 of the CONSORT-NPT statement extension[3]. This checklist was produced with surgical interventions, amongst others, in mind, and raises the issues of surgeons' relevant experience in procedures, as well as differences in procedures in terms of their difficulty. Consequently, it recommends reporting, "Details of how the interventions were standardized"[3]. None of the trials included in this sample actually reference this statement (only eight of the 44 papers published in 2009 or later make any reference to CONSORT at all, and then just in relation to the provision of a flowchart of participants[35-42], but there might be a general, tacit influence of improved reporting standards at play. Finally, there are significant associations between good or poor reporting of the surgeon's experience domain and good or poor reporting of allocation concealment and attrition within this set of RCTs, but there are no other significant associations between the surgeon domain and others. The absence of strong associations might be due to the sample being underpowered, but it might also be due to the fact that existing risk of bias domains in checklists and

tools simply do not adequately capture the issue of comparable levels of relevant experience among trial surgeons.

The meta-analyses did not demonstrate any meaningful differences between the findings of trials at high or unclear risk of bias compared with trials at low risk of bias, except that the latter did tend to produce findings of relatively greater uncertainty across four of the five outcomes in this sample. This outcome accords with other evidence that better-conducted trials can tend to produce less 'positive' findings.[43] However, the results here must be considered very cautiously given the small number of trials within the respective subgroups. Interpretation of these findings is made more difficult still because it must take into account the possibility that, given the technical challenges of some procedures, longer operating time, for example, might indicate a better-performed procedure.[2, 3]

This study must be considered exploratory only, given that the sample of trials included in the systematic review is not large (less than 100). This study also only focused on RCTs of two particular procedures, knee and hip arthroplasty, so its findings might not apply to the reporting and impact of surgeons' relevant experience in trials of other procedures, although the reporting in this sample does compare favourably with the results of a systematic review assessing reporting in a sample of RCTs of various procedures which were published in 2004[10]. It should be noted that even with efforts to control for surgeons' relevant experience, one technique might always be more challenging than another and thus present problems of comparability[2]. However, this issue will only be relevant to particular types of trials, such as those evaluating something new or especially challenging. Its value is more debatable in the assessment of a trial that is comparing two standard, but previously not compared treatments, or just comparing a modification of a standard treatment. Finally, any assessment of risk of bias across many of the domains covered here is in part interpretive and relies on the reporting of these elements in the papers. The reporting of surgeons' experience in these 75 papers is relatively poor, despite evidence of adequate reporting across some other risk of bias domains in these trials. It might therefore be the case that efforts were made to control for this potential confounder, but not fully

reported. After all, it has been suggested elsewhere that authors might fail to report certain elements of surgical trial conduct, even though they were adequately performed[44].

Nevertheless, this is the first systematic review to seek to assess both the reporting and relevance of this variable; the chosen procedures are known to be particularly vulnerable to a learning curve[6, 12]; and the sample is of moderate size and almost certainly includes all known, relevant RCTs comparing relevant procedures from the last 20 years. This study also adhered to published international standards in the conduct and reporting of systematic reviews. The risk of bias explored in this review, i.e. between-arm differentials in surgeons' experience, might be considered to be 'essential information about the intervention'[45], which should be reported if an appropriate assessment of the bias potentially affecting a trial's outcomes is to be conducted. It might even be considered a worthwhile 'clarifying' addition to the intervention description component of the CONSORT-NPT extension statement[3], adherence to which is known to be inadequate but is being strongly encouraged[46, 47].

Differences in the relevant experience of the surgeons performing the procedures being compared in a randomised controlled trial is a recognised potential confounder of a trial's results[3, 6, 12]. This review of trials of total knee and total hip arthroplasty has demonstrated that this variable is often poorly reported, although there is a trend towards improved reporting since 2009, and that it is not adequately captured by any other risk of bias domain. It is therefore worthy of assessment. Future work should concentrate on conducting research on a much larger sample of studies and in procedures other than knee and hip arthroplasty.

ACKNOWLEDGEMENTS:

FM was funded by the University of Sheffield On CampUS Placements programme. Inter-library loans were purchased from Health Economics and Decision Science departmental funds. This work was otherwise unfunded.

REFERENCES:

1. Higgins J, Green S: Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. In: The Cochrane collaboration. http://wwwcochrane-handbookorg 2011. Accessed 18 April 2018.
2. Devereaux PJ, Bhandari M, Clarke M, Montori V, Cook D, Yusuf S et al. Need for expertise based randomised controlled trials. BMJ. 2005; 330(7482):88.
3. Boutron I, Moher D, Altman D, Schulz K, Ravaud P et al: Extending the CONSORT Statement to Randomized Trials of Nonpharmacologic Treatment: Explanation and Elaboration Ann Intern Med. 2008; 148(4):295-309.
4. Masaaki I, Masanori S, Akihiro K, Yusuke N, Yoshiyuki T, Norio S. Influence of learning curve on short-term results after laparoscopic resection for rectal cancer. Surg Endosc. 2009; 23:403-408.
5. Sampatha S, Voona S, Sangstera M, Davies H. The statistical relationship between varus deformity, surgeon's experience, BMI and tourniquet time for computer assisted total knee replacements. Knee. 2009; 16:121-124.
6. King J, Stamper D, Schaad D et al. Minimally invasive total knee arthroplasty compared with traditional total knee arthroplasty. Assessment of the learning curve and the postoperative recuperative period. J Bone Joint Surg Am. 2007; 89A:1497-1503.
7. Cook J, Ramsey C, Fayers P. Statistical evaluation of learning curve effects in surgical trials. Clin Trials. 2004; 1:421-427.
8. Honkavaara P, Pyykkö I. Surgeon's experience as a factor for emetic sequelae after middle ear surgery. Acta Anaesthesiol Scand. 1998; 42:1033-1037.
9. Sosa J, Bowman H, Tielsch J, Powe N, Gordon T, Udelsman R. The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. Ann Surg. 1998; 228(3):320–330.

10. Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. Ann Surg. 2006; 244:677-683.

11. van Oldenrijk J, van Berkel Y, Kerkhoffs G, Bhandari M, Poolman R. Do Authors Report Surgical Expertise in Open Spine Surgery Related Randomized Controlled Trials? A Systematic Review on Quality of Reporting. Spine. 2013; 38(10):857-864.

12. Bridgman SA, Walley G, MacKenzie G, Clement D, Griffiths D, Maffulli N. Sub-vastus approach is more effective than a medial parapatellar approach in primary total knee arthroplasty: A randomized controlled trial. Knee. 2009; 16(3):216-222.

13. Carroll C, Stevenson M, Scope A, Evans P, Buckley S. Hemiarthroplasty and Total Hip Arthroplasty for Treating Primary Intracapsular Fracture of the Hip: A Systematic Review and Cost-Effectiveness Analysis. Health Technol Assess. 2011; 15:36.

14. Liu H-W, Gu W-D, Xu N-W, Sun J-Y. Surgical Approaches in Total Knee Arthroplasty: A Meta-Analysis Comparing the Midvastus and Subvastus to the Medial Peripatellar Approach. J Arthroplasty 2014; 29:2298–2304.

15. Fu PL, Li XH, Wu YL, Xie QY, Sun JY, Wu HS. Comparison of midvastus and standard medial parapatellar approaches in total knee arthroplasty. J Clinical Rehabil Tissue Eng Res. 2008; 12(9):1793-1796.

16. Jenkins D, Rodriguez J, Ranawat A, Alexiades M, Deshmukh A, Fukunaga T, Greiz M et al. Randomized, Controlled, Prospective Study Evaluating the Effect of Patellar Eversion on Functional Outcomes in Primary Total Knee Arthroplasty. J Bone Joint Surg Am. 2014; 96A(10):851-858.

17. Cadossi M, Chiarello E, Savarino L, Tedesco G, Baldini N, Faldini C, Giannini S. A comparison of hemiarthroplasty with a novel polycarbonate–urethane acetabular component for displaced intracapsular fractures of the femoral neck. A randomised controlled trial in elderly patients. J Bone Joint Surg. Brit. 2014; 95B:609-615.

18. Sharma V, Awasthi B, Kumar K, Kohli N, Katoch P. Outcome analysis of hemiarthroplasty vs. Total hip replacement in displaced femoral neck fractures in the elderly. J Clin Diagn Res. 2016; 10(5):RC11-RC13.

19. Macaulay W, Nellans KW, Garvin KL, Iorio R, Healy WL, Rosenwasser MP. Prospective Randomized Clinical Trial Comparing Hemiarthroplasty to Total Hip Arthroplasty in the Treatment of Displaced Femoral Neck Fractures. Winner of the Dorr Award. J Arthroplasty. 2008; 23(6 SUPPL.):2-8.

20. Boutron I, Estelatt C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. J Clinical Epidemiol. 2005; 58:1220-1226.

21. Nestor BJ, Toulson CE, Backus SI, Lyman SL, Foote KL, Windsor RE. Mini-Midvastus vs Standard Medial Parapatellar Approach: A Prospective, Randomized, Double-Blinded Study in Patients Undergoing Bilateral Total Knee Arthroplasty. J Arthroplasty. 2010; 25(SUPPL. 6):5.

22. Keating EM, Faris PM, Meding JB, Ritter MA. Comparison of the midvastus muscle-splitting approach with the median parapatellar approach in total knee arthroplasty. J Arthroplasty. 1999; 14(1):29-32.

23. Engh GA, Holt BT, Parks NL:. A midvastus muscle-splitting approach for total knee arthroplasty. J Arthroplasty. 1997; 12(3):322-331.

24. Koh IJ, Kim MW, Kim MS, Jang SW, Park DC, In Y. The Patient's Perception Does Not Differ Following Subvastus and Medial Parapatellar Approaches in Total Knee Arthroplasty: A Simultaneous Bilateral Randomized Study. J Arthroplasty. 2016; 31(1):112-117.

25. Gelfer Y, Pinkas L, Horne T, Halperin N, Alk D, Robinson D. Symptomatic transient patellar ischemia following total knee replacement as detected by scintigraphy: A prospective, randomized, double-blind study comparing the mid-vastus to the medial para-patellar approach. Knee 2003; 10(4):341-345.

26. Heekin RD, Fokin AA. Mini-Midvastus Versus Mini-Medial Parapatellar Approach for Minimally Invasive Total Knee Arthroplasty Outcomes Pendulum Is at Equilibrium. J Arthroplasty. 2014; 29(2):339-342.

27. Hay GC, Kampshoff J, Kuster MS. Lateral subvastus approach with osteotomy of the tibial tubercle for total knee replacement: A two-year prospective, randomised, blinded controlled trial. J Bone Joint Surg Brit. 2010; 92(6):862-866.

28. Page M, Higgins J. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. Syst Rev. 2016; 5:108.

29. Dorr LD, Glousman R, Hoy ALS. Treatment of femoral neck fractures with total hip replacement versus cemented and noncemented hemiarthroplasty. J Arthroplasty 1986; 1(1):21-28.

30. Cila E, Guzel V, Ozalay M, Tan J, Simsek SA, Kanatli U, Ozturk A. Subvastus versus medial parapatellar approach in total knee arthroplasty. Arch Orthop Traum Surg. 2002;122(2):65-68.

31. Baker RP, Squires B, Gargan MF, Bannister GC. Total hip arthroplasty and hemiarthroplasty in mobile, independent patients with a displaced intracapsular fracture of the femoral neck. A randomized, controlled trial. Journal Bone Joint Surg Am. 2006; 88(12):2583-2589.

32. Bhandari M, Richards R, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the Journal of Bone and Joint Surgery from 1988 through 2000. J Bone Joint Surg Am. 2002; 84:A388-A396.

33. Ellis C, Hall J, Khalil A, Hall J. Evolution of methodological standards in surgical trials. ANZ J Surg. 2005; 75(10):847-877.

34. Blencowe N, Boddy A, Harris A. Systematic review of intervention design and delivery in pragmatic and explanatory surgical randomized clinical trials. Brit J Surg. 2015; 102(9):1037-1047.

35. Wegrzyn J, Parratte S, Coleman-Wood K, Kaufman K, Pagnano M. The John Insall award: no benefit of minimally invasive TKA on gait and strength outcomes: a randomized controlled trial. Clin Otrhop Relat Res. 2013; 471: 46-55.

36. Guy SP, Farndon MA, Conroy JL, Bennett C, Grainger AJ, London NJ. A prospective randomised study of minimally invasive midvastus total knee arthroplasty compared with standard total knee arthroplasty. Knee 2012; 19(6):866-871.

37. Jain S, Wasnik S, Mittal A, Hegde C. Outcome of subvastus approach in elderly nonobese patients undergoing bilateral simultaneous total knee arthroplasty: A randomized controlled study. Indian J Orthop. 2013; 47(1):45-49.

38. Nutton RW, Wade FA, Coutts FJ, Van Der Linden ML. Short term recovery of function following total knee arthroplasty: A randomised study of the medial parapatellar and midvastus approaches. Arthritis. 2014; 7 pages.

39. Verburg H, Mathijssen N, Niesten D-D, Verhaar J, Pilot P. Comparison of mini-midvastus and conventional total knee arthroplasty with clinical and radiographic evaluation a prospective randomized clinical trial with 5-year follow-up. J Bone Joint Surg Am. 2016; 98: 1014-1022.

40. Reid MJ, Booth G, Khan RJK, Janes G. Patellar eversion during total knee replacement a prospective, randomized trial. J Bone Joint Surg Am. 2014; 96(3):207-213.

41. Tomek IM, Kantor SR, Cori LA, Scoville JM, Grove MR, Morgan TS, Swarup I, Moschetti WE, Spratt KF. Early Patient Outcomes After Primary Total Knee Arthroplasty with Quadriceps-Sparing Subvastus and Medial Parapatellar Techniques A Randomized, Double-Blind Clinical Trial. J Bone Joint Surg Am. 2014; 96A(11):907-915.

42. Kim J, Lee S, Ha J, Choi H, Yang S, Lee M. The effectiveness of minimally invasive total knee arthroplasty to preserve quadriceps strength: A randomized controlled trial. Knee. 2011; 18(6):443-447.

43. Schulz K, Chalmers I, Hayes R, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995; 273:408-412.

44. Devereaux PJ, Choi P-L, El-Dika S et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol. 2004; 57:1232-1236.

45. Hoffman T, Erueti C, Glasziou P. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. BMJ. 2013; 347:f3755.

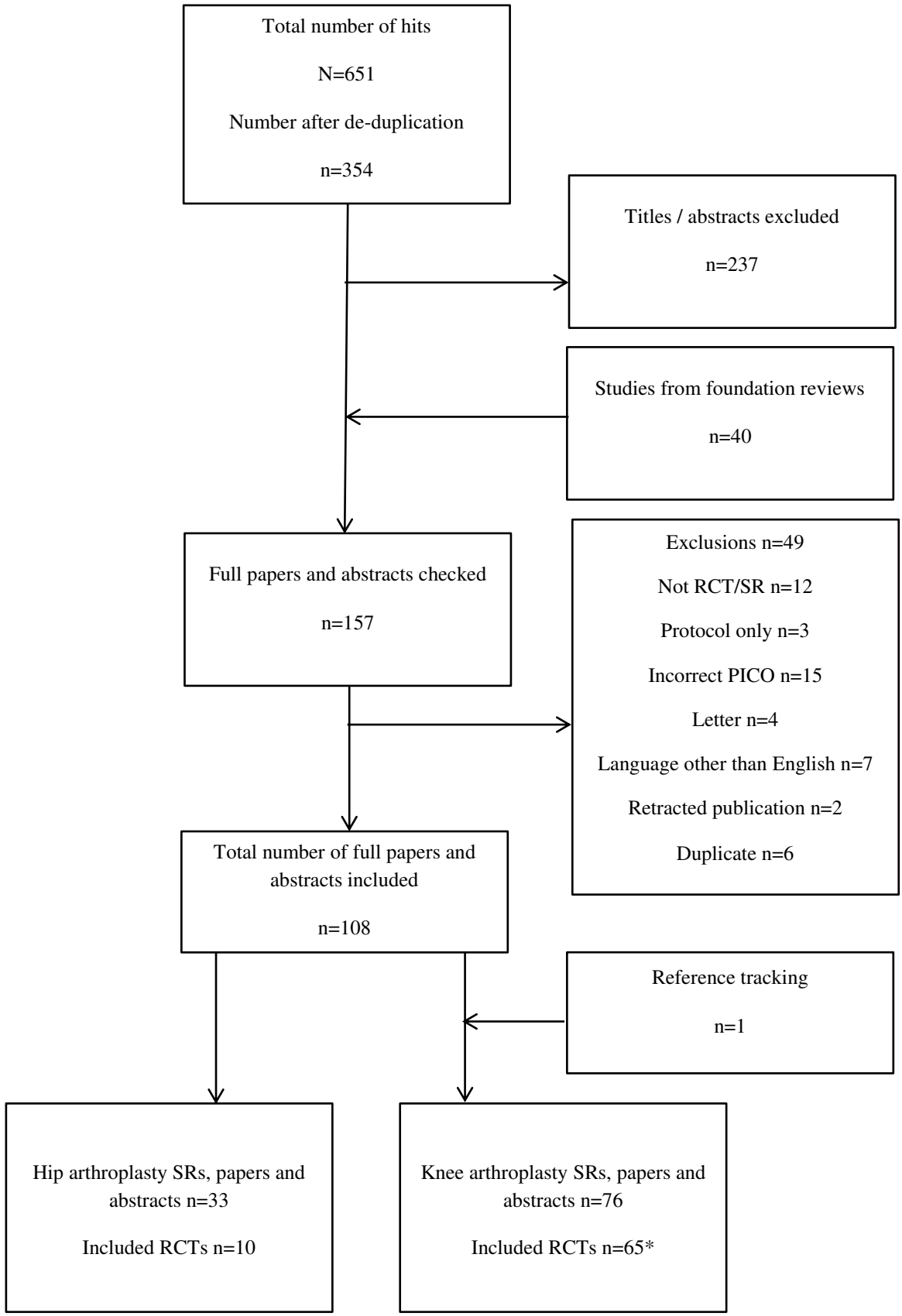
46. Hoffman T, English T, Glasziou P. Reporting of interventions in randomised trials: an audit of journal Instructions to Authors. Trials. 2014; 15:20.

47. Clavien P-A, Lillemoe K. A New Policy to Implement CONSORT Guidelines for Surgical Randomized Controlled Trials. Ann Surg. 2014; 260(6):947-948.

FIGURE LEGENDS:

PRISMA flowchart

Figure: PRISMA flowchart



SR: Systematic Review; PICO: Population, Intervention, Comparator, Outcomes; RCT: Randomised Controlled Trial

*Includes 13 RCTs with TKA comparisons other than MV vs MP or SV vs MP

Table 1: Inclusion criteria

Criteria	Total Knee Arthroplasty review	Total Hip Arthroplasty review
Population	Adult patients eligible for total knee arthroplasty	Adult patients eligible for total hip replacement
Intervention	Total Knee Arthroplasty (TKA)	Total Hip Arthroplasty (THA)
Comparators	Any alternative TKA technique	Hemiarthroplasty (HA)
Outcomes	Primary outcomes: ▪ Pain score by Visual Analog Score (VAS) ▪ Knee Society Score (KSS) ▪ knee range of motion (ROM) Secondary outcomes: ▪ Mortality ▪ Operative time (in minutes) ▪ Blood loss ▪ Length of hospital stay. ▪ Post-operative complications	Primary outcomes: ▪ Dislocation rate ▪ Revision rate Secondary outcomes: ▪ Mortality ▪ Operative time (in minutes) ▪ Blood loss ▪ Length of hospital stay. ▪ Post-operative complications
Study design	Randomised controlled trials (RCTs), systematic reviews or meta-analyses of RCTs	
Follow-up	No minimum duration of follow-up	
Language	Only studies published in English will be included because of the need to read the report in detail	

Table 2: Categories of risk of bias based on reported experience of surgeon(s) performing the procedures

Risk of Bias	Definition	Example
Low	Clear reference is made to the surgeons' relevant experience in the particular procedures and/or an explicit or implicit effort is made to control for this variable between arms	"All operations were performed by the senior surgeon … The senior surgeon had performed more than 100 total knee arthroplasties using each of the two approaches (MV and SV) prior to starting this study"[20] "All operations were performed by the first author … The operating surgeon had performed >1000 total knee arthroplasties using the mini-midvastus approach and >100 procedures using the mini-subvastus approach prior to the start of this study".[21]
High	No reference is made to surgeons' relevant experience in the particular procedures, but the individual(s) performing the surgery is reported	"All the knees were operated on by the same surgeon"[24] "All surgery was performed by the senior author"[25]
Unclear	No details are provided about who conducted the surgery	Nothing reported

TKA: Total Knee Arthroplasty; MV: Midvastus; SV: Subvastus

Table 3: Basic characteristics of included RCTs and risk of bias categorisation according to the reported experience of surgeons

Author, Year	Country	Intervention (n=)	Comparator (n=)	Number of surgeons	Risk of bias
TKA RCTs					
Weinhardt, 2004	Germany	26	26	1	Low
Bathis, 2005	Germany	25	25	1	Low
Seon, 2006	Korea	49	53	1	Low
Chin, 2007	Singapore	30	30	2	Low
Bridgman, 2009	UK	116	115	7	Low
Juosponis, 2009	Lithuania	35	35	2	Low
Lin, 2009	Taiwan	30	30	1	Low
Sastre, 2009	Spain	56	48	2	Low
Bonutti, 2010	USA	51	51	1	Low
Nestor, 2010	USA	27	27	2	Low
Pan, 2010	China	35	33	1	Low
Van Hemert, 2010	Netherlands	20	20	2	Low
Kim, 2011	Korea	25	25	1	Low
Tasker, 2014	UK	46	46	3	Low
Varnell, 2011	USA	20	37	3	Low

Author, Year	Country	Intervention (n=)	Comparator (n=)	Number of surgeons	Risk of bias
Varela-Egocheaga, 2011	Spain	50	50	1	Low
Bourke, 2012	Australia	40	41	6	Low
Masjudin, 2012	Malaysia	23	23	1	Low
Jain, 2013	India	50	50	1	Low
Jarvis, 2013	USA	27	26	3	Low
Wegryzn, 2013	USA	19	18	1	Low
Verburg, 2016	Netherlands	50	50	2	Low
Fezcko, 2016	Netherlands, Germany, Australia	36	33	3	Low
Engh, 1997	USA	57	61	1	High
Dalury, 1999	USA	24	24	1	High
Parentis, 1999	USA	21	21	2	High
Roysam, 2001	UK	46	43	1	High
Komatsu, 2003	Japan	10	10	1	High
Ozkoc, 2005	Turkey	21	21	2	High
Aglietti, 2006	Italy	30	30	1	High

Author, Year	Country	Intervention (n=)	Comparator (n=)	Number of surgeons	Risk of bias
Hart, 2006	Czech Republic	40	40	2	High
Kelly, 2006	USA	20	27	2	High
Kolisek, 2007	USA	40	40	Unclear	High
Dalury, 2008	USA	20	20	1	High
Han, 2008	Korea	15	15	1	High
Karachalios, 2008	Greece	50	50	1	High
Arnout, 2009	Belgium	30	30	1	High
Karpman, 2009	USA	20	19	1	High
Hay, 2010	Australia, Switzerland	16	16	1	High
Dutka, 2011	Poland	97	83	1	High
Lee, 2011	Korea	30	30	1	High
Matsumoto, 2011	Japan	25	25	1	High
Guy, 2012	UK	40	40		High
Pongcharoen, 2013	Thailand	30	30	1	High
Umrani, 2013	Korea	36	36	1	High
Thienpont, 2013	Belgium	150	150	1	High
Zhang, 2013	China	45	44	1	High

Author, Year	Country	Intervention (n=)	Comparator (n=)	Number of surgeons	Risk of bias
Aydogdu, 2014	Turkey	15	11	1	High
Cho, 2014	Korea	33	33	1	High
Heekin, 2014	USA	20	20	1	High
Jenkins, 2014	USA	60	60	3	High
Nutton, 2014	UK	12	16	2	High
Koh, 2016	Korea	51	51	1	High
Aslam, 2017	India	42	42	1	High
Faure, 1993	USA	20	20	NR	Unclear
Keating, 1999	USA	50	50	NR	Unclear
Cila, 2002	Turkey	10	12	NR	Unclear
Gelfer, 2003	Israel	15	15	NR	Unclear
Berth, 2007	Germany	20	20	NR	Unclear
Walter, 2007	USA	61	61	2	Unclear
Hernandez-Vacquero, 2010	Spain	30	40	NR	Unclear
Chiang, 2012	Taiwan	40	40	NR	Unclear
Siramunakul, 2012	Thailand	14	14	NR	Unclear

Author, Year	Country	Intervention (n=)	Comparator (n=)	Number of surgeons	Risk of bias
Reid, 2014	UK, Australia	37	31	2	Unclear
Tomek, 2014	USA	63	66	NR	Unclear
THA vs HA RCTs					
Baker, 2006	UK	41	40	NR	Low
Keating, 2006	UK	69	69	Unclear	Low
Blomfeldt, 2007	Sweden	60	60	9	Low
Van den Bekerom, 2010	Netherlands	137	115	Unclear	Low
Cadossi, 2013	USA	41	42	2	Low
Ravikumar, 2000	UK	91	89	NR	High
Macaulay, 2008	USA	17	23	14	High
Sharma, 2016	India	40	40	2	High
Dorr, 1989	USA	50	39	NR	Unclear
Mouzopou-lous, 2008	Greece	43	43	Unclear	Unclear

NR: Not reported

Table 4: Number of RCTs and risk of bias categorisation by date (median)

Risk of Bias	TKA vs TKA (%)		THA vs HA (%)	
	1997-2008	2009-2017	1997-2008	2009-2017
Low	4 (17)	19 (46)	3 (29)	2 (33)
High	14 (58)	17 (41)	2 (29)	1 (33)
Unclear	6 (25)	5 (12)	2 (29)	0 (0)
Totals	24	41	7	3

TKA: Total Knee Arthroplasty; THA: Total Hip Arthroplasty; HA: Hemiarthroplasty. Percentages might not be 100 due to 'rounding up' and 'rounding down'.

Table 5: Results of meta-analyses of knee and hip arthroplasty based on domain of surgeons' reported experience

Total knee arthroplasty							
Outcome	Comparison	Risk of bias	Number of trials	Mean difference	95% CI	p value *	I^2 statistic (%)
ROM (flexion)	MIS vs MP	High or unclear	10	-1.34	-3.88, 1.19	0.30	57
		Low	5	4.27	-0.46, 9.00	0.08	87
Operative time (minutes)†	MIS vs MP	High or unclear	13	5.87	0.52, 11.22	0.03	93
		Low	7	9.92	1.01, 18.83	0.001	90
Blood loss (ml)	MIS vs MP	High or unclear	8	-51.77	-218.69, 115.15	0.54	99
		Low	5	-0.64	-65.81, 64.54	0.98	69
Total hip arthroplasty							
Outcome	Comparison	Risk of bias	Number of trials	Risk ratio (RR)	95% CI	p value *	I^2 statistic (%)
Dislocation	THA vs HA	High or unclear	3	1.88	1.03, 3.43	0.04	0
		Low	4	5.01	1.33, 18.90	0.32	14
Revision	THA vs HA	High or unclear	4	0.39	0.17, 0.86	0.02	7
		Low	4	0.82	0.21, 3.22	0.77	58

*Test for overall effect; ROM: Range of motion; MIS: Minimally Invasive Surgeries, e.g. MV, SV, QS etc.; CI: Confidence Interval; THA: total hip arthroplasty; HA: Hip arthroplasty. †Tourniquet time if no other time data were provided.

Table 6: Risk of bias for other domains

Risk of bias domain	Randomisation	Allocation	Performance	Detection	Surgeon
Author, Year			**TKA RCTs**		
Weinhardt, 2004	High	High	Unclear	High	Low
Bathis, 2005	Low	Low	Low	Low	Low
Seon, 2006	High	High	Unclear	Low	Low
Chin, 2007	Low	Low	Low	Low	Low
Bridgman, 2009	Low	Low	Low	Low	Low
Juosponis, 2009	Low	Low	Unclear	Low	Low
Lin, 2009	Low	Low	Low	Low	Low
Sastre, 2009	Low	Low	Low	Low	Low
Bonutti, 2010	Low	Low	Unclear	Low	Low
Nestor, 2010	High	High	Low	Low	Low
Pan, 2010	Low	Low	Low	Low	Low
Van Hemert, 2010	High	High	Low	Low	Low
Kim, 2011	Low	Low	Unclear	Low	Low
Tasker, 2014	Low	Low	Unclear	Low	Low
Varnell, 2011	High	High	Unclear	High	Low
Varela-Egocheaga, 2011	Low	High	Unclear	High	Low
Bourke, 2012	Low	Low	Low	Low	Low
Masjudin, 2012	Low	Low	Unclear	Low	Low
Jain, 2013	Low	Low	Unclear	Low	Low
Jarvis, 2013	High	High	Unclear	High	Low
Wegryzn, 2013	Low	High	Low	Low	Low
Verburg, 2016	Low	Low	Unclear	Low	Low
Fezcko, 2016	High	High	High	Low	Low
Engh, 1997	High	High	Low	Low	High
Dalury, 1999	High	High	Low	Low	High

Risk of bias domain	Randomisation	Allocation	Performance	Detection	Surgeon
Parentis, 1999	High	High	High	High	High
Roysam, 2001	Low	Low	Low	Low	High
Komatsu, 2003	High	High	High	High	High
Ozkoc, 2005	High	High	High	High	High
Aglietti, 2006	Low	Low	Low	Low	High
Hart, 2006	High	High	Unclear	Low	High
Kelly, 2006	High	High	High	High	High
Kolisek, 2007	Low	Low	Unclear	High	High
Dalury, 2008	Low	High	Low	Low	High
Han, 2008	Low	High	Unclear	High	High
Karachalios, 2008	Low	Low	Unclear	Low	High
Arnout, 2009	High	High	High	Low	High
Karpman, 2009	Low	High	Low	Low	High
Hay, 2010	Low	Low	Low	Low	High
Dutka, 2011	High	High	Low	Low	High
Lee, 2011	Low	Low	Unclear	Low	High
Matsumoto, 2011	High	High	Low	High	High
Guy, 2012	Low	Low	Unclear	Low	High
Pongcharoen, 2013	Low	High	Unclear	High	High
Umrani, 2013	Low	High	Unclear	Low	High
Thienpont, 2013	High	High	Unclear	High	High
Zhang, 2013	Low	High	Unclear	Low	High
Aydogdu, 2014	High	High	Unclear	Low	High
Cho, 2014	High	High	High	High	High
Heekin, 2014	High	High	Low	High	High
Jenkins, 2014	Low	High	Low	Low	High
Nutton, 2014	Low	Low	Low	Low	High
Koh, 2016	Low	Low	Low	Low	High
Aslam, 2017	Low	High	Low	Low	High

Risk of bias domain	Randomisation	Allocation	Performance	Detection	Surgeon
Faure, 1993	High	High	Low	Low	Unclear
Keating, 1999	High	High	Low	Low	Unclear
Cila, 2002	High	High	Unclear	High	Unclear
Gelfer, 2003	High	High	Low	Low	Unclear
Berth, 2007	Low	Low	Unclear	High	Unclear
Walter, 2007	Low	Low	Unclear	Low	Unclear
Hernandez-Vacquero, 2010	Low	High	High	High	Unclear
Chiang, 2012	Low	High	Low	Low	Unclear
Siramunakul, 2012	High	High	Unclear	High	Unclear
Reid, 2014	Low	Low	Low	Low	Unclear
Tomek, 2014	Low	Low	High	Low	Unclear
THA vs HA RCTs					
Baker, 2006	Low	Low	Unclear	High	Low
Keating, 2006	Low	High	Unclear	Low	Low
Blomfeldt, 2007	Low	Low	Unclear	Low	Low
Van den Bekerom, 2010	Low	Low	Unclear	High	Low
Cadossi, 2013	High	High	Unclear	High	Low
Ravikumar, 2000	High	High	Unclear	High	High
Macaulay, 2008	Low	Low	Unclear	High	High
Sharma, 2016	Low	High	Unclear	High	High
Dorr, 1989	High	High	High	High	Unclear
Mouzopou-lous, 2008	High	High	Unclear	Low	Unclear