Cochrane Library

Cochrane Database of Systematic Reviews

Shoulder replacement surgery for osteoarthritis and rotator cuff tear arthropathy (Review)

Craig RS, Goodier H, Singh JA, Hopewell S, Rees JL

Craig RS, Goodier H, Singh JA, Hopewell S, Rees JL.
Shoulder replacement surgery for osteoarthritis and rotator cuff tear arthropathy.
Cochrane Database of Systematic Reviews 2020, Issue 4. Art. No.: CD012879.
DOI: 10.1002/14651858.CD012879.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Shoulder replacement surgery for osteoarthritis and rotator cuff tear arthropathy

Richard S Craig[1], Henry Goodier[1], Jasvinder A Singh[2], Sally Hopewell[1], Jonathan L Rees[1]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK. [2]Department of Medicine, Birmingham VA Medical Center, Birmingham, AL, USA

Contact: Jonathan L Rees, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Botnar Research Centre, Old Road, Oxford, OX3 7LD, UK. jonathan.rees@ndorms.ox.ac.uk.

Editorial group: Cochrane Musculoskeletal Group.
Publication status and date: Edited (no change to conclusions), published in Issue 4, 2020.

Citation: Craig RS, Goodier H, Singh JA, Hopewell S, Rees JL. Shoulder replacement surgery for osteoarthritis and rotator cuff tear arthropathy. *Cochrane Database of Systematic Reviews* 2020, Issue 4. Art. No.: CD012879. DOI: 10.1002/14651858.CD012879.pub2.

ABSTRACT

Background

Shoulder replacement surgery is an established treatment for patients with end-stage glenohumeral osteoarthritis or rotator cuff tear arthropathy who have not improved with non-operative treatment. Different types of shoulder replacement are commonly used, but their relative benefits and risks compared versus one another and versus other treatments are uncertain. This expanded scope review is an update of a Cochrane Review first published in 2010.

Objectives

To determine the benefits and harms of shoulder replacement surgery in adults with osteoarthritis (OA) of the shoulder, including rotator cuff tear arthropathy (RCTA).

Search methods

We searched the Cochrane Central Register of Controlled Trials, MEDLINE, Embase, CINAHL, SportDiscus, and Web of Science up to January 2019. We also searched clinical trial registers, conference proceedings, and reference lists from previous systematic reviews and included studies.

Selection criteria

We included randomised studies comparing any type of shoulder replacement surgery versus any other surgical or non-surgical treatment, no treatment, or placebo. We also included randomised studies comparing any type of shoulder replacement or technique versus another. Study participants were adults with osteoarthritis of the glenohumeral joint or rotator cuff tear arthropathy.

We assessed the following major outcomes: pain, function, participant-rated global assessment of treatment success, quality of life, adverse events, serious adverse events, and risk of revision or re-operation or treatment failure.

Data collection and analysis

Two review authors independently assessed trial quality and extracted data. We collected trial data on benefits and harms.

Main results

We included 20 studies involving 1083 participants (1105 shoulders). We found five studies comparing one type of shoulder replacement surgery to another type of shoulder replacement surgery, including three studies comparing conventional stemmed total shoulder replacement (TSR) surgery to stemmed humeral hemiarthroplasty. The remaining 15 studies compared one type of shoulder replacement

to the same type of replacement performed with a technical modification or a different prosthetic component. We found no studies comparing shoulder replacement surgery to any other type of surgical treatment or to any type of non-surgical treatment. We found no studies comparing reverse total shoulder replacement surgery to any other type of treatment or to any type of replacement.

Trial size varied from 16 to 161 participants. Participant mean age ranged from 63 to 81 years. 47% of participants were male. Sixteen trials reported participants with a diagnosis of osteoarthritis and intact rotator cuff tendons. Four trials reported patients with osteoarthritis and a rotator cuff tear or rotator cuff tear arthropathy.

All studies were at unclear or high risk of bias for at least two domains, and only one study was free from high risk of bias (included in the main comparison). The most common sources of bias were lack of blinding of participants and assessors, attrition, and major baseline imbalance.

Three studies allowed a comparison of conventional stemmed TSR surgery versus stemmed humeral hemiarthroplasty in people with osteoarthritis. At two years, low-quality evidence from two trials (downgraded for bias and imprecision) suggested there may be a small but clinically uncertain improvement in pain and function. On a scale of 0 to 10 (0 is no pain), mean pain was 2.78 points after stemmed humeral hemiarthroplasty and 1.49 points lower (0.1 lower to 2.88 lower) after conventional stemmed TSR. On a scale of 0 to 100 (100 = normal function), the mean function score was 72.8 points after stemmed humeral hemiarthroplasty and 10.57 points higher (2.11 higher to 19.02 higher) after conventional stemmed TSR. There may be no difference in quality of life based on low-quality evidence, downgraded for risk of bias and imprecision. On a scale of 0 to 100 (100 = normal), mean mental quality of life was rated as 57.4 points after stemmed humeral hemiarthroplasty and 1.0 point higher (5.1 lower to 7.1 higher) after conventional stemmed TSR.

We are uncertain whether there is any difference in the rate of adverse events or the rate of revision, re-operation, or treatment failure based on very low-quality evidence (downgraded three levels for risk of bias and serious imprecision). The rate of any adverse event following stemmed humeral hemiarthroplasty was 286 per 1000, and following conventional stemmed TSR 143 per 1000, for an absolute difference of 14% fewer events (25% fewer to 21% more). Adverse events included fractures, dislocations, infections, and rotator cuff failure. The rate of revision, re-operation, or treatment failure was 103 per 1000, and following conventional stemmed TSR 77 per 1000, for an absolute difference of 2.6% fewer events (8% fewer to 15% more).

Participant-rated global assessment of treatment success was not reported.

Authors' conclusions

Although it is an established procedure, no high-quality randomised trials have been conducted to determine whether shoulder replacement might be more effective than other treatments for osteoarthritis or rotator cuff tear arthropathy of the shoulder. We remain uncertain about which type or technique of shoulder replacement surgery is most effective in different situations. When humeral hemiarthroplasty was compared to TSR surgery for osteoarthritis, low-quality evidence led to uncertainty about whether there is a clinically important benefit for patient-reported pain or function and suggested there may be little or no difference in quality of life. Evidence is insufficient to show whether TSR is associated with greater or less risk of harm than humeral hemiarthroplasty. Available randomised studies did not provide sufficient data to reliably inform conclusions about adverse events and harm. Although reverse TSR is now the most commonly performed type of shoulder replacement, we found no studies comparing reverse TSR to any other type of treatment.

PLAIN LANGUAGE SUMMARY

Shoulder replacement surgery for osteoarthritis and arthritis associated with torn rotator cuff tendons

Background

Osteoarthritis is a condition of the joints. Over time, the cartilage becomes thinner and exposed bone surfaces rub against each other, causing pain and loss of movement. People with torn shoulder tendons can develop a specific type of arthritis, called rotator cuff tear arthropathy. People usually need pain relief medicines and may be offered non-surgical treatments initially, including physiotherapy and injections. Some people with ongoing symptoms from advanced arthritis are offered shoulder replacement surgery. In 'humeral hemiarthroplasty', just the head (ball part) of the humerus is replaced with an artificial one and continues to articulate in the socket. In 'total shoulder replacement', the socket is also replaced with an artificial one. In 'reverse total shoulder replacement', the replacement is intentionally done back-to-front with an artificial ball fixed to the old socket and an artificial socket placed on top of the humerus. The type of replacement performed usually depends on the pattern of joint and tendon damage.

It is not clear when or whether shoulder replacement is the best treatment for people with osteoarthritis or rotator cuff tear arthropathy, or which type of replacement is best for different people. We searched for the best evidence from studies called randomised trials to try to answer these questions.

Study characteristics

This review is current to 31 January 2019 and includes only studies in which treatment was allocated randomly by type. All study participants had osteoarthritis or rotator cuff tear arthropathy of the shoulder and had tried non-surgical treatments already. The average

age of study participants was between 63 and 81 years old. Slightly more than half of the participants were female. We found no studies comparing shoulder replacement surgery to any other type of treatment, including other types of non-replacement surgery, physiotherapy, or no treatment at all. We found five studies comparing one type of shoulder replacement to another type of shoulder replacement. We found 15 studies comparing one type of shoulder replacement technique to the same type, performed with a technical modification or a different prosthetic component. Eight out of 20 studies were funded by a shoulder replacement manufacturer. A further seven out of 20 studies were conducted by researchers who had other financial relationships with shoulder replacement manufacturers.

Key results

Three trials (126 participants) met our inclusion criteria for our main comparison of conventional stemmed total shoulder replacement (TSR) versus stemmed humeral hemiarthroplasty (HA) for treatment of osteoarthritis. TSR may result in less pain and better function compared to HA at two-year follow-up, but this may not be noticeable. We are very uncertain whether there are any differences in the frequency of adverse events and further operations.

TSR resulted in 15% less pain (1% less to 29% less).

• People who had HA rated their pain as 2.8 points (0 to 10 scale).

• People who had TSR rated their pain as 1.29 points.

TSR resulted in 11% better function (2% better to 19% better).

• People who had HA rated their function as 72.8 points (0 to 100 scale).

• People who had TSR rated their function as 83.4 points.

TSR resulted in similar quality of life to HA (5% lower to 7% higher, 5 points lower to 7 points higher (0 to 100 scale)).

• People who had HA rated their quality of life as 57.4 points.

• People who had TSR rated their quality of life as 58.4 points.

TSR resulted in a similar number of adverse events (25% fewer to 21% more) and a similar number of further operations on the same shoulder (8% fewer to 15% more) compared to HA.

• Following HA, 286 per 1000 people experienced an adverse event and 103 per 1000 required further operations.

• Following TSR, 143 per 1000 people experienced an adverse event and 77 per 1000 required further operations.

Quality of the evidence

For the main comparison, the quality of evidence for assessing pain, function, and quality of life was low. For assessment of adverse events and further operations, the quality of evidence was very low. Across the other 12 comparisons, the quality of evidence was also very low.

SUMMARY OF FINDINGS

Summary of findings for the main comparison. One type of shoulder replacement (TSR) to another type of shoulder replacement (hemiarthroplasty)

Conventional stemmed total shoulder replacement compared to stemmed humeral hemiarthroplasty for primary glenohumeral osteoarthritis

Patient or population: adults aged ≥ 18 years with a diagnosis of glenohumeral osteoarthritis who have not responded to non-operative treatments
Setting: secondary care
Intervention: conventional stemmed total shoulder replacement (TSR)
Comparison: stemmed humeral hemiarthroplasty

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with stemmed humeral hemiarthroplasty	Risk with conventional stemmed TSR				
Pain assessed with visual analogue scale (VAS) Scale from 0 to 10, lower = better, MCID 1.5 points Follow-up: range 1 year to 3 years	Mean pain was 2.78 points	MD 1.49 cm lower (0.1 lower to 2.88 lower)	-	92 (2 RCTs)	⊕⊕⊖⊖ LOW[a,b]	Conventional stemmed TSR may reduce pain slightly compared with stemmed hemiarthroplasty and includes both clinically important and unimportant effects[c] Absolute difference 15% lower (1% lower to 29% lower); relative difference 23% lower (2% lower to 44% lower)[d]
Function assessed with WOOS Index Scale from 0 to 100 points, higher = better, MCID 10 points Follow-up: range 1 year to 3 years	Mean function was 72.8 points	MD 10.57 points higher (2.11 higher to 19.02 higher)	-	92 (2 RCTs)	⊕⊕⊖⊖ LOW[a,b]	Conventional stemmed TSR may result in improved function compared with stemmed hemiarthroplasty and includes both clinically important and unimportant effects Absolute difference 11% higher (2% higher to 19% higher); relative difference 32% higher (6% higher to 57% higher).[d] Number needed to achieve 1 additional beneficial outcome (NNTB) = 6 (95% CI 4 to 30)
Participant-rated global assessment of treatment success	See comment					None of the studies measured or reported this outcome

Quality of life assessed with Short Form-12[e] Scale from 0 to 100 points, higher = better, MCID 4 points Follow-up: mean 2 years	Mean quality of life was 57.4 points	MD 1 higher (5.14 lower to 7.14 higher)	-	41 (1 RCT)	⊕⊕⊖⊖ LOW[a,b]	Conventional stemmed TSR probably results in little to no difference in quality of life over stemmed hemiarthroplasty but we are uncertain.[c] Absolute difference 1% higher (5% lower to 7% higher), relative difference 2% higher (9% lower to 13% higher)
Adverse events (total): assessed with number of events within 3 years[f]	286 per 1000	143 per 1000 (40 to 497)	RR 0.50 (0.14 to 1.74)	42 (1 RCT)	⊕⊖⊖⊖ VERY LOW[a,g]	We are uncertain whether there is any difference in the rate of specific adverse events Absolute difference of 14% fewer events with TSR (25% fewer to 21% more); relative difference 50% fewer (86% fewer to 74% more).[c] Includes 1 fatal pulmonary embolus in the TSR group
Adverse events (serious - resulting in hospitalisation or death) Assessed with number of events within 1 year	Only 1 serious adverse event was reported in either arm. Included studies are grossly underpowered for identification of infrequent events			42 (1 RCT)	⊕⊖⊖⊖ VERY LOW[a,g]	We are uncertain whether there is any difference in the rate of serious adverse events[g]
Revision, re-operation, or treatment failure assessed with number of events within 3 years	103 per 1000	77 per 1000 (23 to 254)	RR 0.74 (0.22 to 2.46)	125 (3 RCTs)	⊕⊖⊖⊖ VERY LOW[a,g]	We are uncertain whether there is any difference in the rate of revision, re-operation, or treatment failure[c] Absolute difference of 2.6% fewer events with TSR (8% fewer to 15% more); relative difference 26% fewer (78% fewer to 146% more)

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: confidence interval; MCID: minimum clinically important difference; MD: mean difference; NNTB: number needed to treat for an additional beneficial outcome; NNTH: number needed to treat for an additional harmful outcome; RCT: randomised controlled trial; RR: risk ratio; WOOS: Western Ontario Osteoarthritis of the Shoulder Index.

GRADE Working Group grades of evidence.
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

---

[a]Downgraded one level for risk of bias (at least one trial at high or unclear risk of bias).
[b]Downgraded one level for imprecision: wide confidence intervals due to small sample size from few studies. Confidence intervals include both an important and an unimportant effect.
[c]Downgraded two levels for serious imprecision: very wide confidence intervals. Sample size from few studies grossly underpowered for analysis of infrequent events.
[d]Total adverse events; includes both serious adverse events and local/specific adverse events not requiring further surgery (i.e. further operations are counted in the revision/re-operation section only). Specific adverse events included infections, dislocations, fractures, and rotator cuff failures.
[e]Mental component score.
[f]Numbers needed to achieve on additional beneficial or harmful outcome (NNTB/NNTH) were not calculated in the absence of a clinically important difference.
[g]Relative changes calculated relative to baseline in control group (i.e. absolute change (mean difference) divided by mean at baseline in the placebo group from Lo 2005 (values were 6.52 points on 0 to 10-point VAS Pain Scale; 33.5 points on 0 to 100-point WOOS Score; 55.5 points on 100-point SF-36 mental component score; and 29.5 points on 100-point SF-36 physical component score). Absolute change calculated as mean difference divided by scale of the instrument, expressed as percentage.

BACKGROUND

Description of the condition

Shoulder osteoarthritis (OA) typically results in narrowing of the glenohumeral (shoulder) joint space due to degeneration of the articular cartilage and subchondral bone, and thickening of the joint capsule. The rotator cuff is an important group of four muscles and associated tendons around the shoulder that are vital for shoulder stability, shoulder rotation, initiation of movement, and fine control. People with advanced damage to the rotator cuff tendons around the shoulder commonly develop a specific pattern of arthritis, termed rotator cuff tear arthropathy (RCTA) (Neer 1983; Walch 2005). Shoulder OA and RCTA present primarily with shoulder pain, stiffness, limitation of shoulder function, and disability. These symptoms are common, affecting 5% to 21% of adults in the USA and in Western countries (Bergenudd 1988; Chakravarty 1990; Chard 1991; Breivik 2006; National Center for Health Statistics 2011). Shoulder OA is the underlying cause of shoulder pain in 2% to 5% of this group (Meislin 2005), although few truly population-based studies have been done. Shoulder pain is associated with shoulder-related disability in more than half of the people reporting this pain (Chard 1991; Croft 1996; Pope 1997), and it leads to increased use of healthcare resources (Chalmers 2019; Wofford 1997). Thus, shoulder OA leads to significant morbidity, especially in the ageing population.

Description of the intervention

Current non-surgical treatment options for chronic shoulder pain associated with shoulder OA include oral analgesics, non-steroidal anti-inflammatory drugs (NSAIDs), intra-articular injections (corticosteroids and hyaluronic acid), physical therapy, and acupuncture (Green 2005). NSAIDs can help to alleviate pain but may cause systemic side effects, including renal insufficiency and gastrointestinal problems, especially among the elderly (ACR 2000; Shamoon 2000). Intra-articular corticosteroid injections, electrotherapies (including transcutaneous electrical nerve stimulation), exercise, and physiotherapy may provide benefit as they do for other shoulder conditions (Buchbinder 2003; Page 2016a; Page 2016b; van der Windt 2003), but their benefit in shoulder OA has not been proven. Nor has benefit been proven for intra-articular hyaluronic acid injections for glenohumeral OA over placebo (Blaine 2008; Zhang 2019). If non-operative treatments fail, and there is disabling pain and loss of function, surgery is usually undertaken.

Joint replacement surgery is now the main surgical treatment for shoulder OA. It involves replacement of either the humeral head (hemiarthroplasty) or the humeral head and the glenoid (total shoulder replacement (TSR)) with implants, or replacement of the humeral head and the glenoid with components in a reversed configuration, that is, through insertion of a metal ball where the native socket was and a plastic cup on a metal stem where the native head was (reverse total shoulder replacement (RTSR) (Grammont 1993)). These procedures are now performed more often, among younger people and for those with earlier degrees of OA.

Shoulder joint replacement treatment options for shoulder OA and RCTA are the focus of this review and include all types of humeral hemiarthroplasty, conventional TSR, and RTSR.

How the intervention might work

Joint replacement surgery involves the removal of damaged bone and cartilage, with release of soft tissues that are causing contractures, when necessary. These damaged tissues and the inflammation associated with them contribute to the painful symptoms of arthritis. The bone and cartilage that have been removed are replaced with new, smooth, prosthetic (man-made) materials that try to re-create the anatomy and function of the shoulder joint. The new joint is designed to glide smoothly and restore the centre of rotation of the shoulder joint. The result should be a joint with improved mechanical properties, allowing the muscles to work more easily to move the arm.

The specific reversed geometry design of the reverse TSR is intended to provide the maximum mechanical advantage for the deltoid muscle to move the shoulder and arm in people who do not have intact or functioning rotator cuff muscles (Walker 2011).

Why it is important to do this review

Shoulder pain due to OA is a disabling and common condition. Surgical treatment of shoulder OA with joint replacement has been reported to be associated with significant improvement in pain, function, and quality of life (Fehringer 2002). The previous version of this review found limited, low-quality evidence to inform decision-making for patients with shoulder arthritis (Singh 2010). There has been a rapid expansion in both the number of shoulder replacements available and the number of procedures performed annually for shoulder OA (Dillon 2017; Kim 2011; Lübbeke 2017). Therefore, an up-to-date synthesis of available evidence is needed to assess the effectiveness and safety of different shoulder replacement methods when compared to each other, to placebo, or to other conservative options. A priority setting partnership funded by the National Institute of Health Research and a national surgical society has identified the optimal type of shoulder replacement for OA as an ongoing research uncertainty (JLA 2015), and this topic appears within its top 10 research priorities.

OBJECTIVES

To determine the benefits and harms of shoulder replacement surgery in adults with osteoarthritis (OA) of the shoulder, including rotator cuff tear arthropathy (RCTA).

METHODS

Criteria for considering studies for this review

Types of studies

We considered all randomised controlled trials (RCTs) for inclusion. We excluded non-randomised and quasi-randomised studies to minimise the risk of patient selection bias. We applied no language restriction on included studies, but no articles required translation.

Types of participants

We included studies of adults (aged 18 years and over) with arthritis of the shoulder joint, confirmed by radiographic examination. We included participants with primary osteoarthritis (OA) and OA secondary to rotator cuff tear arthropathy (RCTA).

We excluded studies of adults undergoing surgery for inflammatory arthritis such as rheumatoid arthritis, benign or malignant tumour, adhesive capsulitis, shoulder instability, or fracture.

Types of interventions

We included studies that compared any type of shoulder replacement surgery to any other treatment modality. We specifically included studies that compared shoulder replacement surgery to placebo (i.e. sham surgery), other surgical modalities (e.g. arthroscopic debridement), non-surgical modalities (e.g. intra-articular corticosteroid injections, physiotherapy, acupuncture), or no treatment. In addition, we included studies that compared one type of shoulder replacement to another type of shoulder replacement (e.g. TSR versus RTSR), or one type of shoulder replacement surgical technique to another (e.g. cemented TSR versus uncemented TSR).

Types of outcome measures

Based on the preliminary core domain set described by the OMERACT (Outcome Measures in Rheumatology) Special Interest Group (Buchbinder 2017), we measured the following major outcomes.

Major outcomes

- Pain measured via a visual analogue scale (VAS), a numerical rating scale (NRS), semi-quantitative descriptive scales (e.g. short-form McGill scale (Melzack 1987)), or another instrument
- Function measured with shoulder-specific instruments and analysed according to the following hierarchy:
  - Western Ontario Osteoarthritis of the Shoulder Index (WOOS)
  - American Shoulder and Elbow Surgeons Scale (ASES)
  - Oxford Shoulder Score (OSS)
  - Constant Murley Score
  - Shoulder Pain and Disability Index (SPADI)
  - Disability of the Arm, Shoulder, and Hand (DASH) questionnaire
- Participant-rated global assessment of treatment success
- Quality of life (mental) measured by a generic instrument such as Short-Form 36 (SF-36) and other similar instruments
- Adverse events (total)
- Adverse events (serious): assessed as either serious (death, or requiring hospitalisation) or specific (including shoulder stiffness, instability, infection, and nerve damage)
- Revision or other re-operation, including treatment failure

Several functional outcome scores (e.g. Constant Murley Score) are often reported as subdomains including pain. We report a pain subscale only when there was no other more appropriate continuous pain scale. There is no clear consensus on the most appropriate physical function score to be used to compare shoulder surgery treatments for arthritis. The hierarchy of function scores was chosen to reflect widely used and validated measures in studies of shoulder replacement surgery, prioritising those with no physician-measured component and those with valid international translations. Although the Constant Murley Score is widely used (Page 2015), it is heavily weighted by physician-measured components and was hence downgraded to reduce the potential for risk of bias in this domain.

Death is a rare event in shoulder surgery and therefore was measured within the domain of serious adverse events. Measured range of motion and strength are considered to be of low utility (Buchbinder 2017), and we did not analyse these outcomes separately. We did however assess these outcomes using some functional tools (major outcome 2). For major outcome 6, we defined treatment failure as a procedure for which revision or re-operation was deemed necessary by a clinician but was not performed because the patient declined further surgery or was unfit.

Minor outcomes

Physician-evaluated outcomes, including radiographic assessment of lucency.

Timing of outcome assessment

We collected outcome data for the following time points: short-term (less than one year), intermediate-term (one to three years), and long-term (more than three years). We considered the intermediate time point to be the primary time point for comparisons.

Search methods for identification of studies

Electronic searches

We searched the following databases, from inception, with no date or language restrictions.

- Cochrane Central Register of Controlled Trials (CENTRAL), via the Cochrane Library, Wiley InterScience (www.thecochranelibrary.com).
- MEDLINE (1966 to 31 January 2019).
- Embase (1988 to 31 January 2019).
- Cumulative Index to Nursing and Allied Health Literature (CINAHL) (1937 to 31 January 2019).
- SportDiscus (1985 to 31 January 2019).
- Web of Science (1945 to 31 January 2019).

We conducted searches of ClinicalTrials.gov (www.ClinicalTrials.gov), along with the World Health Organization (WHO) International Clinical Trials Registry Platform (www.who.int/ictrp/en/) (31 January 2019).

See Appendix 1, Appendix 2, Appendix 3, Appendix 4, Appendix 5, Appendix 6, Appendix 7, and Appendix 8 for detailed search strategies.

Searching other resources

We checked that reference lists of all primary studies and review articles for additional references. In addition, we searched for published congress abstracts from, but not limited to, the American Academy of Orthopedic Surgeons (AAOS), the British Orthopaedic Association (BOA), the American Society of Shoulder and Elbow Surgeons, the British Elbow and Shoulder Society (BESS), the European Society of Shoulder and Elbow Surgery (SECEC), and the European Federation of National Associations of Orthopaedics and Traumatology (EFORT), using available archives on relevant society websites up to 31 January 2019. We searched relevant manufacturers' websites for trial information and contacted individuals or organisations when appropriate. We searched for errata or retractions from included studies.

Data collection and analysis

Selection of studies

Independently, two review authors (RC, HG) reviewed the titles and abstracts of studies identified by the searches according to the Criteria for considering studies for this review, and we discarded those that clearly were not relevant. We then retrieved the full text of those remaining potentially eligible studies. Independently, the same review authors (RC, HG) repeated the selection process by screening the full-text versions of these studies to determine which studies should be included and from which data should be extracted. We resolved disagreements by consensus. When consensus was not achieved initially, a third review author (SH or JR) acted as an adjudicator.

We identified and excluded duplicates and collated multiple reports of the same study, so that each study, rather than each report, is the unit of interest in the review. We recorded the selection process in sufficient detail to complete a PRISMA flow diagram (PRISMA Group 2009), as well as the Characteristics of excluded studies tables.

Data extraction and management

We used the online review manager Covidence to create a data collection form for study characteristics and outcome data, which we piloted on one study in the review (Covidence). Independently, two review authors (RC, HG) extracted study characteristics from included studies. Review authors reached consensus for final data extraction by discussion and extracted the following study characteristics.

- Methods: study design, total duration of study, details of any 'run-in' period, number of study centres and locations, study setting, description of eligibility criteria for centres and surgeons, withdrawals, and date of study.
- Participants: N, mean age, age range, sex, sociodemographics, ethnicity, disease duration, severity of condition, diagnostic criteria, important condition-specific baseline data; inclusion criteria, and exclusion criteria.
- Interventions: total number of intervention groups within each trial, specific details of each intervention and comparator (e.g. details of the surgery including number of surgeons in the trial, surgeon experience and duration of operation, descriptions of the procedure for tailoring interventions to individual participants), any co-interventions, and details of rehabilitation following surgery.
- Outcomes: relevant primary and secondary outcomes specified and collected during the trials and time points reported.
- Characteristics of the design of the trial as outlined below in the Assessment of risk of bias in included studies section.
- Notes: funding for trial and notable declarations of interest from trial authors.

Independently, two review authors (RC, HG) extracted outcome data from included studies using Covidence. We extracted the number of events and the number of participants per treatment group for dichotomous outcomes and means and standard deviations, and the number of participants per treatment group for continuous outcomes. We noted in the Characteristics of included studies table whether outcome data were not reported in a usable way, and when data were transformed or estimated from a graph.

We resolved disagreements by reaching consensus or by involving a third person (SH). One review author (RC) transferred data into Review Manager 5 (RevMan 5) (RevMan 2014). We double-checked that data were entered correctly by comparing data presented in the systematic review versus the study reports.

For numerical data presented only in figures or graphs, we contacted the authors of the original report and requested data. When this was not possible, we used software for extraction from graphs (e.g. PlotDigitizer) to extract data from the graphs or figures. We extracted these data in duplicate.

When both final and change from baseline values were reported for a given outcome, we extracted the final value; if both unadjusted and adjusted values for the same outcome were reported, we extracted the unadjusted value. If more than one outcome measure was reported in a trial, we prioritised outcomes based on the hierarchy of major outcomes listed above. When possible, we extracted data based on intention-to-treat analysis.

Main planned comparisons

Our main planned comparisons were as follows.

- Any type of shoulder replacement surgery versus placebo (sham-surgery).
- Any type of shoulder replacement versus any type of non-surgical treatment.
- Any type of shoulder replacement surgery versus any other type of surgery.
- Any one type of shoulder replacement surgery versus any other type of shoulder replacement surgery.
- Any one type of shoulder replacement surgical technique versus any other type of shoulder replacement surgical technique (e.g. cemented versus uncemented implants).

We planned to pool studies of different shoulder replacement types as a single analysis versus a common comparator when techniques were sufficiently similar.

Assessment of risk of bias in included studies

Independently, two review authors (RC, HG) assessed risk of bias for each study using the criteria outlined in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019a). We resolved disagreements by discussion or by consultation with another review author (SH). We assessed risk of bias according to the following domains.

- Random sequence generation.
- Allocation concealment.
- Blinding of participants and personnel.
- Blinding of outcome assessment - self-reported outcomes.
- Blinding of outcome assessment - assessor-reported outcomes.
- Incomplete outcome data.
- Selective outcome reporting.
- Major baseline imbalance.
- Differences in rehabilitation regimen.

We graded each potential source of bias as high, low, or unclear, and we provided a quote from the study report together with a justification for our judgement in the 'Risk of bias' table. We

summarised the 'Risk of bias' judgements across different studies for each of the domains listed. In addition, we considered the impact of missing data by key outcomes.

When information about risk of bias related to unpublished data or correspondence with a triallist, we have noted this in the 'Risk of bias' table. When considering treatment effects, we took into account the risk of bias for studies that contributed to that outcome. We have presented the figures generated by the 'Risk of bias' tool to provide summary assessments of the risk of bias.

Assessment of bias in conducting the systematic review

We have conducted the review according to the published protocol and have reported any deviations from it in the Differences between protocol and review section of the systematic review.

Measures of treatment effect

We analysed dichotomous data as risk ratios or Peto odds ratios when the outcome is a rare event (approximately < 10%) and used 95% confidence intervals (CIs).

We analysed continuous data as mean difference (MD) or standardised mean difference (SMD), with 95% CIs, depending on whether the outcome was measured using the same scale or different scales. We entered data presented as a scale with a consistent direction of effect across studies. When different scales were used to measure the same conceptual outcome, we back-translated SMD to a typical scale (e.g. 0 to 10 for pain), as described in Chapter 12 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2019a). We used a minimal clinically important difference (MCID) of 1.5 points on a 10-point scale for pain on the VAS (Hao 2019). For function, the MCID of the WOOS Index is reported to be 10 points on a 100-point scale (Polk 2013), and for the ASES, it is 13.5 points on a 100-point scale from an anchor-based study (Werner 2016). For the SF-12, the MCID in patients with shoulder pain is 1 point on the physical scale and 4 points on the mental scale (Hao 2019).

For dichotomous outcomes, such as serious adverse events, we calculated the number needed to treat for an additional beneficial outcome (NNTB) and the number needed to treat for an additional harmful outcome (NNTH) from the control group event rate and the risk ratio using the Visual Rx NNT calculator (Cates 2008). We planned to calculate the NNTB for continuous measures using the Wells calculator (available at the Cochrane Musculoskeletal Group (CMSG) Editorial office; musculoskeletal.cochrane.org).

For dichotomous outcomes, we calculated the absolute percent change from the difference in risks between intervention and control groups using GRADEpro and expressed it as a percentage (GRADEpro GDT 2015). The relative percent change was calculated as Risk ratio - 1 and was expressed as a percentage.

For continuous outcomes, we calculated the absolute percent change by dividing the mean difference by the scale of the measure and expressed it as a percentage. The relative difference was calculated as the absolute benefit (mean difference) divided by the baseline mean of the control group, expressed as a percentage.

In the 'Comments' column of Summary of findings for the main comparison, we have provided the absolute per cent difference and the relative per cent change from baseline, along with the NNTB or the NNTH (the NNTB or the NNTH is provided only when the outcome shows a clinically significant difference). When a clinically important difference was present, the MCID values are provided in the Effects of interventions section.

Unit of analysis issues

None of the included trials reported more than two study arms.

When multiple time points were reported, we grouped them into short-term (less than one year), intermediate-term (one to three years), and long-term (more than three years) follow-up. If a single trial reported multiple time points within one of these groups, we extracted the data that related to the latest time point.

The unit of analysis was each shoulder.

Dealing with missing data

We contacted investigators or study sponsors to verify key study characteristics and to obtain missing numerical outcome data (e.g. when a study was identified as an abstract only, when data were not available for all participants). When this was not possible, and the missing data were thought to introduce serious bias, we considered exploring the impact of including such studies in the overall assessment of results by performing a sensitivity analysis; however these studies did not contribute to outcomes suitable for meta-analysis.

For dichotomous outcomes (e.g. number of revision operations), we calculated the event rate using the number of participants randomised in the group as the denominator, unless the number at risk was otherwise clearly stated.

For continuous outcomes (e.g. mean change in pain score), we calculated the MD or the SMD based on the number of participants analysed at that time point. If the number of participants analysed was not presented for each time point, we used the number of randomised participants in each group at baseline.

When possible, we computed missing standard deviations from other statistics such as standard errors, confidence intervals, or P values, according to the methods recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019b).

Assessment of heterogeneity

We assessed clinical and methodological diversity in terms of participants, interventions, outcomes, and study characteristics for the included studies to determine whether a meta-analysis was appropriate. We conducted this by observing these data from the data extraction tables. We assessed statistical heterogeneity by visually inspecting forest plots to assess for obvious differences in results between studies, and by using I^2 and Chi^2 statistical tests.

As recommended in the*Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2019), interpretation of an I^2 value of 0% to 40% might 'not be important'; 30% to 60% may represent 'moderate' heterogeneity; 50% to 90% may represent 'substantial' heterogeneity; and 75% to 100% represents 'considerable' heterogeneity. As noted in the *Cochrane Handbook for Systematic Reviews of Interventions,* we kept in mind that the importance of I^2 depends on (1) the magnitude and direction of effects; and (2) the strength of evidence for heterogeneity.

We interpreted a Chi^2 test P value ≤ 0.10 as indicative of statistical heterogeneity.

When we identified substantial heterogeneity, we reported it and investigated possible causes by following the recommendations in Section 9.6 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2019).

Assessment of reporting biases

We planned to create and examine funnel plots to explore possible small study biases, as outlined in Section 10.4 of the *Cochrane Handbook for Systematic Reviews of Interventions.* We were not able to pool more than 10 trials for meta-analysis; therefore we did not use funnel plots.

To assess outcome reporting bias, we checked trial protocols against their published reports. For studies published after 1 July 2005, we screened the International Clinical Trials Registry Platform of the World Health Organization for the a priori trial protocol (apps.who.int/trialssearch). We evaluated whether selective reporting of outcomes was present.

Data synthesis

We undertook meta-analyses only when this was meaningful, that is, when treatments, participants, and the underlying clinical question were similar enough for pooling to make sense. For clinically homogeneous studies, we pooled outcomes in a meta-analysis using the random-effects model as a default. All included studies were small, hence sensitivity analysis to assess for small study bias was not relevant.

GRADE and 'Summary of findings' tables

We have included 'Summary of findings' (SoF) tables based on the following main comparison.

- Any one type of shoulder replacement surgery versus any other type of shoulder replacement surgery.

This compares two fundamentally different types of shoulder replacement and is a recognised area of research uncertainty (Rangan 2016). Planned comparisons of shoulder replacement surgery to other treatments or sham treatments were not possible due to lack of studies.

We have included the following seven major outcomes in the SoF tables.

- Pain.
- Function.
- Quality of life.
- Participant-rated global assessment of treatment success.
- Adverse events: total.
- Adverse events: serious.
- Revision or re-operation.

No studies were available to include 'Summary of findings' (SoF) tables based on the following main comparisons.

- Any type of shoulder replacement surgery versus placebo (sham surgery).
- Any type of shoulder replacement versus any type of non-surgical treatment.
- Any type of shoulder replacement surgery versus any other type of surgery.

When multiple time points were reported, the SoF reports intermediate outcomes (one to three years post surgery). When multiple time points were recorded within this range by the same study, the latest time point has been used.

Two people (RC, HG) independently assessed the quality of the evidence. We used the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness, and publication bias) to assess the quality of a body of evidence as it relates to studies that contributed data to the meta-analyses for pre-specified outcomes, and we reported the quality of evidence as high, moderate, low, or very low. We used methods and recommendations described in Sections 8.5 and 8.7, and in Chapters 11 and 12, of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2019a; Schünemann 2019a; Schünemann 2019b). We used GRADEpro software to prepare the SoF tables (GRADEpro GDT 2015). We justified all decisions to downgrade the quality of studies by using footnotes, and we made comments to aid the reader's understanding of the review when necessary.

Subgroup analysis and investigation of heterogeneity

Insufficient data were available to carry out the planned subgroup analyses for the following factors thought to influence outcomes (Muh 2013; Simone 2014).

- Age of the participant.
- Presence or absence of significant rotator cuff tear.

Sensitivity analysis

Insufficient data were available to carry out the following sensitivity analyses to investigate the robustness of the treatment effect on pain and function.

- Inclusion of missing data.
- Inclusion of trials identified at risk of selection bias.
- Inclusion of trials with unclear or inadequate blinding of the outcome assessor.
- Selection of the statistical method for pooled data (fixed-effect versus random-effects model).

RESULTS

Description of studies

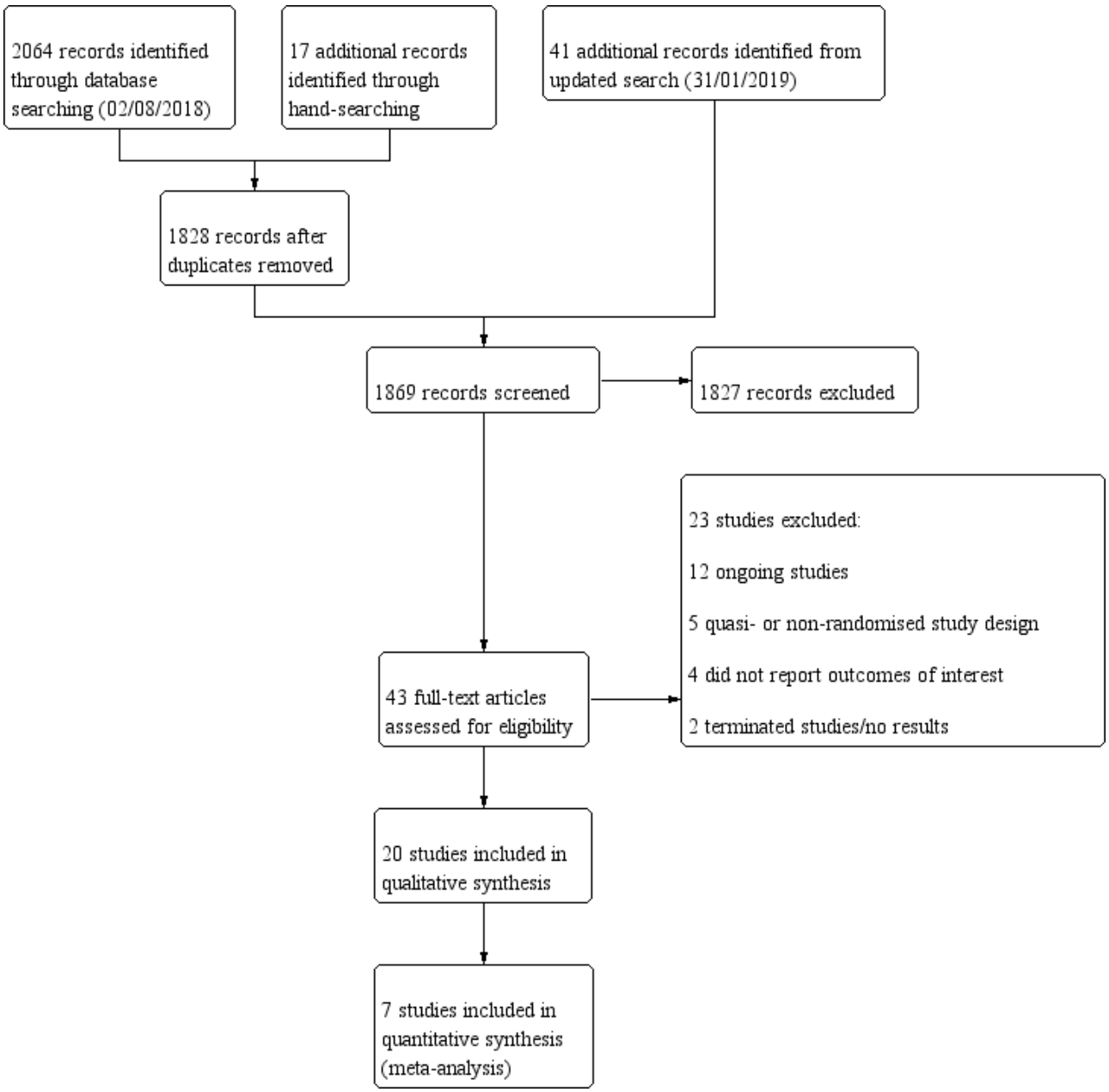
We have summarised the study characteristics under Included studies and Excluded studies. Full details of each study can be found in the Characteristics of included studies, Characteristics of excluded studies, and Characteristics of ongoing studies tables.

Results of the search

Figure 1 details the studies screened and included in the review. The initial searches performed 2 August 2018 yielded 2081 records, with an additional 41 records obtained from the updated search on 31 January 2019. After exclusion of duplicates and screening of abstracts and titles for eligibility, we identified 43 studies for full-

text review. Of the eight studies included in the previous version of this review (Singh 2010), we excluded one due to concerns regarding the study design, and we noted that 13 new studies met the inclusion criteria. In summary, 20 studies were included, 11 studies were excluded, and a further 12 studies were ongoing and were not yet reported.

Figure 1. Study flow diagram.



Included studies

We identified 20 studies for inclusion in the review. These are described in detail in the Characteristics of included studies table and are summarised here.

Trial design

All 20 studies were parallel-group randomised controlled trials with two arms. One study was conducted across seven independent sites, four studies recruited from two sites, and the remaining 15 studies were single-centre trials. Length of follow-up was six weeks in one study, one year in two studies, two years in 12 studies, three years in two studies, and five years in one study. Two studies provided additional data on selected outcomes at five years and at 10 years.

Trial setting

Seven studies were performed in the United States of America, four in Canada, two in Denmark, two in Germany, and one each in the United Kingdom, France, Sweden, New Zealand, and Australia.

Trial size

A total of 1083 participants (1105 shoulders) were randomised. The median number of shoulders analysed in each study was 42 (range 16 to 161). In 14 of 20 studies, each included participant underwent surgery on one shoulder only. In 5 of 20 studies (Boileau 2002; Edwards 2010; Gartsman 2000; Rahme 2009; Rasmussen 2015), some participants underwent two separate shoulder replacement procedures, which were randomised and analysed independently (203 participants, 224 procedures). In Gascoyne 2017, one participant underwent a bilateral procedure. It is unclear whether the shoulders were randomised independently.

Participants

Sixteen studies reported on patients with glenohumeral osteoarthritis (excluding rotator cuff tear arthropathy), seven of which explicitly stated that the diagnosis was primary osteoarthritis (Boileau 2002; Gartsman 2005; Litchfield 2011; Lo 2005; Nuttall 2007; Sandow 2013; Uschok 2017). In these 16 studies, the mean age of participants ranged from 63 to 70 years, and the median proportion of female participants was 55% (range 30% to 75%). Four studies included participants with rotator cuff tear arthropathy (Edwards 2012; Gobezie 2019; Greiner 2015; Poon 2014). In these studies, the mean age of participants ranged from 69 to 81 years, and the median proportion of female participants was 63% (range 56% to 65%).

Interventions

The 13 different comparisons are summarised below. The different study interventions are described in detail in the Characteristics of included studies tables.

- We found no studies comparing any type of shoulder replacement surgery to placebo (sham surgery).
- We found no studies comparing any type of shoulder replacement surgery to any type of non-surgical treatment.
- We found no studies comparing any type of shoulder replacement surgery to any other type of surgery.
- We found five studies comparing one type of shoulder replacement surgery to another type of shoulder replacement surgery.
  - Three compared conventional stemmed TSR to stemmed humeral hemiarthroplasty (Gartsman 2000; Lo 2005; Sandow 2013).
  - One compared re-surfacing humeral hemiarthroplasty to stemmed humeral hemiarthroplasty (Rasmussen 2015).
  - One compared conventional stemless TSR to conventional stemmed TSR (Uschok 2017).
  - No studies compared reverse TSR to any other type of shoulder replacement.
- We found 15 studies comparing one type of shoulder replacement surgical technique versus any other type of shoulder replacement surgical technique.
  - Six compared different fixation methods/materials for glenoid component fixation for conventional stemmed TSR.
    - Metal-backed uncemented versus all-polyethylene keeled glenoid component (Boileau 2002).
    - Pegged versus keeled all-polyethylene glenoid components (Edwards 2010; Gartsman 2005; Gascoyne 2017; Nuttall 2007; Rahme 2009).
  - One compared uncemented to cemented fixation of the humeral stem in conventional stemmed TSR (Litchfield 2011).
  - Three compared different surgical approach techniques for conventional TSR, including the following.
    - Lesser tuberosity osteotomy versus subscapularis peel or tenotomy (Lapner 2012; Levine 2019).
    - Subscapularis sparing versus standard subscapularis tenotomy (Kwon 2019).
  - One compared one brand of re-surfacing humeral hemiarthroplasty to another brand (Mechlenburg 2014).
  - Three compared different glenosphere positioning methods for reverse TSR, including the following.
    - 10-degree inferior inclination versus neutral inclination (Edwards 2012).
    - Bony increased offset versus standard offset (Greiner 2015).
    - Eccentric versus concentric (Poon 2014).
  - One compared a 135-degree humeral neck shaft angle to a 155-degree angle for stemmed reverse TSR (Gobezie 2019).

From the questions with available data, we defined whether to undertake a TSR or a hemiarthroplasty as the main comparison for this review because it is a key uncertainty reported by research priority setting partnerships (Rangan 2016). If studies comparing reverse TSR to other treatments had been available, this would have been rated with high importance. The order of the other comparisons is not specific.

Outcomes

We report the full details of outcomes in the Characteristics of included studies tables. See the summary below.

- Pain: reported by 12 studies.
  - Eight studies reported the VAS (Gartsman 2000; Gobezie 2019; Kwon 2019; Levine 2019; Lo 2005; Nuttall 2007; Poon 2014; Sandow 2013).
  - Five studies reported the pain subdomain of the Constant Murley Score (Boileau 2002; Greiner 2015; Rasmussen 2015; Sandow 2013; Uschok 2017).
- Function: reported on various scales by 19 studies.
  - Seven studies reported the WOOS Index (Edwards 2010; Gascoyne 2017; Lapner 2012; Litchfield 2011; Lo 2005; Mechlenburg 2014; Rasmussen 2015).
  - Twelve studies reported the ASES Score (Edwards 2010; Edwards 2012; Gartsman 2000; Gascoyne 2017; Gobezie 2019; Kwon 2019; Lapner 2012; Levine 2019; Litchfield 2011; Lo 2005; Nuttall 2007; Poon 2014).
  - One study reported the OSS (Poon 2014).
  - Eleven studies reported the Constant Murley Score (Boileau 2002; Edwards 2010; Edwards 2012; Greiner 2015; Lo 2005; Mechlenburg 2014; Nuttall 2007; Rahme 2009; Rasmussen 2015; Sandow 2013; Uschok 2017).
  - One study reported the DASH Score (Greiner 2015).
  - Three studies reported the UCLA Shoulder Rating Scale (Gartsman 2000; Lo 2005; Sandow 2013).
  - Three studies reported the Single Assessment Numerical Evaluation or the Subjective Shoulder Value (Edwards 2010; Gobezie 2019; Rahme 2009).

- Two studies reported the Simple Shoulder Test (Gascoyne 2017; Levine 2019).
- One study reported the McMaster Toronto Arthritis patient preference questionnaire (MACTAR) Score (Litchfield 2011).
- One study reported the Activities of Daily Living and External Rotation (ADLER) Score (Greiner 2015).

- Quality of life (generic): reported by three studies.
  - Two studies reported Short Form-36 (Levine 2019; Lo 2005).
  - One study reported Short Form-12 (Litchfield 2011).
- Participant-rated global assessment of treatment success.
- Adverse events (total): reported by 12 studies (Boileau 2002; Edwards 2010; Edwards 2012; Gartsman 2000; Gobezie 2019; Greiner 2015; Lapner 2012; Levine 2019; Litchfield 2011; Lo 2005; Rasmussen 2015; Sandow 2013).
- Adverse events (serious): reported in two studies (Litchfield 2011; Lo 2005).
- Revision/re-operation/treatment failure: reported by 14 studies (Boileau 2002; Edwards 2012; Gartsman 2000; Gobezie 2019; Kwon 2019; Lapner 2012; Levine 2019; Litchfield 2011; Lo 2005; Mechlenburg 2014; Rahme 2009; Rasmussen 2015; Sandow 2013; Uschok 2017).
- Minor physician-evaluated outcomes.
  - Radiographic classifications of lucency/loosening/notching/muscle atrophy, etc: reported by 10 studies (Boileau 2002; Edwards 2010; Edwards 2012; Gartsman 2005; Gobezie 2019; Lapner 2012; Levine 2019; Poon 2014; Rahme 2009; Uschok 2017).
  - Radiostereometric analysis of component micromotion: reported by five studies (Gascoyne 2017; Mechlenburg 2014; Nuttall 2007; Rahme 2009; Uschok 2017).
  - Range of motion: reported by 15 studies.
    - Nine studies reported measured ranges in degrees (Edwards 2010; Edwards 2012; Gobezie 2019; Greiner 2015; Kwon 2019; Levine 2019; Litchfield 2011; Nuttall 2007; Poon 2014).
    - Six studies reported the subdomain of Constant Murley Score (Boileau 2002; Greiner 2015; Lo 2005; Rasmussen 2015; Sandow 2013; Uschok 2017).
    - One study reported the subdomain of UCLA Score (Gartsman 2000).
  - Strength: reported by nine studies.
    - Two studies reported measurements in pounds (Lapner 2012; Litchfield 2011).
    - Six studies reported the subdomain of Constant Murley Score (Boileau 2002; Greiner 2015; Lo 2005; Rasmussen 2015; Sandow 2013; Uschok 2017).
    - One study reported the subdomain of UCLA Score (Gartsman 2000).
  - Operating time: reported by two studies (Levine 2019; Rasmussen 2015).
  - Humeral head bone mineral density reported by one study (Mechlenburg 2014).

Range of motion and strength are considered to be of low utility and are not included in core outcome sets (Buchbinder 2017). Therefore these have not been explicitly reported in this review.

Excluded studies

We excluded 11 studies following full-text review. Of these, four studies did not report on the outcomes of interest (Ding 2015; Edwards 2007; Hendel 2012; Iannotti 2015), and four studies were quasi-randomised or non-randomised studies (Berth 2013; Hammond 2013; Kasten 2009; Mariotti 2014). Kircher 2009 was described as a randomised study; however significant concerns regarding the allocation process were identified by the review authors, and this study has also been excluded (see Characteristics of excluded studies). Two studies identified in clinical trials registers were excluded: one had been terminated due to slow recruitment and poor follow-up (NCT01884077), and one was listed in 2006 but no results have been published and the inclusion criteria did not match this review (ISRCTN42881741).

Ongoing studies

Twelve ongoing trials were identified in clinical trials registers with planned recruitment of 1533 participants (see Characteristics of ongoing studies). Six are listed as actively recruiting (NCT01697865; NCT02768597; NCT02966886; NCT03111147; NCT03727490; NCT03711175), two are active but not recruiting (NCT01288066; NCT01790113), two have recently been completed but no published results are available yet (NCT01404143; NCT03730597), one study is not yet recruiting (NCT02305966), and the status of one is unknown (NCT01587560). Two ongoing studies are comparing one type of shoulder replacement to another type of shoulder replacement (NCT01288066; NCT01790113). The remaining nine studies are comparing one technique to another for the same type of shoulder replacement. Five studies include participants with glenohumeral osteoarthritis with intact rotator cuff tendons, and six studies include participants with rotator cuff tear arthropathy or glenohumeral osteoarthritis with a large/massive rotator cuff tear. None of the registered ongoing studies compares shoulder replacement surgery to any other form of treatment or to placebo (sham surgery). No ongoing studies are comparing RTSRs to other types of shoulder replacement.

Risk of bias in included studies

See Figure 2 and Figure 3 for a summary of the risk of bias assessments across all included trials and for individual ratings for each trial. Full descriptions and review authors' justifications for the assigned ratings are included in the 'Risk of bias' tables within the Characteristics of included studies section.

Figure 2. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

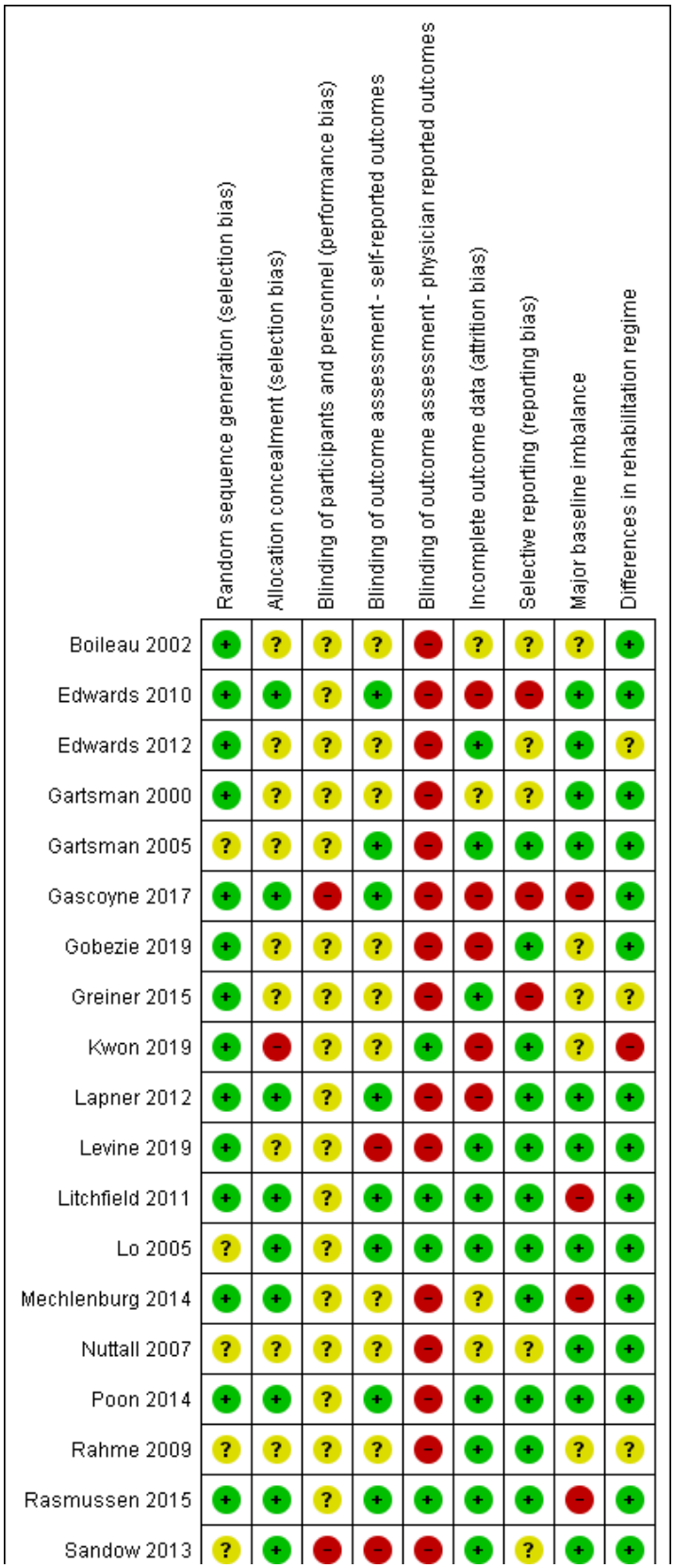


Figure 2. (Continued)

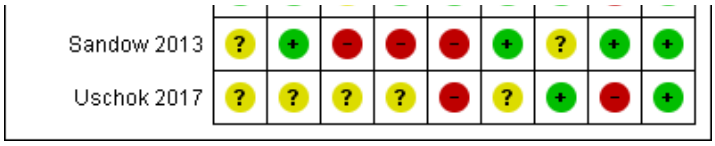
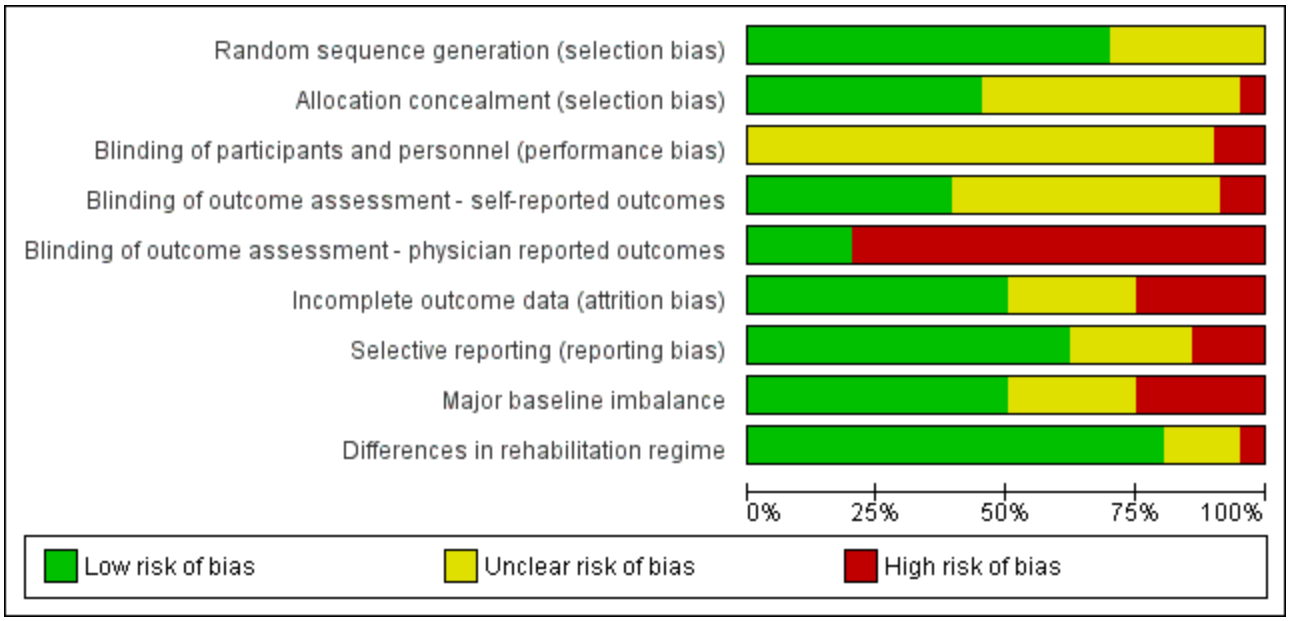


Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



Allocation

Sequence generation

For sequence generation, 14 studies were at low risk (Boileau 2002; Edwards 2010; Edwards 2012; Gartsman 2000; Gascoyne 2017; Gobezie 2019; Greiner 2015; Kwon 2019; Lapner 2012; Levine 2019; Litchfield 2011; Mechlenburg 2014; Poon 2014; Rasmussen 2015), and six studies were at unclear risk of bias (Gartsman 2005; Lo 2005; Nuttall 2007; Rahme 2009; Sandow 2013; Uschok 2017). The most common methods of sequence generation were computer-generated random numbers lists (10 studies) and other unspecified random numbers tables or lists (four studies). One study generated a sequence by drawing lots.

Allocation concealment

Allocation concealment was adequately described and nine studies were at low risk (Edwards 2010; Gascoyne 2017; Lapner 2012; Litchfield 2011; Lo 2005; Mechlenburg 2014; Poon 2014; Rasmussen 2015; Sandow 2013). In 10 studies, allocation concealment was not described or was described in insufficient detail; therefore the risk of bias was unclear (Boileau 2002; Edwards 2012; Gartsman 2000; Gartsman 2005; Gobezie 2019; Greiner 2015; Levine 2019; Nuttall 2007; Rahme 2009; Uschok 2017).

One study was assessed to be at high risk for selection bias (Kwon 2019).

Blinding

Performance bias

All included studies involved surgical interventions for which the surgeon could not be blinded to treatment allocation. The possible effect of this on performance bias (and subsequent outcomes) is unclear, and these studies have all been assessed to be at unclear risk of performance bias. Two studies were rated at high risk of performance bias (Gascoyne 2017; Sandow 2013). In Gascoyne 2017, all procedures were performed by a single surgeon, who terminated the trial early in response to published studies in the literature. In Sandow 2013, study follow-up was performed open-label; therefore the results cannot be assumed to be unbiased.

Outcome assessment (self-reported outcomes)

Follow-up for two studies was performed open-label (Levine 2019; Sandow 2013); these were rated at high risk of bias in this domain. In eight studies, robust blinding of participants throughout follow-up was reported, and these studies were rated at low risk for bias for self-reported outcomes reporting (Edwards 2010; Gartsman 2005; Gascoyne 2017; Lapner 2012; Litchfield 2011; Lo 2005; Poon 2014; Rahme 2009). In the remaining 10 studies, it is unclear how well and how long participants were blinded (Boileau 2002; Edwards 2012; Gartsman 2000; Gobezie 2019; Greiner 2015; Kwon 2019; Mechlenburg 2014; Nuttall 2007; Rahme 2009; Uschok 2017).

Outcome assessment (physician-evaluated outcomes)

Only four studies were rated at low risk of bias for physician-evaluated outcomes (Kwon 2019; Litchfield 2011; Lo 2005; Rasmussen 2015).

Sixteen studies were judged to be at high risk of bias for physician-evaluated outcomes (Boileau 2002; Edwards 2010; Edwards 2012; Gartsman 2000; Gartsman 2005; Gascoyne 2017; Gobezie 2019; Greiner 2015; Lapner 2012; Levine 2019; Mechlenburg 2014; Nuttall 2007; Poon 2014; Rahme 2009; Sandow 2013; Uschok 2017). This most often pertained to radiographic and/or radiostereometric outcomes, for which risk of bias is high because it is not possible to blind assessors to the radiological appearance of different types of shoulder replacement implants. In addition, although some of the scoring systems used for different types of implants were typically the same conceptually, they were not directly comparable nor validated as such.

In Sandow 2013, one participant undergoing revision surgery was excluded from analysis of failures because the surgeon attributed this failure to his own technical error. It is not possible to appreciate whether technical deficiencies contributed to failures in the other arm of the study, nor whether the procedure method was more likely to result in technical error and subsequent early failure. The contribution of surgical technique to failure was not discussed in any other papers. Therefore, for consistency, this technical failure was included in the analysis of revisions for this review.

Incomplete outcome data

We assessed risk of bias due to attrition to be low in 10 studies (Edwards 2012; Gartsman 2005; Greiner 2015; Levine 2019; Litchfield 2011; Lo 2005; Poon 2014; Rahme 2009; Rasmussen 2015; Sandow 2013), unclear in five studies (Boileau 2002; Gartsman 2000; Mechlenburg 2014; Nuttall 2007; Uschok 2017), and high in the remaining five (Edwards 2010; Gascoyne 2017; Gobezie 2019; Kwon 2019; Lapner 2012). Edwards 2010 reported results in two separate papers. The second report included additional procedures but the flow of participants was very unclear between the two, and it is clear how many patients were at risk for different outcomes at different times. Gascoyne 2017 stopped recruiting early and analysed only 9 of 15 randomised participants. Gobezie 2019 did not analyse 32% of patients at the two-year endpoint and showed imbalance between groups. Kwon 2019 reported on only 70 of 107 randomised participants and showed imbalance between groups. Lapner 2012 applied post-randomisation exclusion criteria. Participant flow and loss of follow-up through studies were often poorly reported within the text of study reports.

Selective reporting

We assessed risk of bias due to selective reporting to be low in 12 studies (Gartsman 2005; Gobezie 2019; Kwon 2019; Lapner 2012; Levine 2019; Litchfield 2011; Lo 2005; Mechlenburg 2014; Poon 2014; Rahme 2009; Rasmussen 2015; Uschok 2017), unclear in five studies (Boileau 2002; Edwards 2012; Gartsman 2000; Nuttall 2007; Sandow 2013), and high in the remaining three studies (Edwards 2010; Gascoyne 2017; Greiner 2015). In Edwards 2010, outcomes were inconsistently reported between two papers for the same study. Patient-reported outcomes were included only in the second paper, and the primary study endpoints were not clearly defined. Gascoyne 2017 reported findings without numbers at risk and with no measures of central tendency. Greiner 2015 reported on a subgroup analysis that was not pre-determined using a non-validated measure. This was the only statistically significant study finding. No trials described or referenced a study protocol, and entries for only five studies were identified on clinical trials registers (Kwon 2019; Lapner 2012; Levine 2019; Litchfield 2011; Mechlenburg 2014).

Other potential sources of bias

Major baseline imbalance

Five studies were at high risk of bias from baseline imbalance (Gascoyne 2017; Litchfield 2011; Mechlenburg 2014; Rasmussen 2015; Uschok 2017). These imbalances were seen in participant sex or baseline function scores. In one large multi-centre study (Litchfield 2011), trial authors performed a sensitivity analysis and found that a significant effect on the primary study outcome was highly likely to be attributable to sex imbalance.

In five studies, risk of bias from baseline imbalance was unclear (Boileau 2002; Gobezie 2019; Greiner 2015; Kwon 2019; Rahme 2009); in the remaining 10, it was low (Edwards 2010; Edwards 2012; Gartsman 2000; Gartsman 2005; Lapner 2012; Levine 2019; Lo 2005; Nuttall 2007; Poon 2014; Sandow 2013).

Differences in rehabilitation regimen

Sixteen studies described use of a standard postoperative rehabilitation regimen in both control and comparator arms and were assessed to be at low risk of bias (Boileau 2002; Edwards 2010; Gartsman 2000; Gartsman 2005; Gascoyne 2017; Gobezie 2019; Lapner 2012; Levine 2019; Litchfield 2011; Lo 2005; Mechlenburg 2014; Nuttall 2007; Poon 2014; Rasmussen 2015; Sandow 2013; Uschok 2017). Three studies did not describe the postoperative regimen in the text; therefore the risk of bias was unclear (Edwards 2012; Greiner 2015; Rahme 2009). Kwon 2019 applied different postoperative restrictions to the two study groups, and this domain was rated at high risk for bias.

Funding and financial conflicts of interest

We found a few studies spanning a large number of heterogeneous comparisons. These studies are small and reported a high proportion of industry funding or financial conflict of interest. Eight studies reported funding from industry sources (Gobezie 2019; Greiner 2015; Litchfield 2011; Lo 2005; Mechlenburg 2014; Nuttall 2007; Rahme 2009; Sandow 2013). A further seven studies reported a financial conflict of interest for study authors related to study implants (Edwards 2010; Edwards 2012; Gascoyne 2017; Kwon 2019; Poon 2014; Lapner 2012; Uschok 2017). Only three studies reported freedom from any personal financial or research funding that could be perceived as a conflict of interest (Gartsman 2000; Levine 2019; Rasmussen 2015). Two studies provided no information on funding or conflicts of interest (Boileau 2002; Gartsman 2005).

Effects of interventions

See: **Summary of findings for the main comparison** One type of shoulder replacement (TSR) to another type of shoulder replacement (hemiarthroplasty)

Any type of shoulder replacement surgery compared to placebo (sham surgery)

We identified no randomised controlled trials for this comparison.

Any type of shoulder replacement surgery compared to any type of non-surgical treatment

We identified no randomised controlled trials for this comparison.

Any type of shoulder replacement surgery compared to any other type of surgery

We identified no randomised controlled trials for this comparison.

One type of shoulder replacement surgery compared to another type of shoulder replacement surgery

Conventional stemmed total shoulder replacement (TSR) compared to stemmed humeral hemiarthroplasty

Three studies including a total of 126 participants (130 shoulders) provided the data for this comparison (Gartsman 2000; Lo 2005; Sandow 2013); the main findings are summarised in Summary of findings for the main comparison.

- Pain: conventional stemmed TSR may slightly reduce pain at two years (measured on a visual analogue scale (VAS), 0 to 10 scale) compared with stemmed hemiarthroplasty, and the effect may be clinically uncertain (mean difference (MD) -1.49, 95% confidence interval (CI) -2.88 to -0.10; minimum clinically important difference (MCID) 1.5; absolute difference 15% lower (1% lower to 29% lower); relative difference 23% lower (2% lower to 44% lower); 92 shoulders; 2 studies; I^2 = 55%; Figure 4; low-quality evidence (downgraded for risk of bias and imprecision)). Sandow 2013 also reported a clinically unimportant benefit in favour of conventional stemmed TSR at two years on a VAS (Table 1); however the results were available only as a median with range and were not suitable for inclusion in the meta-analysis.
- Function: conventional stemmed TSR may result in a small, clinically uncertain improvement in shoulder function at two years (measured by different scales represented on WOOS Index, 0 to 100 scale) compared with stemmed hemiarthroplasty (MD 10.57, 95% CI 2.11 to 19.02; MCID 10; absolute difference 11% higher (2% higher to 19% higher); relative difference 32% higher (6% higher to 57% higher); 92 shoulders; 2 studies; I^2 = 0%; Figure 5; low-quality evidence (downgraded for risk of bias and imprecision)). This translates to a number needed to treat for an additional beneficial outcome (NNTB) of 6 (95% CI 4 to 30). Sandow 2013 also reported a clinically unimportant benefit in favour of conventional stemmed TSR at two years on the Constant Murley Score and the UCLS Shoulder Score (Table 1); however the results were available only as a median with range and were not suitable for inclusion in the meta-analysis.
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: TSR probably results in little to no difference in quality of life at two years over hemiarthroplasty, but we are uncertain (MD 1.00, 95% CI -5.11 to 7.14 (mental); MD -0.80, 95% CI -8.2 to 6.6 (physical); 41 shoulders; 1 study; Short Form-12, 0 to 100 scales; Analysis 1.3 and Analysis 1.4; low-quality evidence (downgraded for risk of bias and imprecision)).
- Adverse events (total): we are uncertain whether there is any difference in the rate of specific adverse events occurring within three years of surgery (RR 0.50, 95% CI 0.14 to 1.74; 42 shoulders; 1 study; Analysis 1.5; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Adverse events (serious): we are uncertain whether there is any difference in the rate of serious adverse events within the first year (single event reported in one study arm; 42 shoulders; 1 study; Analysis 1.6; very low-quality evidence (downgraded for risk of bias and serious imprecision)). This is based on one reported fatal pulmonary embolism.
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference in the rate of revision, re-operation, or treatment failure within three years (RR 1.29, 95% CI 0.30 to 5.53; 92 shoulders; 2 studies; Figure 6; very low-quality evidence (downgraded for risk of bias and serious imprecision)). Sandow 2013 noted a trend towards higher revision rates at two years and up to 10 years (RR 0.33, 95% CI 0.07 to 1.53), but we had serious concerns regarding risk of bias for this physician-determined outcome in this study, and we excluded these results from the meta-analysis. Overall, the individual and pooled sample sizes were too small to justify reliable conclusions for this outcome in the absence of a large effect size.
- Physician evaluated: no physician-evaluated outcomes meeting our eligibility criteria were reported.

Figure 4. Forest plot of comparison: 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, outcome: 1.1 Pain: visual analogue scale (0 to 10, lower = better).

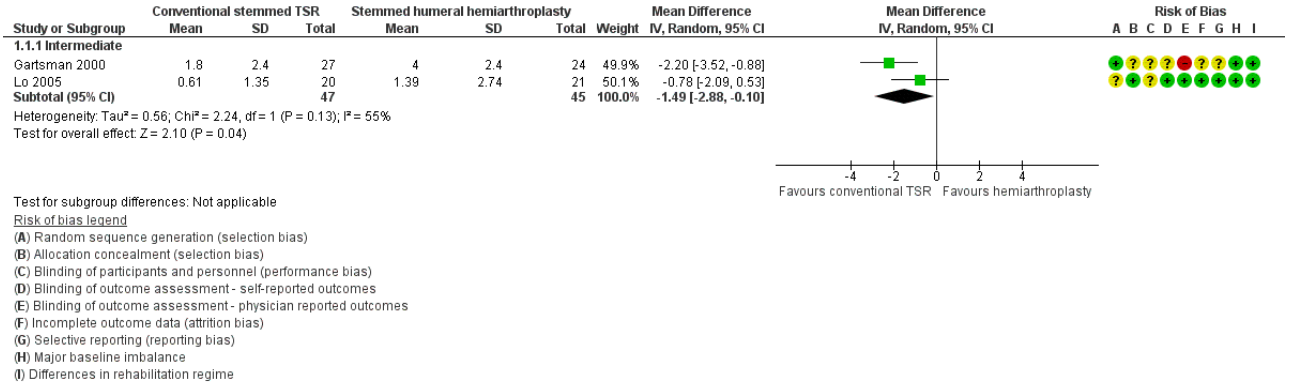


Figure 5. Forest plot of comparison: 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, outcome: 1.2 Disability/Function: WOOS Index (0 to 100, higher = better). Gartsmann 2000 raw data reported as ASES Shoulder Score.

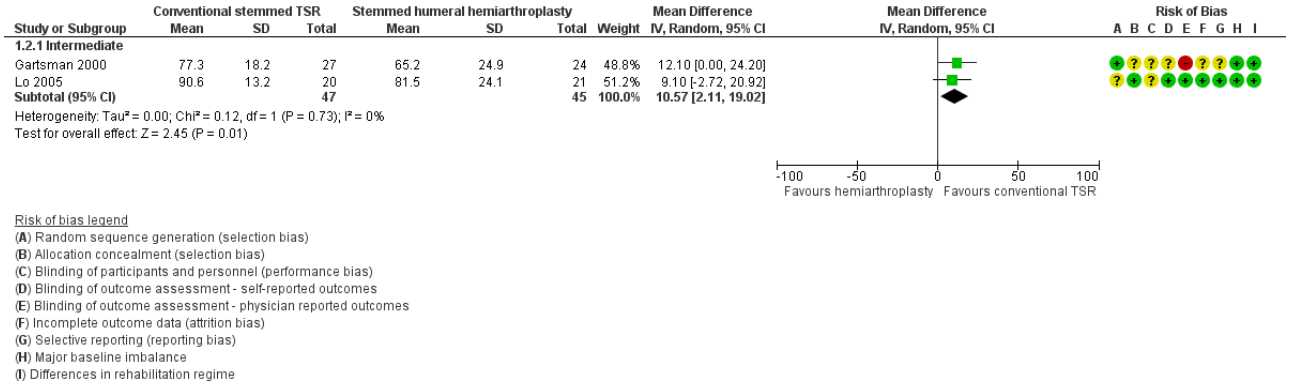
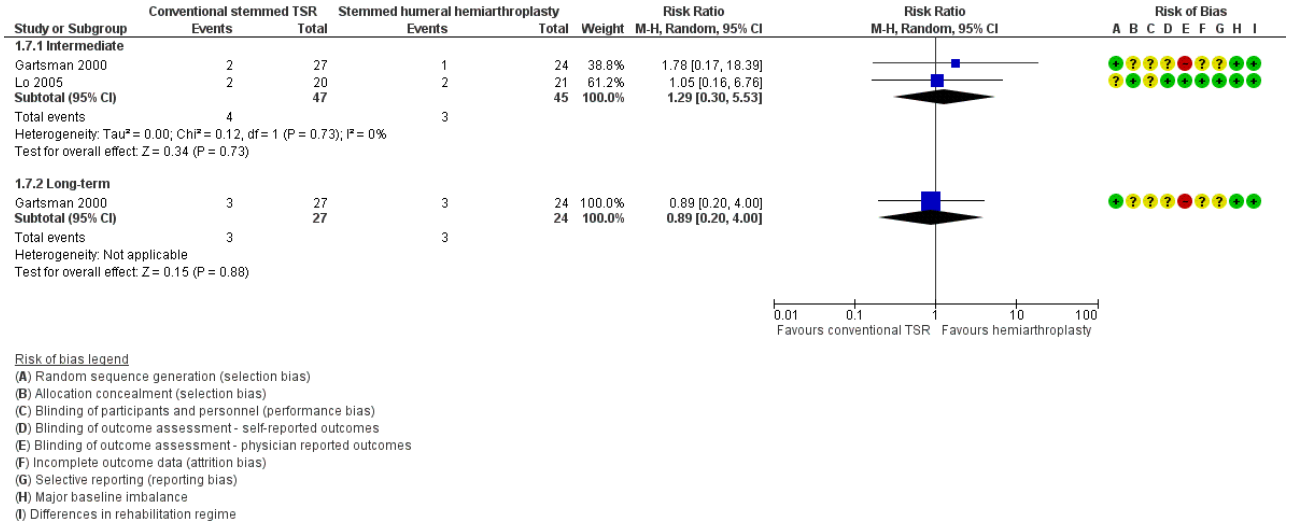


Figure 6. Forest plot of comparison: 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, outcome: 1.7 Revision, re-operation, or treatment failure (cumulative counts). Sandow 2013 excluded from the meta-analysis due to multiple potential sources of bias.



Resurfacing humeral hemiarthroplasty compared to stemmed humeral hemiarthroplasty

One study of 35 participants (40 shoulders) provided the data for this comparison (Rasmussen 2015).

- Pain: we are uncertain whether there is a difference in patient-reported pain between the two treatment groups. This was reported only as a subdomain of the Constant Murley Score. Trial authors reported a small 3.2-point difference in favour of stemmed hemiarthroplasty (Table 1). Interpretation of these subscores is not validated, and the evidence is of very low quality (downgraded for risk of bias and serious imprecision).
- Function: we are uncertain whether there is any effect of resurfacing humeral hemiarthroplasty compared to stemmed humeral hemiarthroplasty on function (WOOS Index, 0 to 100 scale) at one year because the included study reported a large baseline imbalance in this domain, and the confidence interval for the estimated difference was very wide (MD -20.2, 95% CI -36.99 to -3.41; MCID 10; 38 shoulders; Analysis 4.1; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether there is any difference in the rate of adverse events within one year between the two study arms (RR 1.00, 95% CI 0.16 to 6.42; 40 shoulders; Analysis 4.2; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Adverse events (serious): no serious events were reported.
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference in the rate of revision, re-operation, or treatment failure (very low-quality evidence (downgraded for risk of bias and serious imprecision)). Trial

authors stated that were no events in either arm, but the study was very underpowered for rare events.

- Physician evaluated: resurfacing humeral hemiarthroplasty may reduce operating time compared to stemmed humeral hemiarthroplasty (28 minutes shorter, 95% CI 18.7 to 36.7; 40 shoulders). This is an indirect outcome measure with possible relevance to cost-analyses but no direct relevance to efficacy or effectiveness.

Conventional stemless TSR compared to conventional stemmed TSR

One study of 40 participants (40 shoulders) provided data for this comparison (Uschok 2017).

- Pain: we are uncertain whether there is a difference in patient-reported pain between the two treatment groups. This was reported only as a subdomain of the Constant Murley Score. Trial authors report a non-significant difference (MD 2.7 points, 0 to 15 scale, higher = better; Table 1) in favour of stemmed TSR at two years. Interpretation of these subscores is not validated, and the evidence is of very low quality (downgraded for imprecision and two levels for risk of bias).
- Function: we are uncertain of the effects of stemless humeral components compared to stemmed humeral components for conventional TSR on function (Constant Murley Score, 0 to 100 scale; Analysis 3.1; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)). No difference was found at two years (MD -0.2, 95% CI -9.68 to 9.28; 33 shoulders) nor at five years (MD 2.9, 95% CI -7.01 to 12.81; 29 shoulders).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether there is any difference in the rate of adverse events when stemless compared to stemmed humeral components are used for TSR (very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)). The included study was underpowered for rare events, suffered from significant attrition bias, and reported conflicting percentages and counts.
- Adverse events (serious): no serious events were reported. The comparison is underpowered for rare events.
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference in the rate of revision, re-operation, or treatment failure when stemless compared to stemmed humeral components are used for TSR (very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)). The included study was underpowered for rare events, suffered from significant attrition bias, and reported conflicting percentages and counts.
- Physician evaluated: the included study reported on several physician-evaluated radiographic measures, none of which were suitable for inclusion in this review. Lucencies around the humeral components were reported on scales that do not appear to be directly comparable between study arms. The type of glenoid component used changed during the study period (metal-backed uncemented to all-polyethylene cemented), and numbers were not balanced between groups. This has significant potential as a confounder and is the topic of one of the other review comparisons. Several humeral component positioning measurements are reported; these are of unclear/indirect relevance to clinically important outcomes and are beyond the scope of this review.

One type of surgical technique compared to any other type of surgical technique

Conventional stemmed TSR with a metal-backed uncemented component compared to an all-polyethylene keeled cemented glenoid component

One study of 39 participants (40 shoulders) provided the data for this comparison (Boileau 2002).

- Pain: we are uncertain whether there is a difference in patient-reported pain between the two treatment groups. This was reported only as a subdomain of the Constant Murley Score. Trial authors reported no difference between the two groups (Table 1). Interpretation of these subscores is not validated, and the evidence is of very low quality (downgraded for risk of bias and serious imprecision).
- Function: we are uncertain whether there is a difference in function between the two arms (very low-quality evidence (downgraded for risk of bias and serious imprecision)). No clinically important or significant difference was identified between groups for the domain of function measured via the Constant Murley Score at 1, 2, and 3 years, reported as an "average" and a range: cemented 67 (6 to 89), uncemented 75 (17 to 89), 0 to 100 scale.
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether there is a difference in the rate of any adverse events between pegged and keeled components (very low-quality evidence (downgraded for risk of bias and serious imprecision)). Trial authors did not report any adverse events occurring over and above those cases undergoing revision surgery.
- Adverse events (serious): no serious events were reported.
- Revision, re-operation, or treatment failure: there may be a clinically important increased risk of revision at a mean follow-up of 38 months for metal-backed uncemented glenoid implants compared to all-polyethylene keeled cemented components; however we are very uncertain of this effect due to a very wide confidence interval (risk ratio (RR) 11.00, 95% CI 0.65 to 186.62; 40 shoulders; Analysis 2.1; very low-quality evidence (downgraded for risk of bias and serious imprecision)). This is based for 5 of 20 events in the metal-backed uncemented group versus 0 of 20 events in the cemented group.
- Physician evaluated: we are uncertain of the effect or importance of periprosthetic glenoid lucency between the two study arms (very low-quality evidence (downgraded for risk of bias and serious imprecision)). The presence of radiolucent lines was reported narratively. There was a difference in the presence of any radiolucency for cemented versus uncemented components (17/20 versus 5/20); however none of these radiolucencies progressed in the cemented group, and four of five progressed with associated clinical deterioration in the uncemented group.

Conventional stemmed TSR with pegged compared to keeled all-polyethylene cemented glenoid components

Five studies comprising a total of 160 participants (172 shoulders) provided data for this comparison (Edwards 2010; Gartsman 2005; Gascoyne 2017; Nuttall 2007; Rahme 2009).

- Pain: we are uncertain whether there is any difference in self-reported pain at two years. This outcome was reported by only one study, and as mean values only (MD 0, 95% CI not estimable; 20 participants; VAS 0 to 10 scale; very low-quality evidence (downgraded for risk of bias and serious imprecision)) (Nuttall 2007).
- Function: there may be no difference in function between pegged and keeled glenoid components, but we are uncertain (very low-quality evidence (downgraded for risk of bias and serious imprecision)). Four studies reported on this outcome using different outcome scores, different measures of central tendency (means/medians, exact P values/range/no measure of spread), and different time points. No meta-analysis was possible. No clinically important or significant difference was reported by any of these studies at one year, two years, or five years follow-up. The results are summarised in Analysis 5.1 and Table 1.
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether there is a difference in the rate of any adverse events between pegged and keeled components (very low-quality evidence). No studies reported any adverse events occurring over and above those cases undergoing revision surgery, which are reported below.
- Adverse events (serious): no serious events were reported.
- Revision, re-operation, or treatment failure: the effect of using pegged versus keeled glenoid components in conventional stemmed TSR is uncertain at both two-year follow-up (Peto odds ratio (OR) 0.35, 95% CI 0.05 to 2.56; 80 shoulders; 2 studies; I^2 = 0%) and five-year follow-up (Peto OR 0.33, 95% CI 0.08 to 1.46; 59 shoulders; 1 study). Available evidence is of very low quality (downgraded for risk of bias and serious imprecision).
- Physician evaluated: perioprosthetic glenoid lucency on radiographs and radiostereometric evidence of excessive component micromotion are proposed precursors of subsequent component failure by loosening. We are uncertain if there is any difference in rates of periprosthetic glenoid lucency or micromotion. Using a score of 4 or greater (on a 1 to 5 scale) as a cut-off for the scales described by Franklin 1988 and Lazarus 2002, there may be little to no difference in the rate of substantial radiolucency when pegged components are compared to keeled components at two-year follow-up (RR 0.38, 95% CI 0.02 to 8.83; 71 participants; 2 studies; very low-quality evidence (downgraded for risk of bias and serious imprecision)) (Edwards 2010; Rahme 2009), or at five-year follow-up (RR 1.20, 95% CI 0.55 to 2.63; 38 participants; 1 study; very low-quality evidence (downgraded for risk of bias and serious imprecision) (Edwards 2010). Gartsman 2005 reported only on immediate postoperative radiographs. Radiostereometric analysis of glenoid component micromotion was performed in three studies. Results are summarised in Table 1. Two studies of 47 shoulders demonstrated no important difference in component translation and rotation between the two groups (very low-quality evidence (downgraded for risk of bias and serious imprecision)) (Nuttall 2007; Rahme 2009). Gascoyne 2017 found a possible difference in favour of pegged implants; however the study was at high risk of bias across several domains, and the measurement method was changed partway through the study.

Conventional stemmed TSR with uncemented compared to cemented fixation of the humeral stem

One study of 161 participants (161 shoulders) provided the data for this comparison (Litchfield 2011).

- Pain: this outcome was not reported.
- Function: we are uncertain if there is a clinically important improvement in function (WOOS Index, 0 to 100 scale) for cemented fixation versus uncemented fixation of the humeral stem in conventional TSR (MD 8.6, 95% CI 2.4 to 14.8; 152 shoulders; MCID 10 points; Analysis 6.1; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: we are uncertain whether there is a difference in quality of life between cemented and uncemented humeral stem fixation at two years measured via the Short Form-12 mental component (MD 2.59, 95% CI -0.44 to 5.62; MCID 4, scale 0 to 100; Analysis 6.2) and physical component (MD 3.77, 95% CI 0.05 to 7.49; MCID 1, scale 0 to 100; 152 shoulders; Analysis 6.3; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).
- Adverse events (total): we are uncertain if there is any difference in adverse events between cemented and uncemented humeral stem fixation at two years (Peto OR 1.55, 95% CI 0.43 to 5.55; 161 shoulders; Analysis 6.4; very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)).
- Adverse events (serious): we are uncertain whether there is any difference in serious adverse events between cemented and uncemented humeral stem fixation (Peto OR 1.01, 95% CI 0.06 to 16.33; 161 shoulders; Analysis 6.5; very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)).
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference between cemented and uncemented humeral stem fixation at two years (Peto OR 1.27, 95% CI 0.28 to 5.79; 152 shoulders; Analysis 6.6; very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)).
- Physician evaluated: no physician-evaluated outcomes meeting our eligibility criteria were reported.

Conventional stemmed TSR via a subscapularis-sparing approach compared to standard subscapularis tenotomy

One study of 107 participants (107 shoulders) provided the data for this comparison (Kwon 2019).

- Pain: there may be little to no difference in self-reported pain for a subscapularis-sparing versus a standard approach to conventional TSR, but we are uncertain (MD 0.60, 95% CI -0.33 to 1.53; VAS; 70 shoulders; Analysis 7.1; low-quality evidence (downgraded two levels for risk of bias)).

- Function: we are uncertain if there is any difference in levels of function (MD -5.40, 95% CI -14.70 to 3.90; ASES Shoulder Scale; 70 shoulders; Analysis 7.2; very low-quality evidence (downgraded two levels for risk of bias and serious imprecision)).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total and serious): these outcomes were not reported.
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference in rates of revision, re-operation, or treatment failure (Peto OR 3.43, 95% CI 0.46 to 25.67; 70 shoulders; Analysis 7.3; very low-quality evidence (downgraded two levels for risk of bias and imprecision)).
- Physician evaluated: no physician-evaluated outcomes meeting our eligibility criteria were reported.

Conventional stemmed TSR via a lesser tuberosity osteotomy approach compared to subscapularis tenotomy/peel

Two studies of 147 participants (147 shoulders) provided the data for this comparison (Lapner 2012; Levine 2019).

- Pain: we are uncertain whether there is any difference in patient-reported pain at one year (MD -0.1, 95% CI not estimable; 59 shoulders; 1 study; Levine 2019; Table 1; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Function: at two years follow-up, there may be little to no difference in function measured by the WOOS Index, but we are uncertain (MD -1.70, 95% CI -9.16 to 5.76; 87 shoulders; 1 study; Analysis 8.1; low-quality evidence (downgraded for risk of bias and imprecision)) (Lapner 2012). Levine 2019 also reported in this domain; however values reported in the text and in the figures showed inconsistencies that we were unable to reconcile.
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference in the number of events within two years (Peto OR 0.14, 95% CI 0.01 to 2.21; 87 shoulders; 1 study; Analysis 8.3; very low-quality evidence (downgraded for risk of bias and serious imprecision)) (Lapner 2012). Only two failures were reported in the tenotomy/peel group, and none in the osteotomy group. Levine 2019 also reported on this outcome but only up to one year post surgery and revealed no differences (one event in each study arm).
- Adverse events (total): we are uncertain whether there is any difference in the rate of serious adverse events within the first year (Peto OR 3.50, 95% CI 0.57 to 21.54; 59 shoulders; 1 study; Analysis 8.2; very low-quality evidence (downgraded for risk of bias and serious imprecision)) (Levine 2019).
- Adverse events (serious): none were reported.
- Physician evaluated: there may be little to no difference in the likelihood of achieving satisfactory radiological evidence of healing of the repair at one year (RR 0.99, 95% CI 0.87 to 1.13; 140 shoulders; 2 studies; Analysis 8.4; low-quality evidence (downgraded for risk of bias and imprecision)).

Lapner 2012 also reported the degree of fatty infiltration of the rotator cuff tendons (as described by Goutallier 1994; see Table 1) and strength of the subscapularis at one year and described no difference in either outcome. Levine 2019 reported on range of motion and strength and described no difference in either at one year post surgery. Levine 2019 also reported on operative time and described significantly shorter operative time for subscapularis tenotomy (129.3 minutes versus 152.7 minutes) (very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).

Resurfacing humeral hemiarthroplasty with one brand compared to another brand

One study of 32 participants (32 shoulders) provided the data for this comparison and compared the Global C.A.P. shoulder implant with the Copeland shoulder implant (Mechlenburg 2014).

- Pain: this outcome was not reported.
- Function: we are uncertain whether there is any difference in function between study arms at any time point (very low-quality evidence (downgraded for risk of bias and serious imprecision)) measured via the WOOS Index (Global C.A.P. 294 (range 111 to 477); Copeland 128 (range 53 to 550); median/interquartile range (IQR), 0 to 1900 (raw) scale; data at 24 months).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total and serious): no events were reported.
- Revision, re-operation, or treatment failure: we are uncertain whether there is any difference between the two types of implants (RR 2.08, 95% CI 0.40 to 10.72; Analysis 9.1; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Physician evaluated: we are uncertain whether there is any difference between the two types of implants in terms of component micromotion at two years (total translation measured by radiostereometry) (MD -0.16 mm, 95% CI -0.60 to 0.28; Analysis 9.2; very low-quality evidence). This study reported micromotion of the humeral components measured via radiostereometric analysis in three planes of translation, one plane of rotation, and aggregate total translation. No differences among the individual components of the analysis were reported. Bone mineral density of the humeral head at two years may be similar between the two groups (Table 1; very low-quality evidence (downgraded for risk of bias and serious imprecision)).

Reverse stemmed TSR with a 10-degree inferior inclination of the glenosphere compared to neutral inclination

One study of 52 participants (52 shoulders) provided the data for this comparison and compared use of a 10-degree inferior tilted glenosphere with a neutral position for reverse TSR (Edwards 2012).

- Pain: this outcome was not reported.
- Function: there may be no clinically important improvement in function at one year (measured via the ASES Shoulder Score) (MD 7.60, 95% CI 0.83 to 14.37; MCID 13.5; 42 shoulders; Analysis 10.1; low-quality evidence (downgraded for risk of bias and imprecision)).
- Participant-rated global assessment of treatment success: this outcome was not reported.

- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether there is a difference in adverse events between groups (Peto OR 6.75, 95% CI 0.13 to 341.54; 42 shoulders; Analysis 10.2; very low-quality evidence (downgraded for risk of bias and serious imprecision)). Only one event (prosthesis dislocation) was reported in the neutral inclination group at short-term follow-up.
- Adverse events (serious): no serious events were reported.
- Revision, re-operation, or treatment failure: no events were reported in either group.
- Physician evaluated: we are uncertain whether there may be any effect of inferior inclination compared to neutral inclination on the rate of any scapular notching (Nerot grade ≥ 1, as per Valenti 2001) at one year (RR 1.15, 95% CI 0.85 to 1.56; 42 shoulders; Analysis 10.3; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).

Reverse stemmed TSR with bony increased offset (BIO) of the glenosphere compared to standard offset

One study of 34 participants (34 patients) provided the data for this comparison and compared BIO with standard offset technique (Greiner 2015).

- Pain: we are uncertain whether there is a difference in pain between BIO and the standard offset technique (very low-quality evidence). The included study reported on the subdomain of pain from the Constant Murley Score (MD 0 points, 95% CI not estimated; 31 shoulders; Table 1). The subscale is not validated for independent use and analysis as a continuous variable; therefore any interpretation should be done cautiously.
- Function: we are uncertain whether there is any difference in levels of function at two years for BIO compared to the standard technique when measured by the Constant Murley Score (MD 2.60, 95% CI -9.52 to 14.72; 31 shoulders; Analysis 11.1; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Adverse events (total): we are uncertain whether there is any difference in the rate of adverse events (RR 0.94, 95% CI 0.15 to 5.84; 31 shoulders; Analysis 11.2; very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)). Reported complications were acromial stress fractures (two in each arm) seen on a planned computed tomography (CT) scan at one-year follow-up. No other complications were reported.
- Adverse events (serious): no serious adverse events were reported.

This study did not report on the following outcomes quality of life, physician evaluated.

Reverse stemmed TSR via eccentric placement of the glenoid compared to concentric placement

One study of 50 participants (50 shoulders) provided the data for this comparison and compared eccentric with concentric placement of the glenosphere (Poon 2014).

- Pain: there is no clinically important difference between eccentric versus concentric placement of the glenosphere in pain at two years measured on a VAS (MD 0.20, 95% CI -0.63 to 1.03; 50 shoulders; Analysis 12.1; moderate-quality evidence (downgraded for risk of bias)).
- Function: there is no clinically important difference in function at two years measured by the ASES Shoulder Score (MD -2.00, 95% CI -5.17 to 1.17; 50 shoulders; Analysis 12.2; moderate-quality evidence (downgraded for risk of bias)).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether using eccentric versus concentric placement of the glenosphere has any effect on rates of specific adverse events within two years (Peto OR 1.18, 95% CI 0.07 to 19.57; 50 shoulders; Analysis 12.3; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Adverse events (serious): no serious adverse events were reported.
- Revision, re-operation, or treatment failure: we are uncertain whether using eccentric versus concentric placement of the glenosphere has any effect on rates of revision, re-operation, or treatment failure within two years (Peto OR 0.16, 95% CI 0.00 to 8.01; 50 shoulders; Analysis 12.4; very low-quality evidence (downgraded for risk of bias and serious imprecision)).
- Physician evaluated: we are uncertain whether using eccentric versus concentric placement of the glenosphere has any effect on the rate of glenoid notching (defined as a Nerot grade ≥ 1, as per Valenti 2001) (RR 0.29, 95% CI 0.04 to 2.44; 50 shoulders; Analysis 12.5; very low-quality evidence (downgraded for risk of bias and serious imprecision)).

Reverse stemmed TSR via a 135-degree humeral neck-shaft angle compared to a 155-degree neck-shaft angle

One study of 100 participants (100 shoulders) provided the data for this comparison (Gobezie 2019).

- Pain: we are uncertain whether there is any difference in patient-reported pain up to two years between a 135-degree and a 155-degree neck-shaft angle for RTSR, measured on a VAS (MD 1.00, 95% CI -0.13 to 2.13; 68 shoulders; Analysis 13.1; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).
- Function: we are uncertain whether there is any difference in patient-reported pain up to two years between a 135-degree and a 155-degree neck-shaft angle for reverse TSR, measured on the ASES Shoulder Score (MD -4.00, 95% CI -13.54 to 5.54; 68 shoulders; Analysis 13.2; very low-quality evidence (downgraded for imprecision and two levels for risk of bias)).
- Participant-rated global assessment of treatment success: this outcome was not reported.
- Quality of life: this outcome was not reported.
- Adverse events (total): we are uncertain whether using a 135-degree and a 155-degree neck-shaft angle for RTSR has any effect on rates of specific adverse events within two years (RR 1.05, 95% CI 0.31 to 3.57; 68 shoulders; Analysis 13.3; very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)).

- Adverse events (serious): no serious adverse events were reported.
- Revision, re-operation, or treatment failure: we are uncertain whether using a 135-degree and a 155-degree neck-shaft angle for RTSR has any effect on rates of revision, re-operation, or treatment failure (RR 0.42, 95% CI 0.08 to 2.14; 67 shoulders; Analysis 13.4; very low-quality evidence (downgraded for serious imprecision and two levels for risk of bias)).
- Physician evaluated (scapular notching): use of a 135-degree neck-shaft angle humeral component for reverse TSR may be associated with lower rates of scapular notching compared with a 155-degree neck-shaft angle humeral component for reverse TSR (RR 0.37, 95% CI 0.37 to 0.74; 68 shoulders; low-quality evidence (downgraded two levels for risk of bias)).

DISCUSSION

Summary of main results

No randomised trials have compared shoulder replacement surgery versus placebo, non-operative management, or any other type of surgical treatment. No ongoing studies are addressing these comparisons based on the registered descriptions of 12 ongoing trials. Therefore the potential benefits and adverse effects of shoulder replacement surgery for osteoarthritis or rotator cuff tear arthropathy compared to any other treatment modalities are unknown and will not be clarified by current ongoing trials.

A total of 20 trials looked at 13 different comparisons of different types of shoulder replacement and different technical aspects of shoulder replacement surgery. Thus trials on this review topic remain small and diverse - not large or focused on ongoing research uncertainties. Of the 20 identified trials, five addressed pegged versus keeled glenoid components for conventional total shoulder replacement (TSR), three compared conventional TSR to stemmed humeral hemiarthroplasty, two compared lesser tuberosity osteotomy to subscapularis tenotomy, and the remaining 10 comparisons were based on a single trial each. Pooling of data for meta-analysis was possible for only four outcomes in the main comparison of conventional stemmed TSR versus humeral hemiarthroplasty, two outcomes for the comparison of pegged versus keeled glenoid components, and one outcome for the comparison of lesser tuberosity osteotomy versus subscapularis tenotomy. The overall quality of evidence for most comparisons was low or very low; therefore few useful conclusions can be drawn. In particular, for dichotomous outcomes (adverse events and risks of revision, re-operation, or treatment failure), studies were universally too small to be powered for detection and comparison of rare events. The largest study included only 161 participants and was limited to two-year follow-up. Across all comparisons, we are very uncertain whether there is any difference in adverse events between any one comparator group and another.

For the main comparison of conventional stemmed TSR versus stemmed humeral hemiarthroplasty for osteoarthritis, low-quality evidence suggests there may be a clinically unimportant improvement in pain (mean difference (MD) -1.49, 95% confidence interval (CI) -2.88 to -0.10; mean clinically important difference (MCID) 1.5; absolute difference 15% lower (1% lower to 29% lower); relative difference 23% lower (2% lower to 44% lower)) and there may be a clinically unimportant improvement in function (MD 10.57, 95% CI 2.11 to 19.02; MCID 10; absolute difference 11% higher (2% higher to 19% higher); relative difference 32% higher (6% higher to 57% higher)) in favour of TSR. There may be no clinically important difference in overall quality of life measures (MD 1.00, 95% CI -5.11 to 7.14; MCID 4; absolute difference 1% higher (5% lower to 7% higher); relative difference 2% higher (9% lower to 13% higher)). We are uncertain whether there is any difference in rates of adverse events, revision, re-operation, or treatment failure because the evidence is of very low quality. Participant-rated global assessment of treatment success and physician-evaluated outcomes of interest were not reported.

The one study comparing metal-backed uncemented glenoid components to cemented all-polyethylene components for TSR noted that there may be a higher risk of revision or re-operation surgery at a mean of 38 months post surgery (Peto OR 9.29, 95% CI 1.46 to 59.09); however confidence in this estimate is very low based on serious imprecision in the estimate.

For the comparison of a subscapularis-sparing versus a standard approach for TSR, low-quality evidence suggests there may be little or no difference in participant-reported pain at two years (MD 0.6 points, 95% CI -0.33 to 1.53; 0 to 10 scale; MCID 1.5 points).

For the comparison of lesser tuberosity osteotomy versus subscapularis tenotomy/peel for the approach to TSR, low-quality evidence suggests there may be little or no difference in patient-reported function (MD -1.7 points, 95% CI -9.2 to 5.7; 0 to 100 scale).

For reverse shoulder replacement, low-quality evidence from one study comparing a 10-degree inferior inclination position to a neutral glenosphere position suggests there may be a clinically unimportant improvement in participant-reported function at one year (MD 7.60 points, 95% CI 0.83 to 14.37; 0 to 100 scale).

For eccentric versus concentric position of the glenosphere in reverse TSR for cuff tear arthropathy, one RCT provided moderate-quality evidence to show there is little to no difference between the two in terms of pain and function.

Overall completeness and applicability of evidence

Evidence from available studies is inadequate to address the main review objective - to determine the benefits and harms of shoulder replacement surgery in adults with osteoarthritis (OA) of the shoulder, including rotator cuff tear arthropathy (RCTA).

No randomised studies in this field have adequately addressed the fundamental question of the effectiveness and risks of one type of shoulder replacement over another, or the effectiveness and risks of shoulder replacement surgery compared to no treatment, placebo, or any other form of treatment for OA or RCTA. Only three comparisons (five studies) between one class of shoulder replacement and another have been made (stemmed humeral hemiarthroplasty versus TSR, stemmed humeral hemiarthroplasty versus resurfacing humeral hemiarthroplasty, stemless humeral hemiarthroplasty versus stemmed humeral hemiarthroplasty). Important uncertainties regarding the major classes of shoulder replacement remain unanswered by this review, specifically the choice between humeral hemiarthroplasty, conventional TSR, and reverse TSR, which presents far more fundamental questions than those addressed in more specific narrow-subtype studies. Most included studies compared one technique versus another. The importance and generalisability of some of these comparisons are not apparent. Many comparisons were supported by only one

study, and most of the outcomes for comparisons with more than one study were inadequately reported to allow meta-analysis. Therefore, results were for the most part inconclusive.

With regard to serious adverse events and risks of revision surgery, none of the included studies were of sufficient size and length of follow-up to be powered to identify these events reliably. Although three studies reported on revision risk for the main comparison of conventional stemmed TSR versus stemmed hemiarthroplasty, the quality of the evidence is very low. No firm conclusions on these outcomes can be made.

Quality of the evidence

The quality of the review is inherently limited by the low or very low quality of the included studies. Studies were small and covered short periods of follow-up. The quality of evidence for the main comparison is summarised according to GRADE criteria in Summary of findings for the main comparison and is stated for each comparison and outcome in the main results section, together with reasons for downgrading the evidence level.

Evidence was downgraded by at least one level due to bias for all comparisons for patient-reported outcomes including pain, function, and quality of life. Serious concerns for bias were common for reporting of radiological outcome measures (physician evaluated), and risk of performance bias was unclear or high in all studies. The nature of the intervention implies an inherent risk of (physician) performance bias; however, the implications of this are not clear. Imbalances between comparator arms were common at baseline. These imbalances have the potential to significantly distort the results and conclusions of individual studies.

Twenty studies contributed to 13 different comparisons, and few pooled analyses were possible. For the main comparison of conventional stemmed TSR versus stemmed humeral hemiarthroplasty, the direction of treatment effects was consistent for all outcomes.

No studies were identified that directly or indirectly addressed the efficacy of shoulder replacement surgery compared to any other surgical or non-surgical treatments. Populations included in these studies were representative of patients with primary osteoarthritis and rotator cuff tear arthropathy. The evidence may not be applicable to patients with secondary arthritis (e.g. due to sequelae of trauma). Reported physician-evaluated outcomes included radiological and radiostereometric measures of implant loosening. These may have a relationship with future implant failure and performance but are not directly relevant to the patient-experienced outcome.

For the comparison of eccentric versus concentric positioning of the glenoid in reverse stemmed TSR, a precise estimate was made to determine that there was no difference in pain or function. However, for all other outcomes and comparisons, the quality of evidence was downgraded by at least one level for imprecision. For dichotomous outcomes including adverse events and revision/re-operation/treatment failure, the evidence was downgraded by two levels for serious imprecision. The included studies were all too small to be powered to reliably identify these low-frequency events.

We did not identify any systematic evidence of publication bias due to unreported studies. Only one registered study was unpublished due to poor recruitment. However, only five of the included studies were recorded on trials registers, none were referenced in a study protocol, and a large proportion of small studies were industry-funded. We cannot be certain that there have been no unpublished trials.

Potential biases in the review process

This review was conducted according to the previously published protocol. Although this is an update of a previous review on the topic (Singh 2010), the scope was significantly changed to explicitly include trials of participants with rotator cuff tear arthropathy and to restrict the included studies to only those with any type of shoulder replacement as one of the study arms. To reflect this change, all searches were redesigned and run without date limits, and all studies were screened for this review by two independent review authors. These review authors made the decision to include studies and assessments of risk of bias independently of the previous review process with reference to updated Methodological Expectations for Cochrane Intervention Reviews (MECIR). Risk of bias from the review method process is therefore low. However, analysis of revision risk and adverse events is a major limitation of this review, largely due to the inclusion criteria requiring only randomised controlled trials. Much larger studies are needed to identify events that occur at a frequency of between 1 in 1000 and 1 in 100. This information may be better provided by well-designed studies from large registry-based or routinely collected datasets, or by future designed trials that become nested in national registries to monitor this longer-term follow-up.

Agreements and disagreements with other studies or reviews

The overall outcomes reported here are somewhat similar to those described in the reviews of Bryant 2005 and Duan 2013, both of which compared conventional TSR to humeral hemiarthroplasty and performed a meta-analysis using the same four studies (two unpublished and one including rheumatoid arthritis). Both of these reviews concluded that function was superior following TSR compared to humeral hemiarthroplasty.

The previous version of this review included the two published papers focused purely on osteoarthritis (Singh 2010). That version also found that TSR may offer superior function to humeral hemiarthroplasty but, like this updated version of the review, identified that the supporting evidence is of low quality. Singh 2010 included seven randomised controlled trials (RCTs) overall for participants with OA (not RCTA). One of those seven studies was excluded from this version of the review (Kircher 2009; see Excluded studies). Although this updated version of the review has identified an additional 13 studies for inclusion, including nine studies in patients with OA, the new studies are heterogeneous and of low quality. Thus we were unable to draw any new firm conclusions based on these studies.

The remaining published systematic reviews are largely based on low-quality evidence from non-randomised studies and pooled estimates from single-arm studies. Radnay 2007 analysed 1952 patients from 23 studies (only one randomised) and was able to make stronger conclusions in favour of TSR compared to humeral hemiarthroplasty for the outcomes of pain, function, satisfaction, range of motion, and revision surgery. van den Bekerom 2013 performed a systematic review comparing long-term outcomes of TSR to humeral hemiarthroplasty based on 1958 participants

from 19 non-randomised studies. These researchers found that revision rates were higher after humeral hemiarthroplasty, but that complication rates may be higher after TSR.

A systematic review of 14 studies (both randomised and non-randomised) failed to determine that any one method of subscapularis management in shoulder replacement surgery was superior to another (Choate 2018). Papadonikolakis 2014 reviewed 43 studies (only one comparative) and found a higher revision rate following TSR using metal-backed glenoid components versus all-polyethylene components. Vavken 2013 analysed 1460 participants from eight comparative studies of pegged versus keeled glenoid components for TSR (including four RCTs) and reported no difference in rates of glenoid lucency. The main conclusion was a slightly lower revision rate in favour of pegged components, weighted by the results of one large non-randomised study.

Erickson 2016 included 3302 participants from 65 studies to determine the effects of humeral inclination on range of movement achieved after reverse total shoulder replacement (RTSR). These researchers found greater external rotation if a 135-degree inclination was used compared to a 155-degree inclination, but no other differences. The same group used an overlapping set of 2222 shoulders in 38 studies to conclude that there is a higher rate of scapular notching with the 155-degree prosthesis (Erickson 2015), which is supported by the RCT included in this review (Gobezie 2019).

AUTHORS' CONCLUSIONS

Implications for practice

Results from two studies suggest that TSR may provide better function at two years when compared to humeral hemiarthroplasty for glenohumeral osteoarthritis. However, no other important differences were found. This is unchanged from the previous version of this review, as no new high-quality randomised trials have been conducted. For all other comparisons, we found no other important differences because available evidence from randomised controlled trials is generally of low quality; thus this review cannot provide any new guidance and implications for practice. High-quality orthopaedic studies are needed to improve evidence and decision-making for shoulder replacement surgery in relation to shoulder osteoarthritis and rotator cuff tear arthropathy.

Implications for research

High-quality research is clearly needed to determine the benefits and risks of shoulder replacement surgery. Investigators, commissioners, and funders of research must align research questions much more closely to important areas of uncertainty and make comparisons that have the potential to lead to significant changes in practice and patient care. Investigators should engage with trial methodologists to develop high-quality multi-centre studies that are sufficiently robust to reliably answer the study question. The study of adverse events and revision risk remains difficult and not feasible within the context of a surgical RCT with short follow-up, and the numbers needed to identify less frequent events are prohibitive with normal trial designs. If our understanding of this is to improve, RCTs that nest longer-term follow up in prospectively collected national registries must be designed to better address these outcomes.

ACKNOWLEDGEMENTS

We thank Jordi Pardo and Eli Harriss for their advice on constructing the search strategy, and Renea Johnston and Sheila Cyril for their editorial support for the protocol and review. We acknowledge John Sperling, Rachelle Buchbinder, and Kelly McMaken for their work on material included from the original protocol and review, which this version updates. We thank Dr. Bob McCormack for providing additional study data. Thanks also to Mario Lenza, Ann Lyddiatt, and Juha Paloneva for their constructive peer review comments.

REFERENCES

References to studies included in this review

Boileau 2002 *{published data only}*

* Boileau P, Avidor C, Krishnan SG, Walch G, Kempf J-F, Molé D. Cemented polyethylene versus uncemented metal-backed glenoid components in total shoulder arthroplasty: a prospective, double-blind, randomized study. *Journal of Shoulder and Elbow Surgery* 2002;**11**(4):351-9. [DOI: 10.1067/mse.2002.125807; PUBMED: 12195253]

Edwards 2010 *{published data only}*

* Edwards TB, Labriola JE, Stanley RJ, O'Connor DP, Elkousy HA, Gartsman GM. Radiographic comparison of pegged and keeled glenoid components using modern cementing techniques: a prospective randomized study. *Journal of Shoulder and Elbow Surgery* 2010;**19**(2):251-7. [DOI: 10.1016/j.jse.2009.10.013; PUBMED: 20185072]

Kilian CM, Press CM, Smith KM, O'Connor DP, Morris BJ, Elkousy HA. Radiographic and clinical comparison of pegged and keeled glenoid components using modern cementing techniques: midterm results of a prospective randomized study. *Journal of Shoulder and Elbow Surgery* 2017;**26**:2078-85. [DOI: 10.1016/j.jse.2017.07.016; PUBMED: 28918112]

Edwards 2012 *{published data only}*

* Edwards TB, Trappey GJ, Riley C, O'Connor DP, Elkousy HA, Gartsman GM. Inferior tilt of the glenoid component does not decrease scapular notching in reverse shoulder arthroplasty: results of a prospective randomized study. *Journal of Shoulder and Elbow Surgery* 2012;**21**(5):641-6. [DOI: 10.1016/j.jse.2011.08.057; PUBMED: 22079769]

Gartsman 2000 *{published data only}*

* Gartsman GM, Roddey TS, Hammerman SM. Shoulder arthroplasty with or without resurfacing of the glenoid in patients who have osteoarthritis. *Journal of Bone Joint Surgery. American Volume* 2000;**82**(1):26-34. [PUBMED: 10653081]

Gartsman 2005 *{published data only}*

* Gartsman GM, Elkousy HA, Warnock KM, Edwards TB, O'Connor DP. Radiographic comparison of pegged and keeled glenoid components. *Journal of Shoulder and Elbow Surgery* 2005;**14**(3):252-7. [DOI: 10.1016/j.jse.2004.09.006; PUBMED: 15889022]

Gascoyne 2017 *{published data only}*

Gascoyne TC, McRae SMB, Parashin SL, Leiter JRS, Petrak MJ, Bohm ER, MacDonald PB. Radiostereometric analysis of keeled versus pegged glenoid components in total shoulder arthroplasty: a randomized feasibility study. *Canadian Journal of Surgery* 2017;**60**(4):273-9. [PUBMED: 28730988]

Gobezie 2019 *{published data only}*

Gobezie R, Shishani Y, Lederman ES, Denard PJ. Can a functional difference be detected in reverse arthroplasty with 135 degree vs. 155 degree prosthesis for the treatment of rotator cuff arthropathy: a prospective randomized study. Journal of Shoulder and Elbow Surgery 2019; Vol. In press. [DOI: 10.1016/j.jse.2018.11.064; PUBMED: 30773441]

Greiner 2015 *{published data only}*

* Greiner S, Schmidt C, Herrmann S, Pauly S, Perka C. Clinical performance of lateralized versus non-lateralized reverse shoulder arthroplasty: a prospective randomized study. *Journal of Shoulder and Elbow Surgery* 2015;**24**(9):1397-404. [DOI: 10.1016/j.jse.2015.05.041; PUBMED: 26163281]

Kwon 2019 *{published data only}*

Kwon YW, Zuckerman JD. Subscapularis-sparing total shoulder arthroplasty: a prospective, double-blinded, randomized clinical trial. *Orthopedics* 2019;**42**(1):e61-7. [doi: 10.3928/01477447-20181109-02; PUBMED: 30427055]

Lapner 2012 *{published data only}*

* Lapner PLC, Sabri E, Rakhra K, Bell K, Athwal GS. Comparison of lesser tuberosity osteotomy to subscapularis peel in shoulder arthroplasty: a randomized controlled trial. *Journal of Bone and Joint Surgery. American volume* 2012;**94**(24):2239-46. [DOI: 10.2106/JBJS.K.01365; PUBMED: 23318614]

Lapner PLC, Sabri E, Rakhra K, Bell K, Athwal GS. Healing rates and subscapularis fatty infiltration after lesser tuberosity osteotomy versus subscapularis peel for exposure during shoulder arthroplasty. *Journal of Shoulder and Elbow Surgery* 2013;**22**(3):396-402. [DOI: 10.1016/j.jse.2012.05.031; PUBMED: 22944077]

Levine 2019 *{published data only}*

* Levine WN, Munoz J, Hsu S, Byram IR, Bigliani LU, Ahmad CS, et al. Subscapularis tenotomy versus lesser tuberosity osteotomy during total shoulder arthroplasty for primary osteoarthritis: a prospective, randomized controlled trial. Journal of Shoulder and Elbow Surgery 2019; Vol. 28, issue 3:407-14. [DOI: 10.1016/j.jse.2018.11.057; PUBMED: 30771825]

Litchfield 2011 *{published data only}*

* Litchfield RB, McKee MD, Balyk R, Mandel S, Holtby R, Hollinshead R, et al. Cemented versus uncemented fixation of humeral components in total shoulder arthroplasty for osteoarthritis of the shoulder: a prospective, randomized, double-blind clinical trial - a JOINTs Canada Project. *Journal of Shoulder and Elbow Surgery* 2011;**20**(4):529-36. [DOI: 10.1016/j.jse.2011.01.041; PUBMED: 21570660]

Lo 2005 *{published data only}*

* Lo IKY, Litchfield RB, Griffin S, Faber K, Patterson SD, Kirkley A. Quality-of-life outcome following hemiarthroplasty or total shoulder arthroplasty in patients with steoarthritis. *Journal of Bone and Joint Surgery. American volume* 2005;**87**(10):2178-85. [DOI: 10.2106/JBJS.D.02198; PUBMED: 16203880]

Mechlenburg 2014 *{published data only}*

Mechlenburg I, Christiansen T, Stilling M, Klebe T. Comparison of two resurfacing prostheses for treatment of osteoarthritis of the shoulder. Preliminary results. *Osteoarthritis and Cartilage* 2010;**18**(Suppl 2):S138.

* Mechlenburg I, Klebe TM, Dossing KV, Amstrup A, Soballe K, Stilling M. Evaluation of periprosthetic bone mineral

density and postoperative migration of humeral head resurfacing implants: two-year results of a randomized controlled clinical trial. *Journal of Shoulder and Elbow Surgery* 2014;**23**(10):1427-36. [DOI: 10.1016/j.jse.2014.05.012; PUBMED: 25220196]

Nuttall 2007 *{published data only}*

* Nuttall D, Haines JF, Trail II. A study of the micromovement of pegged and keeled glenoid components compared using radiostereometric analysis. *Journal of Shoulder and Elbow Surgery* 2007;**16**(3 Suppl):S65-S70. [DOI: 10.1016/j.jse.2006.01.015; PUBMED: 17493557]

Poon 2014 *{published data only}*

* Poon PC, Chou J, Young SW, Astley T. A comparison of concentric and eccentric glenospheres in reverse shoulder arthroplasty: a randomized controlled trial. *Journal of Bone and Joint Surgery. American volume* 2014;**96**(16):e138. [DOI: 10.2106/JBJS.M.00941; PUBMED: 25143506]

Rahme 2009 *{published data only}*

* Rahme H, Mattsson P, Wikblad L, Nowak J, Larsson S. Stability of cemented in-line pegged glenoid compared with keeled glenoid components in total shoulder arthroplasty. *Journal of Bone and Joint Surgery. American volume* 2009;**91**(8):1965-72. [DOI: 10.2106/JBJS.H.00938; PUBMED: 19651956]

Rasmussen 2015 *{published data only}*

* Rasmussen JV, Olsen BS, Sorensen AK, Hrobjartsson A, Brorson S. Resurfacing hemiarthroplasty compared to stemmed hemiarthroplasty for glenohumeral osteoarthritis: a randomised clinical trial. *International Orthopaedics* 2015;**39**(2):263-9. [DOI: 10.1007/s00264-014-2505-9; PUBMED: 25159010]

Sandow 2013 *{published data only}*

* Sandow MJ, David H, Bentall SJ. Hemiarthroplasty vs total shoulder replacement for rotator cuff intact osteoarthritis: how do they fare after a decade?. *Journal of Shoulder and Elbow Surgery* 2013;**22**(7):877-85. [DOI: 10.1016/j.jse.2012.10.023; PUBMED: 23333174]

Uschok 2017 *{published data only}*

* Uschok S, Magosch P, Moe M, Lichtenberg S, Habermeyer P. Is the stemless humeral head replacement clinically and radiographically a secure equivalent to standard stem humeral head replacement in the long-term follow-up? A prospective randomized trial. *Journal of Shoulder and Elbow Surgery* 2017;**26**(2):225-32. [DOI: 10.1016/j.jse.2016.09.001; PUBMED: 27856267]

References to studies excluded from this review

Berth 2013 *{published data only}*

* Berth A, Pap G. Stemless shoulder prosthesis versus conventional anatomic shoulder prosthesis in patients with osteoarthritis: a comparison of the functional outcome after a minimum of two years follow-up. *Journal of Orthopaedics and Traumatology* 2013;**14**(1):31-7. [DOI: 10.1007/s10195-012-0216-9; PUBMED: 23138538]

Ding 2015 *{published data only}*

* Ding DY, Mahure SA, Akuoko JA, Zuckerman JD, Kwon YW. Total shoulder arthroplasty using a subscapularis-sparing approach: a radiographic analysis. *Journal of Shoulder and Elbow Surgery* 2015;**24**(6):831-7. [DOI: 10.1016/j.jse.2015.03.009; PUBMED: 25979552]

Edwards 2007 *{published data only}*

* Edwards TB, Sabonghy EP, Elkousy H, Warnock KM, Hammerman SM, O'Connor DP, et al. Glenoid component insertion in total shoulder arthroplasty: comparison of three techniques for drying the glenoid before cementation. *Journal of Shoulder and Elbow Surgery* 2007;**16**(3 Suppl):S107-10. [DOI: 10.1016/j.jse.2006.04.006; PUBMED: 17055302]

Hammond 2013 *{published data only}*

* Hammond J, Lin EC, Harwood DP, Juhan TW, Gochanour E, Klosterman EL, et al. Clinical outcomes of hemiarthroplasty and biological resurfacing in patients aged younger than 50 years. *Journal of Shoulder and Elbow Surgery* 2013;**22**(10):1345-51. [DOI: 10.1016/j.jse.2013.04.015; PUBMED: 23796385]

Hendel 2012 *{published data only}*

* Hendel MD, Bryan JA, Barsoum WK, Rodriguez EJ, Brems JJ, Evans PJ, et al. Comparison of patient-specific instruments with standard surgical instruments in determining glenoid component position: a randomized prospective clinical trial. *Journal of Bone and Joint Surgery. American volume* 2012;**94A**(23):2167-75. [DOI: 10.2106/JBJS.K.01209; PUBMED: 23224387]

Iannotti 2015 *{published data only}*

* Iannotti JP, Weiner S, Rodriguez E, Subhas N, Patterson TE, Jun BJ, et al. Three-dimensional imaging and templating improve glenoid implant positioning. *Journal of Bone and Joint Surgery. American volume* 2015;**97**(8):651-8. [DOI: 10.2106/JBJS.N.00493; PUBMED: 25878309]

ISRCTN42881741 *{published data only}*

* de Groot JH. Conform and non conform glenoid components in total shoulder replacements. WHO International Clinical Trials Registry Platform.

Kasten 2009 *{published data only}*

* Kasten P, Maier M, Rettig O, Raiss P, Wolf S, Loew M. Proprioception in total, hemi- and reverse shoulder arthroplasty in 3D motion analyses: a prospective study. *International Orthopaedics* 2009;**33**(6):1641-7. [DOI: 10.1007/s00264-008-0666-0; PUBMED: 18956186]

Kircher 2009 *{published data only}*

* Kircher J, Wiedemann M, Magosch P, Lichtenberg S, Habermeyer P. Improved accuracy of glenoid positioning in total shoulder arthroplasty with intraoperative navigation: a prospective-randomized clinical study. *Journal of Shoulder and Elbow Surgery* 2009;**18**(4):515-20. [DOI: 10.1016/j.jse.2009.03.014; PUBMED: 19559369]

Mariotti 2014 *{published data only}*

* Mariotti U, Motta P, Stucchi A, Ponti di Sant'Angelo F. Stemmed versus stemless total shoulder arthroplasty: a

preliminary report and short-term results. *Musculoskeletal Surgery* 2014;**98**(3):195-200. [DOI: 10.1007/s12306-014-0312-5; PUBMED: 24469705]

NCT01884077 *{published data only}*

* Chamberlain A. Total vs reverse shoulder replacement: a prospective randomized trial. clinicaltrials.gov Accessed 31 Jan 2019. [NCT01884077]

References to ongoing studies

NCT01288066 *{published data only}*

* Suedkamp N. A randomized multicenter study comparing the effectiveness of hemi- versus total shoulder arthroplasty in patients with a degenerative joint disease. clinicaltrials.gov Accessed 31 Jan 2019. [NCT01288066]

NCT01404143 *{published data only}*

* Lapner P. Comparison of two methods of subscapularis management in shoulder arthroplasty: tenotomy versus peel: a multicenter, randomized controlled trial. clinicaltrials.gov Accessed 31 Jan 2019. [NCT01404143]

NCT01587560 *{published data only}*

* Skoldenberg O. A comparison between a pyrocarbon and a CoCr shoulder resurfacing implant. clinicaltrials.gov Accessed 31 Jan 2019. [NCT01587560]

NCT01697865 *{published data only}*

* Hamid N. Reverse shoulder arthroplasty with or without concomitant latissimus and teres major transfer for shoulder pseudoparalysis with teres minor dysfunction. clinicaltrials.gov Accessed 31 Jan 2019. [NCT01697865]

NCT01790113 *{published data only}*

* Hirschberg M. A prospective, randomized, multicenter study comparing the safety and effectiveness of Arthrex's Eclipse™ Shoulder Prothesis to the Univers™ II Shoulder Prosthesis in patients with a degenerative joint disease. clinicaltrials.gov Accessed 31 Jan 2019. [NCT01790113]

NCT02305966 *{published data only}*

* Athwal G. Evaluation of implant fixation in reverse total shoulder arthroplasty. clinicaltrials.gov Accessed 31 Jan 2019. [NCT02305966]

NCT02768597 *{published data only}*

* Morrey M. A prospective, randomized study comparing the outcome of large-diameter vs small-diameter glenospheres in primary reverse shoulder arthroplasty using the ReUnion system. clinicaltrials.gov Accessed 31 Jan 2019. [NCT02768597]

NCT02966886 *{published data only}*

* Lapner P. Comparison of techniques in the management of glenoid deficiencies in shoulder arthroplasty: protocol. clinicaltrials.gov Accessed 31 Jan 2019. [NCT02966886]

NCT03111147 *{published data only}*

* Wiater JM. The impact of humeral component version on outcomes following reverse total shoulder arthroplasty: a prospective randomized controlled trial. clinicaltrials.gov Accessed 31 Jan 2019. [NCT03111147]

NCT03711175 *{published data only}*

* Holt L. The Influence of Repairing the Sub-scapularis on Outcomes After Reverse Arthroplasty. clinicaltrials.org Accessed 31 Jan 2019. [NCT03711175]

NCT03727490 *{published data only}*

* Aleem A. The role of subscapularis repair in reverse shoulder arthroplasty. clinicaltrials.gov Accessed 31 Jan 2019. [NCT03727490]

NCT03730597 *{published data only}*

* Torrens C. Influence of glenosphere size in scapular notch development. Randomized study comparing 42 to 38 ECC. clinicaltrials.gov Accessed 31 Jan 2019. [NCT03730597]

Additional references

ACR 2000

American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. *Arthritis and Rheumatism* 2000;**43**(9):1905-15.

Bergenudd 1988

Bergenudd H, Lindgarde F, Nilsson B, Petersson CJ. Shoulder pain in middle age. A study of prevalence and relation to occupational work load and psychosocial factors. *Clinical Orthopaedics and Related Research* 1988;**231**:234-8.

Blaine 2008

Blaine T, Moskowitz R, Udell J, Skyhar M, Levin R, Friedlander J, et al. Treatment of persistent shoulder pain with sodium hyaluronate: a randomized, controlled trial. A multicenter study. *Journal of Bone and Joint Surgery. American volume* 2008;**90**(5):970-9.

Breivik 2006

Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *European Journal of Pain* 2006;**10**(4):287-333.

Bryant 2005

Bryant D, Litchfield R, Sandow M, Gartsman GM, Guyatt G, Kirkley A. A comparison of pain, strength, range of motion, and functional outcomes after hemiarthroplasty and total shoulder arthroplasty in patients with osteoarthritis of the shoulder. *Journal of Bone and Joint Surgery. American volume* 2005;**87**(9):1947-56. [DOI: 10.2106/JBJS.D.02854; PUBMED: 16140808]

Buchbinder 2003

Buchbinder R, Green S, Youd JM. Corticosteroid injections for shoulder pain. *Cochrane Database of Systematic Reviews* 2003, Issue 1. [DOI: 10.1002/14651858.CD004016]

Buchbinder 2017

Buchbinder R, Page MJ, Huang H, Verhagen AP, Beaton D, Kopkow C, et al. A preliminary core domain set for clinical trials of shoulder disorders: a report from the OMERACT 2016 Shoulder Core Outcome Set Special Interest Group. *Journal of Rheumatology* 2017;**[Epub ahead of print]**:jrheum.161123.

Cates 2008 [Computer program]

Dr Christopher Cates EBM website URL: www.nntonline.net. Visual Rx. Version 3. Dr Christopher Cates EBM website URL: www.nntonline.net, 2008.

Chakravarty 1990

Chakravarty KK, Webley M. Disorders of the shoulder: an often unrecognised cause of disability in elderly people. *BMJ* 1990;**300**(6728):848-9.

Chalmers 2019

Chalmers PN, Kahn T, Broschinsky K, Ross H, Stertz I, Nelson R, et al. An analysis of costs associated with shoulder arthroplasty. *Journal of Shoulder and Elbow Surgery* 2019;**28**(7):1334-40. [DOI: 10.1016/j.jse.2018.11.065; PUBMED: 30827836]

Chard 1991

Chard MD, Hazleman R, Hazleman BL, King RH, Reiss BB. Shoulder disorders in the elderly: a community survey. *Arthritis and Rheumatism* 1991;**34**(6):766-9.

Choate 2018

Choate WS, Kwapisz A, Momaya AM, Hawkins RJ, Tokish JM. Outcomes for subscapularis management techniques in shoulder arthroplasty: a systematic review. *Journal of Shoulder and Elbow Surgery* 2018;**27**(2):363-70. [DOI: 10.1016/j.jse.2017.08.003; PUBMED: 29195900]

Covidence [Computer program]

Veritas Health Innovation. Available at www.covidence.org (accessed 18 July 2017). Covidence. Melbourne, Australia: Veritas Health Innovation. Available at www.covidence.org (accessed 18 July 2017).

Croft 1996

Croft P, Pope D, Silman A. The clinical course of shoulder pain: prospective cohort study in primary care. Primary Care Rheumatology Society Shoulder Study Group. *BMJ* 1996;**313**(7057):601-2.

Deeks 2019

Deeks JJ, Higgins JPT, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA editor(s). Cochrane Handbook for Systematic Reviews of Interventions. 6.0. Cochrane, July 2019.

Dillon 2017

Dillon MT, Chan PH, Inacio MCS, Singh A, Yian EH, Navarro RA. Yearly trends in elective shoulder arthroplasty, 2005 through 2013. *Arthritis Care and Research* 2017;**69**:1574–81. [DOI: 10.1002/acr.23167; PUBMED: 27992683]

Duan 2013

Duan X, Zhang W, Dong X, Liu M, Gao Y, Huang F, et al. Total shoulder arthroplasty versus hemiarthroplasty in patients with shoulder osteoarthritis: a meta-analysis of randomized controlled trials. *Semianrs in Arthritis and Rheumatism* 2013;**43**:297-302. [DOI: 10.1016/j.semarthrit.2013.04.002; PUBMED: 23735559]

Erickson 2015

Erickson BJ, Frank RM, Harris JD, Mall N, Romeo AA. The influence of humeral head inclination in reverse total shoulder arthroplasty: a systematic review.. *Journal of Shoulder and Elbow Surgery* 2015;**24**(6):988-93. [DOI: 10.1016/j.jse.2015.01.001; PUBMED: 25725965]

Erickson 2016

Erickson BJ, Harris JD, Romeo AA. The effect of humeral inclination on range of motion in reverse total shoulder arthroplasty: a systematic review. *American Journal of Orthopedics* 2016;**45**(4):E174-9. [PUBMED: 27327922]

Fehringer 2002

Fehringer EV, Kopjar B, Boorman RS, Churchill RS, Smith KL, Matsen FA 3rd. Characterizing the functional improvement after total shoulder arthroplasty for osteoarthritis. *Journal of Bone and Joint Surgery. American volume* 2002;**84-A**(8):1349-53.

Franklin 1988

Franklin JL, Barrett WP, Jackins SE, Matsen FA III. Glenoid loosening in total shoulder arthroplasty. Association with rotator cuff deficiency. *Journal of Arthroplasty* 1988;**3**:39-46.

Goutallier 1994

Goutallier D, Postel J, Berganeau J, Lavau L, Voisin MC. Fatty muscle degeneration in cuff ruptures. Pre- and postoperative evaluation by CT scan. *Clinical Orthopaedics and Related Research* 1994;**304**:78-83. [PUBMED: 8020238]

GRADEpro GDT 2015 [Computer program]

McMaster University (developed by Evidence Prime). GRADEpro GDT. Version accessed 18 July 2017. Hamilton (ON): McMaster University (developed by Evidence Prime), 2015.

Grammont 1993

Grammont PM, Baulot E. Delta shoulder prosthesis for rotator cuff rupture. *Orthopedics* 1993;**16**:65-8.

Green 2005

Green S, Buchbinder R, Hetrick SE. Acupuncture for shoulder pain. *Cochrane Database of Systematic Reviews* 2005, Issue 2. [DOI: 10.1002/14651858.CD005319]

Hamada 1990

Hamada K, Fukuda H, Mikasa M, Kobayashi Y. Roentgenographic findings in massive rotator cuff tears. A long-term observation. *Clinical Orthopaedics and Related Research* 1990;**254**:92-6. [PUBMED: 2323152]

Hao 2019

Hao Q, Devji T, Zeraatkar D, Wang Y, Qasim A, Siemeiniuk RA, et al. Minimal important differences for improvement in

shoulder condition patient-reported outcomes: a systematic review to inform a BMJ Rapid Recommendation. *BMJ Open* 2019;**9**:e028777. [doi: 10.1136/bmjopen-2018-028777; PUBMED: 30787096]

Higgins 2019a

Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA editor(s). Cochrane Handbook for Systematic Reviews of Interventions. 6.0. Cochrane, July 2019.

Higgins 2019b

Higgins JPT, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA editor(s). Cochrane Handbook for Systematic Reviews of Interventions. 6.0. Cochrane, July 2019.

JLA 2015

Rangan A, Upadhyaya S, Regan S, Toye F, Rees J. Research priorities for shoulder surgery: results of the 2015 James Lind Alliance patient and clinician priority setting partnership. *BMJ Open* 2016;**6**(4):e010412.

Kilian 2017

Kilian CM, Press CM, Smith KM, O'Connor DP, Morris BJ, Elkousy HA. Radiographic and clinical comparison of pegged and keeled glenoid components using modern cementing techniques: midterm results of a prospective randomized study. *Journal of Shoulder and Elbow Surgery* 2017;**26**:2078-85. [DOI: 10.1016/j.jse.2017.07.016; PUBMED: 28918112]

Kim 2011

Kim SH, Wise BL, Zhang Y, Szabo RM. Increasing incidence of shoulder arthroplasty in the United States. *Journal of Bone and Joint Surgery. American volume* 2011;**93**(24):2249-54. [DOI: 10.2106/JBJS.J.01994; PUBMED: 22258770]

Lazarus 2002

Lazarus MD, Jensen KL, Southworth C, Matsen FA III. The radiographic evaluation of keeled and pegged glenoid component insertion. *Journal of Bone and Joint Surgery. American volume* 2002;**84A**:1174-82.

Lübbeke 2017

Lübbeke A, Rees JL, Barea C, Combescure C, Carr AJ, Silman AJ. International variation in shoulder arthroplasty. *Acta Orthopaedica* 2017;**88**(6):592-9. [DOI: 10.1080/17453674.2017.1368884; PUBMED: 28880117]

Meislin 2005

Meislin RJ, Sperling JW, Stitik TP. Persistent shoulder pain: epidemiology, pathophysiology, and diagnosis. *American Journal of Orthopedics* 2005;**34**(12 Suppl):5-9.

Muh 2013

Muh SJ, Streit JJ, Wanner JP, Lenarz CJ, Shishani Y, Rowland DY, et al. Early follow-up of reverse total shoulder arthroplasty in patients sixty years of age or younger. *Journal of Bone and Joint Surgery. American volume* 2013;**95**(20):1877-83.

National Center for Health Statistics 2011

Bernstein AB, Makuc DM, Bilheimer LT. Health, United States, 2010: with special feature on death and dying. Hyattsville, MD: National Center for Health Statistics, 2011.

Neer 1983

Neer CS, Craig EV, Fukuda H. Cuff-tear arthropathy. *Journal of Bone and Joint Surgery. American volume* 1983;**65**(9):1232-44.

Page 2015

Page MJ, McKenzie JE, Green SE, Beaton DE, Jain NB, Lenza M, et al. Core domain and outcome measurement sets for shoulder pain trials are needed: systematic review of physical therapy trials. *Journal of Clinical Epidemiology* 2015;**68**:1270-81. [DOI: 10.1016/j.jclinepi.2015.06.006]

Page 2016a

Page MJ, Green S, McBain B, Surace SJ, Deitch J, Lyttle N, et al. Manual therapy and exercise for rotator cuff disease. *Cochrane Database of Systematic Reviews* 2016, Issue 6. [DOI: 10.1002/14651858.CD012224]

Page 2016b

Page MJ, Green S, Mrocki MA, Surace SJ, Deitch J, McBain B, et al. Electrotherapy modalities for rotator cuff disease. *Cochrane Database of Systematic Reviews* 2016, Issue 6. [DOI: 10.1002/14651858.CD012225]

Papadonikolakis 2014

Papadonikolakis A, Matsen FA. Metal-backed glenoid components have a higher rate of failure and fail by different modes in comparison with all-polyethylene components. *Journal of Bone and Joint Surgery. American volume* 2014;**96**:1041-7. [DOI: 10.2106/JBJS.M.00674; PUBMED: 24951741]

Polk 2013

Polk A, Rasmussen JV, Brorson S, Olsen BS. Reliability of patient-reported functional outcome in a joint replacement registry. *Acta Orthopaedica* 2013;**84**(1):12-7. [DOI: 10.3109/17453674.2013.765622; PUBMED: 23343374]

Pope 1997

Pope DP, Croft PR, Pritchard CM, Silman AJ, Macfarlane GJ. Occupational factors related to shoulder pain and disability. *Occupational and Environmental Medicine* 1997;**54**(5):316-21.

PRISMA Group 2009

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;**339**:b2535. [DOI: 10.1136/bmj.b2535]

Radnay 2007

Radnay CS, Setter KJ, Chambers L, Levine WN, Bigliani LU, Ahmad CS. Total shoulder replacement compared with humeral head replacement for the treatment of primary glenohumeral osteoarthritis: a systematic review. *Journal of Shoulder and Elbow Surgery* 2007;**16**(4):396-402. [DOI: 10.1016/j.jse.2006.10.017; PUBMED: 17582789]

Rangan 2016

Rangan A, Upadhaya S, Regan S, Toye F, Rees JL. Research priorities for shoulder surgery: results of the 2015 James Lind Alliance patient and clinician priority setting partnership. *BMJ Open* 2016;**6**:e010412. [DOI: 10.1136/bmjopen-2015-010412]

RevMan 2014 [Computer program]

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Schünemann 2019a

Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH. Chapter 15: Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA editor(s). Cochrane Handbook for Systematic Reviews of Interventions. 6.0. Cochrane, July 2019.

Schünemann 2019b

Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, Guyatt GH. Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA editor(s). Cochrane Handbook for Systematic Reviews of Interventions. 6.0. Cochrane, July 2019.

Shamoon 2000

Shamoon M, Hochberg MC. Treatment of osteoarthritis with acetaminophen: efficacy, safety, and comparison with nonsteroidal anti-inflammatory drugs. *Current Rheumatology Reports* 2000;**2**(6):454-8.

Simone 2014

Simone JP, Streubel PH, Sperlin JW, Schleck CD, Cofield RH, Athwal GS. Anatomica total shoulder replacement with rotator cuff repair for osteoarthritis of the shoulder. *Bone and Joint Journal* 2014;**96-B**(2):244-8.

Valenti 2001

Valenti P, Boutens D, Nerot C. Delta 3 reversed prosthesis for osteoarthritis with massive rotator cuff tear: long term results (>5 years). In: Walch G, Boileau P, Molé D editor(s). 2000 Prothèses d'Epaule: Recul de 2 à 10 Ans. Montpellier, France: Sauramps Médical, 2001. [ISBN-10:2840232731]

van den Bekerom 2013

van den Bekerom MP, Geervliet PC, Somford MP, van den Borne MP, Boer R. Total shoulder arthroplasty versus hemiarthroplasty for glenohumeral arthritis: a systematic review of the literature at long-term follow-up. *International Journal of Shoulder Surgery* 2013;**7**(3):110-5. [DOI: 10.4103/0973-6042.118915; PUBMED: 24167403]

van der Windt 2003

van der Windt DA, Bouter LM. Physiotherapy or corticosteroid injection for shoulder pain?. *Annals of the Rheumatic Diseases* 2003;**62**(5):385-7.

Vavken 2013

Vavken P, Sadoghi P, von Keudell A, Rosso C, Valderrabano V, Müller AM. Rates of radiolucency and loosening after total shoulder arthroplasty with pegged or keeled glenoid components. *Journal of Bone and Joint Surgery. American volume* 2013;**95**(3):215-21. [DOI: 10.2106/JBJS.L.00286; PUBMED: 23389784]

Walch 1999

Walch G, Badet R, Boulahia A, Khoury A. Morphological study of the glenoid in primary glenohumeral osteoarthritis. *Journal of Arthroplasty* 1999;**14**(6):756-60. [PUBMED: 10512449]

Walch 2005

Walch G, Edwards TB, Boulahia A, Nové-Josserand L, Neyton L, Szabo I. Arthroscopic tenotomy of the long head of the biceps in the treatment of rotator cuff tears: clinical and radiographic results of 307 cases. *Journal of Shoulder and Elbow Surgery* 2005;**3**:238-46.

Walker 2011

Walker M, Brooks J, Willis M, Frankle M. How reverse shoulder arthroplasty works. *Clinical Orthopaedics and Related Research* 2011;**469**(9):1440-51. [PUBMED: 21484471]

Werner 2016

Werner BC, Chang B, Nguyen JT, Dines DM, Gulotta LV. What change in American shoulder and elbow surgeons score represents a clinically important change after shoulder arthroplasty?. *Clinical Orthopaedics and Related Research* 2016;**474**(12):2672-81. [DOI: 10.1007/s11999-016-4968-z; PUBMED: 27392769]

Wofford 1997

Wofford JL, Mansfield RJ, Watkins RS. Patient characteristics and clinical management of patients with shoulder pain in U.S. primary care settings: secondary data analysis of the National Ambulatory Medical Care Survey. *BMC Musculoskeletal Disorders (available from www.biomedcentral.com/1471-2474/6/4)* 2005;**6**:4. [DOI: 10.1186/1471-2474-6-4]

Zhang 2019

Zhang B, Thayaparan A, Horner N, Bedi A, Alolabi B, Khan M. Outcomes of hyaluronic acid injections for glenohumeral osteoarthritis: a systematic review and meta-analysis. *Journal of Shoulder and Elbow Surgery* 2019;**28**(3):596-606. [DOI: 10.1016/j.jse.2018.09.011; PUBMED: 30502030]

References to other published versions of this review

Singh 2010

Singh JA, Sperling J, Buchbinder R, McMaken K. Surgery for shoulder osteoarthritis. *Cochrane Database of Systematic Reviews* 2010, Issue 10. [DOI: 10.1002/14651858.CD008089.pub2]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Boileau 2002

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: September 1995 to September 1996 Setting: single centre, France Length of follow-up: minimum 36 months (mean 38 months)

Participants	Number randomised: 39 patients, 40 shoulders
	
	Number analysed: 35 shoulders (at 3 years)
	
	Number lost to follow-up: 5 shoulders
	
	Baseline characteristics
	
	• Age (mean/range): cemented 69 years (59 to 77); cementless 68 years (55 to 85)
	• Sex (male:female): cemented 3:17; cementless 7:13
	• Constant-Murley Score (mean/range): cemented 25.2 (6 to 41); cementless 27.5 (6 to 51)
	• Glenoid morphology (graded as per Walch 1999)
	  ○ Cemented: 7× A1, 3× A2, 5× B1, 2× B2, 1× C
	  ○ Cementless: 10× A1, 1× A2, 6× B1, 2× B2, 1× C
	• Partial cuff tears (counts)
	  ○ Cemented: 3 supraspinatus and 1 infraspinatus
	  ○ Cementless: 1 supraspinatus
	• Fatty degeneration of cuff (graded as per Goutallier 1994)
	  ○ Cemented: supraspinatus - 4× grade 0, 10× grade 1, 4× grade 2, 2× grade 3; subscapularis 5× grade 0, 10× grade 1, 5× grade 2
	  ○ Cementless: supraspinatus - 6× grade 0, 9× grade 1, 4× grade 2, 1× grade 3; subscapularis 7× grade 0, 7× grade 1, 6× grade 2
	
	Inclusion criteria
	
	• Patients scheduled to undergo TSR for primary glenohumeral OA with intact rotator cuff (fatty degeneration permitted)
	• Disabling pain and poor function
	• Failure of non-operative conservative management for longer than 6 months
	
	Exclusion criteria
	
	• Another shoulder disease (inflammatory arthritis, AVN, cuff tear arthritis, fracture sequelae)
	• Previous shoulder surgery, evidence of infection or neurological disease
	• Could not be followed up for at least 3 years after TSR
	• Severe glenoid deficiency that would need bone graft

Interventions	Conventional total shoulder replacement with cemented humeral stem (Aequalis; Tornier Inc., St. Ismier, France) in 2 arms: • Cemented all-polyethylene keeled glenoid component (n = 20) • Cementless metal-backed glenoid component with polyethylene bearing (n = 20)
Outcomes	Outcomes reported at 1, 2, and 3 years • Pain: subdomain of the Constant-Murley Score (0/5/10/15 scale, higher = better, reported as "average" and range) - included in Table 1 but not suitable for meta-analysis

Boileau 2002 *(Continued)*

- Function
  - Constant Murley Score (0 to 100 scale, higher = better, reported as "average" and range) - included in Table 1 but not suitable for meta-analysis
- Revision, re-operation, or treatment failure
  - Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
  - Study includes 1 patient scheduled for revision, hence identified for this review as a treatment failure
- Adverse events
  - Specific: all recorded as revisions/failures; therefore not double-counted by the review as separate adverse events
- Physician-evaluated
  - Radiographic evidence of glenoid loosening (0 to 4, defined as "(1) a complete radiolucent line that extended the length of the bone-cement interface greater than 2 mm in width, (2) a tilt or migration of the glenoid component, (3) a fracture of the cement or a breakage of the metallic petals, and/or (4) a revision to change or remove the glenoid component")
  - This endpoint is not reported with sufficient clarity to be analysed formally in this review but is recorded in Table 1
  - Range of motion: measured as subdomain of Constant Murley Score (not included in this review)
  - Strength: measured as subdomain of Constant Murley Score (not included in this review)
  - Activity: measured as subdomain of Constant Murley Score (not included in this review)

Notes	Source of funding: not reported Conflicts of interest: not reported Continuous measures are not suitable for meta-analysis due to inexplicit reporting

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "the choice of glenoid component (cemented polyethylene or uncemented metal-backed) was decided by use of a random table after humeral preparation at the time of glenoid preparation"
Allocation concealment (selection bias)	Unclear risk	Quote: "both patients and surgeons were blinded until glenoid preparation with regard to which prosthesis would be implanted" Comment: method of allocation concealment not reported
Blinding of participants and personnel (performance bias)	Unclear risk	It is unclear when/if patients were unblinded to their allocation. The surgeon could not be blinded. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Unclear risk	Quote: "both patients and surgeons were blinded until glenoid preparation with regard to which prosthesis would be implanted" Comment: it is unclear when/if patients were unblinded to their allocation
Blinding of outcome assessment - physician reported outcomes	High risk	Personnel were unblinded after sequence allocation; therefore there is potential for bias in assessor-measured functional domains within the Constant Score Radiographic assessors could not be blinded to the type of implant used
Incomplete outcome data (attrition bias)	Unclear risk	Quote: "the 3 cases of metal-backed glenoid loosening were revised between the third and fifth postoperative year" Comment: outcome tables report the number at risk for years 1 to 3; however completeness of follow-up is not reported beyond 3 years. Therefore the num-

Boileau 2002 *(Continued)*

		ber at risk of revision is not clear. Additionally, withdrawals due to death were not reported by implant type
Selective reporting (reporting bias)	Unclear risk	Quote: "all patients were prospectively followed up and underwent radiography at regular intervals after surgery at 3, 6, 12, 24, 36, and 48 months" Comment: outcome scores are reported only for 1, 2, and 3 years. Although the mean follow-up was only 38 months, no data for the 4-year endpoint were provided (including completeness of follow-up)
Major baseline imbalance	Unclear risk	Quote: "the 2 groups were statistically comparable with regard to age, sex, hand dominance, and pre-operative functional score" Comment: there is an imbalance in sex distribution between the 2 groups (15% male in cemented group and 35% male in uncemented group). Although not reaching statistical significance, the possible effect on fixation outcomes and strength is unclear
Differences in rehabilitation regime	Low risk	Quote: "a standard physiotherapy regimen was used. Passive motion was started the day after surgery. The arm was placed in a sling for 4 to 6 weeks to aid healing of the repaired subscapularis. Strengthening and stretching exercises were then progressively added"

Edwards 2010

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: December 2004 to December 2005 Setting: single centre, United States of America Length of follow-up: 2 years (primary radiographic endpoint) and 5 years

Participants	This study was updated by a second publication (Killian 2017), which included additional procedures on the contralateral shoulders of previously recruited patients. The flow of patients is unclear and demographics are reported only for those procedures analysed in each paper, hence patient groups at intermediate- and long-term follow-up are reported separately below
	
	Number randomised: 50 patients, 59 shoulders. 29 received pegged glenoid, 30 keeled glenoid (totals as reported by second study paper, Kilian 2017)
	
	Number analysed
	
	• Intermediate-term (mean 26 months): 47 shoulders (21 pegged, 26 keeled)
	• Long-term (minimum 5 years): 38 shoulders (16 pegged, 22 keeled)
	
	Number lost to follow-up
	
	• Intermediate-term: 2 died, 4 did not attend, 2 cases in keeled revised before clinical and radiographic review (included in demographic tables; included in this review in revision analyses)
	• Long-term: 10 died, 8 did not attend, 3 cases revised before clinical review (excluded from demographic tables; included in this review in revision analyses)
	
	Baseline characteristics
	
	• Age (mean ± SD)
	  ◦ Intermediate-term: pegged 71.8 ± 10.4 years, keeled 66.3 ± 11.6 years
	  ◦ Long-term: pegged 68.0 ± 10.8 years, keeled 68.0 ± 12.2 years
	• Sex (male:female)

Edwards 2010 *(Continued)*

	○ Intermediate-term: pegged 13:8, keeled 12:14 ○ Long-term: pegged 12:4, keeled 7:15 Inclusion criteria • Patients undergoing total shoulder replacement with a diagnosis of primary glenohumeral osteoarthritis • Glenoid that did not require bone grafting Exclusion criteria • Patients with an inflammatory arthropathy in the shoulder (rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis), osteochondromatosis, acromegaly, Paget's disease, postinfectious arthropathy, skeletal dysplasia, neurological problems (Charcot arthropathy, Parkinson's disease), osteonecrosis • History of shoulder trauma (fracture or soft tissue injury), instability (surgically or non-surgically treated), prior shoulder surgery • Marked rotator cuff disorders of the shoulder, as indicated by acromiohumeral arthritis, a massive rotator cuff tear, or a rotator cuff tear involving the infraspinatus or subscapularis
Interventions	Conventional total shoulder replacement with uncemented stem and cemented glenoid (Aequalis; Tornier, Mont Bonnot, France) in 2 arms • Pegged glenoid component (n = 29) • Keeled glenoid component (n = 30)
Outcomes	Outcomes reported at 2 and 5 years, as above • Function ○ WOOS Index (0 to 100 scale, higher = better, reported as mean and P value) ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean and P value) ○ Constant Murley Score (0 to 100 scale, higher = better, reported as mean and P value) ○ Single-Assessment Numerical Evaluation Score (0 to 100, higher = better, reported as mean and P value) • Revision, re-operation, or treatment failure: ○ Counts extracted and reported in this review as cumulative totals up to and including the relevant time point • Adverse events - specific ○ Counts ○ Reported events included posterior dislocation and component fracture • Physician evaluated ○ Glenoid lucency graded as follows ■ Pegged (as per Lazarus 2002): (1) incomplete radiolucency around 1 or 2 pegs, (2) complete radiolucency (2 mm wide) around 1 peg only, with or without incomplete radiolucency around 1 other peg, (3) complete radiolucency (2 mm wide) around 2 or more pegs, (4) complete radiolucency (2 mm wide) around 2 or more pegs, (5) gross loosening ■ Keeled (as per Franklin 1988): (1) radiolucency at inferior and/or superior flange, (2) incomplete radiolucency at keel, (3) complete radiolucency (2 mm wide) around keel, (4) complete radiolucency (2 mm wide) around keel, (5) gross loosening ○ Range of motion not included in this review
Notes	Source of funding: no outside grants received Conflicts of interest: 2 trial authors disclosed royalties/consulting fees and research support from Tornier. One trial author disclosed royalties/consulting fees from Elsevier
Risk of bias	

Edwards 2010 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "a simple randomization technique using a random numbers table with glenoid component type placed in sealed envelopes was employed"
Allocation concealment (selection bias)	Low risk	Quote: "the design of the glenoid component, pegged versus keeled, was determined by opening a randomly selected envelope immediately preoperatively without any specific indication"
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "the patients were not told which glenoid component design they had received" Comment: the surgeon could not be blinded. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "the patients were not told which glenoid component design they had received"
Blinding of outcome assessment - physician reported outcomes	High risk	Comment: it would not be possible to blind the assessors to radiographic outcomes
Incomplete outcome data (attrition bias)	High risk	The study was updated by a secondary publication with additional randomised procedures within the same study population. The flow of patients through the study is very unclear, particularly with regard to numbers of revised cases and numbers at risk at different time points
Selective reporting (reporting bias)	High risk	Planned endpoints are not stated, and patient-reported measures are available only at late follow-up in the secondary paper. Demographic tables include those revised in the first paper but excluded from the second paper
Major baseline imbalance	Low risk	No significant differences in reported demographics
Differences in rehabilitation regime	Low risk	Quote: "postoperatively, the patients were placed in a standard sling. After 1 week, aquatic therapy rehabilitation was initiated to begin shoulder range of motion in elevation, extension, horizontal adduction, internal rotation, and external rotation. External rotation was limited to neutral for 4 weeks. After at least 5 weeks of hydrotherapy, if acceptable range of motion was gained, patients graduated to a self-directed land based program. Strengthening exercises were not prescribed"

Edwards 2012

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: November 2005 to August 2008 Setting: single centre, United States of America Length of follow-up: 1 year
Participants	Number randomised: 52 Number analysed: 42 (22 in control group (neutral glenosphere) and 20 in intervention group (inferior tilted glenosphere) Number lost to follow-up: 10

Edwards 2012 *(Continued)*

	Baseline characteristics • Age (mean ± SD): neutral 66.3 ± 9.8 years, tilted 71.8 ± 8.0 years • Sex (male:female): neutral 9:13, tilted 10:10 • ASES Shoulder Score (mean ± SD): neutral 59.6 ± 5.5, tilted 56.3 ± 10.6 • Constant Murley Score (mean ± SD): neutral 15.7 ± 19.8, tilted 13.1 ± 9.2 • Age- and gender-adjusted Constant Murley Score (mean ± SD): neutral 21.2 ± 14.8, tilted 17.6 ± 11.9 Inclusion criteria • Diagnosis of rotator cuff tear arthropathy for which the patient elected to undergo reverse total shoulder replacement Exclusion criteria • Glenoid requiring grafting
Interventions	Reverse total shoulder replacement (Aequalis Reverse System; Tornier Inc., Edina, MN, USA) in 2 arms • Neutral glenosphere inclination (n = 26) • Inferior tilted glenosphere (10 degrees) (n = 26) Glenoid reaming orientation was controlled via the NaviPro shoulder computer navigation system to within 1.5 degrees. All procedures were completed by a single surgeon

Outcomes	Radiological outcomes reported at 1 year. Clinical outcomes assumed same time point, although this was not specified directly
	• Function
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)
	  ○ Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
	  ○ Age- and gender-adjusted Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
	• Adverse events - specific
	  ○ Counts
	  ○ Reported events included 1 dislocation
	• Revision, re-operation, or treatment failure
	  ○ None occurred during the study
	• Physician evaluated
	  ○ Radiographic evidence of scapular notching graded 0 to 4 according to the Nerot classification (Valenti 2001)
	  ○ Range of motion: measured in degrees (not included in this review)

Notes	Source of funding: not reported Conflicts of interest: 2 trial authors reported royalties for consulting for implant manufacturers including Tornier Inc.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "subjects were masked to group assignment and randomized in the preoperative area, using simple randomization and a random numbers table"
Allocation concealment (selection bias)	Unclear risk	Quote: "subjects were masked to group assignment and were randomized in the preoperative area, using simple randomization and a random numbers table"

Edwards 2012 *(Continued)*

		Comment: although participants were masked, it is not clear at what point the surgeon was unmasked in relation to the timing of randomisation
Blinding of participants and personnel (performance bias)	Unclear risk	It is unclear when/if patients were unblinded to allocation. The surgeon could not be blinded. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Unclear risk	It is not clear when/if patients were unblinded to allocation
Blinding of outcome assessment - physician reported outcomes	High risk	Two orthopaedic surgeons independently rated the postoperative radiographs. Neither rater was involved in the surgical procedure, but it would not be possible to blind them to the glenoid inclination
Incomplete outcome data (attrition bias)	Low risk	19% missing outcomes at 1 year with no imbalance between groups Neutral group: 26 randomised, 22 analysed Tilted group: 26 randomised, 20 analysed
Selective reporting (reporting bias)	Unclear risk	All outcomes were reported, but planned time points were unclear both in methods and results, and no protocol was published
Major baseline imbalance	Low risk	Quote: "no statistically significant differences in demographic data or length of follow-up existed between groups"
Differences in rehabilitation regime	Unclear risk	Not reported

Gartsman 2000

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: December 1992 to December 1996 Setting: single centre, United States of America Length of follow-up: mean 34 months (range 24 to 72 months)
Participants	Number randomised: 51 patients, 55 shoulders Number analysed: 47 patients, 51 shoulders Number lost to follow-up: 4 patients, 4 shoulders Baseline characteristics • Age (mean ± SD): TSR 63.5 ± 8.4, hemiarthroplasty 64.6 ± 6.3 • Sex (male:female): TSR 15:10, hemiarthroplasty 13:9 • ASES Shoulder Score (mean ± SD): TSR 22.7 ± 14.4, hemiarthroplasty 22.6 ± 15.1 • UCLA Shoulder Rating Scale (mean ± SD): TSR 8.1 ± 2.8, hemiarthroplasty 8.2 ± 3.5 Inclusion criteria • Diagnosis of osteoarthritis, an intact rotator cuff, and a concentric glenoid • Unresponsive to medical treatment and interfering with activities of daily living • Degeneration of articular cartilage confirmed intraoperatively Exclusion criteria

Gartsman 2000 *(Continued)*

- Diagnosis other than osteoarthritis
- Uneven bone loss

Interventions	Shoulder replacement with uncemented humeral stem (Global; DePuy, Warsaw, IN, USA) in 2 arms • Hemiarthroplasty (n = 24 - analysed) • Conventional total shoulder replacement (TSR) via cemented all-polyethylene keeled glenoid component (n = 27 - analysed) All procedures performed by or under the direct supervision of a single surgeon

Outcomes	Followed up at 2 weeks, 6 weeks, 3 months, 6 months, and 1 year after the operation and yearly thereafter. Timing of outcome reporting not clearly stated. Included as an intermediate time point in this review (1 to 3 years)
	• Pain: visual analogue scale reported as subdomain ASES Shoulder Score (50 to 0 scale translates to 0 to 10 scale, lower = better, reported as mean and exact P value)
	• Function
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)
	  ○ UCLA Shoulder Rating Scale (0 to 35 scale, higher = better, reported as mean ± SD)
	• Revision, re-operation, or treatment failure
	  ○ Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
	• Adverse events - specific
	  ○ Counts
	  ○ Types of events included severe stiffness, severe unexplained pain, and glenoid erosion

Notes	Source of funding: research fellowship provided by HCA/Columbia and Texas Orthopaedic Hospital Conflicts of interest: none reported Trial authors performed a sample size calculation indicating that 35 participants per group were required. They did not recruit this number

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "randomization was accomplished with use of a random-numbers list generated with Microsoft Excel (Redmond, Washington) and kept by the operating room nurse. After completion of the humeral osteotomy, anterior and inferior capsular release, mobilization of the subscapularis, and exposure of the glenoid, the circulating nurse reviewed the random-numbers list. Patients who had been assigned an odd number were subsequently managed with a hemiarthroplasty, and those who had been assigned an even number were managed with a total shoulder arthroplasty"
Allocation concealment (selection bias)	Unclear risk	Allocation concealment method was not stated
Blinding of participants and personnel (performance bias)	Unclear risk	Blinding of personnel and patients was not stated
Blinding of outcome assessment - self-reported outcomes	Unclear risk	Blinding of patients was not stated

Gartsman 2000 *(Continued)*

Blinding of outcome assessment - physician reported outcomes	High risk	Quote 1: "patients' names and identification numbers were kept separately, so that the analysis of the data was performed without any knowledge of the patient's identity" Quote 2: "we recorded all measurements during the initial physical examination and during subsequent visits to the clinic. No attempt was made to increase the precision of the measurement with use of such techniques as blinding of the examiner" Comment: although blinding of the statistical analysis is an appropriate step, it does not appear that physician observers were blinded; therefore the level of bias may be high
Incomplete outcome data (attrition bias)	Unclear risk	4 shoulders in 4 patients (8%) were lost to follow-up. It is not clear which groups the patients were lost from. If there were significant unexplained imbalance here, this may be more important. However, the overall number remains low and it would still be unlikely to have any serious effect on the study results
Selective reporting (reporting bias)	Unclear risk	Quote: "the patients were evaluated at two weeks, six weeks, three months, six months, and one year after the operation and yearly thereafter. At each annual visit to the clinic, before the examination, the patients completed self-assessment forms to allow tabulation of the shoulder scores described earlier" Comment: only a single value is reported for each measure per patient. It is unclear at which time point results were extracted
Major baseline imbalance	Low risk	There were no apparent differences in the baseline data presented
Differences in rehabilitation regime	Low risk	Quote: "all patients were managed with the same postoperative regimen, including administration of antibiotics and physical therapy"

Gartsman 2005

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: September 2000 to October 2002 Setting: single centre, United States of America Length of follow-up: 6 weeks
Participants	Number randomised: 47 patients, 47 shoulders Number analysed: 43 patients, 43 shoulders Number lost to follow up: nil lost, 4 excluded (see below) Baseline characteristics • Age (mean ± SD): mean age of men 67.8 ± 8.1 years, mean age of women 69.9 ± 7.2 years • Sex (male:female): pegged glenoid 12:8, keeled glenoid 15:8 • Rotator cuff tear: pegged glenoid 5/20, keeled glenoid 1/23 Inclusion criteria • Primary osteoarthritis of the glenohumeral joint Exclusion criteria • Radiographic evidence of osteopenia or other metabolic bone disease

Gartsman 2005 *(Continued)*

	• Two patients in the keeled group and 2 in the pegged group were excluded because raters were unable to grade the postoperative radiographs because of poor image quality
Interventions	Total shoulder replacement with uncemented humeral stem and all-polyethylene cemented glenoid components (Cofield 2 Modular Arthroplasty; Smith and Nephew, Memphis, TN, USA) in 2 arms • Pegged glenoid component (n = 20 - analysed) • Keeled glenoid component (n = 23 - analysed)

Outcomes	Followed up at 6 weeks only
	• Physician evaluated: glenoid lucency graded as follows
	  ○ Pegged: (1) incomplete radiolucency around 1 or 2 pegs, (2) complete radiolucency (2 mm wide) around 1 peg only, with or without incomplete radiolucency around 1 other peg, (3) complete radiolucency (2 mm wide) around 2 or more pegs, (4) complete radiolucency (2 mm wide) around 2 or more pegs, (5) gross loosening (as per Lazarus 2002)
	  ○ Keeled: (1) radiolucency at inferior and/or superior flange, (2) incomplete radiolucency at keel, (3) complete radiolucency (2 mm wide) around keel, (4) complete radiolucency (2 mm wide) around keel, 5. Gross loosening (as per Franklin 1988)
	  ○ Included in the meta-analysis as dichotomous outcome - patients with grade ≥4 vs <4

Notes	Source of funding: not stated Conflicts of interest: not stated

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "the circulating nurse consulted a random number list to select the implant" Comment: the method used to generate and allocate the random number list is not stated
Allocation concealment (selection bias)	Unclear risk	Quote: "the circulating nurse consulted a random number list to select the implant" Comment: it is not clear how the random number list was stored and what measures where taken to prevent allocation disclosure
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "the patients were not told which glenoid component design they had received" Comment: the surgeon could not be blinded. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Low risk	Patients were blinded; however no patient-reported outcomes were recorded
Blinding of outcome assessment - physician reported outcomes	High risk	Quote: "the radiographs were evaluated by 3 raters (2 orthopaedic surgeons and 1 radiologist). The orthopaedic surgeon who had performed the arthroplasties was not a rater. These raters had participated in a training session in which they graded the radiographs of 30 patients with total shoulder arthroplasty who had not been enrolled in this study. The training session allowed the raters to confer during the grading to ensure that each rater was using consistent criteria"

Gartsman 2005 *(Continued)*

		Comment: it is not possible to blind radiographic outcome assessors to the type of glenoid component used
Incomplete outcome data (attrition bias)	Low risk	4/47 patients (8.5%) not analysed due to inadequate follow-up radiographs; balanced between the 2 groups
Selective reporting (reporting bias)	Low risk	Planned outcomes reported
Major baseline imbalance	Low risk	There was an imbalance between groups in the number of patients with a rotator cuff tear (pegged: 5/20, keeled: 1/23). This is a reported risk factor for early wear and loosening of glenoid components; however, it would not be expected to have any impact on early postoperative radiographic appearances. Trial authors performed a sensitivity analysis to confirm the absence of any effect
Differences in rehabilitation regime	Low risk	Quote: "postoperative care consisted of sling wear for 2 weeks. We instructed the patients in active range-of-motion exercises 2 weeks after surgery and began passive range-of-motion stretching exercises 6 weeks after surgery. At 12 weeks after surgery, patients were advised to resume normal activities gradually, using caution to avoid sudden large or painful joint movements"

Gascoyne 2017

Methods	Study design: randomised feasibility study in 2 parallel arms Recruitment: August 2008 to March 2011 Setting: single centre, Canada Length of follow-up: up to 24 months
Participants	Number randomised: 15 patients (16 shoulders) Number analysed: 11 shoulders at 12 months, 9 shoulders at 24 months Number lost to follow-up: 1 patient withdrew consent, 2 patients died, 2 did not attend, and 1 patient was assessed but the RSA data were "problematic" Baseline characteristics • Age (mean (range)): 64 (46 to 75) • Sex (male:female): 10:5 • WOOS Index (median): keeled 80.2, pegged 59.3 • ASES Shoulder Score (median): keeled 22.5, pegged 47.0 • Simple Shoulder Test (median): keeled 2.0, pegged 4.5 Inclusion criteria • Patients aged 18 years and older undergoing total shoulder replacement Exclusion criteria • Patients requiring a reverse total shoulder replacement • Presence of a rotator cuff tear or rotator cuff tear arthropathy • Active Workers' Compensation claim regarding the affected shoulder • Previous shoulder joint replacement surgery

Gascoyne 2017 *(Continued)*

Interventions	Conventional total shoulder replacement with a stemmed humeral component and a cemented all-polyethylene glenoid component in 2 arms (manufacturer/brand line not stated) • Pegged glenoid component (n = 8) • Keeled glenoid component (n = 7)

Outcomes	Outcomes reported at 6, 12, and 24 months
	• Function/disability:
	  ○ WOOS Index (0 to 100 scale, higher = better, reported as median only)
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as median only)
	  ○ Simple Shoulder Test (0 to 10 scale, higher = better, reported as median only)
	• Physician evaluated
	  ○ Radiostereometric analysis (RSA) of glenoid component micromotion
	  ■ Coronal plane translation (distance/mm, lower = better, reported as median only)
	  ■ Coronal plane rotation (in degrees, lower = better, reported as median only)

Notes	Source of funding: American Shoulder and Elbow Surgeons Research Grant and Dr. Paul H.T. Thoralkson Foundation Fund Conflicts of interest: E. Bohm declared speaking fees from Depuy, Zimmer, and Stryker, and institutional support from Smith and Nephew, Depuy, and Stryker

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "a series of sequentially numbered opaque envelopes were created by a research assistant before study recruitment based on a computer-based randomization list in blocks of 10"
Allocation concealment (selection bias)	Low risk	Quote 1: "a series of sequentially numbered opaque envelopes were created by a research assistant..." Quote 2: "each envelope held an allocation to 1 of the 2 study arms, and the envelope was opened once the surgeon determined the patient was eligible based on intraoperative findings"
Blinding of participants and personnel (performance bias)	High risk	Procedures were carried out by a single surgeon, who could not be blinded. The study was ceased by the surgeon due to loss of equipoise between the procedures based on interpretation of published studies at the time
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "the patient was not informed which implant was used"
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "the patient was not informed which implant was used"
Blinding of outcome assessment - physician reported outcomes	High risk	The primary study outcome was radiographic (physical reported). It is not possible to blind evaluators to treatment allocation, which is evident on the imaging. It is not clear who performed the RSA evaluations. The RSA technique was changed partway through the study

Gascoyne 2017 *(Continued)*

Incomplete outcome data (attrition bias)	High risk	The study was ceased before the planned sample size was achieved. Of 15 patients randomised to treatment, only 9 were available for analysis of the primary endpoint at 24 months
Selective reporting (reporting bias)	High risk	Although all outcomes were reported at 6, 12, and 24 months, the numbers at risk for each outcome at each time point are not clear. In addition, only median values are provided with no measures of central tendency
Major baseline imbalance	High risk	There are significant differences in preoperative ASES and WOOS Scores. One bilateral case was included in a small sample
Differences in rehabilitation regime	Low risk	Quote: "there were no appreciable differences in postoperative care between study groups. Patients' shoulders were immobilized in a sling for 3 weeks, and patients underwent postplication rehabilitation"

Gobezie 2019

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: August 2013 to July 2014 Setting: single centre, United States of America Length of follow-up: 2 years
Participants	Number randomised: 100 patients (100 shoulders) Number analysed: 68 patients (68 shoulders) Number lost to follow-up: 32 patients (group balance unclear) Baseline characteristics • Age (mean, range): 73 (43 to 94), 135° - 71, 155° - 73 • Male (%): 135° - 38%, 155° - 29% • Follow-up (mean months, range): 38 (29 to 45) • ASES Shoulder Score: 135° - 36.9 ± 22.6, 155° - 26.9 ± 18.8 • SANE Score: 135° - 32.5 ± 19.6, 155° - 35.9 ± 20.5 Inclusion criteria • Patients undergoing primary reverse total shoulder replacement for an irreparable rotator cuff tear and/or severe glenohumeral joint osteoarthritis Exclusion criteria • Previous shoulder replacement surgery • Preoperative infection • Preoperative fracture
Interventions	Reverse total shoulder replacement in 2 arms • Humeral implant with 135° neck-shaft angle (n = 37 analysed) • Humeral implant with 155° neck-shaft angle (n = 31 analysed)
Outcomes	Outcomes reported at 2 years post surgery • Pain

Gobezie 2019 *(Continued)*

- Visual analogue scale (0 to 10 scale, higher = worse)
- Function/Disability
  - ASES Shoulder Score (0 to 100 scale, higher = better)
  - SANE Score (0 to 100 scale, higher = better)
- Adverse events (specific) (cumulative counts)
- Revision, re-operation, or treatment failure (cumulative counts)
- Physician evaluated
  - Radiographic evidence of scapular notching (dichotomous, counts)
  - Range of motion (not an outcome of interest for this review)

Notes	Source of funding: Arthrex, Inc. Conflicts of interest: 3 of the study authors reported consultancy fees/royalties from Arthrex, Inc. Comments: complication and re-operation reporting is conflicting in this study. The numbers included in this review have been extracted from the tables

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "randomization based on even and odd numbers (1-10) was performed using a random number generator, placing our patients in 2 treatment groups (135 [even] vs 155 [odd])"
Allocation concealment (selection bias)	Unclear risk	It is unclear whether treatment allocation was concealed from patients or surgeons before treatment
Blinding of participants and personnel (performance bias)	Unclear risk	No blinding of participants or personnel was reported. It is unclear what effect this may have
Blinding of outcome assessment - self-reported outcomes	Unclear risk	No blinding procedures were reported
Blinding of outcome assessment - self-reported outcomes	Unclear risk	No blinding procedures were reported
Blinding of outcome assessment - physician reported outcomes	High risk	No blinding procedures were reported. For radiographic outcomes, the implant type would be obvious to the assessor
Incomplete outcome data (attrition bias)	High risk	32% of participants were not included in the outcomes analysis at 2 years, with imbalance between groups. The study paper indicates that 9 patients may have been revised before final follow-up and were excluded but then goes on to report additional revisions in the text and in the tables. It is very unclear whether there is any overlap in these numbers and how many cases may have been lost to follow-up.
Selective reporting (reporting bias)	Low risk	Planned outcomes are reported clearly
Selective reporting (reporting bias)	Low risk	Planned outcomes are reported clearly

Gobezie 2019 *(Continued)*

Major baseline imbalance	Unclear risk	Patients had well-balanced demographics and preoperative scores. However, more patients in the 135-degree group had the largest glenosphere size. The possible effect of this is unclear
Differences in rehabilitation regime	Low risk	The same postoperative rehabilitation protocol was used for all patients

Greiner 2015

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: not stated Setting: single centre, Germany Length of follow-up: 22 ± 8.1 months (mean ± SD)
Participants	Number randomised: 34 patients, 34 shoulders Number analysed: 31 patients, 31 shoulders Number lost to follow-up: 3 (2 in "STD" group, 1 in "BIO" group) Baseline characteristics • Age (mean/range): 75.4 years (66 to 88) • Sex (male:female): STD 9:8, BIO 3:14 • Constant Murley Score (mean ± SD): STD 26.1 ± 15.1, BIO 28.1 ± 13.2 • DASH Score (mean ± SD): STD 66.5 ± 11.4, BIO 60.6 ± 13.0 • ADLER Score (mean ± SD): STD 12.7 ± 6.5, BIO 13.7 ± 5.7 Inclusion criteria • Aged between 65 and 100 years with CTA > grade 2 according to the Hamada classification (Hamada 1990) • Range of motion of active abduction and flexion < 90° • Severe pain and failure of conservative treatment for a minimum of 6 months Exclusion criteria • Any relevant glenoid bone loss in the horizontal plane (types B2 and C according to Walch) or in the vertical plane (stage E3 according to Sirveaux) • Post-traumatic conditions including humeral head necrosis or other conditions precluding harvesting of a bone graft of the humeral head
Interventions	Reverse total shoulder replacement with cemented humeral components, 36 mm glenosphere, and 29-mm-diameter uncemented baseplate (Aequalis Reversed Shoulder Prosthesis; Tornier, Houston, TX, USA) Two arms • Standard (STD) offset glenosphere implanted with routine glenoid preparation and 15 mm baseplate peg (n = 17) • Bony increased offset (BIO) glenosphere implanted with 1 cm autologous bone graft block and 25 mm baseplate peg (to allow lateralisation of the centre of rotation) (n = 17)
Outcomes	Outcomes collected at 12 and 24 months (only 1 time point reported)

Greiner 2015 *(Continued)*

- Pain: subdomain of the Constant Murley Score (0/5/10/15 scale, higher = better, reported as mean ± SD) - included in additional tables but not suitable for meta-analysis
- Function
  - Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
  - Age- and gender-adjusted Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
  - DASH Score (0 to 100 scale, higher = worse, reported as mean ± SD)
  - ADLER Score (0 to 30 scale, higher = better, reported as mean ± SD) - "Activities of Daily Living requiring active External Rotation" (not validated and not a predetermined outcome of interest for this review)
- Adverse events - specific
  - Counts
  - Events reported were all acromial stress fractures

Notes	Source of funding: Tornier Quote: "the company was not involved in data collection, data analysis, or preparation or editing of the manuscript" Conflicts of interest: trial authors report no financial conflicts of interest

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "preoperative randomization was carried out using an online tool for randomization (Research Randomizer, version 3.0 [http://www.randomizer.org/])"
Allocation concealment (selection bias)	Unclear risk	The method of allocation concealment was not reported in the text
Blinding of participants and personnel (performance bias)	Unclear risk	Patients were blinded, but it is not clear if this lasted for the study duration. The surgeon could not be blinded. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Unclear risk	Quote: "patients were informed about both surgical techniques, consented to take part in the study, and were blinded to the treatment" Comment: patients were blinded, but it is not clear if this lasted for the study duration
Blinding of outcome assessment - physician reported outcomes	High risk	It would not be possible to blind assessors to radiographic appearances. For physician-measured components of function scores, trial authors report "highly standardized" methods but do not report any blinding process
Incomplete outcome data (attrition bias)	Low risk	3/34 patients lost to follow-up with no imbalance between groups
Selective reporting (reporting bias)	High risk	Scores collected at 12 and 24 months but reported only as a single "post-op" time point; we assume this is 24 months. A subgroup analysis was performed for just patients with intact teres minor tendons based on a non-validated score of specific shoulder movements. This was the only significant difference and does not appear to be a pre-determined study question
Major baseline imbalance	Unclear risk	Quote: "group characteristics were comparable in both groups, with no significant differences in the preoperative CS"

Greiner 2015 *(Continued)*

		Comment: contrary to the quoted text, there was obvious imbalance in gender between the 2 study arms (male:female = 9:8 vs 3:14). It is unclear whether this may impact measures of strength and function
Differences in rehabilitation regime	Unclear risk	Rehabilitation regimen not stated

Kwon 2019

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: 2010 to 2014 Setting: single centre, United States of America Length of follow-up: minimum 2 years
Participants	Number randomised: 107 shoulders (57 "sparing" group, 50 "standard" group) Number analysed: 70 (32 sparing, 38 standard) Number not completed as randomised: 14 Number lost to follow-up: 23 Baseline characteristics (of analysed patients only) • Age (mean ± SD): sparing 69.5 ± 9.5, standard 69.1 ± 8.2 • Sex (male:female): sparing 11:21, standard 22:16 • Pain - visual analogue scale (mean ± SD): sparing 6.6 ± 2.1, standard 6.3 ± 2.4 • ASES Shoulder Score (mean ± SD): sparing 29.3 ± 12.5, standard 32.8 ± 16.1 • Glenoid type (keeled:pegged): sparing 19:13, standard 10:28 Inclusion criteria • Patients with end-stage osteoarthritis scheduled for total shoulder replacement Exclusion criteria • Prior shoulder replacement • Full-thickness rotator cuff tear • Medial erosion of the glenoid • Morbid obesity (BMI > 40 kg/m^2)
Interventions	Conventional total shoulder replacement in 2 arms (manufacturer/brand line not stated) • "Sparing": surgical approach performed via the rotator interval, without violation of the subscapularis tendon (n = 57 randomised) • "Standard": surgical approach performed via a subscapularis tenotomy (n = 50 randomised)

Outcomes	Outcomes reported at 2 years
	• Pain: measured on a visual analogue scale (0 to 10 scale, lower = better, reported as mean ± SD)
	• Function
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)
	• Revision, re-operation, or treatment failure
	  ○ Counts extracted and reported in this review as cumulative totals up to and including the relevant time point

Kwon 2019 *(Continued)*

- Physician evaluated
  - Range of forward flexion and external rotation
  - This was not an outcome of interest for this review

Notes	Funding: not stated Conflicts of interest: one of the trial authors reported implant design royalties from Exactech Comment: identified on the clinicaltrials.gov register (NCT01961986) with a stated completion date of 2022 and planned recruitment of 120 patients

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "for each enrolled patient, a number was created by a random number generator. Patients with even numbers were treated with the traditional TSR procedure (STANDARD), and patients with odd numbers were treated with the SSC-sparing procedure (SPARING)"
Allocation concealment (selection bias)	High risk	Final decisions on patient eligibility for treatment were reliant on the adequacy of surgical exposure achieved and the status of the rotator cuff. However, random allocation was revealed to the surgeon in advance of this on the morning of surgery. There was a high rate of patients not receiving treatment as allocated
Blinding of participants and personnel (performance bias)	Unclear risk	Although patients and evaluating personnel were reported to be blinded throughout, the surgeon could not be blinded and specific rehabilitation restrictions were applied to one of the groups. This therefore invalidates the blinding statement, but any possible effect is unclear
Blinding of outcome assessment - self-reported outcomes	Unclear risk	Although patients and evaluating personnel were reported to be blinded throughout, the surgeon could not be blinded and specific rehabilitation restrictions were applied to one of the groups. This therefore invalidates the blinding statement, but any effect is unclear
Blinding of outcome assessment - self-reported outcomes	Unclear risk	Although patients and evaluating personnel were reported to be blinded throughout, specific rehabilitation restrictions were applied to one of the groups. This therefore invalidates the blinding statement, but any effect is unclear
Blinding of outcome assessment - physician reported outcomes	Low risk	Quote: "all outcome data collection was performed by a study coordinator who also remained blinded to the surgical technique"
Incomplete outcome data (attrition bias)	High risk	Out of 107 randomised patients, only 70 were included in the final analysis and there was considerable imbalance. 12 patients in the "sparing" study arm and 2 patients in the "standard" study arm did not receive treatment as allocated. 13 patients in the "sparing" arm and 10 patients in the "standard" arm were lost to follow-up for 2-year outcomes
Selective reporting (reporting bias)	Low risk	Outcomes in methods were fully and clearly reported
Major baseline imbalance	Unclear risk	For patients available for analysis, there was imbalance in the proportion of patients who were male (34% "sparing" group, 58% "standard" group) and in the proportion of keeled glenoid implants used (59% "sparing" group, 26% "standard" group). It is unclear whether this could have affected these results

Kwon 2019 *(Continued)*

Differences in rehabilitation regime	High risk	For the STANDARD group, a maximum limit on external rotation was determined during surgery and was set for the first 4 weeks after surgery. All patients were allowed to use a sling for 2 to 4 weeks. In the SPARING group, use of a sling was discretionary, but the STANDARD group was instructed to use a sling more regularly during ambulation and general daily activities

Lapner 2012

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: November 2006 to June 2009 Setting: 2 centres, Canada Length of follow-up: 2 years
Participants	Number randomised: 87 patients, 87 shoulders Number analysed: 79 for radiological outcomes at 6 months (41 osteotomy, 40 peel), 73 for clinical outcome scores (36 osteotomy, 37 peel) and clinical evaluations Lost to follow-up: 7 (5 osteotomy, 2 peel) Further exclusions: 7 (2 osteotomy, 5 peel) Baseline characteristics • Age (mean): osteotomy 70.4, peel 65.3, overall 67.8 (range 34 to 90) • Sex: 39% male • WOOS Index (mean ± SD/range): osteotomy 28.3 ± 17.6 (4 to 79), peel 27.0 ± 13.7 (8 to 62) • ASES Shoulder Score (mean ± SD/range): osteotomy 25.5 ± 12.7 (4 to 61), peel 22.6 ± 12.2 (0.2 to 48) • Strength in kg (mean ± SD): osteotomy Inclusion criteria • Unsuccessful standard non-operative treatment of osteoarthritis or inflammatory arthritis • Failed medical treatment defined as persistent pain and disability despite adequate non-operative treatment for 6 months • Medical treatment defined as (1) use of medications, including analgesics and non-steroidal anti-inflammatory drugs, (2) physiotherapy consisting of stretching, strengthening, and local modalities, and (3) activity modification Exclusion criteria • Active joint or systemic infection, rotator cuff arthropathy, muscle paralysis, neuropathic arthropathy, major medical illness (life expectancy less than one year or unacceptably high operative risk), inability to speak or read English or French, psychiatric illness that precluded informed consent, inability to be followed for 2 years
Interventions	Conventional total shoulder replacement with an uncemented stemmed humeral component and a cemented keeled all-polyethylene glenoid component (Aequalis; Tornier, Montbonnot, France). Surgical exposure performed in 2 arms • Lesser tuberosity osteotomy (n = 43 - randomised) • Subscapularis peel (n = 44 - randomised)
Outcomes	Outcomes reported at 3, 6, 12, and 24 months

Lapner 2012 *(Continued)*

- Function
  - WOOS Index (0 to 100 scale, higher = better, reported as mean ± SD and range)
  - ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD and range)
- Revision, re-operation, or treatment failure
  - Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
- Adverse events - specific
  - Counts
- Physician evaluated
  - Subscapularis muscle strength measure in the belly-press position with an electronic hand-held dynamometer (microFET2; Hoggan Health Industries, West Jordan, Utah)
  - Radiological evidence of healing seen on computed tomography (CT) scans (reported as counts)
  - Grade of fatty infiltration of subscapularis muscle (as per Goutallier 1994) (reported as counts)

Notes	Source of funding: Physicians' Services Incoporated Foundation (independent funding agency) Conflicts of interest: "one or more of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of an aspect of this work" Comments: there are numerical inconsistencies between study flowcharts and results reported in the tables and in the text

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "allocation was carried out with use of computer-generated blocked randomization, with the surgeon masked to block size"
Allocation concealment (selection bias)	Low risk	Quote 1: "final eligibility for the study was determined intraoperatively following visual inspection of the subscapularis tendon to ensure that it was intact" Quote 2: "treatment allocation was printed on cards that were inserted into sealed, opaque envelopes. The envelopes were opened by the circulating nurse once patient eligibility was confirmed"
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "because of the nature of the surgical trial, it was not possible to blind the surgeon to the surgical intervention" Comment: it is unclear whether this may have any effect
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "a trained research assistant performed the follow-up assessments and was blinded to the surgical procedure. The assessor did not have access to the patient chart prior to the evaluation. The patient was also blinded to the treatment assignment"
Blinding of outcome assessment - physician reported outcomes	High risk	Quote: "a trained research assistant performed the follow-up assessments and was blinded to the surgical procedure. The assessor did not have access to the patient chart prior to the evaluation" Comment: it is not possible to blind for the radiological outcomes assessed because the surgical technique would be apparent from appearances. The risk of bias for muscle strength measures would be assessed as low
Incomplete outcome data (attrition bias)	High risk	Overall 84% of those randomised were analysed. However, 7 patients with rheumatoid arthritis or a history of previous surgery were randomised to treatment but then subsequently were excluded from analyses. There is no protocol to refer to, and this post hoc exclusion is not justified in the methods

Lapner 2012 *(Continued)*

Selective reporting (reporting bias)	Low risk	Outcomes are comprehensively reported at 4 separate time points
Major baseline imbalance	Low risk	No significant difference is noted
Differences in rehabilitation regime	Low risk	Quote: "postoperative care and physiotherapy were identical in both groups"

Levine 2019

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: December 2009 to March 2012 Setting: single centre, United States of America Length of follow-up: 2 years
Participants	Number randomised: 60 patients (60 shoulders) Number analysed: 59 patients (59 shoulders) Number lost to follow-up: 1 patient (osteotomy group) Baseline characteristics • Trial authors state that there were no differences in demographic characteristics but provide no information at all on patient groups Inclusion criteria • Patients of any age or sex undergoing primary total shoulder replacement for advanced glenohumeral osteoarthritis • Failed non-operative treatment for minimum of 1 year Exclusion criteria • Prior subscapularis injury • Previous rotator cuff surgery or massive tear • Previous lesser tuberosity fracture/deformity • Inflammatory arthropathy • History of shoulder infection or active scute systemic infection
Interventions	Conventional total shoulder replacement in 2 arms • Surgical approach via lesser tuberosity osteotomy (n = 30 randomised) • Surgical approach via subscapularis tenotomy (n = 30 randomised)

Outcomes	Outcomes reported at 3, 6, and 12 months post surgery
	• Pain
	  ○ Visual analogue scale (0 to 10 scale, higher = worse)
	• Function
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better)
	  ○ Simple Shoulder Test Score (0 to 10 scale, higher = better)
	• Quality of life
	  ○ Short Form-36 (0 to 100 scale, higher = better)

Levine 2019 *(Continued)*

- Adverse events (specific)
- Revision, re-operation, or treatment failure
- Physician evaluated
  - Radiographic or ultrasound evidence of a healed repair
  - Range of motion (not an outcome of interest for this review)
  - Operative time (minutes)

Some results were reported in figure form only and required data extraction from the plots

Notes	Source of funding: not stated Conflicts of interest: trial authors report no relevant conflicts of interest Comments: outcomes reported in the text are not consistent with those in the figures. Request for raw data was not acknowledged; therefore when doubt existed, outcomes were not included in the meta-analysis

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "randomization was performed by a random number list generated by SPSS software"
Allocation concealment (selection bias)	Unclear risk	It is not clear whether the patient remained blinded to allocation before undergoing surgery
Blinding of participants and personnel (performance bias)	Unclear risk	There is no discussion of blinding of anyone involved in the study. Based on the different measures of outcome assessment (ultrasound vs X-ray) used, it would not be possible to blind anyone. It is unclear what effect this could have
Blinding of outcome assessment - self-reported outcomes	High risk	There is no discussion of blinding of anyone involved in the study. Based on the different measures of outcome assessment (ultrasound vs X-ray) used, it would not be possible to blind anyone. Patients would therefore be aware of their study allocation based on their scan type
Blinding of outcome assessment - physician reported outcomes	High risk	There is no discussion of blinding of anyone involved in the study. Based on the different measures of outcome assessment (ultrasound vs X-ray) used, it would not be possible to blind anyone
Incomplete outcome data (attrition bias)	Low risk	30 patients were randomised to each arm. One patient was lost to follow-up in the osteotomy group and was followed up at another centre
Selective reporting (reporting bias)	Low risk	All outcomes stated in the clinical trials registry are reported, in addition to patient-reported function
Major baseline imbalance	Low risk	Groups were well matched. Although patients in the tenotomy group had slightly lower strength of forward elevation at baseline, all other measures were similar, and this is unlikely to impact the main outcome measures recorded in this review
Differences in rehabilitation regime	Low risk	Quote: "patients in both groups followed the same postoperative rehabilitation protocol. A 30° abduction sling was used for 6 weeks postoperatively"

Litchfield 2011

Methods	Study design: randomised controlled trial in 2 parallel arms

Litchfield 2011 *(Continued)*

	Recruitment: June 2002 to August 2006 Setting: 7 centres, Canada Length of follow-up: 24 months (final follow-up)
Participants	Number randomised: 161 patients, 161 shoulders Number analysed: 152 (78 cemented, 74 uncemented) Lost to follow-up: 9 patients Baseline characteristics • Age (mean ± SD): cemented 69.4 ± 7.8, uncemented 68.4 ± 10.7 • Sex (male:female): cemented 43:35, uncemented 26:48 • WOOS Index (mean ± SD): cemented 28 ± 18.6, uncemented 24 ± 14.3 • ASES Shoulder Score (mean ± SD): cemented 26.7 ± 14.9, uncemented 23.12 ± 13.53 • Short Form-12 mental component (mean ± SD): cemented 49.8 ± 11.4, uncemented 45.9 ± 11.4 • Short Form-12 physical component (mean ± SD): cemented 31.9 ± 6.9, uncemented 31.7 ± 6.0 • MACTAR Score (mean ± SD): cemented 257.4 ± 31.2, uncemented 248.8 ± 42.2 Inclusion criteria • Diagnosis of primary osteoarthritis of the shoulder defined as shoulder pain and radiographic evidence of grade III or higher osteoarthritic changes (joint space narrowing, osteophytes, and subchondral sclerosis at the glenohumeral joint) with absence of findings or a history to indicate an aetiology of trauma, infection, avascular necrosis, inflammatory arthropathy, or previous reconstructive shoulder surgery • Failure of 6 months of medical management of the shoulder, including analgesics, non-steroidal anti-inflammatory drugs, physiotherapy, and activity modification Exclusion criteria • Evidence of major joint trauma, infection, avascular necrosis, cuff tear arthropathy, chronic dislocation, massive rotator cuff tear, inflammatory arthropathy, or previous shoulder surgery (other than arthroscopic debridement) • Preoperative computed tomography scans of the shoulder that showed insufficient glenoid bone stock to allow for implantation of a glenoid prosthesis • Active joint or systemic infection, significant muscle paralysis, or Charcot arthropathy • Life expectancy < 2 years or unacceptably high operative risk • Inability to speak or read English/French • Psychiatric illness or cognitive deficit that precluded informed consent • Unwillingness to be followed up for 2 years
Interventions	Conventional total shoulder replacement via stemmed humeral component and cemented polyethylene glenoid component (Bigliani/Flatow Total Shoulder Solution; Zimmer, Warsaw, IN, USA) 2 arms • Cemented humeral stem (n = 80 - randomised) • Uncemented humeral stem (n = 81 - randomised)

Outcomes	Outcomes reported at 6, 12, 18, and 24 months
	
	• Function
	  ◦ WOOS Index (0 to 100 scale, higher = better, reported as mean ± SD)
	  ◦ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)
	  ◦ MACTAR Score (0 to 500 scale, lower = better, reported as mean ± SD) - McMaster-Toronto Arthritis Patient Preference Disability Questionnaire

Litchfield 2011 *(Continued)*

- Adverse events: serious and specific
  - Counts
- Quality of life
  - Short Form-12 (0 to 100 scale, higher = better, reported as mean ± SD) - physical and mental component scores reported independently
- Revision/re-operation
  - Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
- Physician evaluated
  - Neither of these measures are included in this review because they have limited utility and are outside the scope of the published protocol
  - Measured strength in pounds, as per the measurement method of Constant Murley scoring (continuous scale, higher = better, reported as mean ± SD)
  - Range of motion in flexion and external rotation (higher = better, reported as mean ± SD)

Notes	Source of funding: Canadian Institutes of Health Research and Zimmer (USA and Canada) Conflicts of interest: none

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "randomization was done by a computer-generated, stratified randomization procedure, by use of variable block sizes of 2 and 4, and was stratified by surgeon"
Allocation concealment (selection bias)	Low risk	Quote: "upon verifying that the patient met the eligibility criteria, the surgeon was provided with the patient's assigned treatment by phoning a centralized data center"
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "patients were also blinded to their group allocation and were not told of their group assignment until their final follow-up at 2 years after surgery" Comment: surgeons cannot be blinded to the procedure performed. It is unclear whether this may have any effect
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "patients were also blinded to their group allocation and were not told of their group assignment until their final follow-up at 2 years after surgery"
Blinding of outcome assessment - physician reported outcomes	Low risk	Quote: "baseline and follow-up evaluations were performed by a trained research coordinator who was blinded to the group assignment"
Incomplete outcome data (attrition bias)	Low risk	Only 5.5% (9/161) lost to final follow-up with no significant imbalance between groups
Selective reporting (reporting bias)	Low risk	Comment: all outcomes from methods appear to be reported at the planned time points
Major baseline imbalance	High risk	Comment: there was significant imbalance in the gender breakdown between groups (male:female - cemented 43:35, uncemented 26:48) Quote: "the ancillary analysis for gender did suggest that the difference between study groups may be due to differences perceived by men rather than women"

Litchfield 2011 *(Continued)*

Differences in rehabilitation regime	Low risk	Quote: "after surgery, all patients were required to attend physiotherapy sessions. The physiotherapy protocol was standardized for all centers" Comment: all patients and assessors were blinded. There is unlikely to be any difference in postsurgical care provided

Lo 2005

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: not stated Setting: single centre, Canada Length of follow-up: 24 months
Participants	Number randomised: 42 patients (42 shoulders, 21 in each arm) Number analysed: 41 (21 hemiarthroplasty, 20 conventional TSR) Number lost to follow-up: none (1 death recorded in TSR group at 2 days) Baseline characteristics • Age (mean ± SD): hemiarthroplasty 70.3 ± 7.3, TSR 70.4 ± 9.0 • Sex (male:female): hemiarthroplasty 8:13, TSR 10:10 • McGill Pain Questionnaire (mean ± SD): hemiarthroplasty 16.0 ± 10.6, TSR 12.5 ± 9.4 • McGill Pain Visual Analogue Scale (mean ± SD): hemiarthroplasty 65.2 ± 24.3, TSR 65.0 ± 20.9 • WOOS Index (mean ± SD): hemiarthroplasty 33.5 ± 19.7, TSR 31.4 ± 17.7 • ASES Shoulder Score (mean ± SD): hemiarthroplasty 31.1 ± 16.6, TSR 30.7 ± 19.5 • UCLA Shoulder Rating Scale (mean ± SD): hemiarthroplasty 12.6 ± 3.5, TSR 13.2 ± 3.9 • Constant Score (mean ± SD): hemiarthroplasty 30.7 ± 14.2, TSR 28.7 ± 16.4 • Short Form-36 (SF-36) mental component scale (mean ± SD): hemiarthroplasty 55.5 ± 11.8, TSR 51.4 ± 14.7 • Short Form-36 (SF-36) physical component scale (mean ± SD): hemiarthroplasty 29.5 ± 7.6, TSR 31.3 ± 8.4 • Range of motion, domain extracted from Constant Score (mean ± SD): hemiarthroplasty 13.7 ± 7.2, TSR 13.4 ± 9.5 Inclusion criteria • Primary osteoarthritis of the shoulder • Failure of a minimum of 6 months of non-operative treatment (including analgesics, anti-inflammatory medication, and physiotherapy), desire to have surgical intervention • Primary osteoarthritis of the shoulder was defined as shoulder pain; no history of major trauma, infection, osteonecrosis, cuff tear arthropathy, chronic dislocation, or a secondary cause of osteoarthritis; and radiographic evidence of joint space narrowing, osteophyte formation, and/or subchondral sclerosis Exclusion criteria • Condition other than shoulder osteoarthritis that would substantially contribute to shoulder dysfunction (e.g. cervical spine disease), a rotator cuff tear (> 1 cm), inflammatory arthritis, or post-capsulorrhaphy osteoarthritis • Major medical illness that would substantially influence quality of life (e.g. unstable angina) • Active infection • Substantial muscle paralysis

Lo 2005 *(Continued)*

	• Lack of fitness for surgery or unwillingness to be followed for 2 years
Interventions	Shoulder replacement in 2 arms (Neer Series II modular shoulder implants; 3M Canada, London, Ontario, Canada) • Stemmed hemiarthroplasty (n = 21 - randomised) • Conventional stemmed total shoulder replacement (n = 21 - randomised)

Outcomes	Followed up at 6 weeks and at 3, 6, 12, 18, and 24 months
	Outcomes reported at 24 months
	• Pain
	  ○ McGill Pain Visual Analogue Scale (0 to 100 scale, lower = better, reported as mean ± SD)
	  ○ McGill Pain Questionnaire (lower = better, reported as mean ± SD, unclear what scale was used)
	• Function
	  ○ WOOS Index (0 to 100 scale, higher = better, reported as mean ± SD)
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)
	  ○ UCLA Shoulder Rating Scale (0 to 35 scale, higher = better, reported as mean ± SD)
	  ○ Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
	• Adverse events
	  ○ Counts: serious and specific
	• Revision, re-operation, or treatment failure
	  ○ Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
	• Quality of life
	  ○ Short Form-36 scale (0 to 100 scale, higher = better, reported as mean ± SD) - physical and mental component scores reported independently
	• Physician evaluated
	  ○ Range of motion score - domain extracted from Constant Murley Score. Limited utility; not included in this review

Notes	Source of funding: 3M, Canada Conflicts of interest: no other conflicts were reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "a sealed envelope containing the randomly assigned treatment group allocation was opened by the circulating nurse" Comment: it is not clear how the random sequence was generated
Allocation concealment (selection bias)	Low risk	Quote: "after it was confirmed that the patient had primary osteoarthritis as well as good-quality glenoid bone stock that was adequate for the performance of either a hemiarthroplasty or a total shoulder arthroplasty, a sealed envelope containing the randomly assigned treatment group allocation was opened by the circulating nurse"
Blinding of participants and personnel (performance bias)	Unclear risk	Although patients and independent evaluators were blinded for the duration of the study, surgeons were not. It is unclear what effect this may have
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "both the patient and an independent evaluator remained blinded to the group assignment for the duration of the study"

Lo 2005 *(Continued)*

Blinding of outcome assessment - physician reported outcomes	Low risk	A research assistant who was blinded to the treatment group performed a standardised assessment of all patients preoperatively; at 6 weeks postoperatively; and at 3, 6, 12, 18, and 24 months postoperatively
Incomplete outcome data (attrition bias)	Low risk	Only 1 patient was excluded from the primary outcome analysis and was recorded as having a serious adverse event. Outcomes analysed were complete based on a conservative efficacy analysis to account for patients crossing over between groups. The last score recorded before cross-over was carried through
Selective reporting (reporting bias)	Low risk	All planned measures were reported in tables I through IV
Major baseline imbalance	Low risk	There were no significant differences between groups in terms of age, gender, preoperative range of motion, or functional status
Differences in rehabilitation regime	Low risk	Quote: "a sling was applied with the arm at the side. Active-assisted range-of-motion exercises were begun on the first postoperative day in the hospital, with emphasis on forward elevation and external rotation. External rotation was limited according to the intraoperative assessment of the tension of the subscapularis repair and was usually 30°. An active range of motion was allowed at four weeks postoperatively, and strengthening exercises were begun at eight weeks postoperatively. The patient's return to normal activities progressed as tolerated over three to six months" Comment: the same regimen was applied to both

Mechlenburg 2014

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: 2007 to 2010 Setting: 2 centres, Denmark Length of follow-up: 2 years
Participants	Number randomised: 32 patients, 32 shoulders (14 Copeland, 18 Global C.A.P.) Number analysed: varied by outcome type • Baseline demographics: 13 Copeland, 18 Global C.A.P. (1 patient excluded from the outset due to failure to deliver adequate RSA beads) • WOOS Index and Constant Score: 10 Copeland, 15 Global C.A.P. • Radiostereometric analysis: 8 Copeland, 15 Global C.A.P. • Length of glenohumeral offset: 12 Copeland, 17 Global C.A.P. • Humeral head bone mineral density: 9 Copeland, 15 Global C.A.P. Number lost to follow-up: 2 (1 Copeland, 1 Global C.A.P.), in addition to the aforementioned excluded case Baseline characteristics • Age (median/range): Copeland 61 (40 to 82), Global C.A.P. 63 (53 to 83) • Sex (male:female): Copeland 8:5, Global C.A.P. 10:8 • Pain on a visual analogue scale (median/range): Copeland 53 (15 to 97), Global C.A.P. 44 (8 to 99) • WOOS Index (raw score on scale 0 to 2100, median/range): Copeland 939 (441 to 1574), Global C.A.P. 1088 (504 to 1870)

Mechlenburg 2014 *(Continued)*

- Constant Murley Score (median/range): Copeland 57 (9 to 70), Global C.A.P. 35 (10 to 65)
- Osteopenia (defined as T-score -1 to -2.5): Copeland 3/13, Global C.A.P. 10/18

Inclusion criteria

- Individuals aged 18 to 85 years with shoulder osteoarthritis and cartilage defects involved on the humeral rather than the glenoid side of the joint

Exclusion criteria

- Previous shoulder replacement or other major shoulder surgery
- Severe shoulder instability with a large rotator cuff defect
- Rheumatoid arthritis or metabolic bone disease
- Patients unable to avoid non-steroidal anti-inflammatory drugs after surgery
- Patients requiring regular systemic steroid treatment
- Women in the fertile age range who did not use safe contraception, and women taking hormone replacement therapy

Interventions	Humeral head resurfacing surgery (also known as resurfacing humeral hemiarthroplasty performed in 2 arms using 1 of 2 implants) • Copeland cementless humeral head resurfacing implant (Biomet Inc., Warsaw, IN, USA) (n = 14 - randomised) • Global C.A.P. cementless humeral head resurfacing (Depuy Int., Warsaw, IN, USA) (n = 18 - randomised)

Outcomes	Outcomes reported at 3, 6, 12, and 24 months
	• Function
	  ○ WOOS Index (0 to 1900 raw scale, lower = better, reported in box plots as median plus 10th/25th/75th/90th centiles)
	  ○ Constant Murley Score (0 to 100 scale, higher = better, reported in box plots as median plus 10th/25th/75th/90th centiles)
	• Revision, re-operation, or treatment failure
	  ○ Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
	• Physicial evaluated
	  ○ Radiostereometric analysis
	  ■ Study reports translation in 3 planes and rotation in 1 plane
	  ■ Extracted for this review: total translation of humeral component in mm (continuous scale, reported as mean ± SD and in box plots as median plus 10th/25th/75th/90th centiles)
	  ○ Humeral head bone mineral density in g/cm^3 (continuous scale, reported in box plots as median plus 10th/25th/75th/90th centiles)
	  ○ Length of glenohumeral offset in cm at 6 months - not extracted for this review

Notes	Source of funding: "this study was financially supported by the Danish Rheumatism Association, The Aase and Ejnar Danielsen Foundation, The AP Møller Foundation, The Danish Medical Association, Cooperative Organizations Humanitarian and Cultural Foundation, The Hede Nielsen Family Foundation, Jacob Madsen & Olga Madsen's Foundation, Protesekompagniet/DePuy Denmark, and Biomet Denmark" Conflicts of interest: none reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "the random allocation sequence was generated by the first and last authors by drawing labels from a box"

Mechlenburg 2014 *(Continued)*

Allocation concealment (selection bias)	Low risk	Quote: "labels were then concealed in sequentially numbered closed envelopes" Comment: groups were assigned only after confirmation in theatre that the patient was suitable for humeral head resurfacing
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "except for the 2 surgeons and the observers evaluating RSA and LGHO, all other assessors, care providers, and physiotherapists were blinded to the patient's implant assignment" Comment: it is unclear whether patients were blinded
Blinding of outcome assessment - self-reported outcomes	Unclear risk	It is not clear whether patients were blinded
Blinding of outcome assessment - physician reported outcomes	High risk	Quote: "(RSA). Except for the 2 surgeons and the observers evaluating RSA and LGHO, all other assessors, care providers, and physiotherapists were blinded to the patient's implant assignment" Comment: the primary outcome for this study was radiological, for which the assessor could not be blinded
Incomplete outcome data (attrition bias)	Unclear risk	There is lack of clarity regarding the number of patients included at the interim time points in the RSA outcomes. Exclusion of data for revised patients may lead to the potential to underestimate the mean translation
Selective reporting (reporting bias)	Low risk	All outcomes appear to be fully reported
Major baseline imbalance	High risk	There is imbalance in the preoperative Constant Scores (57 vs 35) and in the number of patients with osteopenia (3/13 vs 10/18). These may have an impact on both patient-reported and radiological outcomes
Differences in rehabilitation regime	Low risk	Quote: "the postoperative course was as at present after insertion of HHRI, with an arm sling for 6 weeks. After the first postoperative day, unweighted, passive movement, supervised by a physiotherapist, was allowed to a maximum of 60 outward rotation, whereas only pain limited abduction and flexion. After 6 weeks, free, active movement, respecting the pain threshold, was encouraged under supervision of a physiotherapist"

Nuttall 2007

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: 2000 to 2004 Setting: single centre, United Kingdom Length of follow-up: 2 years
Participants	Number randomised: 20 (10 pegged, 10 keeled) Number analysed: 16 (however, see comment below on lost to follow-up) Number lost to follow-up: 2 died (1 unstable marker-bead placement, 1 difficult to visualise the glenoid). Dropouts are not reported with sufficient clarity to determine the number remaining at risk for different outcomes

Nuttall 2007 *(Continued)*

	Baseline characteristics • Age (mean ± SD): pegged 63 ± 14, keeled 71 ± 8 • Sex (male:female) 7:3 both groups • Pain on a visual analogue scale (mean): pegged 5.6, keeled 6.6 • ASES Shoulder Score (mean): pegged 25, keeled 13 • Constant Score (mean): pegged 32, keeled 20 Inclusion criteria • Primary osteoarthritis • Good glenoid bone stock, sufficient to introduce a glenoid component • An intact and functioning rotator cuff. Exclusion criteria • None specified
Interventions	Conventional total shoulder replacement (Global Shoulder; Depuy International, Leeds, England) using all polyethylene cemented glenoid components in 2 arms • Pegged glenoid component (n = 20 - randomised) • Keeled glenoid component (n = 20 - randomised)

Outcomes	RSA outcomes reported at 3, 6, 12, and 24 months; clinical outcomes at 24 months only
	• Pain
	  ○ Visual analogue scale (0 to 10 cm, lower = better, reported as mean only)
	• Function
	  ○ Constant Murley Score (0 to 100 scale, higher = better, reported as mean only)
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean only)
	• Physician evaluated
	  ○ Radiostereometric analysis of glenoid component micromotion
	  ■ Translation and rotation reported in 3 planes
	  ■ Included in this review only: maximum total point movement in mm (continuous scale, reported as mean only)
	  ○ Range of motion (not included in this review)

Notes	Source of funding: Depuy International (Leeds, England) Conflicts of interest: no additional conflicts reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Trial authors report that the selection was made randomly but do not describe the randomisation process
Allocation concealment (selection bias)	Unclear risk	Quote: "the selection of a keeled or pegged implant was made randomly by use of a sealed-envelope technique" Comment: it is not clear when the allocation was unmasked
Blinding of participants and personnel (performance bias)	Unclear risk	The paper does not comment on blinding of patients or personnel

Nuttall 2007 *(Continued)*

Blinding of outcome assessment - self-reported outcomes	Unclear risk	The paper does not comment on blinding of patients or personnel
Blinding of outcome assessment - physician reported outcomes	High risk	The paper does not mention blinding of observers. It would not be possible to blind for implant type for radiographic/RSA analysis
Incomplete outcome data (attrition bias)	Unclear risk	Quote: "two patients died before the end of the study, another had unstable marker-bead placement, and the glenoid component in a fourth patient could not be clearly visualized" Comment: although 4 patients were lost to follow-up, tables continue to report 20 patients at risk
Selective reporting (reporting bias)	Unclear risk	No explicit mention of presence or absence of adverse events. No measures of central tendency are provided for clinical outcomes
Major baseline imbalance	Low risk	No obvious imbalance in age, sex, and glenoid morphology
Differences in rehabilitation regime	Low risk	Quote: "postoperatively, all patients began a passive and supervised active mobilization program"

Poon 2014

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: August 2007 to January 2011 Setting: single centre, New Zealand Length of follow-up: minimum 2 years
Participants	Number randomised: 50 (23 eccentric glenosphere, 27 concentric glenosphere) Number analysed: 50 Number lost to follow-up: none at 2 years Baseline characteristics • Age (mean/range): eccentric 81 (67 to 91), concentric 81 (65 to 91) • Sex (male:female): eccentric 7:16, concentric 8:19 • Dominant arm affected: eccentric 16/23, concentric 11/27 • Pain - VAS (mean/range): eccentric 6.6 (3 to 9), Concentric 6.6 (1 to 10) • ASES Shoulder Score: eccentric 30 (5 to 48), concentric 28 (5 to 58) • Oxford Shoulder Score: eccentric 20 (6 to 33), concentric 20 (4 to 37) Inclusion criteria • Patients over the age of 65 years • Referred to the tertiary orthopaedic surgical centre with clinical and radiographic evidence of cuff tear arthropathy Exclusion criteria • Deficient glenoid bone stock • Abnormal deltoid muscle

Poon 2014 *(Continued)*

Interventions	Reverse total shoulder replacement using Shoulder Modular Replacement (SMR; Lima Corporate, San Daniele del Friuli, Italy) in 2 arms • Eccentric glenosphere position - 4 mm inferior offset relative to the glenoid baseplate (n = 23 - randomised) • Concentric glenosphere position - centred on the glenoid baseplate (n = 27 - randomised)

Outcomes	Outcomes reported at primary 24-month endpoint
	
	• Pain on a visual analogue scale (0 to 10 scale, lower = better, reported as mean/range with P value)
	• Function
	  ○ ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean/range with P value)
	  ○ Oxford Shoulder Score (0 to 48 scale, higher = better, reported as mean/range with P value)
	• Physician evaluated
	  ○ Inferior notching of scapula
	  ■ Nerot grading system (grades I to IV, counts)
	  ■ Any scapula notching (≥Nerot grade I, dichotomous)
	  ○ Inferior overhang and prosthesis-scapular neck angle achieved (mm) - not reported in this review
	  ○ Range of motion - not reported in this review

Notes	Source of funding: no external funding source is reported Conflicts of interest: 1 or more of the review authors reported financial relationships that could have the potential to influence the work of the study (not specifically started in the journal report)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "randomization involved the use of computer-generated, sequentially numbered, and sealed opaque envelopes"
Allocation concealment (selection bias)	Low risk	Quote: "randomization involved the use of computer-generated, sequentially numbered, and sealed opaque envelopes"
Blinding of participants and personnel (performance bias)	Unclear risk	Quote: "patients were blinded to the type of glenosphere and were informed that the type to be used had been randomly allocated in a concealed envelope. The surgeon was blinded to the type of glenosphere until after implantation of the glenoid baseplate" Comment: it is not possible to blind the surgeon from the point of implantation onwards. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote 1: "secondary outcomes were assessed by an independent research nurse blinded to the treatment groups" Quote 2: "patients were blinded to the type of glenosphere and were informed that the type to be used had been randomly allocated in a concealed envelope"
Blinding of outcome assessment - physician reported outcomes	High risk	Quote: "radiographs were independently measured by two trained assessors (J.C. and S.W.Y.) who were not blinded because of the distinctive radiographic appearances of the glenospheres (Fig. 4)"
Incomplete outcome data (attrition bias)	Low risk	Quote: "all patients completed the minimum two-year follow-up"

Poon 2014 *(Continued)*

Selective reporting (reporting bias)	Low risk	Ouctomes fully reported at the primary endpoint
Major baseline imbalance	Low risk	No major imbalances in baseline demographics reported
Differences in rehabilitation regime	Low risk	Quote: "postoperatively, all patients underwent a standardized physiotherapy rehabilitation program. For the first six weeks, the patients wore a sling and were allowed passive range-of-motion exercises only. Between six and twelve weeks after surgery, active range-of-motion exercise began, supervised by physiotherapists. Three months after surgery, gradual strengthening exercises were encouraged"

Rahme 2009

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: 2001 to 2004 Setting: 2 centres, Sweden Follow-up: 24 months
Participants	Number randomised: 30 shoulders in 28 patients (15 pegged, 15 keeled) Number analysed: 27 shoulders in 25 patients (14 pegged, 13 keeled) Number lost to follow-up: 3 shoulders in 3 patients due to instability of the tantalum markers (1 pegged, 2 keeled); 1 further patient died before final 2-year follow-up (keeled) Baseline characteristics (reported as number of shoulders analysed (n = 27)) • Age (mean ± SD): pegged 63.6 ± 11.1, keeled 64.4 ± 7.1 • Sex (male:female): pegged 4:10, keeled 5:8 • Constant Score (mean ± SD): pegged 21.7 ± 14.7, keeled 25.0 ± 10.0 • Subjective shoulder value (mean ± SD): pegged 39.1 ± 24.2, keeled 35.4 ± 17.8 • Primary/secondary osteoarthritis (n): pegged 12/2, keeled 10/3 • Reasons for secondary osteoarthritis: 3 osteonecrosis, 1 (sequelae of) fracture, 1 recurrent dislocation Inclusion criteria • Patients with primary or secondary osteoarthritis, all of whom were scheduled for total shoulder replacement Exclusion criteria • Instability of RSA markers preventing analysis
Interventions	Conventional total shoulder replacement with uncemented humeral component and all-polyethylene cemented glenoid technique (Bigliani/Flatow; Zimmer, Warsaw, IN, USA) in 2 arms • Pegged glenoid component (n = 15 - randomised) • Keeled glenoid component (n = 15 - randomised)
Outcomes	Outcomes reported at 4, 12, and 24 months • Function ○ Constant Murley Score (0 to 100 scale, higher = better, reported as mean only)

Rahme 2009 *(Continued)*

- Subjective shoulder value (0 to 100 scale, self-evaluated % of normal shoulder function, higher = better, reported as mean only)

- Revision, re-operation, or treatment failure
  - Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
- Physician evaluated
  - Radiostereometric analysis
    - Translation (mm) and rotation (degrees) in 3 axes (reported as median and interquartile range) - reported in this review within narrative text only
  - Glenoid lucency grade (1 to 5 scale, lower = better, reported as number graded ≥ 4)

Notes	Source of funding: Zimmer (Warsaw, Indiana, USA) Conflicts of interest: no additional conflicts reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "during surgery, the patients were allocated by block randomization, with use of a closed envelope technique" Comment: it is not clear what method was actually used to generate the sequence
Allocation concealment (selection bias)	Unclear risk	Quote: "during surgery, the patients were allocated by block randomization, with use of a closed envelope technique" Comment: the method of concealment and timing of allocation are not reported with sufficient clarity
Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind the surgeon to treatment allocation. The possible effect of this is unclear. It is not clear whether patients were blinded to their treatment allocation
Blinding of outcome assessment - self-reported outcomes	Unclear risk	It is not clear whether patients were blinded to their treatment allocation
Blinding of outcome assessment - physician reported outcomes	High risk	Details of who performed the RSA were not stated. Although the radiographic assessor was not aware of the matching clinical and radiostereometric outcomes, it is not possible to blind for the type of glenoid, which is obvious on plain radiographs
Incomplete outcome data (attrition bias)	Low risk	26 of 30 shoulders were available for analysis at the final follow-up with adequate explanation of excluded patients
Selective reporting (reporting bias)	Low risk	Planned outcomes are reported at the primary endpoint
Major baseline imbalance	Unclear risk	Groups are comparable with respect to baseline demographics. However, given that the main outcomes were micromotion and radiolucency, some baseline information regarding glenoid morphology and bone stock would be required to make an informed judgement
Differences in rehabilitation regime	Unclear risk	Details of the postoperative regimen are not stated

Rasmussen 2015

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: September 2009 to August 2012 Setting: single centre, Denmark Length of follow-up: 12 months
Participants	Number randomised: 40 shoulders in 35 patients (20 resurfacing, 20 stemmed hemiarthroplasty) Number analysed: 38 shoulders (19 resurfacing, 19 stemmed hemiarthroplasty) Number lost to follow-up: 2 shoulders. An additional patient (stemmed hemiarthroplasty) declined follow-up beyond 6 months. For this patient, 3-month scores were carried through Baseline characteristics • Age (mean/range): resurfacing 65.6 (40 to 88), stemmed hemiarthroplasty 69.1 (46 to 87) • Sex (male:female): resurfacing 7:13, stemmed hemiarthroplasty 6:14 • Constant Murley Score (mean/range): resurfacing 21.5 (0 to 42), stemmed hemiarthroplasty 26.7 (6 to 50) • WOOS Index (mean/range): resurfacing 24.9 (5.5 to 64.7), stemmed hemiarthroplasty 38.4 (7.8 to 74.6) Inclusion criteria • Patients diagnosed with glenohumeral osteoarthritis scheduled for shoulder replacement • Osteoarthritis was defined by subchondral sclerosis, joint space narrowing or osteophytes visualised on plain radiographs with an anteroposterior and a lateral view combined with a history of pain, stiffness of the shoulder, and physical examination showing decreased range of motion • The indication for operation was pain with limited response to non-surgical treatment with analgesics and physiotherapy for at least 6 months Exclusion criteria • Presence of any other pathological conditions affecting function of the upper extremity • Symptomatic rotator cuff pathology: the integrity of the rotator cuff was evaluated by a clinical examination including inspection, palpation, range of motion, and strength. Magnetic resonance imaging was conducted if rotator cuff pathology was suspected • Rheumatoid arthritis • American Society of Anesthesiologists (ASA) Score ≥ 3 • Non-concentric glenoid: concentricity of the glenoid was evaluated by a preoperative radiograph with an anteroposterior and a lateral view and was confirmed during the operation • Less than 60% of the native humeral head intact • Cognitive difficulties • Absence of co-morbidity and other pathological conditions in the shoulder was confirmed by the patient and through review of medical records
Interventions	Shoulder hemiarthroplasty in 2 arms • Cementless humeral head resurfacing hemiarthroplasty implant with hydroxyapatite coating (n = 20 - randomised) • Cemented stemmed modular hemiarthroplasty implant (n = 20 - randomised) Manufacturers not stated
Outcomes	Followed up and outcomes reported at 3 and 12 months • Function

Rasmussen 2015 *(Continued)*

- WOOS Index (0 to 100 scale, higher = better, reported as mean, range, and difference in means with 95% confidence interval)
- Constant Murley Score (0 to 100 scale, higher = better, reported as mean, range, and difference in means with 95% confidence interval)

- Adverse events
  - Specific (counts)
- Revision, re-operation, or treatment failure
  - No events occurred
- Physician evaluated
  - Operating time in minutes (continuous scale, reported as mean, range, difference in means with 95% confidence interval)

Notes	Source of funding: not reported Conflicts of interest: none reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "the patients were randomly assigned to treatment with one of two arthroplasty designs using a random numbers list generated with Microsoft Excel (Redmond, WA, USA)"
Allocation concealment (selection bias)	Low risk	Quote: "the assignment group for each patient was kept in sealed, opaque and consecutively numbered envelopes and revealed to the surgeon in the operating theatre just before surgery"
Blinding of participants and personnel (performance bias)	Unclear risk	Patients were kept blinded to treatment allocation; however it is not possible to blind the operating surgeon. The possible effect of this is unclear
Blinding of outcome assessment - self-reported outcomes	Low risk	Quote: "the patient remained blinded to the randomisation for the duration of the study. There were no cases of accidental loss of blinding"
Blinding of outcome assessment - physician reported outcomes	Low risk	Quote: "the first author, not involved in the surgical procedure but non-blinded to the randomisation, evaluated CMS preoperatively, three months postoperatively and at one year and a senior author, not involved in the surgical procedure and blinded to the randomisation, evaluated CMS at one year" Comment: the Constant Murley Score includes a number of physician-measured domains and as such, cannot be assumed to be free of bias from a non-blinded assessor. However, for the reported primary endpoint, these assessments were blinded. In addition, blinded and non-blinded assessments were carried out and no significant difference was seen between recorded observations
Incomplete outcome data (attrition bias)	Low risk	Patients lost to follow-up adequately reported
Selective reporting (reporting bias)	Low risk	Planned outcomes presented at the primary endpoint
Major baseline imbalance	High risk	There was a statistically significant difference between groups of 13.5 points for the WOOS Index (this exceeds the quoted minimal clinically important difference)

Rasmussen 2015 *(Continued)*

Differences in rehabilitation regime	Low risk	Quote: "all patients received identical postoperative treatment with a sling and swathe for two weeks followed by a simple sling for another two weeks. Active range of motion was allowed at two weeks with protection of the subscapularis muscle. Strengthening exercises were allowed at six weeks"

Sandow 2013

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: commenced 1994 Setting: single centre, Australia Length of follow-up: primary endpoints at 2 years, long-term 10 years
Participants	Number randomised: 33 shoulders in 33 patients (20 conventional total shoulder replacement, 13 hemiarthroplasty) Number analysed (at 2 years): 31 shoulders (18 TSR, 13 hemiarthroplasty) Number lost to follow-up (at 2 years): 2 shoulders Number analysed (at 10 years): 18 shoulders (11 TSR, 7 hemiarthroplasty) Baseline characteristics • Age (median): TSR 72, hemiarthroplasty 68 • Sex (male:female): TSR 5:15, hemiarthroplasty 6:7 • Pain on visual analogue scale (median/range, 0 to 10 scale, higher = worse): TSR 7.3 (4.5 to 9.7), hemiarthroplasty 7 (3.8 to 9.5) • UCLA Shoulder Rating Scale (median/range, 0 to 35 scale, higher = better): TSR 10 (7 to 20), hemiarthroplasty 12 (8 to 16) • Constant Murley Score (median/range, 0 to 100 scale, higher = better): TSR 25.5 (9 to 55), hemiarthroplasty 31 (22 to 52) Inclusion criteria • Advanced osteoarthritis of the shoulder with an intact rotator cuff and no evidence of infection, inflammatory disease, or previous fracture Exclusion criteria • Inadequate surgical exposure, apparent inflammatory process, significant rotator cuff tear (such that a major defect would remain after subscapularis tendon repair and rotator interval closure), proximal humeral or glenoid deformity
Interventions	Shoulder replacement with a cementless stemmed humeral component (Global Shoulder Arthroplasty System; Depuy Orthopaedics, Warsaw, IL, USA) in 2 arms • Conventional TSR using cemented all-polyethylene pegged glenoid component (n = 20 - randomised) • Hemiarthroplasty (without glenoid replacement) (n = 13 - randomised)

Outcomes	Outcomes reported at 6, 12, 24, and 36 months. Primary endpoint was 24 months; 36-month data incomplete. Revisions reported to 10 years
	
	• Pain
	  ○ Visual analogue scale (0 to 10 scale, higher = worse, reported as median/range)
	• Function
	  ○ UCLA Shoulder Rating Scale (0 to 35 scale, higher = better, reported as median/range)

Sandow 2013 *(Continued)*

- Constant Murley Score (0 to 100 scale, higher = better, reported as median/range)
- Revision, re-operation, or treatment failure
  - Counts extracted and reported in this review as cumulative totals up to and including the relevant time point
- Adverse events
  - Serious and specific
  - Counts

Notes	Source of funding: educational grant from Depuy Conflicts of interest: none declared Early termination of study: during follow-up, 2 patients underwent early revision of hemiarthroplasty and 2 further patients had deteriorating pain. The institutional review board therefore suspended the trial before completion of patient recruitment

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "[patients were] intraoperatively randomized to HA or TSR using the randomization method of sequentially numbered, opaque, sealed envelopes" Comment: the technique for sequence generation is not stated
Allocation concealment (selection bias)	Low risk	Quote: "[patients were] intraoperatively randomized to HA or TSR using the randomization method of sequentially numbered, opaque, sealed envelopes" Comment: allocation was revealed intraoperatively only once adequate exposure to permit either operation had been achieved
Blinding of participants and personnel (performance bias)	High risk	Quote: "neither the surgeon nor the patients were blinded to the final procedure" Comment: all procedures were performed by the same surgeon who was fully involved in the follow-up process. Because follow-up was performed open-label, this process cannot be assumed to be unbiased
Blinding of outcome assessment - self-reported outcomes	High risk	Quote: "neither the surgeon nor the patients were blinded to the final procedure due to difficulty in avoiding viewing of the X-ray images by medical or paramedic practitioners treating the study participants"
Blinding of outcome assessment - physician reported outcomes	High risk	Quote: "one TSR patient required revision of the humeral component alignment within 1 week due to malposition and incorrect version of the humeral prosthesis. Owing to the issue of the surgeon's technical error and the need for perioperative management, this was regarded as a complication rather than as fulfilling the criteria of prosthesis revision for pain, loosening, or infection" Comment: the surgeon was not blinded and was the primary assessor for measured outcomes, adverse events, and decisions to re-operate. The above quote suggests there is risk of bias in the allocation of adverse events
Incomplete outcome data (attrition bias)	Low risk	The only patients lost to follow-up were those who had died
Selective reporting (reporting bias)	Unclear risk	Quote 1: "all patients who were still alive at 10 years were reviewed by questionnaire or clinic review, as was possible, using the same set of questions" Quote 2: "in the initial 2-year review period, primary outcome measures related to pain relief, motion, and possible implant revision. The secondary ques-

Sandow 2013 *(Continued)*

		tions were related to early pain relief, recovery of motion, and functional outcome. Late review questions related primarily to pain relief and the occurrence of revision or further treatment Comment: late reviews do not include the use of a validated functional scale. The potential effect of this is unclear
Major baseline imbalance	Low risk	No significant differences shown for baseline demographics and functional scores
Differences in rehabilitation regime	Low risk	Quote: "postoperative mobilization for the 2 groups was identical"

Uschok 2017

Methods	Study design: randomised controlled trial in 2 parallel arms Recruitment: November 2005 to May 2008 Setting: 2 centres, Germany Length of follow-up: 5 years
Participants	Number randomised: 40 shoulders in 40 patients (20 stemless, 20 stemmed) Number analysed: 33 shoulders at 2 years, 29 shoulders at 5 years Number lost to follow-up: at 5 years, 7 patients were uncontactable, 2 were too unwell to attend, 1 had died, and 1 declined further participation Baseline characteristics • Age (mean): stemless 65, stemmed 69 • Sex (male:female): stemless 10:10, stemmed 7:13 • Constant Murley Score (mean ± SD): stemless 53.9 ± 11.3, stemmed 25.6 ± 15.2 • Age- and gender-adjusted Constant Murley Score (mean ± SD): stemless 70.7 ± 18.8, stemmed 34.8 ± 19 • Glenoid morphology (watch grade): stemless 4× A2, 5× B1, 5× B2, stemmed 4× A2, 8× B1, 3× B2 • Type of glenoid component (metal-backed:keeled): stemless 9:5, stemmed 13:2 (of patients followed up to 5 years) Inclusion criteria • Primary osteoarthritis of the shoulder • No diagnostic, severity, or pretreatment criteria provided Exclusion criteria • Prior surgery of the affected shoulder, lesions of the rotator cuff, osteoporosis, formation of subchondral cysts, prior infection; secondary arthritis due to instability, fracture sequelae, or rheumatoid arthritis
Interventions	Total shoulder replacement in 2 arms • Stemless cementless humeral component (Eclipse; Arthrex, Freiham, Germany) (n = 20 - randomised) • Stemmed humeral component (Univers II; Arthrex, Freiham, Germany) (n = 20 - randomised) The glenoid was resurfaced using a cementless metal-backed (Univers II, Arthrex) component or a cemented all-polyethylene keeled component (Arthrex). This was not part of the randomisation and was

Uschok 2017 *(Continued)*

	due to a change in the authors' philosophy regarding glenoid components, based on poorly reported results of uncemented metal-backed components

Outcomes	Followed up at 2 and 5 years
	
	Clinical outcomes reported at 2 and 5 years, radiological outcomes at 5 years only
	
	• Pain: subdomain of Constant Murley Score (0, 5, 10, 15 scale, higher = better, reported as mean ± SD)
	• Function
	  ○ Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
	  ○ Age- and gender-adjusted Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)
	• Revision, re-operation, or treatment failure
	  ○ Reported as percentages - the overall percentage revision rate for the study is reported as the sum of the percentage revision rates for the 3 arms. This is an incongruous statement, and the raw numbers were reported very ambiguously in the published study, so the outcome has not been included in the review
	• Physician evaluated
	  ○ Radiographic superior migration of humeral head (change in height preoperative to 5-year follow-up, mm, mean ± SD)
	  ○ Radiographic lucency around humeral components (not included in this review: measured via not directly comparable scales)
	  ○ Radiographic lucency around glenoid components (not included in this review: confounded by the use of different glenoid components)
	  ○ Range of motion: measured in degrees and subdomain of Constant Murley Score (not included in this review)
	  ○ Strength: measured as subdomain of Constant Murley Score (not included in this review)
	  ○ Activity: measured as subdomain of Constant Murley Score (not included in this review)
	
	Additional radiographic parameters are reported regarding placement and inclination of the prosthesis. These do not explicitly represent outcomes and have not been reported in this review. The quoted paper (Takase 2004) also describes high levels of correlation between head measurements and offset measurements. Because the size of humeral head components is not stated, they cannot be assumed to be balanced

Notes	Source of funding: not stated Conflicts of interest: 2 of the trial authors report a financial interest in the study implants through patent fees and consulting fees

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "the patients were randomized into 2 groups" Comment: the method used to generate the sequence was not stated
Allocation concealment (selection bias)	Unclear risk	Allocation concealment is not described
Blinding of participants and personnel (performance bias)	Unclear risk	Blinding of patients and evaluators is not reported
Blinding of outcome assessment - self-reported outcomes	Unclear risk	Blinding of patients and evaluators is not reported

Uschok 2017 *(Continued)*

Blinding of outcome assessment - physician reported outcomes	High risk	Evaluators cannot be blinded for radiographic assessments because the implant type is apparent from the radiographs. In addition, the scales used to assess radiographic lucency between different humeral components are not directly comparable
Incomplete outcome data (attrition bias)	Unclear risk	33 of 40 patients (83%) were available at 2 years and 29 at 5 years. However, glenoid component allocation is described only for the 29 patients analysed at 5 years. This has the potential to be an important confounding factor and should be more clearly reported
Selective reporting (reporting bias)	Low risk	All outcomes reported at the stated endpoints
Major baseline imbalance	High risk	Baseline Constant Murley Scores were higher in the stemless group; this group also contained a higher proportion of metal-backed glenoid components. Both of these factors have been shown to affect outcomes and may bias this study
Differences in rehabilitation regime	Low risk	Quote: "the patients were immobilized postoperatively in a 30° abduction brace for 3 weeks. The rehabilitative program began directly postoperative with passive movement and a limited range of motion for 6 weeks. Starting from the seventh postoperative week, active and passive movements without a restriction to the range of motion and additional strengthening exercises were allowed"

ADLER: Activities of Daily Living and External Rotation.
ASA: American Society of Anesthesiologists.
ASES: American Shoulder and Elbow Surgeons Scale.
AVN: avascular necrosis.
BIO: bony increased offset.
BMI: body mass index.
CTA: cuff tear arthropathy.
DASH: Disability of the Arm, Shoulder, and Hand questionnaire.
MACTAR: McMaster Toronto Arthritis patient preference questionnaire.
OA: osteoarthritis.
RSA: radiostereometric analysis.
SANE: single-assessment numerical evaluation.
SD: standard deviation.
STD: standard offset.
TSR: total shoulder replacement.
VAS: visual analogue scale.
WOOS: Western Ontario Osteoarthritis of the Shoulder Index.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Berth 2013	Quasi-randomised study
Ding 2015	Study did not measure the outcomes of interest
Edwards 2007	Study did not measure the outcomes of interest
Hammond 2013	Non-randomised comparative study
Hendel 2012	Did not measure the outcomes of interest

Study	Reason for exclusion
Iannotti 2015	Did not measure the outcomes of interest
ISRCTN42881741	This study comparing the conformity of the glenoid and humeral bearing components in conventional TSR was registered with a start date of 2006. No results are available. Inclusion criteria also included some patients with rheumatoid arthritis, which does not fit with the criteria for this review
Kasten 2009	Non-randomised comparative study
Kircher 2009	This study does not appear to be truly randomised; there is no description of any randomised method used to allocate groups. The stated intention was to randomise into 2 groups of 10 patients; however 6 of the 10 patients in the control group had actually crossed over from the intervention group, replaced by 6 new patients in the intervention group Quote: "to gain the full number of n=10 patients for group 1, the intraoperative navigation was started in n=16 patients. The data of the 6 patients with aborted intraoperative navigation were included in group 2" Although this study was included in the previous version of this review (Singh 2010), the inclusion criteria in the protocol for this updated version specifically exclude quasi-randomised studies
Mariotti 2014	Quasi-randomised study
NCT01884077	This study comparing conventional TSR to reverse TSR for treatment of osteoarthritis (intact cuff) in older adults (aged 70 to 95 years) was terminated due to slow recruitment and poor follow-up

TSR: total shoulder replacement.

Characteristics of ongoing studies *[ordered by study ID]*

NCT01288066

Trial name or title	A Randomized Multicenter Study Comparing the Effectiveness of Hemi- Versus Total Shoulder Arthroplasty in Patients With a Degenerative Joint Disease
Methods	Randomised controlled trial
Participants	76 patients aged 18 years and older with primary or secondary shoulder arthrosis. Glenoid
Interventions	Standard hemiarthroplasty vs total shoulder replacement
Outcomes	Constant Score, Shoulder Pain and Disability Index, duration of surgery, adverse events, revision rate, quality of life (EQ-5D)
Starting date	September 2011
Contact information	Norbert Suedkamp, Universitätsklinikum Freiburg, Freiburg, Germany
Notes	Estimated study completion date February 2022

NCT01404143

Trial name or title	Comparison of Two Methods of Subscapularis Management in Shoulder Arthroplasty: Tenotomy Versus Peel: A Multicenter, Randomized Controlled Trial

NCT01404143 *(Continued)*

Methods	Randomised controlled trial
Participants	100 patients aged 18 to 90 years with imaging and intraoperative findings confirming advanced humeral head cartilage loss and/glenoid cartilage loss. Persistent pain and functional disability for at least 6 months and failure of 6 months of conservative treatment
Interventions	Subscapularis tenotomy vs subscapularis peel for shoulder replacement
Outcomes	Subscapularis strength, WOOS, ASES, Constant Score, tendon-healing rate on ultrasound
Starting date	August 2011
Contact information	Peter Lapner, The Ottawa Hospital, Ottawa, Ontario, Canada, K1H 8L6
Notes	Anticipated study completion September 2017. Results not available at the time of searches and review completion

NCT01587560

Trial name or title	A Comparison Between a Pyrocarbon and a CoCr Shoulder Resurfacing Implant
Methods	Randomised controlled trial
Participants	80 patients aged 40 to 75 years with primary or secondary osteoarthritis of the shoulder
Interventions	Pyrocarbon resurfacing shoulder replacement vs cobalt chrome resurfacing shoulder replacement
Outcomes	Fixation to bone measure by radiostereometric analysis, EuroQol-5D (EQ-5D), American Shoulder and Elbow Surgeons Score (ASES), Constant Score, and Western Ontario Osteoarthritis of the Shoulder index (WOOS)
Starting date	January 2012
Contact information	Olof Skoldenberg, Director of Research, Orthopedic Department, Danderyd Hospital, Stockholm, Sweden 18288
Notes	Estimated study completion date January 2018. Results not yet publicly available

NCT01697865

Trial name or title	Reverse Shoulder Arthroplasty With or Without Concomitant Latissimus and Teres Major Transfer for Shoulder Pseudoparalysis With Teres Minor Dysfunction: A Prospective, Randomized Investigation
Methods	Randomised controlled trial
Participants	42 patients with • Shoulder pseudoparalysis due to chronic rotator cuff dysfunction with or without glenohumeral arthritis • Chronic rotator cuff tear with severe retraction, atrophy, fatty infiltration • Active forward elevation < 90 degrees • Teres minor dysfunction

NCT01697865 *(Continued)*

	• Positive lag and hornblower sign • Grade 2 or greater fatty infiltration of the teres minor and infraspinatus seen on MRI • Meeting all criteria to have a latissimus and teres major transfer
Interventions	Reverse total shoulder replacement in 2 arms: (1) includes a concomitant latissimus and teres major transfer (transfer group), (2) does not include a concomitant latissimus and teres major transfer (control group)
Outcomes	Activities of Daily Living and External Rotation (ADLER) Score, DASH Score, ASES Score, SF-12 Score, range of motion, X-ray measures
Starting date	October 2012
Contact information	Susan Odum, OrthoCarolina Research Institute, Charlotte, North Carolina, United States 28207 Susan.Odum@Orthocarolina.com
Notes	Estimated study completion in October 2019

NCT01790113

Trial name or title	A Prospective, Randomized, Multicenter Study Comparing the Safety and Effectiveness of Arthrex's Eclipse™ Shoulder Prothesis to the Univers™ II Shoulder Prosthesis in Patients With a Degenerative Joint Disease
Methods	Randomised open-label study
Participants	350 adults > 21 years with primary or secondary osteoarthritis, avascular necrosis, or rheumatoid arthritis
Interventions	Conventional total shoulder replacement with stemless vs stemmed humeral component
Outcomes	Composite Score using Constant Murley Score and radiological outcome
Starting date	January 2013
Contact information	Melissa Hirschberg, Arthrex, Inc.
Notes	Estimated study completion December 2020. From the clinicaltrials.gov entry, it is ambiguous whether this will be a randomised study due to contradictory statements

NCT02305966

Trial name or title	Evaluation of Implant Fixation in Reverse Total Shoulder Arthroplasty
Methods	Randomised controlled trial
Participants	40 patients aged 50 to 85 years with rotator cuff tear arthropathy requiring reverse total shoulder replacement
Interventions	Reverse total shoulder replacement in four arms: (1) Pressfit humerus and non-lateralised glenoid, (2) Pressfit humerus and lateralised glenoid, (3) cemented humerus and non-lateralised glenoid, (4) cemented humerus and lateralised glenoid

NCT02305966 *(Continued)*

Outcomes	Component migration measured by radiostereometric analysis
Starting date	August 2017
Contact information	Associate Professor George Athwal, Lawson Health Research Institute, London, Ontario, Canada
Notes	Estimated study completion date November 2019

NCT02768597

Trial name or title	A Prospective, Randomized Study Comparing the Outcome of Large-Diameter vs Small-Diameter Glenospheres in Primary Reverse Shoulder Arthroplasty Using the ReUnion System
Methods	Randomised controlled trial
Participants	220 participants aged 50 to 90 requiring a primary reverse total shoulder replacement for a diagnosis of cuff tear arthropathy (CTA), massive irreparable rotator cuff tear (MRCT), or osteoarthritis (OA) with marked posterior subluxation or bone loss
Interventions	Reverse total shoulder replacement in 4 arms: (1) large-diameter glenosphere +2 mm offset, (2) small-diameter glenosphere +2 mm offset, (3) large-diameter glenosphere +6 mm offset, (4) small-diameter glenosphere +6 mm offset
Outcomes	Shoulder range of motion, Oxford Shoulder Score, American Shoulder and Elbow Surgeons Scale, Quick-DASH questionnaire
Starting date	April 2016
Contact information	Mark Morrey, Mayo Clinic, Rochester, MN, United States 55905
Notes	Estimated study completion date February 2022

NCT02966886

Trial name or title	Comparison of Techniques in the Management of Glenoid Deficiencies in Shoulder Arthroplasty
Methods	Randomised controlled trial (3 parallel arms)
Participants	160 patients aged 18 years and older with shoulder osteoarthritis and B2-type glenoid retroversion
Interventions	Conventional total shoulder replacement with glenoid technique in 3 arms: (1) eccentric reaming, (2) augmented glenoid component, (3) posterior glenoid bone grafting
Outcomes	At 24 months: WOOS Index, Constant Murley Shoulder Score, ASES Shoulder Score, EQ-5D-5L, Shoulder Health Utilization Assessment Form
Starting date	January 2017
Contact information	Peter Lapner, The Ottawa Hospital, Ontario, Canada 6137378899, ext 78377; plapner@toh.on.ca
Notes	Estimated study completion date January 2021

NCT03111147

Trial name or title	The Impact of Humeral Component Version on Outcomes Following Reverse Total Shoulder Arthroplasty: A Prospective Randomized Controlled Trial
Methods	Randomised controlled trial
Participants	70 adults aged 18 years and older undergoing reverse total shoulder replacement for cuff tear arthropathy or osteoarthritis associated with a rotator cuff tear
Interventions	Reverse total shoulder replacement with the stemmed humeral component implanted in 0 degrees version vs 30 degrees retroversion
Outcomes	Range of motion at 2 years
Starting date	April 2017
Contact information	Lisa Motowski, Beaumont Helth, Royal Oak, MI, United States 248-551-6679; lisa.stellon@beaumont.org
Notes	Estimated study completion date April 2020

NCT03711175

Trial name or title	The Influence of Repairing the Sub-scapularis on Outcomes After Reverse Arthroplasty
Methods	Randomised controlled trial
Participants	200 adults aged 21 years and older with severe glenohumeral arthropathy and a grossly deficient rotator cuff
Interventions	Reverse total shoulder replacement (AltiVate system) with repair of subscapularis vs no repair of subscapularis
Outcomes	Physician-evaluated strength, range of motion, pain/function using the ASES Shoulder Score/SST, general health using the VR-12, radiographic assessments, adverse events
Starting date	September 2018
Contact information	Lisa Holt, Encore Medical; lisa.holt@djoglobal.com
Notes	Estimated study completion date December 2028

NCT03727490

Trial name or title	The Role of Subscapularis Repair in Reverse Shoulder Arthroplasty
Methods	Randomised controlled trial
Participants	100 adults undergoing reverse total shoulder replacement

NCT03727490 *(Continued)*

Interventions	Reverse total shoulder replacement with bone-to-bone repair of subscapularis tendon vs no repair of subscapularis tendon
Outcomes	Function measured using shoulder score and Simple Shoulder Test Score at 2 years
Starting date	January 2017
Contact information	Julianne Sefko, 314-747-2496; sefkoj@wudosis.wustl.edu
Notes	Estimated study completion date January 2021

NCT03730597

Trial name or title	Influence of Glenosphere Size in Scapular Notch Development. Randomized Study Comparing 42 to 38 ECC
Methods	Randomised controlled trial
Participants	95 patients undergoing reverse shoulder replacement for rotator cuff disorders and acute fractures
Interventions	Reverse total shoulder replacement using 42-mm-diameter glenosphere vs 38-mm-diameter glenosphere
Outcomes	Radiological evidence of scapular notching
Starting date	September 2014
Contact information	Carlos Torrens, Head of Shoulder Unit, Hospital del Mar
Notes	Study completed in September 2016 but submitted in October 2018 Results not yet available. Inclusion in subsequent versions of this review will depend on availability of subgroup data (excluding trauma cases)

ADLER: Activities of Daily Living and External Rotation.
ASES: American Shoulder and Elbow Surgeons Scale.
DASH: Disability of the Arm, Shoulder, and Hand questionnaire.
EQ-5D: EuroQoL Group Quality of Life Questionnaire based on five dimensions.
EQ-5D-5L: EuroQoL Group Quality of Life Questionnaire based on five-level scale.
SF-12: Short Form-12.
SST: Simple Shoulder Test.
VR-12: Veterans RAND 12-item health survey.
WOOS: Western Ontario Osteoarthritis of the Shoulder Index.

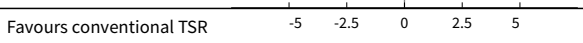
DATA AND ANALYSES

Comparison 1. Conventional stemmed TSR vs stemmed humeral hemiarthroplasty

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain: Visual Analogue Scale (0 to 10, lower = better)	2		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Intermediate	2	92	Mean Difference (IV, Random, 95% CI)	-1.49 [-2.88, -0.10]
2 Disability/Function: WOOS Index (0 to 100, higher = better)	2		Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Intermediate	2	92	Mean Difference (IV, Random, 95% CI)	10.57 [2.11, 19.02]
3 Quality of life: Short Form-12 mental component (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
4 Quality of life: Short Form-12 physical component (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
5 Adverse events: total (cumulative counts)	2		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
5.1 Short-term	2	75	Risk Ratio (M-H, Random, 95% CI)	1.60 [0.36, 7.05]
5.2 Intermediate	1	42	Risk Ratio (M-H, Random, 95% CI)	0.5 [0.14, 1.74]
6 Adverse events: serious (counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
6.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
7 Revision, re-operation, or treatment failure (cumulative counts)	2		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
7.1 Intermediate	2	92	Risk Ratio (M-H, Random, 95% CI)	1.29 [0.30, 5.53]
7.2 Long-term	1	51	Risk Ratio (M-H, Random, 95% CI)	0.89 [0.20, 4.00]

Analysis 1.1. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 1 Pain: Visual Analogue Scale (0 to 10, lower = better).

Study or subgroup	Conventional stemmed TSR N	Mean(SD)	Stemmed humeral hemiarthroplasty N	Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
1.1.1 Intermediate							
Gartsman 2000	27	1.8 (2.4)	24	4 (2.4)		49.88%	-2.2[-3.52,-0.88]
Lo 2005	20	0.6 (1.4)	21	1.4 (2.7)		50.12%	-0.78[-2.09,0.53]
Subtotal ***	**47**		**45**			**100%**	**-1.49[-2.88,-0.1]**
Heterogeneity: Tau^2=0.56; Chi^2=2.24, df=1(P=0.13); I^2=55.27%							
Test for overall effect: Z=2.1(P=0.04)							



Analysis 1.2. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 2 Disability/Function: WOOS Index (0 to 100, higher = better).

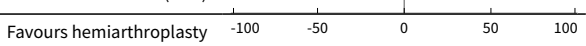
Study or subgroup	Conventional stemmed TSR N	Mean(SD)	Stemmed humeral hemiarthroplasty N	Mean(SD)	Mean Difference Random, 95% CI	Weight	Mean Difference Random, 95% CI
1.2.1 Intermediate							
Gartsman 2000	27	77.3 (18.2)	24	65.2 (24.9)		48.84%	12.1[0,24.2]
Lo 2005	20	90.6 (13.2)	21	81.5 (24.1)		51.16%	9.1[-2.72,20.92]
Subtotal ***	**47**		**45**			**100%**	**10.57[2.11,19.02]**
Heterogeneity: Tau^2=0; Chi^2=0.12, df=1(P=0.73); I^2=0%							
Test for overall effect: Z=2.45(P=0.01)							

Analysis 1.3. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 3 Quality of life: Short Form-12 mental component (0 to 100 scale, higher = better).

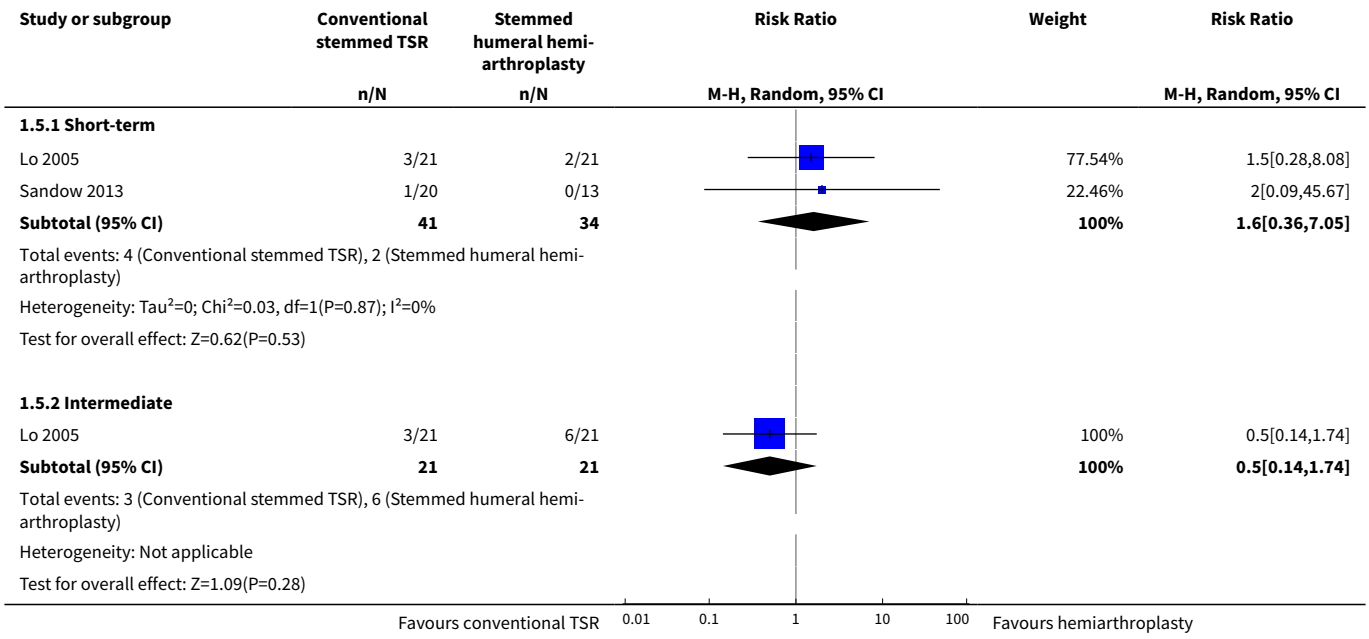
Study or subgroup	Conventional stemmed TSR N	Mean(SD)	Stemmed humeral hemiarthroplasty N	Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
1.3.1 Intermediate						
Lo 2005	20	58.4 (9.1)	21	57.4 (10.9)		1[-5.14,7.14]

Analysis 1.4. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 4 Quality of life: Short Form-12 physical component (0 to 100 scale, higher = better).

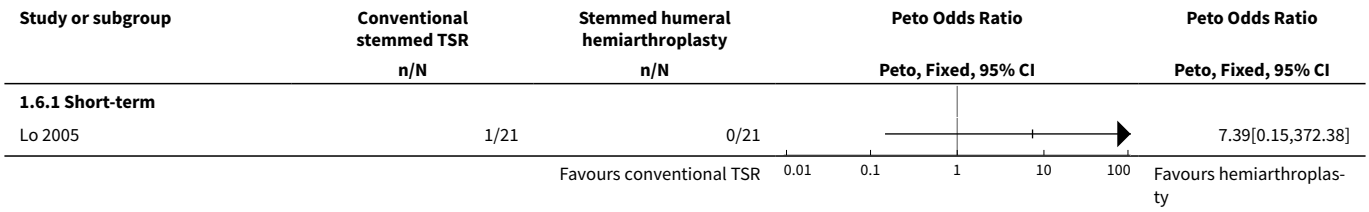
Study or subgroup	Conventional stemmed TSR N	Mean(SD)	Stemmed humeral hemiarthroplasty N	Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
1.4.1 Intermediate						
Lo 2005	20	42.1 (13.2)	21	42.9 (10.9)		-0.8[-8.23,6.63]



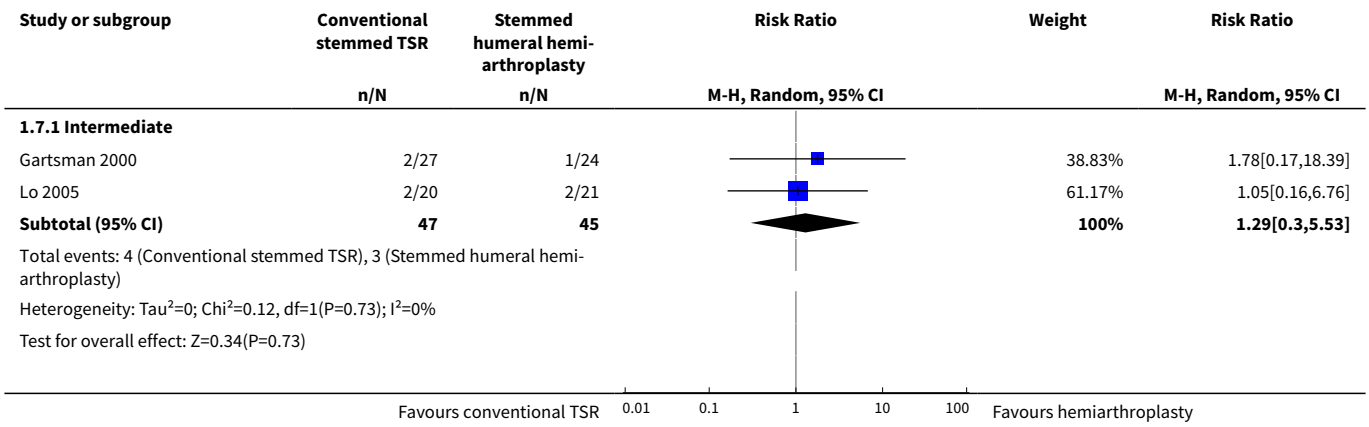
Analysis 1.5. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 5 Adverse events: total (cumulative counts).

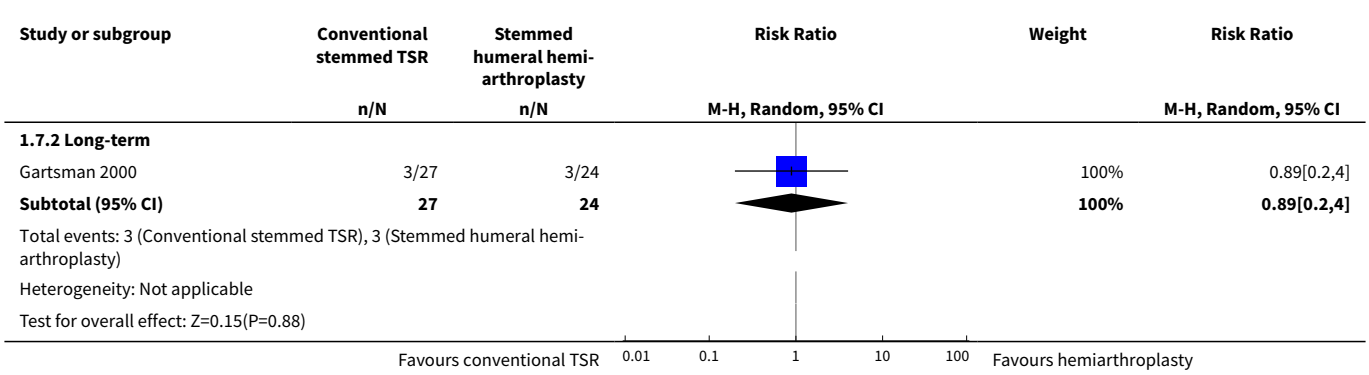


Analysis 1.6. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 6 Adverse events: serious (counts).



Analysis 1.7. Comparison 1 Conventional stemmed TSR vs stemmed humeral hemiarthroplasty, Outcome 7 Revision, re-operation, or treatment failure (cumulative counts).

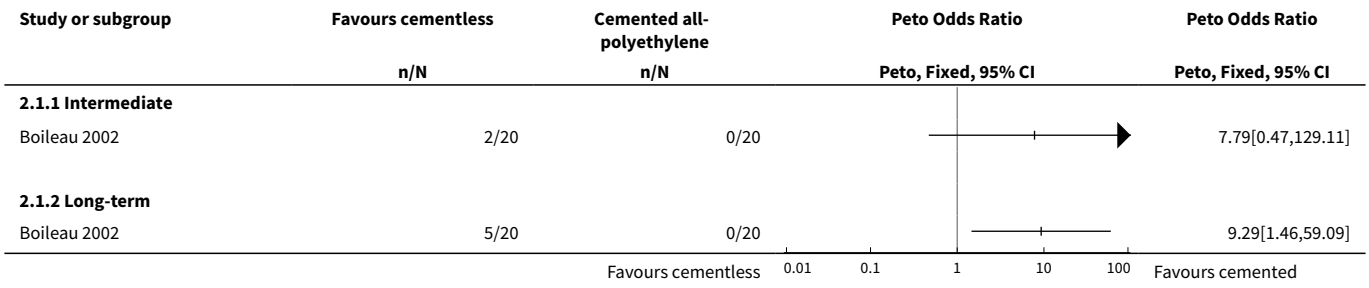




Comparison 2. Conventional stemmed TSR with cemented polyethylene glenoid component vs uncemented metal-backed glenoid component

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Revision, re-operation, or treatment failure (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
1.1 Intermediate	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
1.2 Long-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]

Analysis 2.1. Comparison 2 Conventional stemmed TSR with cemented polyethylene glenoid component vs uncemented metal-backed glenoid component, Outcome 1 Revision, re-operation, or treatment failure (cumulative counts).

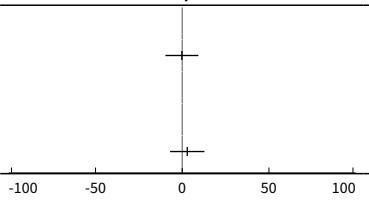


Comparison 3. Conventional stemless TSR vs conventional stemmed TSR

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: Constant Murley Score (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Long-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]

Analysis 3.1. Comparison 3 Conventional stemless TSR vs conventional stemmed TSR, Outcome 1 Disability/Function: Constant Murley Score (0 to 100 scale, higher = better).

Study or subgroup	Conventional stemless TSR N	Conventional stemless TSR Mean(SD)	Conventional stemmed TSR N	Conventional stemmed TSR Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
3.1.1 Intermediate						
Uschok 2017	15	65.5 (15.4)	18	65.7 (11.7)		-0.2[-9.68,9.28]
3.1.2 Long-term						
Uschok 2017	14	72.8 (11.8)	15	69.9 (15.3)		2.9[-7.01,12.81]

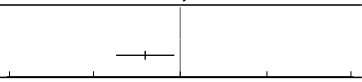


Comparison 4. Resurfacing humeral hemiarthroplasty vs stemmed humeral hemiarthroplasty

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: WOOS Index (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Short-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Adverse events: total (cumulative counts)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Short-term	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

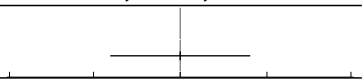
Analysis 4.1. Comparison 4 Resurfacing humeral hemiarthroplasty vs stemmed humeral hemiarthroplasty, Outcome 1 Disability/Function: WOOS Index (0 to 100 scale, higher = better).

Study or subgroup	Resurfacing humeral hemiarthroplasty N	Resurfacing humeral hemiarthroplasty Mean(SD)	Stemmed humeral hemiarthroplasty N	Stemmed humeral hemiarthroplasty Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
4.1.1 Short-term						
Rasmussen 2015	19	59.2 (26.4)	19	79.4 (26.4)		-20.2[-36.99,-3.41]



Analysis 4.2. Comparison 4 Resurfacing humeral hemiarthroplasty vs stemmed humeral hemiarthroplasty, Outcome 2 Adverse events: total (cumulative counts).

Study or subgroup	Resurfacing humeral hemiarthroplasty n/N	Stemmed humeral hemiarthroplasty n/N	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
4.2.1 Short-term				
Rasmussen 2015	2/20	2/20		1[0.16,6.42]

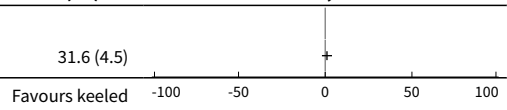


Comparison 5. Conventional stemmed TSR with pegged glenoid component vs keeled glenoid component

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: WOOS Index (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Long-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Revision, re-operation, or treatment failure (cumulative counts)	2		Peto Odds Ratio (Peto, Fixed, 95% CI)	Subtotals only
2.1 Intermediate	2	80	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.35 [0.05, 2.56]
2.2 Long-term	1	59	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.33 [0.08, 1.46]
3 Physician-evaluated: glenoid lucency grade (0 to 5 grade, higher = worse, reported as count graded ≥ 4)	2		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
3.1 Intermediate	2	71	Risk Ratio (M-H, Random, 95% CI)	0.38 [0.02, 8.83]
3.2 Long-term	1	38	Risk Ratio (M-H, Random, 95% CI)	1.20 [0.55, 2.63]

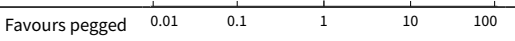
Analysis 5.1. Comparison 5 Conventional stemmed TSR with pegged glenoid component vs keeled glenoid component, Outcome 1 Disability/Function: WOOS Index (0 to 100 scale, higher = better).

Study or subgroup	Pegged glenoid N	Pegged glenoid Mean(SD)	Keeled glenoid N	Keeled glenoid Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
5.1.1 Long-term						
Edwards 2010	16	32.6 (4.5)	22	31.6 (4.5)		1[-1.9,3.9]



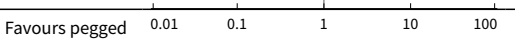
Analysis 5.2. Comparison 5 Conventional stemmed TSR with pegged glenoid component vs keeled glenoid component, Outcome 2 Revision, re-operation, or treatment failure (cumulative counts).

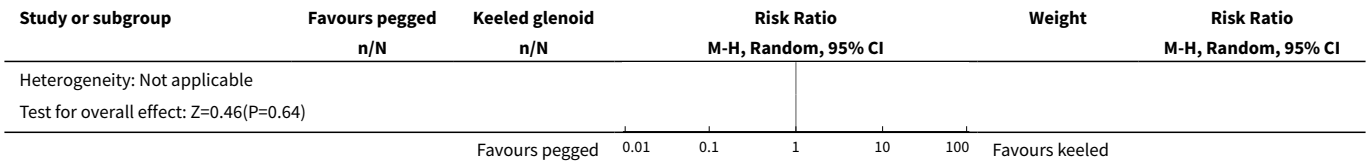
Study or subgroup	Pegged glenoid n/N	Keeled glenoid n/N	Peto Odds Ratio Peto, Fixed, 95% CI	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
5.2.1 Intermediate					
Edwards 2010	0/26	2/27		50.52%	0.14[0.01,2.22]
Rahme 2009	1/14	1/13		49.48%	0.93[0.05,15.67]
Subtotal (95% CI)	**40**	**40**		**100%**	**0.35[0.05,2.56]**
Total events: 1 (Pegged glenoid), 3 (Keeled glenoid)					
Heterogeneity: Tau^2=0; Chi^2=0.9, df=1(P=0.34); I^2=0%					
Test for overall effect: Z=1.03(P=0.3)					
5.2.2 Long-term					
Edwards 2010	2/29	6/30		100%	0.33[0.08,1.46]
Subtotal (95% CI)	**29**	**30**		**100%**	**0.33[0.08,1.46]**
Total events: 2 (Pegged glenoid), 6 (Keeled glenoid)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.46(P=0.15)					



Analysis 5.3. Comparison 5 Conventional stemmed TSR with pegged glenoid component vs keeled glenoid component, Outcome 3 Physician-evaluated: glenoid lucency grade (0 to 5 grade, higher = worse, reported as count graded ≥ 4).

Study or subgroup	Favours pegged n/N	Keeled glenoid n/N	Risk Ratio M-H, Random, 95% CI	Weight	Risk Ratio M-H, Random, 95% CI
5.3.1 Intermediate					
Edwards 2010	0/21	1/24		100%	0.38[0.02,8.83]
Rahme 2009	0/14	0/12			Not estimable
Subtotal (95% CI)	**35**	**36**		**100%**	**0.38[0.02,8.83]**
Total events: 0 (Favours pegged), 1 (Keeled glenoid)					
Heterogeneity: Not applicable					
Test for overall effect: Z=0.6(P=0.55)					
5.3.2 Long-term					
Edwards 2010	7/16	8/22		100%	1.2[0.55,2.63]
Subtotal (95% CI)	**16**	**22**		**100%**	**1.2[0.55,2.63]**
Total events: 7 (Favours pegged), 8 (Keeled glenoid)					





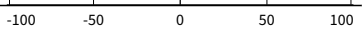
Comparison 6. Conventional stemmed TSR with cemented humeral component vs uncemented humeral component

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: WOOS Index (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Short-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Quality of life: Short Form-12 mental component (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 Short-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2.2 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Quality of life: Short Form-12 physical component (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1 Short-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3.2 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
4 Adverse events: total (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
4.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
4.2 Intermediate	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
5 Adverse events: serious (counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
5.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6 Revision, re-operation, or treatment failure (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
6.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
6.2 Intermediate	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]

Analysis 6.1. Comparison 6 Conventional stemmed TSR with cemented humeral component vs uncemented humeral component, Outcome 1 Disability/Function: WOOS Index (0 to 100 scale, higher = better).

Study or subgroup	Cemented humeral stem		Uncemented humeral stem		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI	Random, 95% CI
6.1.1 Short-term						
Litchfield 2011	78	88.5 (15.1)	74	79.6 (22.2)		8.9[2.83,14.97]
6.1.2 Intermediate						
Litchfield 2011	78	87.9 (15.4)	74	79.3 (22.8)		8.6[2.38,14.82]

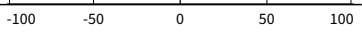


Analysis 6.2. Comparison 6 Conventional stemmed TSR with cemented humeral component vs uncemented humeral component, Outcome 2 Quality of life: Short Form-12 mental component (0 to 100 scale, higher = better).

Study or subgroup	Cemented humeral stem		Uncemented humeral stem		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI	Random, 95% CI
6.2.1 Short-term						
Litchfield 2011	78	54.2 (9.5)	74	50.6 (10.2)		3.6[0.46,6.74]
6.2.2 Intermediate						
Litchfield 2011	78	54.1 (9.1)	74	51.5 (9.9)		2.59[-0.44,5.62]

Analysis 6.3. Comparison 6 Conventional stemmed TSR with cemented humeral component vs uncemented humeral component, Outcome 3 Quality of life: Short Form-12 physical component (0 to 100 scale, higher = better).

Study or subgroup	Cemented humeral stem		Uncemented humeral stem		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI	Random, 95% CI
6.3.1 Short-term						
Litchfield 2011	78	42.1 (11.4)	74	42 (10.6)		0.1[-3.4,3.6]



Study or subgroup	Cemented humeral stem		Uncemented humeral stem		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI	Random, 95% CI
6.3.2 Intermediate						
Litchfield 2011	78	43 (11.8)	74	39.2 (11.6)		3.77[0.05,7.49]

Favours uncemented stem -100 -50 0 50 100 Favours cemented stem

Analysis 6.4. Comparison 6 Conventional stemmed TSR with cemented humeral component vs uncemented humeral component, Outcome 4 Adverse events: total (cumulative counts).

Study or subgroup	Cemented humeral stem	Uncemented humeral stem	Peto Odds Ratio	Peto Odds Ratio
	n/N	n/N	Peto, Fixed, 95% CI	Peto, Fixed, 95% CI
6.4.1 Short-term				
Litchfield 2011	6/80	4/81		1.55[0.43,5.55]
6.4.2 Intermediate				
Litchfield 2011	6/80	4/81		1.55[0.43,5.55]

Favours cemented stem 0.01 0.1 1 10 100 Favours uncemented stem

Analysis 6.5. Comparison 6 Conventional stemmed TSR with cemented humeral component vs uncemented humeral component, Outcome 5 Adverse events: serious (counts).

Study or subgroup	Cemented humeral stem	Uncemented humeral stem	Peto Odds Ratio	Peto Odds Ratio
	n/N	n/N	Peto, Fixed, 95% CI	Peto, Fixed, 95% CI
6.5.1 Short-term				
Litchfield 2011	1/80	1/81		1.01[0.06,16.33]

Favours cemented stem 0.01 0.1 1 10 100 Favours uncemented stem

Analysis 6.6. Comparison 6 Conventional stemmed TSR with cemented humeral component vs uncemented humeral component, Outcome 6 Revision, re-operation, or treatment failure (cumulative counts).

Study or subgroup	Cemented humeral stem	Uncemented humeral stem	Peto Odds Ratio	Peto Odds Ratio
	n/N	n/N	Peto, Fixed, 95% CI	Peto, Fixed, 95% CI
6.6.1 Short-term				
Litchfield 2011	1/79	2/75		0.48[0.05,4.71]
6.6.2 Intermediate				
Litchfield 2011	4/78	3/74		1.27[0.28,5.79]

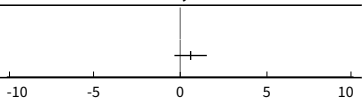
Favours cemented stem 0.01 0.1 1 10 100 Favours uncemented stem

Comparison 7. Conventional stemmed TSR via subscapularis-sparing approach ("sparing") vs standard approach ("standard")

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain: visual analogue scale (0 to 10 scale, lower = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Disability/Function: ASES Shoulder Score (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Revision, re-operation, or treatment failure (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
3.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]

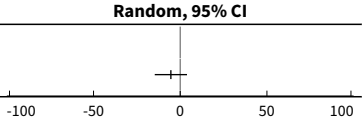
Analysis 7.1. Comparison 7 Conventional stemmed TSR via subscapularis-sparing approach ("sparing") vs standard approach ("standard"), Outcome 1 Pain: visual analogue scale (0 to 10 scale, lower = better).

Study or subgroup	Sparing N	Sparing Mean(SD)	Standard N	Standard Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
7.1.1 Intermediate						
Kwon 2019	32	1.6 (2.2)	38	1 (1.7)		0.6[-0.33,1.53]

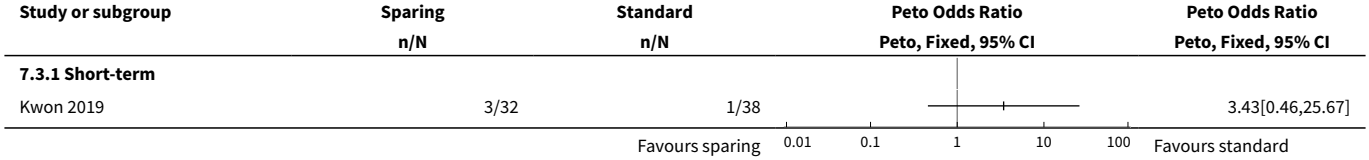


Analysis 7.2. Comparison 7 Conventional stemmed TSR via subscapularis-sparing approach ("sparing") vs standard approach ("standard"), Outcome 2 Disability/Function: ASES Shoulder Score (0 to 100 scale, higher = better).

Study or subgroup	Sparing N	Sparing Mean(SD)	Standard N	Standard Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
7.2.1 Intermediate						
Kwon 2019	32	81.7 (23.3)	38	87.1 (14.5)		-5.4[-14.7,3.9]



Analysis 7.3. Comparison 7 Conventional stemmed TSR via subscapularis-sparing approach ("sparing") vs standard approach ("standard"), Outcome 3 Revision, re-operation, or treatment failure (cumulative counts).

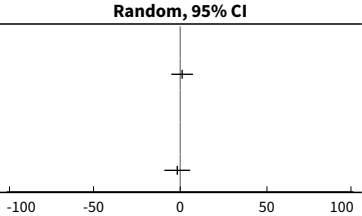


Comparison 8. Conventional stemmed TSR via a lesser tuberosity osteotomy approach compared to subscapularis tenotomy/peel

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: WOOS Index (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Short-term	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Adverse events: total (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Subtotals only
2.1 Short-term	1	59	Peto Odds Ratio (Peto, Fixed, 95% CI)	3.50 [0.57, 21.54]
3 Revision, re-operation, or treatment failure (cumulative counts)	2		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
3.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
3.2 Intermediate	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
4 Physician evaluated: radiographic evidence of healing of repair confirmed by CT (counts)	2		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
4.1 Short-term	2	140	Risk Ratio (M-H, Random, 95% CI)	0.99 [0.87, 1.13]

Analysis 8.1. Comparison 8 Conventional stemmed TSR via a lesser tuberosity osteotomy approach compared to subscapularis tenotomy/peel, Outcome 1 Disability/Function: WOOS Index (0 to 100 scale, higher = better).

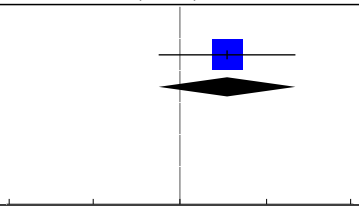
Study or subgroup	Lesser tuberosity osteotomy N	Lesser tuberosity osteotomy Mean(SD)	Tenotomy/peel N	Tenotomy/peel Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
8.1.1 Short-term						
Lapner 2012	36	88.2 (12.3)	37	87.1 (14.8)		1.1[-5.14,7.34]
8.1.2 Intermediate						
Lapner 2012	36	86.5 (16)	37	88.2 (16.5)		-1.7[-9.16,5.76]



Favours osteotomy — Favours tenotomy/peel

Analysis 8.2. Comparison 8 Conventional stemmed TSR via a lesser tuberosity osteotomy approach compared to subscapularis tenotomy/peel, Outcome 2 Adverse events: total (cumulative counts).

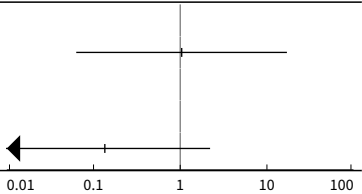
Study or subgroup	Lesser tuberosity osteotomy n/N	Tenotomy/peel n/N	Peto Odds Ratio Peto, Fixed, 95% CI	Weight	Peto Odds Ratio Peto, Fixed, 95% CI
8.2.1 Short-term					
Levine 2019	4/30	1/29		100%	3.5[0.57,21.54]
Subtotal (95% CI)	**30**	**29**		**100%**	**3.5[0.57,21.54]**
Total events: 4 (Lesser tuberosity osteotomy), 1 (Tenotomy/peel)					
Heterogeneity: Not applicable					
Test for overall effect: Z=1.35(P=0.18)					



Favours tenotomy/peel — Favours osteotomy

Analysis 8.3. Comparison 8 Conventional stemmed TSR via a lesser tuberosity osteotomy approach compared to subscapularis tenotomy/peel, Outcome 3 Revision, re-operation, or treatment failure (cumulative counts).

Study or subgroup	Lesser tuberosity osteotomy n/N	Tenotomy/peel n/N	Peto Odds Ratio Peto, Fixed, 95% CI	Peto Odds Ratio Peto, Fixed, 95% CI
8.3.1 Short-term				
Levine 2019	1/29	1/30		1.04[0.06,16.96]
8.3.2 Intermediate				
Lapner 2012	0/36	2/37		0.14[0.01,2.21]

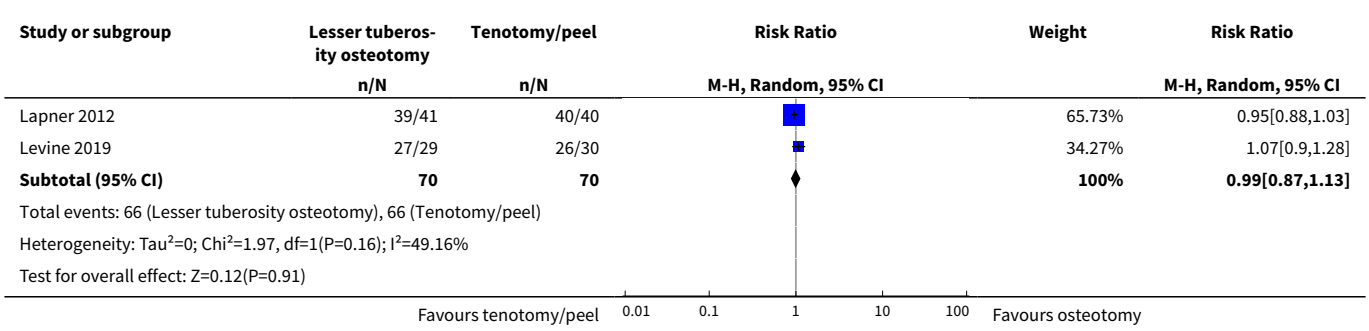


Favours osteotomy — Favours tenotomy/peel

Analysis 8.4. Comparison 8 Conventional stemmed TSR via a lesser tuberosity osteotomy approach compared to subscapularis tenotomy/peel, Outcome 4 Physician evaluated: radiographic evidence of healing of repair confirmed by CT (counts).

Study or subgroup	Lesser tuberosity osteotomy n/N	Tenotomy/peel n/N	Risk Ratio M-H, Random, 95% CI	Weight	Risk Ratio M-H, Random, 95% CI
8.4.1 Short-term					

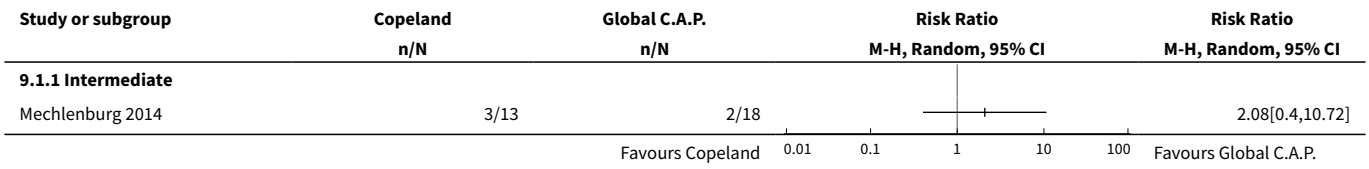
Favours tenotomy/peel — Favours osteotomy



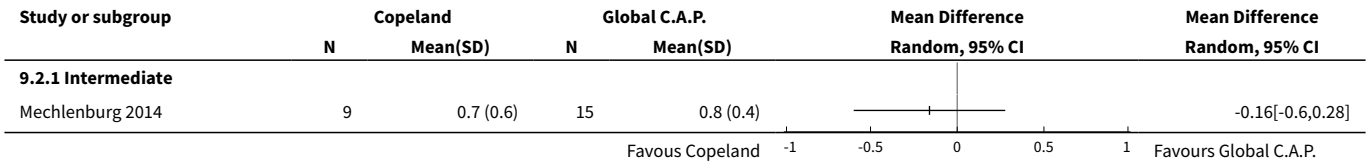
Comparison 9. Resurfacing humeral hemiarthroplasty with Copeland implant vs Global C.A.P. implant

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Revision, re-operation, or treatment failure (cumulative counts)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
2 Physician-evaluated: radiostereometric analysis total translation (mm)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]

Analysis 9.1. Comparison 9 Resurfacing humeral hemiarthroplasty with Copeland implant vs Global C.A.P. implant, Outcome 1 Revision, re-operation, or treatment failure (cumulative counts).



Analysis 9.2. Comparison 9 Resurfacing humeral hemiarthroplasty with Copeland implant vs Global C.A.P. implant, Outcome 2 Physician-evaluated: radiostereometric analysis total translation (mm).

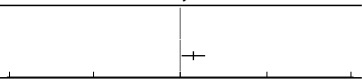


Comparison 10. Reverse polarity stemmed TSR via neutral glenosphere position vs inferior tilted glenosphere

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: ASES Shoulder Score (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Adverse events: total (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
2.1 Short-term	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
3 Physician evaluated: radiographic evidence of glenoid notching (Nerot grade ≥ 1 count)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

Analysis 10.1. Comparison 10 Reverse polarity stemmed TSR via neutral glenosphere position vs inferior tilted glenosphere, Outcome 1 Disability/Function: ASES Shoulder Score (0 to 100 scale, higher = better).

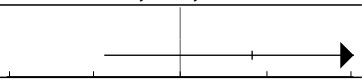
Study or subgroup	Neutral glenosphere		Inferior tilted glenosphere		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	Random, 95% CI	Random, 95% CI
10.1.1 Intermediate						
Edwards 2012	22	86.5 (11.6)	20	78.9 (10.8)		7.6[0.83,14.37]



Favours neutral glenosphere | Favours inferior tilted glenosphere

Analysis 10.2. Comparison 10 Reverse polarity stemmed TSR via neutral glenosphere position vs inferior tilted glenosphere, Outcome 2 Adverse events: total (cumulative counts).

Study or subgroup	Neutral glenosphere	Inferior tilted glenosphere	Peto Odds Ratio	Peto Odds Ratio
	n/N	n/N	Peto, Fixed, 95% CI	Peto, Fixed, 95% CI
10.2.1 Short-term				
Edwards 2012	1/22	0/20		6.75[0.13,341.54]



Favours neutral glenosphere | Favours inferior tilted glenosphere

Analysis 10.3. Comparison 10 Reverse polarity stemmed TSR via neutral glenosphere position vs inferior tilted glenosphere, Outcome 3 Physician evaluated: radiographic evidence of glenoid notching (Nerot grade ≥ 1 count).

Study or subgroup	Neutral glenosphere n/N	Inferior tilted glenosphere n/N	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
10.3.1 Intermediate				
Edwards 2012	19/22	15/20		1.15[0.85,1.56]
		Favours neutral glenosphere	0.01 0.1 1 10 100	Favours inferior tilted glenosphere

Comparison 11. Reverse polarity stemmed TSR via bony increased offset vs standard offset for glenoid component

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Disability/Function: Constant Murley Score (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Adverse events: total (cumulative counts)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

Analysis 11.1. Comparison 11 Reverse polarity stemmed TSR via bony increased offset vs standard offset for glenoid component, Outcome 1 Disability/Function: Constant Murley Score (0 to 100 scale, higher = better).

Study or subgroup	Bony increased offset N	Mean(SD)	Standard offset N	Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
11.1.1 Intermediate						
Greiner 2015	16	64.1 (18.4)	15	61.5 (16)		2.6[-9.52,14.72]
				Favours standard offset	-100 -50 0 50 100	Favours bony increased offset

Analysis 11.2. Comparison 11 Reverse polarity stemmed TSR via bony increased offset vs standard offset for glenoid component, Outcome 2 Adverse events: total (cumulative counts).

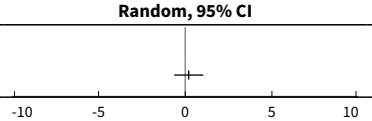
Study or subgroup	Bony increased offset n/N	Standard offset n/N	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
11.2.1 Intermediate				
Greiner 2015	2/16	2/15		0.94[0.15,5.84]
		Bony increased offset	0.01 0.1 1 10 100	Standard offset

Comparison 12. Reverse polarity stemmed TSR via eccentric glenosphere position vs concentric position

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain: visual analogue scale (0 to 10, lower = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Disability/Function: ASES Shoulder Score (0 to 100, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Adverse events: total (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
3.1 Intermediate	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
4 Revision, re-operation, or treatment failure (cumulative counts)	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	Totals not selected
4.1 Intermediate	1		Peto Odds Ratio (Peto, Fixed, 95% CI)	0.0 [0.0, 0.0]
5 Physician evaluated: radiographic evidence of glenoid notching (Nerot grade ≥ 1 count)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
5.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

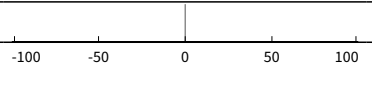
Analysis 12.1. Comparison 12 Reverse polarity stemmed TSR via eccentric glenosphere position vs concentric position, Outcome 1 Pain: visual analogue scale (0 to 10, lower = better).

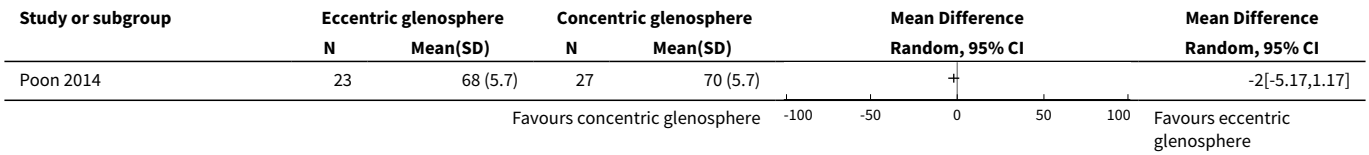
Study or subgroup	Eccentric glenosphere N	Eccentric glenosphere Mean(SD)	Concentric glenosphere N	Concentric glenosphere Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
12.1.1 Intermediate						
Poon 2014	23	1.8 (1.5)	27	1.6 (1.5)		0.2[-0.63,1.03]



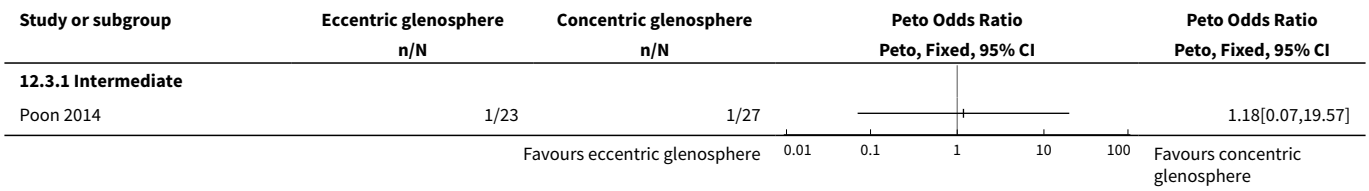
Analysis 12.2. Comparison 12 Reverse polarity stemmed TSR via eccentric glenosphere position vs concentric position, Outcome 2 Disability/Function: ASES Shoulder Score (0 to 100, higher = better).

Study or subgroup	Eccentric glenosphere N	Eccentric glenosphere Mean(SD)	Concentric glenosphere N	Concentric glenosphere Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
12.2.1 Intermediate						

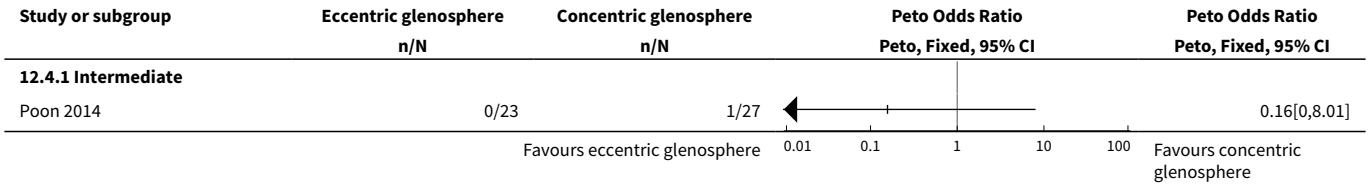




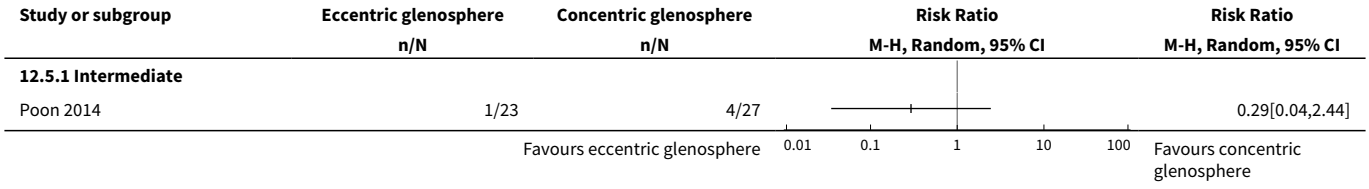
Analysis 12.3. Comparison 12 Reverse polarity stemmed TSR via eccentric glenosphere position vs concentric position, Outcome 3 Adverse events: total (cumulative counts).



Analysis 12.4. Comparison 12 Reverse polarity stemmed TSR via eccentric glenosphere position vs concentric position, Outcome 4 Revision, re-operation, or treatment failure (cumulative counts).



Analysis 12.5. Comparison 12 Reverse polarity stemmed TSR via eccentric glenosphere position vs concentric position, Outcome 5 Physician evaluated: radiographic evidence of glenoid notching (Nerot grade ≥ 1 count).



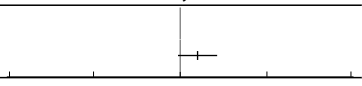
Comparison 13. Reverse polarity stemmed TSR via 135-degree vs 155-degree neck-shaft angle humeral component

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain: visual analogue scale (0 to 10, lower = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Disability/Function: ASES Shoulder Score (0 to 100 scale, higher = better)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Intermediate	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Adverse events: total (cumulative counts)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
4 Revision, re-operation, or treatment failure (cumulative counts)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
5 Physician evaluated: radiographic evidence of glenoid notching (Nerot grade ≥ 1 count)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
5.1 Intermediate	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

Analysis 13.1. Comparison 13 Reverse polarity stemmed TSR via 135-degree vs 155-degree neck-shaft angle humeral component, Outcome 1 Pain: visual analogue scale (0 to 10, lower = better).

Study or subgroup	135-degrees N	135-degrees Mean(SD)	155-degrees N	155-degrees Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
13.1.1 Intermediate						
Gobezie 2019	37	2 (2.9)	31	1 (1.8)		1[-0.13,2.13]



Analysis 13.2. Comparison 13 Reverse polarity stemmed TSR via 135-degree vs 155-degree neck-shaft angle humeral component, Outcome 2 Disability/Function: ASES Shoulder Score (0 to 100 scale, higher = better).

Study or subgroup	135-degrees N	135-degrees Mean(SD)	155-degrees N	155-degrees Mean(SD)	Mean Difference Random, 95% CI	Mean Difference Random, 95% CI
13.2.1 Intermediate						
Gobezie 2019	37	74 (24.6)	31	78 (15.1)		-4[-13.54,5.54]

Analysis 13.3. Comparison 13 Reverse polarity stemmed TSR via 135-degree vs 155-degree neck-shaft angle humeral component, Outcome 3 Adverse events: total (cumulative counts).

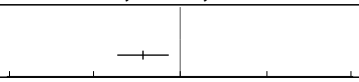
Study or subgroup	135-degrees n/N	155-degrees n/N	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
13.3.1 Intermediate				
Gobezie 2019	5/37	4/31		1.05[0.31,3.57]

Analysis 13.4. Comparison 13 Reverse polarity stemmed TSR via 135-degree vs 155-degree neck-shaft angle humeral component, Outcome 4 Revision, re-operation, or treatment failure (cumulative counts).

Study or subgroup	135-degrees n/N	155-degrees n/N	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
13.4.1 Intermediate				
Gobezie 2019	2/37	4/31		0.42[0.08,2.14]

Analysis 13.5. Comparison 13 Reverse polarity stemmed TSR via 135-degree vs 155-degree neck-shaft angle humeral component, Outcome 5 Physician evaluated: radiographic evidence of glenoid notching (Nerot grade ≥ 1 count).

Study or subgroup	135-degrees n/N	155-degrees n/N	Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
13.5.1 Intermediate				
Gobezie 2019	8/37	18/31		0.37[0.19,0.74]



ADDITIONAL TABLES

Table 1. Additional study data

Comparison	Study ID	Outcome	Measure	Time-point	Arm 1			Arm 2			Notes
					Description	n	Outcome	Description	n	Outcome	
Conventional TSR vs stemmed hemi-arthroplasty	Gartsman 2000	Function	UCLA Shoulder Rating Scale (0 to 35 scale, higher = better, reported as mean ± SD)	Intermediate	TSR	27	27.4 ± 4.9	Stemmed hemi-arthroplasty	24	23.2 ± 5.9	P = 0.008
	Lo 2005	Pain	McGill Pain Questionnaire (lower = better, reported as mean ± SD)	Intermediate	TSR	20	0.9 ± 1.4	Stemmed hemi-arthroplasty	21	2.7 ± 6.8	Scale used is unclear. Original questionnaire reference uses 0 to 78. This is incompatible with the values in this study
		Function	ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)	Intermediate	TSR	20	91.1 ± 14.3	Stemmed hemi-arthroplasty	21	83.1 ± 25.6	P = 0.25
			UCLA Shoulder Rating Scale (0 to 35 scale, higher = better, reported as mean ± SD)	Intermediate	TSR	20	26.7 ± 3.8	Stemmed hemi-arthroplasty	21	24.2 ± 5.0	P = 0.10
			Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)	Intermediate	TSR	20	70.8 ± 17.2	Stemmed hemi-arthroplasty	21	67.1 ± 19.6	P = 0.55
	Sandow 2013	Pain	Visual analogue scale (0 to 10 scale, higher = worse, reported as median and range)	Short-term	TSR	16	1 (0 to 2.8)	Stemmed hemi-	13	2 (0 to 8.8)	P < 0.05

Table 1. Additional study data *(Continued)*

							arthroplasty			
				Intermediate	TSR	11	0.2 (0 to 4)	Stemmed hemiarthroplasty	7	4.6 (0.4 to 8.5)	P < 0.05
		Function	UCLA Shoulder Rating Scale (0 to 35 scale, higher = better, reported as median and range)	Short-term	TSR	18	30 (21 to 35)	Stemmed hemiarthroplasty	13	29 (12 to 33)	P < 0.05
				Intermediate	TSR	11	33 (24 to 34)	Stemmed hemiarthroplasty	6	18.5 (10 to 25)	P < 0.05
			Constant Murley Score (0 to 100 scale, higher = better, reported as median and range)	Short-term	TSR	15	68 (48 to 89)	Stemmed hemiarthroplasty	10	59.5 (30 to 85)	
				Intermediate	TSR	6	77 (67 to 95)	Stemmed hemiarthroplasty	4	54.5 (43 to 59)	P < 0.05
Humeral head resurfacing vs stemmed hemiarthroplasty	Rasmussen 2015	Pain	Subdomain of Constant Murley Score (0, 5, 10, 15 scale, higher = better, reported as mean and range)	Intermediate	Humeral head resurfacing	19	11.1 (0 to 15)	Stemmed hemiarthroplasty	19	8.0 (0 to 15)	MD 3.2 (95% CI 0.1 to 6.2), P = 0.04
		Function	Constant Murley Score (0 to 100 scale, higher = better, reported as mean and range)	Short-term	Humeral head resurfacing	19	48.9 (6 to 80)	Stemmed hemiarthroplasty	19	59.1 (0 to 88)	P = 0.14
		Physician evaluated: op-	In minutes (continuous scale, lower presumed better, reported as mean and range)	Short-term	Humeral head resurfacing	20	52 (34 to 80)	Stemmed hemiarthroplasty	20	80 (56 to 103)	P < 0.001

Table 1. Additional study data *(Continued)*

		erating time									
Conventional stemless TSR vs conventional stemmed TSR	Uschok 2017	Pain	Subdomain of Constant Murley Score (0, 5, 10, 15 scale, higher = better, reported as mean ± SD)	Intermediate	Stemless TSR	15	10.9 ± 4.4	Stemmed TSR	18	13.6 ± 2.9	P = 0.136
				Long-term	Stemless TSR	14	12.7 ± 2.4	Stemmed TSR	15	12.4 ± 2.1	P = 0.590
Conventional TSR with cemented polyethylene glenoid component vs uncemented metal-backed glenoid component	Boileau 2002	Pain	Subdomain of Constant Murley Score (0, 5, 10, 15 scale, higher = better, reported as "average" and range)	Intermediate	Cemented glenoid	20	12.5 (4 to 15)	Uncemented glenoid	20	12 (5 to 15)	
				Long-term	Cemented glenoid	17	12 (0 to 15)	Uncemented glenoid	18	13 (3 to 15)	
		Function	Constant Murley Score (0 to 100, higher = better, reported as average and range)	Intermediate	Cemented glenoid	20	67 (6 to 89)	Uncemented glenoid	20	75 (17 to 89)	
				Long-term	Cemented glenoid	17	68 (6 to 92	Uncemented glenoid	18	73 (42 to 89)	
		Physician evaluated: Glenoid lucency	Novel 4-level grading system described but reported only as dichotomous outcome	Long-term	Cemented glenoid	20	17	Uncemented glenoid	20	5	Progression over time observed in only 4 cases in the uncemented group. None in the cemented group
Conventional TSR with	Nuttall 2007	Pain	Visual analogue scale (0 to 10, lower = better, reported as mean only)	Intermediated	Pegged	10	0.6	Keeled	10	0.6	

Table 1. Additional study data *(Continued)*

pegged glenoid component vs keeled glenoid component	Function	ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean only)	Intermediate	Pegged	10	78	Keeled	10	84	
		Constant Murley Score (0 to 100 scale, higher = better, reported as mean only)	Intermediate	Pegged	10	65	Keeled	10	62	
	Edwards 2010	Function	ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean with exact P value)	Long-term	Pegged	16	68	Keeled	22	67	P = 0.635 WOOS Index reported in main analyses
			Constant Murley Score (0 to 100 scale, higher = better, reported as mean with exact P value)	Long-term	Pegged	16	59.7	Keeled	22	58.9	P = 0.728 WOOS Index reported in main analyses
			Single-assessment numerical evaluation (0 to 100%, higher = better, reported as mean only)	Long-term	Pegged	16	58.7	Keeled	22	66.6	P = 0.247 WOOS Index already included in function analysis
	Rahme 2009	Function	Constant Murley Score (0 to 100 scale, higher = better, reported as mean only)	Intermediate	Pegged	14	70	Keeled	12	70	
			Subjective shoulder value (0 to 100%, higher = better, reported as mean only)	Intermediate	Pegged	14	80	Keeled	12	80	
	Gascoyne 2017	Function	WOOS Index (0 to 100 scale, reversed from normal here - higher = worse, reported as median only)	Short-term	Pegged	5	7.15	Keeled	6	34.7	Authors have used WOOS Index in op-

Table 1. Additional study data *(Continued)*

				Interme-diate	Pegged	4	22.3	Keeled	5	18.5	posite di-rection to the usu-al conven-tion
			ASES Shoulder Score (0 to 100 scale, higher = better, reported as median only)	Short-term	Pegged	5	97.1	Keeled	6	72.5	
				Interme-diate	Pegged	4	96.4	Keeled	5	73.5	
			Simple Shoulder Test Score (0 to 12, higher = better, reported as me-dian only)	Short-term	Pegged	5	11.0	Keeled	6	7.0	
				Interme-diate	Pegged	4	10.5	Keeled	5	6.0	
		Physi-cian evaluat-ed: ra-diostere-ometric analysis (RSA)	Coronal plane translation (mm, lower = better, reported as median only)	Short-term	Pegged	5	0.267	Keeled	6	1.518	P < 0.05
				Interme-diate	Pegged	4	0.235	Keeled	5	0.990	P < 0.05
			Coronal plane rotation (degrees, lower = better, reported as median only)	Short-term	Pegged	5	0.601	Keeled	6	0.307	
				Interme-diate	Pegged	4	1.074	Keeled	5	-0.624	
Conven-tion-al TSR with ce-mented stemmed humer-al com-ponent vs unce-mented stemmed humeral	Litch-field 2011	Function	ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)	Short-term	Cement-ed	78	70.2 ± 10.3	Unce-mented	74	66.2 ± 13.9	P = 0.09
				Interme-diate	Cement-ed	78	69.2 ± 13.3	Unce-mented	74	64.74 ± 15.7	P = 0.2
			MACTAR Score (0 to 500, lower = better)	Short-term	Cement-ed	78	50.6 ± 59.1	Unce-mented	74	70.1 ± 74.1	P = 0.19
				Interme-diate	Cement-ed	78	56.1 ± 76.6	Unce-mented	74	69.2 ± 77.7	P = 0.49

Table 1. Additional study data *(Continued)*

component											
Conventional stemmed TSR via lesser tuberosity osteotomy approach vs subscapularis tenotomy/peel	Lapner 2012	Function	ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean ± SD)	Short-term	Osteotomy	36	77.1 ± 23.7	Peel	37	81.3 ± 18.7	
				Intermediate	Osteotomy	36	79.4 ± 24.6	Peel	37	83.3 ± 19.0	
		Physician evaluated: fatty infiltration of rotator cuff muscles	Goutallier grade (0 to 4 grades, higher = worse, reported as mean ± SD)	Short-term	Osteotomy	41	0.90 ± 0.89	Peel	41	0.95 ± 0.85	
	Levine 2019	Pain	Visual analogue scale (0 to 10 scale, higher = worse, reported as mean)	Short-term	Osteotomy	29	1.8	Tenotomy	30	1.9	Inconsistencies between text and figures. Numbers reported here are from the text
		Function	ASES Shoulder Score (0 to 100 scale, higher = better, reported as mean)	Short-term	Osteotomy	29	75.6	Tenotomy	30	74.6	
			Simple Shoulder Test Score (0 to 10 scale, higher = better, reported as mean)	Short-term	Osteotomy	29	9.1	Tenotomy	30	7.6	
		Quality of life	Short Form-36 (0 to 100 scale, higher = better, reported as mean)	Short-term	Osteotomy	29	71.1	Tenotomy	30	64.9	
		Operative time	Minutes (mean, lower better)	Short-term	Osteotomy	29	152.7	Tenotomy	30	129.3	
Humeral head resurfacing with Copeland implant vs Glob-	Mechlenburg 2014	Function	WOOS Index (raw scale 0 to 1900, lower = better, reported in box plots as median plus 10th/25th/75th/90th centiles) - presented here as median (IQR)	Short-term	Copeland	10	298 (81 to 788)	Global C.A.P.	15	383 (115 to 822)	
				Intermediate	Copeland	10	128 (53 to 550)	Global C.A.P.	15	294 (111 to 477)	

Table 1. Additional study data *(Continued)*

al C.A.P. implant			Constant Murley Score (0 to 100 scale, higher = better, reported in box plots as median plus 10th/25th/75th/90th centiles) - presented here as median (IQR)	Short-term	Copeland	10	71.6 (59.6 to 87.7)	Global C.A.P.	15	72.7 (58.8 to 88.2)	
				Intermediate	Copeland	10	76.9 (61.1 to 81.2)	Global C.A.P.	15	72.6 (64.6 to 85.7)	
		Physician evaluated: bone mineral density of humeral head	Measured in g/cm^3 (continuous scale, higher = better, reported in box plots as median plus 10th/25th/75th/90th centiles) - presented here as median (IQR)	Short-term	Copeland	9	0.81 (0.62 to 0.97)	Global C.A.P.	15	0.83 (0.60 to 1.04)	
				Intermediate	Copeland	9	0.59 (0.50 to 0.65)	Global C.A.P.	15	0.57 (0.47 to 0.73)	
Reverse polarity TSR with neutral glenosphere vs inferior tilted glenosphere	Edwards 2012	Function	Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)	Intermediate	Neutral	22	71.4 ± 14.9	Tilted	20	63.6 ± 12.3	P = 0.136 ASES Shoulder Score reported in main analyses
			Age- and gender-adjusted Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)	Intermediate	Neutral	22	92.6 ± 18.9	Tilted	20	87.7 ± 23.6	P = 0.129 ASES Shoulder Score reported in main analyses
Reverse polarity TSR with eccentric glenosphere	Poon 2014	Function	Oxford Shoulder Score (0 to 48 scale, higher = better, reported as mean with range and P value - back-translated to SD)	Intermediate	Eccentric	23	35 ± 10.5	Concentric	27	38 ± 10.5	P = 0.32

Table 1. Additional study data *(Continued)*

position vs concentric position											
Reverse polarity TSR with bony increased offset vs standard offset for glenoid component	Greiner 2015	Pain	Subdomain of Constant Murley Score (0, 5, 10, 15 scale, higher = better, reported as mean ± SD)	Intermediate	BIO	16	12.7 ± 2.8	STD	15	12.7 ± 3.2	Not included in meta-analysis: categorical scale may not behave in same manner as a VAS or NPS
		Function	Age- and gender-adjusted Constant Murley Score (0 to 100 scale, higher = better, reported as mean ± SD)	Intermediate	BIO	16	83.3 ± 23.4	STD	15	89.4 ± 20.8	Study also reports unadjusted score
			ADLER Score (0 to 30 scale, higher = better, reported as mean ± SD)	Intermediate	BIO	16	25.7 ± 6.9	STD	15	26.1 ± 5.0	"Activities of Daily Living requiring External Rotation"
			DASH Score (0 to 100 scale, higher = worse, reported as mean ± SD)	Intermediate	BIO	16	40.9 ± 23.7	STD	15	34.2 ± 20.2	
Reverse polarity TSR with 135° humeral neck-shaft angle vs 155° humeral neck-	Gobezie 2019	Function	SANE Score (0 to 100 scale, higher = better, reported as mean ± SD)	Intermediate	135° neck-shaft angle	37	74 ± 24.4	155° neck-shaft angle	31	76 ± 16.8	
			Simple Shoulder Test Score (0 to 10 scale, higher = better, reported as mean ± SD)	Intermediate	135° neck-shaft angle	37	8 ± 3.0	155° neck-shaft angle	31	7 ± 2.2	

Table 1. Additional study data *(Continued)*

shaft angle	

ADLER: Activities of Daily Living and External Rotation.
ASES: American Shoulder and Elbow Surgeons Scale.
DASH: Disability of the Arm, Shoulder, and Hand questionnaire.
IQR: interquartile range.
MACTAR: McMaster Toronto Arthritis patient preference questionnaire.
RSA: radiostereometric analysis.
SANE: single-assessment numerical evaluation.
SD: standard deviation.
TSR: total shoulder replacement.
WOOS: Western Ontario Osteoarthritis of the Shoulder Index.

APPENDICES

Appendix 1. MEDLINE (Ovid) search strategy

1 randomized controlled trial.pt.
2 controlled clinical trial.pt.
3 randomized.ab.
4 placebo.ab.
5 drug therapy.fs.
6 randomly.ab.
7 trial.ab.
8 groups.ab.
9 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10 exp animals/ not humans.sh.
11 9 not 10
12 exp osteoarthritis/
13 osteoarthr$.tw.
14 (degenerative adj2 arthritis).tw.
15 arthrosis.tw.
16 arthropat$.tw.
17 rotator cuff arthro$.tw.
18 or/12-17
19 Shoulder/
20 shoulder joint/
21 shoulder pain/
22 shoulder$.tw.
23 or/19-22
24 exp surgical Procedures, Operative/
25 su.fs.
26 (surgery$ or surgeries or surgical or operat$).tw.
27 (arthroplast$ or hemiarthroplast$ or (joint$ adj2 replace$)).tw.
28 (surface$ adj replace$).tw.
29 resurfac$.tw.
30 RTSA.tw.
31 glenoid.tw.
32 glenosphere.tw.
33 exp "Prostheses and Implants"/
34 (glenoid adj2 component).tw.
35 (humor$ adj2 component).tw.
36 endopro$.tw.
37 reverse.tw.
38 or/24-37
39 and/11,18,23,38

Appendix 2. Embase (Ovid) search strategy

1 random$.ti,ab.
2 factorial$.ti,ab.
3 (crossover$ or cross over$ or cross-over$).ti,ab.
4 placebo$.ti,ab.
5 (doubl$ adj blind$).ti,ab.
6 (singl$ adj blind$).ti,ab.
7 assign$.ti,ab.
8 allocat$.ti,ab.
9 volunteer$.ti,ab.
10 crossover procedure.sh.
11 double blind procedure.sh.
12 randomized controlled trial.sh.
13 single blind procedure.sh.
14 or/1-13
15 exp osteoarthritis/
16 osteoarthr$.tw.

17 (degenerative adj2 arthritis).tw.
18 arthrosis.tw.
19 arthropat$.tw.
20 rotator cuff arthro$.tw.
21 or/15-20
22 Shoulder/
23 Shoulder Pain/
24 shoulder$.tw.
25 or/22-24
26 exp Surgery/
27 su.fs.
28 (surgery$ or surgeries or surgical or operat$).tw.
29 RTSA.tw.
30 (arthroplast$ or hemiarthroplast$ or (joint$ adj2 replace$)).tw.
31 (surface$ adj replace$).tw.
32 resurfac$.tw.
33 glenoid.tw.
34 glenosphere.tw.
35 (glenoid adj2 component).tw.
36 (humor$ adj2 component).tw.
37 endopro$.tw.
38 reverse.tw.
39 or/26-38
40 14 and 21 and 25 and 39

Appendix 3. CENTRAL search strategy

1 MeSH descriptor: [Osteoarthritis] explode all trees
2 osteoarth*:ti,ab
3 degenerative near/2 arthritis:ti,ab
4 arthrosis:ti,ab
5 arthropat*:ti,ab
6 (#1 or #2 or #3 or #4 or #5)
7 MeSH descriptor: [Shoulder] explode all trees
8 MeSH descriptor: [Shoulder Joint] explode all trees
9 MeSH descriptor: [Shoulder Pain] explode all trees
10 shoulder*:ti,ab
11 (#7 or #8 or #9 or #10)
12 MeSH descriptor: [Surgical Procedures, Operative] explode all trees
13 Any MeSH descriptor with qualifier(s): [Surgery - SU]
14 RTSA or reverse or glenoid or glenosphere:ti,ab
15 (surgery* or surgeries or surgical or operat*):ti,ab
16 (arthroplast* or hemiarthroplast* or (joint* near/2 replace*)):ti,ab
17 (glenoid near/2 component) or (humor* near/2 component):ti,ab
18 surface* replace*:ti,ab
19 resurfac*:ti,ab
20 endopro*:ti,ab
21 reverse:ti,ab
22 (#12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21)
23 (#6 and #11 and #22)

Appendix 4. CINAHL search strategy

S1 (MH "Osteoarthritis+")

S2 TI osteoarthr* OR AB osteoarthr*

S3 TI (degenerative n2 arthritis) OR AB (degenerative n2 arthritis)

S4 TI arthrosis OR AB arthrosis

S5 TI arthropat* OR AB arthropat*

S6 TI "rotator cuff arthro*" OR AB "rotator cuff arthro*"

S7 S1 OR S2 OR S3 OR S4 OR S5 OR S6

S8 (MH "Shoulder")

S9 (MH "Shoulder Joint")

S10 (MH "Shoulder Pain")

S11 TI shoulder* OR AB shoulder*

S12 S8 OR S9 OR S10 OR S11

S13 (MH "Surgery, Operative+")

S14 TI (surgery* OR surgeries OR surgical OR operat*) OR AB (surgery* OR surgeries OR surgical OR operat*)

S15 TI (RTSA OR reverse) OR AB (RTSA OR reverse)S16 TI (arthroplast* OR hemiarthroplast* OR (joint* N2 replace*)) OR AB (arthroplast* OR hemiarthroplast* OR (joint* N2 replace*))

S17 TI surface* replace* OR AB surface* replace*

S18 TI resurfac* OR AB resurfac*

S19 TI glenoid OR AB glenoid

S20 TI glenosphere OR AB glenosphere

S21 (MH "Prostheses and Implants+")

S22 TI (glenoid N2 component) OR AB (glenoid N2 component)

S23 TI (humor* N2 component) OR AB (humor* N2 component)

S24 TI endopro* OR AB endopro*

S25 S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24

S26 (MH "Clinical Trials+")

S27 PT Clinical Trial

S28 TI clinical* trial* OR AB clinical* trial*

S29 TI singl* blind* or singl* mask* or doub* blind* or doubl* mask* or trebl* blind* or trebl* mask* or tripl* blind* or tripl* mask*

S30 AB singl* blind* or singl* mask* or doub* blind* or doubl* mask* or trebl* blind* or trebl* mask* or tripl* blind* or tripl* mask*

S31 TI randomi?ed control* trial* OR AB randomi?ed control* trial*

S32 (MH "Random Assignment")

S33 TI random* allocat* OR AB random* allocat*

S34 TI placebo* OR AB placebo*

S35 (MH "Placebos")

S36 (MH "Quantitative Studies")

S37 TI allocat* random* OR AB allocat* random*

S38 S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37

S39 S7 AND S12 AND S25 AND S38

Appendix 5. SportDiscus search strategy

S1 TI (osteoarthr* OR (degenerative n2 arthritis) OR arthrosis OR arthopat* OR "rotator cuff arthro*") OR AB (osteoarthr* OR (degenerative n2 arthritis) OR arthrosis OR arthopat* OR "rotator cuff arthro*") OR SU (osteoarthr* OR (degenerative n2 arthritis) OR arthrosis OR arthopat* OR "rotator cuff arthro*") OR KW (osteoarthr* OR (degenerative n2 arthritis) OR arthrosis OR arthopat* OR "rotator cuff arthro*")

S2 TI shoulder* OR AB shoulder* OR SU shoulder* OR KW shoulder*

S3 TI (surgery* OR surgeries OR surgical OR operat* OR arthroplast* OR hemiarthroplast* OR joint* replace* OR surface* replace* OR resurfac* OR RTSA OR reverse OR glenoid OR glenosphere OR (glenoid n2 component) OR (humor* n2 component) OR endopro*) OR AB (surgery* OR surgeries OR surgical OR operat* OR arthroplast* OR hemiarthroplast* OR joint* replace* OR surface* replace* OR resurfac* OR RTSA OR reverse OR glenoid OR glenosphere OR (glenoid n2 component) OR (humor* n2 component) OR endopro*) OR SU (surgery* OR surgeries OR surgical OR operat* OR arthroplast* OR hemiarthroplast* OR joint* replace* OR surface* replace* OR resurfac* OR RTSA OR reverse OR glenoid OR glenosphere OR (glenoid n2 component) OR (humor* n2 component) OR endopro*) OR KW (surgery* OR surgeries OR surgical OR operat* OR arthroplast* OR hemiarthroplast* OR joint* replace* OR surface* replace* OR resurfac* OR RTSA OR reverse OR glenoid OR glenosphere OR (glenoid n2 component) OR (humor* n2 component) OR endopro*)

S5 S1 AND S2 AND S3 AND S4 (24)

Appendix 6. Web of Science search strategy

#1 TOPIC: (osteoarthr*)

#2 TOPIC: (degenerative near/2 arthritis)

#3 TOPIC: (arthrosis OR arthropat* OR "rotator cuff arthro*")

#4 #1 or #2 or #3

#5 TOPIC: (shoulder*)

#6 TOPIC: (surgery* OR surgeries OR surgical OR operat* OR arthroplast* OR hemiarthroplast* OR joint* replace* OR surface* replace* OR resurfac* OR RTSA OR reverse OR glenoid OR glenosphere OR (glenoid near/2 component) OR (humor* near/2 component) OR endopro*)

#7 TOPIC: (trial* or random* or placebo* or control* or double or treble or triple or blind* or mask* or allocat* or prospective* or volunteer*or comparative or evaluation or follow-up or followup)

#8 #4 and #5 and #6 and #7

Appendix 7. clinicaltrials.org search strategy

shoulder AND (arthroplasty OR hemiarthroplasty OR replacement OR resurfacing OR surface)

Appendix 8. WHO ICTRP search strategy

1 shoulder AND replacement

2 shoulder AND arthroplasty

3 shoulder and hemiarthroplasty

4 shoulder AND reverse

5 shoulder and surface

WHAT'S NEW

Date	Event	Description
21 April 2020	Amended	DoI updated

HISTORY

Protocol first published: Issue 11, 2017

Review first published: Issue 4, 2020

Date	Event	Description
9 March 2017	Amended	Change of scope
8 April 2009	Amended	CMSG ID A044-P

CONTRIBUTIONS OF AUTHORS

RC: screening, data extraction, assessment of risk of bias, analysis, drafting of review, editing of review.

HG: screening, data extraction, assessment of risk of bias.

SH, JR: adjudication of screening/RoB, analysis, editing of review.

JS: author of previous version of the review, editing of review.

DECLARATIONS OF INTEREST

RC: clinical research fellow funded by the National Joint Registry/Royal College of Surgeons of England during the conduct of this study.

HG: none known.

JS: JAS has received consultant fees from Crealta/Horizon, Medisys, Fidia, UBM LLC, Trio health, Medscape, WebMD, Clinical Care options, Clearview healthcare partners, Putnam associates, Focus forward, Navigant consulting, Spherix, Practice Point communications, the National Institutes of Health and the American College of Rheumatology. JAS owns stock options in Amarin pharmaceuticals and Viking therapeutics. JAS is on the speaker's bureau of Simply Speaking. JAS is a member of the executive of OMERACT, an organization that develops outcome measures in rheumatology and receives arms-length funding from 12 companies. JAS serves on the FDA Arthritis Advisory Committee. JAS is the chair of the Veterans Affairs Rheumatology Field Advisory Committee. JAS is the editor and the Director of the UAB Cochrane Musculoskeletal Group Satellite Center on Network Meta-analysis. .

SH: none known.

JR: has received grants from NIHR HTA and NIHR PGfAR. He works within an NIHR BRC and currently holds other grants from the Royal College of Surgeons of England, the Dinwoodie Foundation, McLaren Applied Technologies, and the National Joint Registry (NJR). He sits on committees at NJR, NICE, and ODEP (Orthopaedic Data Evaluation Panel); advises the MHRA; and is Council Member of the British Elbow and Shoulder Society. JR is a minority share holder in a new university spin-out software company PROMAPP LTD. The company utilises web-based solutions to monitor patient outcomes.

SOURCES OF SUPPORT

Internal sources

- National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), UK.

  Support for academic research

External sources

- Royal College of Surgeons and National Joint Registry, UK.

  Research Fellowship (RC)

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Several sensitivity and subgroup analyses were planned. Searches did not yield sufficient results to warrant these analyses. Overall, very scant meta-analysable data were available. We also planned to report on participant-perceived success of treatment, but this was not reported by any studies. It has been renamed in the review as "Participant-rated global assessment of treatment success" to align with recent consensus work on core outcome sets. Quality of life is reported preferentially via mental components of available scales because physical function is reported as a separate domain. Serious adverse events are reported as total and serious. Physician-evaluated

measures of success, including radiographic evaluations, were downgraded to a minor outcome. These changes ensured consistency with the reporting style of the Cochrane Musculoskeletal Group.

NOTES

This is an update of a previous review (Singh 2010). The original review sought to assess the benefits and adverse events of any surgical treatment for shoulder OA, but only identified eligible studies looking at joint replacement surgery. The focus of this updated review has been narrowed to assess the effects of surgical shoulder joint replacement treatments only. The scope of the review has been explicitly expanded to ensure that patients with OA secondary to rotator cuff tear arthropathy and patients treated with RTSRs are captured by the inclusion criteria. Some of the quality assessments and tabulated data intentionally vary from the previous review. This reflects adherence to the strengthened guidance for reporting methodology from Cochrane and the GRADE group.

INDEX TERMS

Medical Subject Headings (MeSH)

Arthroplasty, Replacement, Shoulder [adverse effects] [*methods]; Osteoarthritis [*surgery]; Pain Measurement; Postoperative Complications [etiology]; Quality of Life; Randomized Controlled Trials as Topic; Reoperation [statistics & numerical data]; Rotator Cuff [*surgery]; Rotator Cuff Tear Arthropathy [*surgery]

MeSH check words

Aged; Aged, 80 and over; Humans; Middle Aged