Open access Review

BMJ Surgery, Interventions, & Health Technologies

An umbrella review comparing computer-assisted and conventional total joint arthroplasty: quality assessment and summary of evidence

Mohamed Mosaad Hasan ,[1] Manrui Zhang,[2] Matthew Beal,[3] Hassan M K Ghomrawi[4]

To cite: Hasan MM, Zhang M, Beal M, *et al*. An umbrella review comparing computer-assisted and conventional total joint arthroplasty: quality assessment and summary of evidence. *BMJ Surg Interv Health Technologies* 2020;**2**:e000016. doi:10.1136/bmjsit-2019-000016

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjsit-2019-000016).

Received 21 June 2019
Revised 22 November 2019
Accepted 18 December 2019

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

[1]Institute of Public Health and Medicine, Northwestern University, Chicago, Illinois, USA
[2]Department of Medical Social Sciences, Northwestern University, Chicago, Illinois, USA
[3]Orthopedic Surgery, Northwestern University, Chicago, Illinois, USA
[4]Department of Surgery, Northwestern University, Chicago, Illinois, USA

Correspondence to
Dr Mohamed Mosaad Hasan; mohamedhasan2020@u.northwestern.edu

ABSTRACT

Background Systematic reviews (SRs) of computer-assisted (CA) total knee arthroplasty (TKA) and total hip arthroplasty (THA) report conflicting evidence on its superiority over conventional surgery. Little is known about the quality of these SRs; variability in their methodological quality may be a contributing factor. We evaluated the methodological quality of all published SRs to date, summarized and examined the consistency of the evidence generated by these SRs.

Methods We searched four databases through December 31, 2018. A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2) was applied to assess the methodological quality. Evidence from included meta-analyses on functional, radiological and patient-safety outcomes was summarized. The corrected covered area was calculated to assess the overlap between SRs in including the primary studies.

Results Based on AMSTAR 2, confidence was critically low in 39 of the 42 included SRs and low in 3 SRs. Low rating was mainly due to failure in developing a review protocol (90.5%); providing a list of excluded studies (81%); accounting for risk of bias when discussing the results (67%); using a comprehensive search strategy (50%); and investigating publication bias (50%). Despite inconsistency between SR findings comparing functional, radiological and patient safety outcomes for CA and conventional procedures, most TKA meta-analyses favored CA TKA, whereas most THA meta-analyses showed no difference. Moderate overlap was observed among TKA SRs and high overlap among THA SRs.

Conclusions Despite conclusions of meta-analyses favoring CA arthroplasty, decision makers adopting this technology should be aware of the low confidence in the results of the included SRs. To improve confidence in future SRs, journals should consider using a methodological assessment tool to evaluate the SRs prior to making a publication decision.

Key messages

What is already known on the subject?

- Systematic reviews of computer-assisted total knee and hip arthroplasty report conflicting evidence on its superiority over conventional arthroplasty.
- Little is known about the quality of these Systematic reviews (SRs); variability in their methodological quality may be a contributing factor.

What are the new findings?

- Most of the SRs showed that computer-assisted (CA) is equivalent to or better than conventional knee and hip arthroplasty; however, the confidence in the included SRs ranges from critically low to low.
- There is a plethora of outcomes measures and inconsistency in reporting outcomes in SRs.

How might these results affect future research or surgical practice?

- They highlight the need to conduct a high-quality SR to inform the decision on adopting CA knee and hip arthroplasty.
- Journals should consider using a methodological assessment tool (eg, A MeaSurement Tool to Assess systematic Reviews 2) to assess the quality of SRs.
- To strengthen evidence synthesis related to Total Knee Arthroplasty and Total Hip Arthroplasty outcomes, standardized outcome measures such as those recommended by the Outcome Measures in Rheumatology Trials Total Joint Replacement Working Group should be used and reported.

BACKGROUND

Instability and loosening of the implant are among the most common reasons for revisions of total knee arthroplasty (TKA),[1] and total hip arthroplasty (THA),[2] and are mainly due to inaccurate positioning of the implant and malalignment of the limb.[3] Computer-assisted (CA) arthroplasty, whether navigation or robotic systems, is proposed as an alternative to improve the accuracy of implant positioning and reduce malalignment[4] through providing intraoperative feedback to the surgeons before cutting the bones.[5] The navigation system guides the surgeon during the operation,[6] whereas the robot system operates on patients to insure precise cutting of the bones.[6]

Utilization of CA arthroplasty has been steadily increasing over the past few years in USA. For example, CA TKA has increased from 0.37% in 2005 to 2.32% in 2012 with average increase of 0.26% per year.[7] CA arthroplasty is associated with a steep learning curve (10–20 cases) for the surgeon, and significant costs for equipment and continuous maintenance for hospitals.[8 9] With concerns about overutilization of joint replacement,[10] investment in new technologies should be supported by high quality evidence to justify societal resources use.[11]

Multiple systematic reviews (SRs) have been conducted to compare CA TKA and THA to conventional approaches; however, the results of these SRs are conflicting.[12] For example, Shi and colleagues conducted a meta-analysis on the alignment outcomes of conventional versus CA TKA and suggested no difference,[13] whereas Rebal and colleagues found improved alignment outcomes with CA TKA.[14] Both were published in the same year, suggesting potential inconsistency in the methodology of conducting these SRs.

SRs and meta-analyses provide the highest level of evidence[15] and should be of high quality. However, little is known about the quality of SR comparing CA and conventional approaches. We conducted an umbrella review to (1) Evaluate the methodological quality of SRs. (2) Summarize and examine the consistency of the evidence generated by these SRs.

METHODS

Structure of the umbrella review

An umbrella review systematically evaluates and collects evidence from multiple SRs on all outcomes for which these have been conducted.[16] To develop our umbrella review, we followed the steps outlined in the Cochrane Handbook and other methodological papers on conducting umbrella reviews.[17–19] A protocol has been developed prior to the conduction of this review. We developed a comprehensive search strategy to include all SRs and meta-analyses comparing CA to conventional TKA and THA. We included both TKA and THA SRs because both procedures are elective orthopaedic procedures on the rise,[20] and provide long-lasting joints that are effective in alleviating pain and regaining function for patients with end-stage osteoarthritis.[21 22] Moreover, TKA and THA are often considered together in reimbursement policies. However, we summarized the results separately for TKA and THA because surgical outcomes may differ by joint type. We executed the study selection, data extraction and quality assessment of the SRs in duplicate. We used the validated AMSTAR 2 tool to assess the methodological quality of the included SRs.[23] To summarize and examine the consistency of the evidence, we compared conclusions from meta-analyses for outcomes common across more than one meta-analysis. We also calculated the corrected covered area (CCA)[24] to assess the level of overlap between meta-analyses in including the same pool of primary studies, since high levels of overlap should produce more consistent conclusions. We used Covidence SR software, Veritas Health Innovation, Melbourne, Australia (available at www.covidence.org).

Search strategy

We searched MEDLINE, EMBASE, the Cochrane Database and Epistemonikos to identify SRs published through May 2017 comparing CA-TKA and THA versus conventional TKA and THA. The search strategy combined keywords (eg, knee arthroplasty, hip arthroplasty) with subject heading terms (eg, surgery, CA, arthroplasty, replacement, knee, hip), and specialized clinical queries for SRs. We also searched the gray literature (eg, conference proceedings, reports, and doctoral theses). We reran the search strategy to include the rest of 2017 and the whole year of 2018. See online supplementary appendix A for details.

Screening and selection of SRs

To exclude irrelevant citations, one reviewer (MMH) screened all citations for their titles and abstracts. Full-text articles of the remaining citations were retrieved and assessed independently by two reviewers (MMH and MZ). Included reviews satisfied the following inclusion criteria: they were SRs as defined by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement,[25 26] and explicitly compared CA to conventional procedures. To identify any potential studies not identified in the database searches, we contacted authors of the included studies, and searched the bibliographies. Two reviewers (MMH and MZ) independently extracted data from the included SRs. The extracted data were general information about the SR (eg, year of publication, journal, and sources of funding) as well as details about the interventions, design, and main findings of the studies included in the reviews.

Assessment of methodological quality

After agreement on study inclusion, two reviewers (MMH, MZ) independently assessed the methodological quality of the included reviews using 'A MeaSurement Tool to Assess systematic Reviews 2' (AMSTAR 2). In case of disagreement, consensus was reached by discussions mediated by the senior author (HMKG).

First developed in 2007 (as AMSTAR) to only evaluate the methodological quality of SRs that synthesize evidence from randomized trials, this appraisal tool has been further developed, as AMSTAR 2, in 2017 to expand its use to SRs of randomized trials and non-randomized studies.[23] Since its release, AMSTAR 2 has been used widely in many umbrella reviews.[27–33] AMSTAR 2 comprises 16 domains, 7 of them are critical domains as they strongly undermine the confidence in the conclusions of the SRs: 1 domain is related to protocol registration, 2 are related to search strategy (adequacy and justifying studies' exclusion), 2 are related to the assessment of risk of bias of the included studies and its effect on SR conclusions, 1 is related to the method of evidence synthesis, and 1 is

Table 1 Critical and non-critical domains of A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2) quality assessment tool

Critical domains	Non-critical domains
Protocol development/registration before commencement of the review	Satisfying the components of PICO (population, intervention, comparison, and outcome)
Comprehensiveness of the literature search	Clarification of the reasons for selection of the study designs for inclusion in the review.
Justification for studies' exclusion	Study selection is done in duplicate
Assessment of the risk of bias	Extraction is done in duplicate
Appropriateness of meta-analysis	Detailed description of the included studies
Accounting for risk of bias in the discussion	Report on the sources of funding for the primary studies
Assessment of publication bias	Assessment of the potential impact of risk of bias on the results of the evidence synthesis
	Satisfactory explanation for any heterogeneity
	Report of any potential sources of conflict of interest

related to the publication bias (table 1).[23] The overall confidence in the results of the SRs is rated into four categories: high (no or one non-critical weakness), moderate (more than one non-critical weakness), low (one critical flaw with or without non-critical weaknesses), and very low (more than one critical flaw with or without non-critical weaknesses).[23] AMSTAR 2 is a valid and reliable instrument, similar to other appraisal tools of SRs.[34 35]

Summary and assessment of the consistency of the evidence

First, we summarized the evidence out of the SRs that conducted meta-analyses. We categorized the outcomes reported by SRs into functional, radiological, and patient safety related and others. Then, we assessed the overlap between those meta-analyses in using the same primary studies by calculating the CCA.[24] CCA assesses over-representation bias induced by using the same primary studies in different meta-analyses. As such, higher CCA suggests that the evidence summarized in an umbrella review is more likely to support the results of the primary studies included in multiple meta-analyses.[24] CCA uses the number of the included meta-analyses, the number of the primary publications including the duplications, and the number of the primary publications after removing the duplications.[24] CCA value ≤5 indicates slight overlap, 6–10 indicates moderate overlap, 11–15 indicates high overlap, and >15 indicates very high overlap.[24]

RESULTS

After deduplication, our initial search yielded 442 citations (figure 1). After screening the titles and abstracts, we excluded 330 citations not meeting the inclusion criteria. We retrieved the full texts of the remaining 112 citations for detailed full-text screening. After examining the full texts, we excluded 73 articles for not meeting our inclusion criteria (online supplementary appendix B). We also searched the gray literature and screened the references of the included studies and added two articles not captured by the original search strategy. Also, we contacted experts in the field resulting in one more article eligible for inclusion. As a result, we included 42 SRs.[12–14 36–74]

Description of the included SRs

The publication years of the included SRs ranged from 2004 to 2018 with most of the SRs (78%) published between 2011 and 2019. Four SRs were published in languages other than English: one German,[40] one Korean,[62] and two Mandarin.[54 68] Of all the 42 SRs, 3 compared conventional to CA modalities of both TKA and THA,[38 52 71] 9 addressed THA,[36 42 49 53 55 59 63 70 74] and the rest addressed TKA. The approach to evidence synthesis was as follows: 7 SRs synthesized the evidence qualitatively,[38 40 44 52 56 63 66] 7 SRs conducted meta-analysis and qualitative evidence synthesis,[37 45 67 70 71 73 74] and the remaining 28 SRs conducted only meta-analysis. Regarding the intervention, four SRs compared minimally invasive (MI) CA TKA to MI TKA,[39 64–66] one SR compared MI THA to CA THA to conventional THA,[63] four SRs compared robotic THA to conventional THA,[38 52 70 71] five SRs compared robotic TKA to conventional TKA.[38 52 66 71 73] The remaining SRs compared CA navigation arthroplasty versus conventional surgery.

Methodological quality of the included SRs

Based on AMSTAR 2, confidence was rated critically low in the results of 39 studies and low in 3 studies. Low confidence was attributed to reasons such as 38 (90.5%) SRs not reporting development of a protocol; 34 (81%) SRs not providing a list of the excluded studies and not justifying the exclusion. In addition, of the 28 SRs that included non-randomized primary studies, 24 (85.%) SRs did not account for the risk of bias when interpreting the results. Figure 2 shows the prevalence of critical flaws and non-critical weaknesses across the included SRs. Table 2 shows a detailed rating of the critical flaws and non-critical weaknesses for each SR.

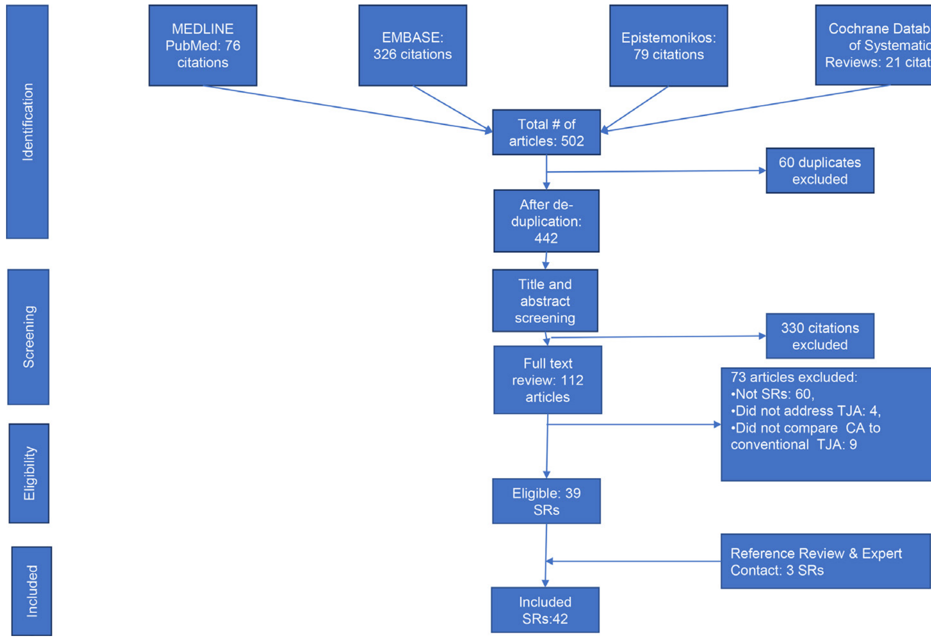

Figure 1 Study selection process: describing the steps of the study selection process. The process is divided into four phases: identification of all references from the consulted databases, screening of the titles and abstracts, eligibility determination by reviewing the full texts, and finally including the selected final list in the review. SR, systematic review.

Summary and consistency of the evidence

Consistency of the conclusions from meta-analyses.

Functional outcomes

Three functional outcomes were compared for TKA.[14 45 60 64 67 69 71 72] Knee Society Scores: Two out of eight SRs showed that CA TKA had superior scores,[14 60] while six out of eight SRs showed no difference.[45 64 67 69 71 72] Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): One out of three SRs showed slightly improved scores with the CA TKA,[60] while two out of three SRs showed no difference.[71 72] Range of motion: two out of two SRs showed postoperative improvement with CA TKA.[64 69] Two SRs reported meta-analysis results for THA and found no significant difference in the functional scores (Harris Hip Score (HHS), Merle d'Aubigne Hip Score, and Japanese Orthopedic Association Score) between CA and conventional THA.[70 71] (figure 3)

Radiological outcomes

Six radiological outcomes were compared for TKA (figure 3). Mechanical axis malalignment: 12/15 SRs showed significantly less malalignment with CA TKA[12 14 37 43 46–48 51 54 57 60 66 73] whereas 3/15 SRs showed no significant difference.[13 64 65] Coronal plane femoral malalignments: six out of eight SRs showed significantly

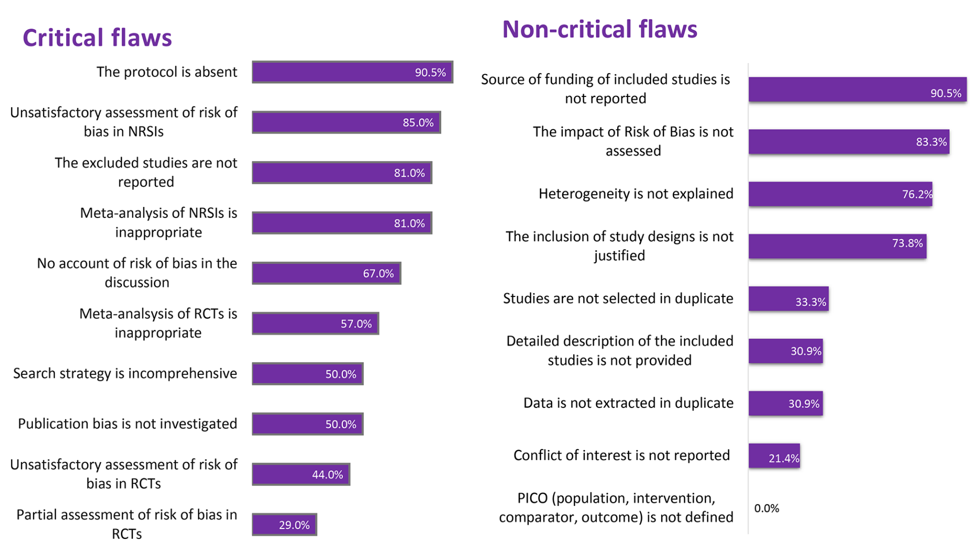

Figure 2 Quality assessment of the included systematic reviews: summarizing the results of the quality assessment via A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2). The methodological flaws are classified into critical flaws and non-critical weaknesses. Each bar represents the percentage of the occurrence of the flaw in the systematic reviews.

Table 2 Critical flaws and non-critical weaknesses of the included SRs

Article	Overall	Q1	Q2*	Q3	Q4*	Q5	Q6	Q7*	Q8	Q9 -RCT*	Q9-NRSI*	Q 10	Q11-RCT*	Q11-NRSI*	Q 12	Q 13*	Q 14	Q 15*	Q 16	No. of critical flaws	No. of non-critical weaknesses
Oh 2014	Low	Y	Y	N	P Y	Y	Y	Y	P Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	1	3
Xie, 2012	Low	Y	P Y	N	Y	Y	Y	N	Y	Y	NA	N	Y	NA	Y	Y	N	Y	Y	1	3
Karunaratne 2018	Low	Y	Y	N	Y	Y	Y	Y	PY	Y	Y	N	Y	Y	N	Y	Y	N	Y	1	3
Gandhi 2009	Critically low	Y	N	Y	P Y	Y	Y	N	P Y	P Y	NA	N	Y	NA	Y	Y	Y	Y	N	2	2
Xu 2014	Critically low	Y	P Y	N	N	Y	Y	N	P Y	Y	NA	N	Y	NA	Y	Y	N	Y	Y	2	3
Beckmann, 2009	Critically low	Y	N	Y	P Y	N	Y	Y	N	Y	NA	N	N	NA	N	N	N	Y	N	3	6
Li 2014	Critically low	Y	N	N	P Y	Y	Y	N	P Y	Y	NA	N	Y	NA	Y	Y	Y	N	Y	3	2
Liu 2014	Critically low	Y	N	N	N	Y	N	N	N	P Y	NA	N	Y	NA	Y	Y	Y	Y	Y	3	4
Smith 2012	Critically low	Y	N	N	Y	Y	Y	N	P Y	P Y	NA	N	N	NA	N	Y	N	Y	N	3	5
Hetaimish 2012	Critically low	Y	N	Y	P Y	N	Y	N	N	Y	NA	N	Y	NA	N	N	N	N	Y	4	5
Reininga 2010	Critically low	Y	N	N	N	Y	N	N	P Y	P Y	N	N	NA	NA	NA	Y	Y	NA	Y	4	3
Thieopont 2013	Critically low	Y	N	N	P Y	Y	N	Y	P Y	N	N	N	NA	NA	NA	N	N	NA	Y	4	4
Mannan 2018	Critically low	Y	N	Y	PY	Y	Y	N	PY	Y	Y	N	Y	Y	Y	Y	N	N	Y	4	3
Alcelik 2016	Critically low	Y	N	Y	P Y	Y	Y	N	P Y	P Y	N	N	Y	N	N	N	N	Y	Y	5	3
Cheng 2012	Critically low	Y	N	N	P Y	N	Y	N	N	P Y	N	N	Y	N	N	N	Y	Y	Y	5	5
Cheng,2012	Critically low	Y	N	N	P Y	N	Y	N	N	P Y	N	N	N	N	N	Y	N	Y	Y	5	6
Computer-assisted hip and knee arthroplasty	Critically low	Y	N	N	N	N	N	N	Y	N	N	N	NA	NA	NA	Y	N	NA	Y	5	5
Fu, 2012	Critically low	Y	N	Y	P Y	Y	Y	N	N	P Y	NA	N	N	NA	N	N	N	N	Y	5	4
Han 2016	Critically low	Y	N	N	P Y	Y	Y	N	N	NA	Y	N	NA	N	N	N	N	N	N	5	6
Karthik 2015	Critically low	Y	N	N	N	Y	N	N	P Y	N	N	N	NA	NA	NA	Y	N	NA	Y	5	4
Rebal 2014	Critically low	Y	N	Y	N	N	Y	Y	N	P Y	NA	N	N	NA	N	N	N	N	Y	5	5
Shi 2014	Critically low	Y	N	N	N	Y	N	N	P Y	N	NA	N	Y	NA	N	N	Y	Y	Y	5	4
Shin 2016	Critically low	Y	N	N	P Y	Y	Y	N	P Y	Y	N	N	Y	N	N	N	N	Y	N	5	5
Zamora 2013	Critically low	Y	N	N	N	N	Y	N	P Y	P Y	NA	N	Y	NA	N	N	Y	N	Y	5	4
Snijders 2017	Critically low	Y	N	N	PY	N	N	N	PY	N	NA	N	N	NA	N	N	N	Y	N	5	8
Panjwani 2019	Critically low	Y	N	Y	PY	Y	Y	PY	Y	Y	N	N	N	N	N	N	N	Y	Y	5	3
Bathis 2006	Critically low	Y	N	N	N	N	N	N	P Y	N	N	N	NA	NA	NA	N	N	NA	Y	6	5
Bauwens 2007	Critically low	Y	N	N	N	N	N	N	P Y	P Y	N	N	N	Y	N	N	N	Y	Y	6	6
Burnett 2013	Critically low	Y	N	N	N	N	N	N	N	N	N	N	NA	NA	NA	N	N	NA	Y	6	6
Nvicoff 2010	Critically low	Y	N	N	N	N	Y	Y	P Y	N	NA	N	N	NA	N	N	N	N	Y	6	5
Wang 2014	Critically low	Y	N	N	N	Y	Y	N	P Y	Y	NA	N	N	NA	N	N	N	N	Y	6	4

Continued

Table 2 Continued

Article	Overall	Q1	Q2*	Q3	Q4*	Q5	Q6	Q7*	Q8	Q9-RCT*	Q9-NRSI*	Q 10	Q11-RCT*	Q11-NRSI*	Q 12	Q 13*	Q 14	Q 15*	Q 16	No. of critical flaws	No. of non-critical weaknesses
Cheng 2011	Critically low	Y	N	N	PY	N	Y	N	N	N	N	N	N	N	N	N	Y	Y	N	7	6
Mason 2007	Critically low	Y	PY	N	N	Y	Y	N	N	N	N	Y	Y	N	N	N	Y	N	Y	7	3
Moskal 2011	Critically low	Y	PY	N	N	Y	Y	PY	PY	N	N	Y	N	N	N	N	N	N	Y	7	3
Chen 2018	Critically low	Y	N	N	PY	N	Y	N	Y	N	N	N	N	N	N	N	N	Y	Y	7	4
Cheng 2011	Critically low	Y	N	Y	PY	N	Y	N	PY	N	N	N	N	N	N	N	N	N	Y	8	4
Liu 2015	Critically low	Y	N	N	N	N	Y	N	N	PY	N	N	N	N	N	N	N	N	Y	8	6
Mejer 2014	Critically low	Y	N	Y	N	Y	Y	N	PY	N	N	N	N	N	N	Y	N	N	Y	8	3
VanderList 2016	Critically low	Y	N	N	N	Y	N	N	PY	N	N	N	N	N	N	N	N	Y	Y	8	5
Brin, 2011	Critically low	Y	N	Y	N	N	N	N	PY	N	N	N	N	N	N	N	N	N	Y	9	5
Lüring 2006	Critically low	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	9	7
Moskal 2014	Critically low	Y	N	N	N	N	Y	N	PY	N	N	Y	N	N	N	N	N	N	N	9	5

*Critical domains.
Y, yes; N, no; PY, partial yes. NA, not applicable.

less malalignments with CA TKA,[12 14 43 47 57 60] whereas two out of eight SRs showed no difference.[13 64] Coronal plane tibial component outliers: five out of six SRs showed significantly fewer outliers,[12 14 43 57 60] while one out of six SRs showed no difference.[13] Sagittal femoral component malalignment: two out of two SRs showed significantly less malalignment with CA TKA.[12 57] Femoral slope malalignment: one out of one SR showed significantly less slope with CA TKA.[60] Tibial slope malalignment: two out of three SRs showed significant difference in favor of CA TKA,[12 60] while one out of three SRs showed no difference.[57]

Seven radiological outcomes were compared for THA. Cup positioning outside the safe zone: significantly reduced with CA THA in five out of five SRs.[36 42 55 59 70] Number of outliers of acetabular cups outside the desired alignment range: two out of two SRs showed no significant difference.[49 55] Cup inclination: one out of four SRs showed improved inclination in the navigated group,[74] while three out of four SRs reported no significant difference.[36 42 55] Cup anteversion: one out of four SRs reported improved anteversion,[74] while three SRs reported no difference.[36 42 55] Postoperative dislocation: one out of two SRs reported significant reduction with CA THA,[59] while one out of two SRs reported insignificant difference.[36] Reduction in the leg length discrepancy: one out of two SRs showed significant reduction within the navigated group,[36] while one out of two SRs showed no significant difference.[70] Heterotopic ossification: one out of one SR reported a higher rate in patients that underwent conventional THA.[70]

Patient safety related and other outcomes

Six patient safety related outcomes were compared for TKA (figure 3). Complications and adverse events: two out of three SRs showed no difference,[39 46] while one out of three SRs showed that CA TKA is associated with fewer complications and adverse events.[60] Postoperative blood loss and calculated blood loss: one out of one SR showed less blood loss with CA TKA.[50] Allogenic blood transfusion rate: one out of one SR showed no difference.[50] Operative blood loss: two out of two SRs showed no difference[60 69] Hematocrit value after surgery: one out of one SR showed no difference.[50] Tourniquet time: one out of one SR showed decreased tourniquet time with conventional TKA.[60] Table 3 provides details on TKA outcome measures.

Four patient safety related outcome measures were compared for THA (figure 3). Operative time: one out of two SRs reported a significantly longer time with the navigated procedures,[36] another SR showed no significant difference.[70] Deep venous thrombosis (DVT): one out of one SR concluded no significant difference.[36] Joint infection: one out of one SR concluded no difference.[59] Total complication rate: one out of one SR showed higher rate in patients who underwent conventional THA.[70] Table 4 provides details on THA outcome measures.

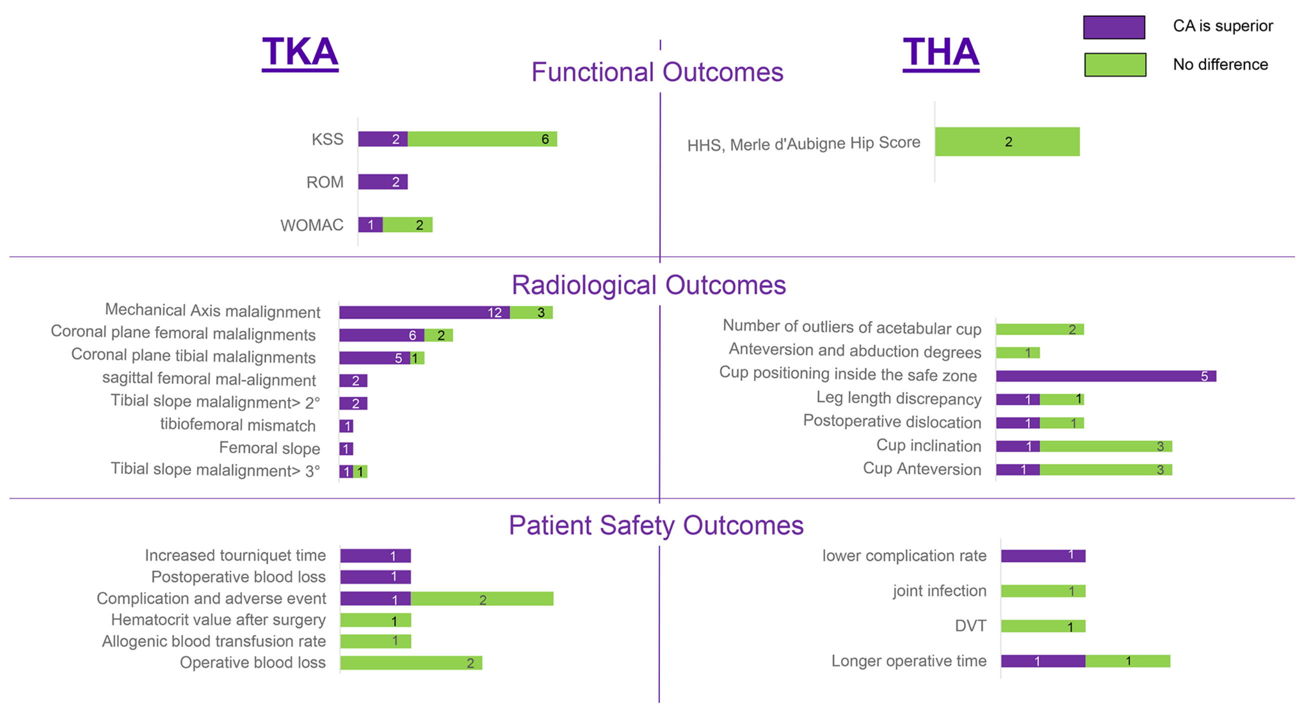

Figure 3 Consistency of the evidence: describing the consistency of the evidence of the included systematic reviews that conducted meta-analyses. The outcome measures are categorized into three categories: functional, radiological, and patient-safety outcomes and . Each bar represents the number of meta-analyses that concluded that computer-assisted (CA) surgery is superior (in purple), and those that concluded no difference (in green). KSS, Knee Society Score; ROM, range of motion; THA, total hip arthroplasty; TKA, total knee arthroplasty; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

Overlap between SRS

For TKA, the number of primary studies included in the meta-analyses was 180 (468 before removal of duplicates) resulting in a CCA of 8%, which indicates moderate overlap (online supplementary appendix C). For THA, the number of primary studies included in the meta-analyses was 23 (36 before removal of duplicates), resulting in a CCA of 13.2%, which indicates high overlap (online supplementary appendix D).

DISCUSSION

We aimed to evaluate the methodological quality of SRs comparing CA and conventional arthroplasty using the AMSTAR 2 tool and summarize and examine the consistency of the evidence generated by these SRs. Our umbrella review identified 42 SRs. We found low confidence in the evidence provided by 3 SRs, and very low confidence in the evidence provided by the remaining 39 SRs. Most SRs concluded that CA procedures had generally better radiological and similar functional outcomes compared with conventional procedures. However, depending on the outcome, discrepancy in the conclusions of the SRs varied significantly. Patient safety related outcomes were infrequently reported in the included SRs. Over-representation of the primary studies was shown by the moderate overlap among TKA SRs, and high overlap among THA SRs. These conclusions have implications for policy makers evaluating and adopting this technology, and for journals considering future SRs for publication.

We found that most of the included SRs showed that CA procedures are equivalent or better than conventional ones, which may have been used to support the increase in utilization of CA THA and TKA.[7] However, given that these SRs are inflicted by the very low confidence in their conclusions, we caution that these findings should not be used to support further adoption of this technology. Moreover, the published SRs included little data on patient related safety outcomes, which creates a major gap in the assessment of the technology, especially knowing that THA and TKA are among the top seven orthopaedic procedures with the highest complication rates.[75] While the US Food and Drug Administration approved the use of navigation systems, postmarket surveillance is still needed to minimize unintended consequences, as is the case with metal-on-metal hip resurfacing, which proved costly and unsafe.[76 77]

There is a plethora of outcome measures and inconsistency in reporting outcomes in SRs. This finding highlights the need to standardize the outcomes reported by both the primary studies and SRs,[78–80] in order to synthesize the evidence more comprehensively and meaningfully for technology assessment and guidelines development. To address this, core domains have been developed for clinical trials of TKA and THA developed by the Outcome Measures in Rheumatology Trials (OMERCAT) Total Joint Replacement Working Group;[81 82] however, those core domains are not yet fully represented in trials and SRs of CA TKA and THA. For example, the included SRs

Table 3 Outcome measures of SRS that addressed total knee arthroplasty

Functional outcomes of SRs that addressed TKA	
Knee Society Scores (KSS)	
SR	**Results**
Xie, 2012	Within 6 months: mean standard difference: 4.47; (95% CI 21.05 to 9.99, p=0.36)
Cheng, 2012	At 3 months (WMD=1.11, 95% CI −6.33 to 8.56) and 6 months (WMD=2.13, 95% CI −2.53 to 6.79) follow-up
Rebal, 2014	3 months postoperative change: CA TKA had a mean score increase of 68.5 (52.9–75.0), significantly superior to the mean score of 58.1 (47.3–64.0) for knees performed with conventional guides (p=0.03, 95% CI 1.13 to 19.78) Postoperative change at 12–32 months follow-up: knees in the CA TKA groups improved by a mean of 53.1 (37.7–96.6), significantly superior to the mean of 45.8 (32–89.5) for the CONV group (p<0.01, 95% CI 2.87 to 11.90)
Shin, 2016	The pooled data showed that the mean difference in the postoperative KSS was 11.15 points higher with the MINA approach than the CONV approach, but this difference was not statistically significant (95 % CI −8.55 to 30.84; N.S.; I 2=98%
VanderList, 2016	6 months: pooled mean difference=5.2 (3.41, 7.00). 1 year: (8.46, 90.65, 16.28). 2 years: 1.97 (−1.91, 5.84). More than 4 years: 2.65 (0.84, 4.46). Total=2.86 (0.96, 4.76) Navigated vs conventional TKA surgery with navigation systems aiming to control for alignment and component position: pooled mean difference=0.66 (−2.06 to 3.38) Navigated vs conventional TKA surgery with navigation systems aiming to control for alignment and component position with controlling for soft tissue balance: pooled mean difference=4.84 (1.61 to 8.07)
Moskal, 2014	KSS were slightly larger for NAV, demonstrating some improvement compared with CONV. Knee Society Function Scores: standardized difference in means=−0.341 95% CI (−0.532 to −0.150) p=0.000 Knee Society Knee Scores: standardized difference in means=−0.294 95% CI (−0.489 to −0.100) p=0.003 Knee Society total Scores: standardized difference in means=−0.623 95% CI (−0.940 to −0.307) p=0.000
Panjwani, 2019	Follow-up ≥ 2 years and <5 years postoperatively: no difference (p=0.13; pooled mean difference=−0.86; CI −1.96 to 0.25) Follow-up >5 years and <8 years postoperatively: no significant difference in the two groups (p=0.09) Follow-up >8 years postoperatively: no significant difference in the two groups (p=0.91)
WOMAC Scores	
Moskal 2014	WOMAC Scores (Pain Score, Stiffness Score, and Physical Function Score) were slightly lower for NAV, showing some improvement compared with CONV. WOMAC Pain Scores: standardized difference in means=−0.472; 95% CI (−0.826 to −0.117), p<0.009 WOMAC Stiffness Scores: standardized difference in means=−0.274 95% CI (−0.538 to −0.010), p<0.042 WOMAC Physical Function Scores: standardized difference in means=−0.369 95% CI (−0.619 to −0.173), p<0.001
Panjwani, 2019	Follow-up ≥2 years and <5 years postoperatively: no difference between the two groups (p=0.60; pooled mean difference=0.86; CI −2.32 to 4.04) Follow-up >5 years and <8 years postoperatively: significantly better in the CAS-TKA group (p<0.0001; pooled mean difference=−2.05; CI −2.82 to −1.28) Follow-up >8 years postoperatively: no difference in the two groups (p=0.94; pooled mean difference= −0.08; CI −2.10 to 1.95)
Karunaratne, 2019	Low quality of evidence of no difference (mean difference= −0.51; 95% CI −1.95 to 0.94)
Knee range of motion (ROM)	
Xie, 2012	Mean standard difference: 1.38; 95% CI 21.43 to 4.18, p=0.34)
Shin 2016	The pooled mean difference in postoperative flexion ROM was 16.64 (95 % CI 14.26 to 19.01, p<0.001; I2=0%)
Hospital for special surgery (HSS)	
Karunaratne, 2019	Medium-term: low quality of evidence of no difference between groups (mean difference=0.04 (−2.94 to 3.01)) Long-term: low quality of evidence of no difference between groups (mean difference=−0.51 (−1.83 to 0.82))
Radiological outcomes of SRs that addressed TKA	
Mechanical axis malalignment	
Smith, 2012	No difference
Hetaimish, 2012	Malalignment >3°: RR=0.37 (95% CI 0.24 to 0.58, p=0.00001), MAM>2°: 0.54 (95% CI 0.42 to 0.69, p=0.004)
Thieopont, 2013	Malalignment >2°, the effect measures (ie, OR and risk ratio) ranged from 0.21 to 0.76 Malalignment >3°, the reduction was comparable, and effect measures ranged from 0.19 to 0.79
Cheng, 2012	Mechanical axis malalignment >3°: OR=0.4 (95% CI (0.31 to 0.51)
Fu, 2012	Malalignment of >3°: a meta-analysis OR of 0.26 (95% CI 0.17 to 0.38) Malalignment at >2°: a meta-analysis OR of 0.33 (95% CI 0.26 to 0.42)
Rebal, 2014	The risk difference of alignment within 3° of ideal is 0.14 (CI 0.1 to 1.18) The absolute value of degrees deviation: for the CAS group the mean deviation was 1.3° (1.0° to 1.9°), significantly less than the CONV group deviation of 2.4° (1.8° to 3.2°) (P < .01, 95% CI −1.38 to −0.67 The mean average of the degrees of deviation of the mechanical axis: The CAS group had a mean deviation of 0.3° (−0.6° to 1.0°) while the CONV group had a mean deviation of 0.5° (−2.4° to 1.2) (p=0.33, 95% CI −0.65 to 0.22) Note: positive values for varus deviation and negative values for valgus deviation

Continued

Table 3 Continued

Functional outcomes of SRs that addressed TKA	
Shi, 2014	The pooled OR for overall outliers in mechanical axis showed no difference between the two groups; no heterogeneity was observed (p=1.000; I2=0.0%)
Shin 2016	No statistical difference is present (95 % CI 1.01 to 0.54; N.S.; I2=64%)
Zamora, 2013	OR of postoperative alignment of the mechanical axis in the frontal plane (postoperative deviation of 3° from target angle of 180°=2.32. (95% CI 1.77 to 3.04)
Cheng, 2011 (A)	RR=0.4; 95% CI 0.31 to 0.5
Mason, 2007	Malalignment >3 °: a meta-analytic mean OR 0.22 (95% CI 0.16 to 0.29) Malalignment >2 °: a meta-analytic mean OR of 0.35 (95% CI 0.28 to 0.43)
Cheng, 2011 (B)	Malalignment >3° (RR=0.19, 95% CI 0.11 to 0.32, p<0.00001, I2=10%
Liu, 2014	Malalignment >3°(OR=0.55; 95% CI 0.44 to 0.68, p<0.001)
Brin, 2011	Malalignment >3°: (prospective randomized studies alone): OR=0.03 (95% CI 0.15 to 0.52), (prospective randomized and retrospective studies): OR=0.21 (95% CI 0.12 to 0.33)
Moskal, 2014	Femoral flexion angle deviation from neutral: standardized difference in means: −0.606. 95% CI −0.856 to −0.356, p=0.000 Anatomic axis outliers: standardized difference in means: 0.242; 95% CI 0.098 to 0.593, p=0.002 Mechanical axis outliers: OR=0.356; 95% CI 0.237 to 0.536, p=0.000 Tibial component alignment outliers: OR= 0.356; 95% CI 0.237 to 0.536,) p =0.000 Femoral component alignment outliers: OR=0.387; 95% CI 0.254 to 0.589, p=0.000
Mannan, 2018	Weighted mean differences of postoperative alignment to be more accurate in the robotic knee group: mean difference=−0.63; 95% CI −1.18 to 0.08, z=2.25, p=0.02
Coronal plane femoral component outliers	
Fu, 2012	The OR of malalignment of >3° was estimated at 0.33 (95% CI 0.14 to 0.75)
Rebal 2014	Within 3° of ideal (90°): 97.6% (94%–100%) in the CAS groups, significantly more than the 87.4% (81%–97%) in the CONV group (P b .01, 95% CI 0.05 to 0.14)
Shi 2014	The pooled data in the random-effects model showed no difference between the two groups. No heterogeneity was observed
Shin 2016	The pooled mean difference was similar between the MINA and CONV approaches (95 % CI −0.91 to 2.97; N.S.; I2=95%)
Cheng 2011 (A)	RR=0.37; 95% CI 0.22 to 0.64
Mason 2007	Malalignment >3°: a meta-analytic mean OR 0.34 (95% CI 0.24 to 0.48) Malalignment >2°: a meta-analytic mean OR 0.29 (95% CI 0.14 to 0.56)
Brin, 2011	Femoral angle (prospective randomized and retrospective studies) OR=0.19 (95% CI 0.08 to 0.39)
Moskal 2014	Standardized difference in means: −0.663; 95% CI −0.929 to −0.397, p=0.000
Coronal plane tibial component outliers	
Fu, 2012	Malalignment of >3°: the OR was estimated at 0.29 (95% CI 0.16 to 0.50)
Rebal 2014	Within 3° of ideal (90°) demonstrated equivalent results in the CAS group (92.1% (83–100%)) and the CONV group (91.7% (82–97%)) (p=0.73; 95% CI −0.06 to 0.09)
Shi 2014	No difference between the two groups.
Mason 2007	Malalignment >3°: a meta-analytic mean OR 0.36; 95% CI 0.23 to 0.57 Malalignment >2°: a meta-analytic mean OR 0.26; 95% CI 0.17 to 0.40
Brin, 2011	Tibial angle (prospective randomized and retrospective studies): OR=0.19 (0.07 to 0.41)
Moskal 2014	Standardized difference in means: −0.268; 95% CI −0.350 to −0.185, p=0.000
Sagittal femoral component malalignment	
Fu, 2012	Malalignment of >3°: OR=0.35; 95% CI 0.17 to 0.74 Malalignment at >2°: OR=0.22; 95% CI 0.06 to 0.76
Mason, 2007	Malalignment >3°: a meta-analytic mean OR=0.39; 95% CI 0.11 to 1.34 Malalignment >2°: a meta-analytic mean OR=0.13; 95% CI 0.03 to 0.54
Femoral slope	
FMoskal, 2014	Femoral slope outliers: OR=0.465; 95% CI 0.303 to 0.712, p=0.000
Tibial slope	
Fu, 2012	Malalignment of >3°: OR=0.43; 95% CI 0.30 to 0.61 Malalignment at >2°: OR=0.42; 95% CI 0.23 to 0.76
Mason 2007	Malalignment >3°: a meta-analytic mean OR=0.43; 95% CI 0.13 to 1.39 Malalignment >2°: a meta-analytic mean OR=0.31; 95% CI 0.16 to 0.61

Continued

Table 3 Continued

Functional outcomes of SRs that addressed TKA	
Moskal 2014	Standardized difference in means: −0.268; 95% CI −0.350 to −0.185, p=0.000 Tibial slope outliers: OR=0.474; 95% CI 0.309 to 0.729, p=0.001
Tibiofemoral mismatch	
Mejer, 2014	Standardized mean difference=−0.37 (−1.67 to −0.08)
Component axial rotation	
Mejer 2014	Postoperative rotation of the femoral component: standardized mean difference = −7° (−0.19 to 0.04) Postoperative rotation of the tibial component=0.110.010.24) Number of femoral rotational: pooled OR=1.05 (0.78 to 1.43) Number of tibial rotational: pooled OR=1.12 (0.68 to 1.47)
Patient safety outcomes of SRs that addressed TKA	
Allogenic blood transfusion rate	
Han 2016	The difference is not statistically significant OR 0.70; 95 % CI 0.49 to 1.01; I 2=0%
Operative blood loss	
Xie, 2012	No significant difference. Mean standard difference=-54.38; 95% CI -119.76 to 11.00; p=0.10)
Moskal 2014	No difference
Change in hemoglobin concentration/hematocrit before and after surgery	
Han, 2016	The pooled mean difference in change of hemoglobin was −0.39 g/ dL (95 % CI −0.67 to −0.11, p=0.006; I2=75 %) The pooled mean difference of change of hematocrit was similar in the two groups (0.24%; 95 % CI −0.89% to 0.41%; N.S.; I2=25%)
Postoperative blood loss via drainage	
Han, 2016	The pooled standard mean difference in drainage blood loss was −83.1 mL (95 % CI −159.0 to −7.1, p=0.03; I2=75 %)
Calculated total blood loss	
Han, 2016	The pooled standard mean difference in calculated total blood loss was −185.4 mL (95 % CI −303.3 to −67.5 mL; p=0.002)
Tourniquet time	
Moskal 2014	CONV had significantly lower tourniquet times: standardized difference in means: 0.993; 95% CI 0.567 to 1.419, p=0.000
Allogenic blood transfusion rate	
Han, 2016	The difference is not statistically significant OR 0.70; 95 % CI 0.49 to 1.01; I2=0%
Complications/adverse events	
Alcelik, 2016	The OR between the MIS CA and the MIS group was 1.31; 95% CI 0.47 to 3.65, p=0.61
Cheng, 2011 (B)	No significant difference (RR=1.50; 95% CI 0.44 to 5.11, p=0.51)
Bauwens 2007	Risk ratio, 0.69; 95% CI 0.44 to 1.08. There was no evidence of a difference in infection rates (risk ratio, 0.97; 95% CI 0.33 to 2.85) or the onset of thromboembolic events (risk ratio, 0.64; 95% CI 0.31 to 1.34)

A) Cheng T, Zhang G, Zhang X. Imageless navigation system does not improve component rotational alignment in total knee arthroplasty. *J Surg Res* 2011;171(2):590–600. doi: 10.1016/j.jss.2010.05.006.
B) Cheng T, Zhang G, Zhang X. Clinical and radiographic outcomes of image-based computer-assisted total knee arthroplasty: an evidence-based evaluation. *Surg Innov* 2011;18(1):15–20. doi: 10.1177/1553350610382012.
CA, computer assisted; CAS, computer-assisted surgery; CONV, conventional; KSS, Knee Society Score; MIS, Minimally invasive; NAV, navigation; ROM, range of motion; RR, relative risk; SR, systematic review; TKA, total knee arthroplasty; WMD, Weighted mean difference; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

in this review did not report measures related to patient satisfaction, revision, and death domains and only few reported on adverse events domains.

By synthesizing evidence from RCTs and other comparative non-randomized studies,[83] SRs provide much needed data for the evaluation of medical devices. The Idea Development Evaluation Assessment and Long-term (IDEAL) framework, allows robust evaluation of surgical innovations based on its stage of development.[84 85] Although 35 of the 42 SRs were published after the publication of the IDEAL framework in 2009, none reported the IDEAL stage of the primary studies. We attribute the under-reporting of the IDEAL framework in the included SRs mainly to the lack of awareness of its existence and its value, but partly also because SRs are perceived as outside of the scope of the framework.[86] We suggest including SRs in the IDEAL framework as they have the potential to inform the evaluation and assessment phases depending on the robustness of the SRs and the quality of the primary studies.

Our findings also have important implications for journals considering SRs, in general, and on this topic in particular. Since the availability of the Quality of Reporting of Meta-analyses (QUOROM) statement in 1999 and the PRISMA statement in 2009,[25 87] most journals require adherence to these guidelines to improve the reporting quality of SRs. Despite the enforcement of these reporting requirements, confidence was low in all included SRs in our study. Therefore, to enhance the

Table 4 Outcome measures of SRS that addressed total hip arthroplasty

SR	Results
Functional outcomes	
Harris Hip Score (HHS)dd	
Karunaratne, 2019	Long term: low-quality evidence of no difference between groups (mean difference=−2.90 (−9.04 to 3.24))
Chen, 2018	Pooled analysis of functional scores found no significant differences between robotic-assisted and conventional THA (weighted mean difference=0.12; 95% CI −0.09 to 0.34)
Merle d'Aubigne	
Karunaratne, 2019	Short-term: no significant differences between groups (mean difference=−0.41 (−5.31 to 4.48)) Long-term: low-quality evidence of no difference between groups (mean difference=−1.25 (−3.90 to 1.41))
Chen, 2018	Pooled analysis of functional scores found no significant differences between robotic-assisted and conventional THA (weighted mean difference=0.12, 95% CI −0.09 to 0.34)
Radiological outcomes	
The number of outliers of acetabular cups	
Gandhi 2009	The statistically significant beneficial OR for the number of outliers was 0.285 (95% CI 0.143 to 0.569, p<0.001). There was no evidence of statistical heterogeneity between studies (p=0.88)
Liu 2015	No significant difference between the two groups (−1.46; CI −3.00 to −0.08, p=0.06)
Cup inclination	
Xu 2014	No significant difference between the groups (MD=−0.93°; 95% CI −3.88 to 2.02, p=0.54)
Snijders 2017	Significantly better accuracy for the NAV group than for the freehand group (mean difference= −1.87; 95% CI −3.31 to −0.44)
Liu 2015	The weighted mean difference in inclination between conventional and imageless navigation groups was not statistically significant (0.2; 95% CI −1.69 to 2.09, p=0.83)
Beckmann, 2009	The weighted mean difference in inclination between conventional and computer-assisted positioning was not statistically significant (−0.89°; 95% CI −4.2 to 2.4)
Cup anteversion	
Xu 2014	No significant difference between the groups (MD=−0.96°; 95% CI −4.29 to 2.37, p=0.57)
Snijders 2017	Significantly better accuracy for the NAV group than for the freehand group (mean difference= −3.95 (95% CI −5.06 to −1.42))
Liu 2015	No significant difference between the two groups in respect of mean anteversion (−0.19; 95% CI −2.98 to 2.60, p=0.89)
Beckmann, 2009	No statistically significant difference in mean anteversion of cups placed with and without navigational support −1.7% and 95% CI −4.8 to 1.5
Cup positioning outside the safe zone	
Xu 2014	This difference was significant (RR=0.13; 95% CI 0.08 to 0.22, p<0.00001)
Liu 2015	RR with navigation was statistically significantly reduced (RR=0.31; CI 0.17 to 0.55, p<0.0001)
Beckmann, 2009	Statistically significantly reduced RR of cup positioning outside the safe zone with navigation: the pooled RR=0.2% and 95% CI 0.1% to 0.3%
Moskal 2011	Abduction safe zone: OR=1.53; 95% CI 1.01 to 2.33, p=0.0444 Anteversion safe zone: OR=1.96; 95% CI 1.33 to 2.88, p=0.0005 In both safe zones OR= 2.48; 95% CI 1.51 to 4.06, p=0.0003
Abduction (degrees)	
Moskal 2011	Mean=43.08; 95% CI 41.46 to 44.71; Fisher's p=0.5686
Anteversion (degrees)	
Moskal 2011	Mean=20.17; 95% CI 16.98 to 23.36, Fisher's p=0.9672
Leg length discrepancy	
Xu 2014	The pooled results show a significant difference between the groups (Mean difference =4.61 mm; 95% CI 7.74 to 1.48, p=0.004).
Chen, 2018	Not significantly different (weighted mean difference: −0.24; 95% CI −0.61 to 0.12, p=0.19)
Postoperative dislocation	
Moskal 2011	OR=0.41; 95% CI 0.17 to 0.95, p=0.0317
Xu 2014	The difference between the groups was not significant (RR=1.44; 95% C I0.04 to 56.79, p=0.85)

Continued

Table 4 Continued

Functional outcomes	
Patient safety and other outcomes	
Operative time	
Xu, 2014	Duration of navigated procedures was significantly longer (MD=19.87 min, 95% CI 14.04 to 24.35, p<0.00001) than that of conventional surgery
Chen, 2018	No significant difference between patients who underwent robotic-assisted and conventional THA: (weighted mean difference=23.21 min, 95% CI −3.76 to 50.09)
Deep venous thrombosis (DVT)	
Xu, 2014	There was no significant difference between the groups (RR=1.21; 95% CI 0.30 to 4.98, p=0.79)
Infections	
Moskal, 2011	OR=0.29; 95% CI 0.03 to 2.81, p=0.2569
Intraoperative complications	
Chen, 2018	The intraoperative complication rate was significantly higher in patients who underwent conventional compared with robotic-assisted THA
Xu 2014	Significantly higher in patients who underwent conventional compared with robotic-assisted THA (OR=0.12; 95% CI 0.05 to 0.34, p<0.0001)

MD, Mean difference; NAV, navigation; RR, relative risk; SR, systematic review; THA, total hip arthroplasty; WMD, Weighted Mean Difference.

confidence in the evidence synthesized by SRs, journals may consider requiring authors to abide by a methodological assessment tool (eg, AMSTAR 2) in addition to PRISMA guidelines. We suspect that many of these additional requirements will not be burdensome to authors. In our umbrella review, many of the unfulfilled requirements for AMSTAR 2 were administrative in nature (eg, presence of a protocol, availability of a list of excluded studies and reasons for exclusion) and can easily be addressed to increase the transparency and raise the confidence level in future SRs.

To our knowledge, this is the first umbrella review evaluating all published SRs comparing CA and conventional total joint arthroplasty. We followed the umbrella review guidelines strictly, and conducted the study selection, data extraction, and quality appraisal in duplicates. We then summarized the evidence in a structured way. We also assessed the overlap bias, an important step usually underreported by umbrella reviews.[24] Nevertheless, we must mention the limitations of our study. First, although we developed a protocol to help plan for our review, we did not register it, a step that would have provided more methodological strength for our review. Second, despite extensive efforts to identify all relevant SRs without language restrictions, it is still possible that we missed some SRs. Third, due to the absence of a reliable method of quantitatively synthesizing the evidence from multiple meta-analyses, we narratively summarized the evidence. Fourth, our extraction and assessment relied on the available manuscripts and supplemental materials. While we tried to contact the journals and authors inquiring for specific missing information, not all of them responded with clarification or additional information. Therefore, we cannot eliminate the possibility of underestimating the methodological quality for some studies because of the lack of access to relevant information.

Based on the findings of this review, we call for high quality SRs that can be used with great confidence to inform the decision on using CA TKA and THA. In addition, we encourage journals publishing SRs to use a methodological assessment tool to assess the quality of SRs. Finally, we advocate for standardization of the reported outcome measures for CA TKA and THA to facilitate evidence synthesis and outcome research.

CONCLUSIONS

Our umbrella review of 42 SRs found low methodological quality of the SRs undermining the confidence in the evidence synthesized by those reviews. Despite fairly high levels of overlap between the SRs in the primary studies examined, we found inconsistency in the results of the SRs tackling TKA and THA. Our findings suggest the need to improve the methodological quality of studies synthesizing evidence in this area to better inform clinical practice.

Contributors MMH and HMKG conceived and designed the study. MMH and MZ acquired the data. MZ and MMH performed the analyses. MMH, MZ and HMKG drafted the manuscript. All authors critically reviewed the manuscript and approved the final submitted version.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests MB reports personal fees and other from Medacta, other from Zimmer Biomet, outside the submitted work. All other authors have no competing interests to declare.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Mohamed Mosaad Hasan http://orcid.org/0000-0003-3128-3982

REFERENCES

1 Dalury DF, Pomeroy DL, Gorab RS, *et al*. Why are total knee arthroplasties being revised? *J Arthroplasty* 2013;28:120–1.
2 Barnett AJ, Toms AD. Revision total hip and knee replacement. *Clin Geriatr Med* 2012;28:431–46.
3 Stulberg SD, Loan P, Sarin V. Computer-Assisted navigation in total knee replacement: results of an initial experience in thirty-five patients. *J Bone Joint Surg Am* 2002;::90–8. 84-A Suppl 2 (published Online First: 2002/12/14).
4 Health Quality O. Computer-Assisted hip and knee arthroplasty. navigation and active robotic systems: an evidence-based analysis. *Ont Health Technol Assess Ser* 2004;4:1–39.
5 Goradia VK. Computer-Assisted and robotic surgery in orthopedics: where we are in 2014. *Sports Med Arthrosc Rev* 2014;22:202–5.
6 Davies BL. Robotic control in knee joint replacement surgery proceedings of the institution of mechanical engineers part H. *Journal of engineering in medicine*2007;221:71–80.
7 Bala A, Penrose CT, Seyler TM, *et al*. Computer-Navigated total knee arthroplasty utilization. *J Knee Surg* 2016;29:430–5.
8 Calliess T, Ettinger M, Windhagen H. Computer-assisted systems in total knee arthroplasty. Useful aid or only additional costs]. *Der Orthopade* 2014;43:529–33.
9 Quack VM, Kathrein S, Rath B, *et al*. Computer-Assisted navigation in total knee arthroplasty: a review of literature. *Biomed Tech* 2012;57:269–75.
10 Riddle DL, Jiranek WA, Hayes CW. Use of a validated algorithm to judge the appropriateness of total knee arthroplasty in the United States: a multicenter longitudinal cohort study. *Arthritis Rheumatol* 2014;66:2134–43.
11 Blom AW, Artz N, Beswick AD, *et al*. Improving patients' experience and outcome of total joint replacement: the RESTORE programme. *Programme Grants Appl Res* 2016;4:1–508.
12 Fu Y, Wang M, Liu Y, *et al*. Alignment outcomes in navigated total knee arthroplasty: a meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2012;20:1075–82.
13 Shi J, Wei Y, Wang S, *et al*. Computer navigation and total knee arthroplasty. *Orthopedics* 2014;37:e39–43.
14 Rebal BA, Babatunde OM, Lee JH, *et al*. Imageless computer navigation in total knee arthroplasty provides superior short term functional outcomes: a meta-analysis. *J Arthroplasty* 2014;29:938–44.
15 Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg* 2011;128:305–10.
16 Ioannidis JPA. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *Can Med Assoc J* 2009;181:488–93.
17 Aromataris E, Fernandez R, Godfrey CM, *et al*. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc* 2015;13:132–40.
18 Smith V, Devane D, Begley CM, *et al*. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol* 2011;11:15.
19 Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. *Evid Based Ment Health* 2018;21:95–100.
20 March LM, Bagga H. Epidemiology of osteoarthritis in Australia. *Med J Aust* 2004;180:S6–10.
21 March L, Cross M, Tribe K, *et al*. Cost of joint replacement surgery for osteoarthritis: the patients' perspective. *J Rheumatol* 2002;29:1006–14.
22 Dailiana ZH, Papakostidou I, Varitimidis S, *et al*. Patient-Reported quality of life after primary major joint arthroplasty: a prospective comparison of hip and knee arthroplasty. *BMC Musculoskelet Disord* 2015;16:366.
23 Shea BJ, Reeves BC, Wells G, *et al*. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
24 Pieper D, Antoine S-L, Mathes T, *et al*. Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol* 2014;67:368–75.
25 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
26 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
27 He Y, Li X, Gasevic D, *et al*. Statins and multiple Noncardiovascular outcomes: umbrella review of meta-analyses of observational studies and randomized controlled trials. *Ann Intern Med* 2018;169:543–53.
28 D'Auria E, Bergamini M, Staiano A, *et al*. Baby-led weaning: what a systematic review of the literature adds on. *Ital J Pediatr* 2018;44:49.
29 Farragher JF, Seaton SE, Stewart KE, *et al*. Not all systematic reviews are created equal. *Can J Occup Ther* 2018;85:180–2.
30 Pahwa M, Labrèche F, Demers PA. Night shift work and breast cancer risk: what do the meta-analyses tell us? *Scand J Work Environ Health* 2018;44:432–5.
31 Moore PA, Ziegler KM, Lipman RD, *et al*. Benefits and harms associated with analgesic medications used in the management of acute dental pain: an overview of systematic reviews. *J Am Dent Assoc* 2018;149:256–65.
32 Churuangsuk C, Kherouf M, Combet E, *et al*. Low-Carbohydrate diets for overweight and obesity: a systematic review of the systematic reviews. *Obesity Reviews* 2018;19:1700–18.
33 Wolfe D, Yazdi F, Kanji S, *et al*. Incidence, causes, and consequences of preventable adverse drug reactions occurring in inpatients: a systematic review of systematic reviews. *PLoS One* 2018;13:e0205426.
34 Lorenz RC, Matthias K, Pieper D, *et al*. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *J Clin Epidemiol* 2019;114:133–40.
35 Pieper D, Puljak L, González-Lorenzo M, *et al*. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. *J Clin Epidemiol* 2019;108:26–33.
36 Xu K, Li Y-min, Zhang H-feng, *et al*. Computer navigation in total hip arthroplasty: a meta-analysis of randomized controlled trials. *Int J Surg* 2014;12:528–33.
37 Zamora LA, Humphreys KJ, Watt AM, *et al*. Systematic review of computer-navigated total knee arthroplasty. *ANZ J Surg* 2013;83:22–30.
38 Computer-Assisted hip and knee arthroplasty. navigation and active robotic systems: an evidence-based analysis. *Ont Health Technol Assess Ser* 2004;4:1–39.
39 Alcelik IA, Blomfield MI, Diana G, *et al*. A comparison of short-term outcomes of minimally invasive computer-assisted vs minimally invasive conventional instrumentation for primary total knee arthroplasty. *J Arthroplasty* 2016;31:410–8.
40 Bathis H, Shafizadeh S, Paffrath T, *et al*. Are computer assisted total knee replacements more accurately placed? A meta-analysis of comparative studies]. *Orthopade* 2006;35:1056–65.
41 Bauwens K, Matthes G, Wich M, *et al*. Navigated total knee replacement. A meta-analysis. *J Bone Joint Surg Am* 2007;89:261–9.
42 Beckmann J, Stengel D, Tingart M, *et al*. Navigated cup implantation in hip arthroplasty. *Acta Orthop* 2009;80:538–44.
43 Brin YS, Nikolaou VS, Joseph L, *et al*. Imageless computer assisted versus conventional total knee replacement. A Bayesian meta-analysis of 23 comparative studies. *Int Orthop* 2011;35:331–9.
44 Burnett RSJ, Barrack RL. Computer-Assisted total knee arthroplasty is currently of NO proven clinical benefit: a systematic review. *Clin Orthop Relat Res* 2013;471:264–76.
45 Cheng T, Pan X-Y, Mao X, *et al*. Little clinical advantage of computer-assisted navigation over conventional instrumentation in primary total knee arthroplasty at early follow-up. *Knee* 2012;19:237–45.
46 Cheng T, Zhang G, Zhang X. Clinical and radiographic outcomes of image-based computer-assisted total knee arthroplasty: an evidence-based evaluation. *Surg Innov* 2011;18:15–20.
47 Cheng T, Zhang G, Zhang X. Imageless navigation system does not improve component rotational alignment in total knee arthroplasty. *J Surg Res* 2011;171:590–600.
48 Cheng T, Zhao S, Peng X, *et al*. Does computer-assisted surgery improve postoperative leg alignment and implant positioning following total knee arthroplasty? A meta-analysis of randomized controlled trials? *Knee Surg Sports Traumatol Arthrosc* 2012;20:1307–22.
49 Gandhi R, Marchie A, Farrokhyar F, *et al*. Computer navigation in total hip replacement: a meta-analysis. *Int Orthop* 2009;33:593–7.
50 Han S-B, Kim H-J, Kim T-K, *et al*. Computer navigation is effective in reducing blood loss but has no effect on transfusion requirement following primary total knee arthroplasty: a meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2016;24:3474–81.

51 Hetaimish BM, Khan MM, Simunovic N, *et al*. Meta-Analysis of navigation vs conventional total knee arthroplasty. *J Arthroplasty* 2012;27:1177–82.
52 Karthik K, Colegate-Stone T, Dasgupta P, *et al*. Robotic surgery in trauma and orthopaedics: a systematic review. *Bone Joint J* 2015;97-B:292–9.
53 YL L, Jia J, Wu Q, *et al*. Evidence-Based computer-navigated total hip arthroplasty: an updated analysis of randomized controlled trials. *Eur J Orthop Surg Traumatol* 2014;24:531–8.
54 Liu BG, Pang QJ. Meta-Analysis of therapeutic effects of computer-assisted navigation versus conventional total knee arthroplasty. *Chinese J Tissue Eng Res* 2014;18:6542–7.
55 Liu Z, Gao Y, Cai L. Imageless navigation versus traditional method in total hip arthroplasty: a meta-analysis. *Int J Surg* 2015;21:122–7.
56 Lüring C, Bäthis H, Tingart M, *et al*. Computer assistance in total knee replacement - a critical assessment of current health care technology. *Comput Aided Surg* 2006;11:77–80.
57 Mason JB, Fehring TK, Estok R, *et al*. Meta-Analysis of alignment outcomes in computer-assisted total knee arthroplasty surgery. *J Arthroplasty* 2007;22:1097–106.
58 Meijer MF, Reininga IHF, Boerboom AL, *et al*. Does imageless computer-assisted TKA lead to improved rotational alignment or fewer outliers? A systematic review. *Clin Orthop Relat Res* 2014;472:3124–33.
59 Moskal JT, Capps SG. Acetabular component positioning in total hip arthroplasty: an evidence-based analysis. *J Arthroplasty* 2011;26:1432–7.
60 Moskal J, Capps S, Mann J, *et al*. Navigated versus conventional total knee arthroplasty. *J Knee Surg* 2014;27:235–48.
61 Novicoff WM, Saleh KJ, Mihalko WM, *et al*. Primary total knee arthroplasty: a comparison of computer-assisted and manual techniques. *Instr Course Lect* 2010;59:109–17.
62 KJ O, Jang EJ, Kim YJ, *et al*. [Comparative effectiveness research of computer-assisted navigation in knee arthroplasty]. *HTA Database* 2014.
63 Reininga IHF, Zijlstra W, Wagenmakers R, *et al*. Minimally invasive and computer-navigated total hip arthroplasty: a qualitative and systematic review of the literature. *BMC Musculoskelet Disord* 2010;11:92.
64 Shin Y-S, Kim H-J, Ko Y-R, *et al*. Minimally invasive navigation-assisted versus conventional total knee arthroplasty: a meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2016;24:3425–32.
65 Smith TO, King JJ, Hing CB. A meta-analysis of randomised controlled trials comparing the clinical and radiological outcomes following minimally invasive to conventional exposure for total knee arthroplasty. *Knee* 2012;19:1–7.
66 Thienpont E, Fennema P, Price A. Can technology improve alignment during knee arthroplasty. *Knee* 2013;20:S21–8.
67 van der List JP, Chawla H, Joskowicz L, *et al*. Current state of computer navigation and robotics in unicompartmental and total knee arthroplasty: a systematic review with meta-analysis. *Knee Surg Sports Traumatol Arthrosc* 2016;24:3482–95.
68 Wang ZL, Zhao L, Zhao JG. Meta-Analysis of limb and prosthesis alignment restoration after navigated total knee arthroplasty versus conventional total knee arthroplasty. *Chinese J Tissue Eng Res* 2014;18:5707–14.
69 Xie C, Liu K, Xiao L, *et al*. Clinical outcomes after computer-assisted versus conventional total knee arthroplasty. *Orthopedics* 2012;35:e647–53.
70 Chen X, Xiong J, Wang P, *et al*. Robotic-Assisted compared with conventional total hip arthroplasty: systematic review and meta-analysis. *Postgrad Med J* 2018;94:335.1–41.
71 Karunaratne S, Duan M, Pappas E, *et al*. The effectiveness of robotic hip and knee arthroplasty on patient-reported outcomes: a systematic review and meta-analysis. *Int Orthop* 2019;43:1283–95.
72 Panjwani TR, Mullaji A, Doshi K, *et al*. Comparison of functional outcomes of computer-assisted vs conventional total knee arthroplasty: a systematic review and meta-analysis of high-quality, prospective studies. *J Arthroplasty* 2019;34:586–93.
73 Mannan A, Vun J, Lodge C, *et al*. Increased precision of coronal plane outcomes in robotic-assisted total knee arthroplasty: A systematic review and meta-analysis. *The Surgeon* 2018;16:237–44.
74 Snijders T, van Gaalen SM, de Gast A. Precision and accuracy of imageless navigation versus freehand implantation of total hip arthroplasty: a systematic review and meta-analysis. *Int J Med Robotics Comput Assist Surg* 2017;13:e1843.
75 Molina CS, Thakore RV, Blumer A, *et al*. Use of the National surgical quality improvement program in orthopaedic surgery. *Clin Orthop Relat Res* 2015;473:1574–81.
76 Kramer DB, Xu S, Kesselheim AS. How does medical device regulation perform in the United States and the European Union? A systematic review. *PLoS Med* 2012;9:e1001276.
77 Marshall DA, Pykerman K, Werle J, *et al*. Hip resurfacing versus total hip arthroplasty: a systematic review comparing standardized outcomes. *Clin Orthop Relat Res* 2014;472:2217–30.
78 Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Rheum* 2008;59:876–83.
79 Riddle DL, Stratford PW, Singh JA, *et al*. Variation in outcome measures in hip and knee arthroplasty clinical trials: a proposed approach to achieving consensus. *J Rheumatol* 2009;36:2050–6.
80 Wall PDH, Richards BL, Sprowson A, *et al*. Do outcomes reported in randomised controlled trials of joint replacement surgery fulfil the OMERACT 2.0 filter? A review of the 2008 and 2013 literature. *Syst Rev* 2017;6:106.
81 Singh JA, Dohm M, Sprowson AP, *et al*. Outcome domains and measures in total joint replacement clinical trials: can we harmonize them? an OMERACT collaborative initiative. *J Rheumatol* 2015;42:2496–502.
82 Singh JA, Dowsey MM, Dohm M, *et al*. Achieving consensus on total joint replacement trial outcome reporting using the OMERACT filter: endorsement of the final core domain set for total hip and total knee replacement trials for endstage arthritis. *J Rheumatol* 2017;44:1723–6.
83 Moher D, Tetzlaff J, Tricco AC, *et al*. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
84 McCulloch P, Altman DG, Campbell WB, *et al*. No surgical innovation without evaluation: the ideal recommendations. *The Lancet* 2009;374:1105–12.
85 Hirst A, Philippou Y, Blazeby J, *et al*. No surgical innovation without evaluation: evolution and further development of the ideal framework and recommendations. *Ann Surg* 2019;269:211–20.
86 Allin B, Aveyard N, Campion-Smith T, *et al*. What evidence underlies clinical practice in paediatric surgery? A systematic review assessing choice of study design. *PLoS One* 2016;11:e0150864.
87 Moher D, Cook DJ, Eastwood S, *et al*. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. quality of reporting of meta-analyses. *Lancet* 1999;354:1896–900.