National Institute for Health and Care Excellence

Final

Hip fracture: management (update)

[A] Evidence reviews for femoral component design used for hemiarthroplasties

NICE guideline CG124

Evidence reviews underpinning recommendations 1.6.5 to 1.6.7 and a research recommendation in the NICE guideline

January 2023

Final

Commissioned by the National Institute for Health and Care Excellence

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-4917-5

Contents

Review question

In adults undergoing surgery for displaced intracapsular hip fracture what is the most clinically effective and cost-effective femoral component design used for hemiarthroplasties?

1.1.1 Introduction

Hemiarthroplasty is a procedure that involves replacing the femoral head with a prothesis. There are several different types of prothesis available for this procedure and guidance is required on which are the most clinically and cost-effective. The previous guideline recommends using a proven femoral stem design rather than Austin Moore or Thompson designs, however, a NICE surveillance review has indicated there is new evidence that could have an impact on this.

1.1.2 Summary of the protocol

Table 1: Summary of protocol

Population	• Adults presenting to the health service with a firm or provisional clinical diagnosis of fragility fracture of the hip. • Adults with displaced intracapsular hip fracture.

Interventions	• Femoral component designs used for cemented hemiarthroplasties, for example:
	  • Exeter Trauma stem (ETS) monoblock
	  • Exeter V40 stem
	  • Unitrax stem
	  • Austin Moore stem
	  • Thompson stem
	  • C stem
	  • Corail stem

Comparator	• Femoral component designs used for cemented hemiarthroplasties, for example:
	  • Exeter Trauma stem (ETS) monoblock
	  • Exeter V40 stem
	  • Unitrax stem
	  • Austin Moore stem
	  • Thompson stem
	  • C stem
	  • Corail stem

Outcomes	• All-cause mortality • Unplanned return to theatre (including number of reoperations or surgical revisions) • Functional status (using any validated measure such as the Barthel Index, mobility component of the EQ5D, Nottingham Extended Activities of Daily Living, WOMAC score, Harris hip score) • Pain (measured by any validated scale) • Health-related quality of life (measured by any validated scale)

	• Length of stay in an acute trust • Place of residence at 120 days • Periprosthetic fracture • Surgical site infection (grouped by SSIs up to 30 days and 1 year) • Number of adverse events (if data is available this will be grouped by those related to the femoral component (e.g. loosening of prosthesis, dislocation, leg length discrepancy, etc.) and those unrelated to the femoral component (e.g. thrombosis, embolism, neurological adverse events) Except where stated, outcomes will be reported at 30 days, 90 days, 1 year and >1 year

1.1.3 Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and the methods in appendix K.

During development of the review question, a Cochrane systematic review was identified that included RCT comparisons relevant to this review question. The meta-analyses for these comparisons were used for this review, with an additional RCT search performed by NICE to identify any RCTs published after the Cochrane review's final search date (July 2020). Mortality at 30 and 90 days were not reported in the Cochrane review, but these were added to this review from one of the primary RCT papers as they matched this review's protocol. The Cochrane review did not include non-randomised or observational studies and so a separate search was conducted for these study types which did not have a date limit. Please see table 2 for a summary on what has been included from the Cochrane systematic review and the further work done by NICE for this evidence review.

Table 2: Summary of work from Cochrane and NICE

NICE	Cochrane
• RCT results for early mortality at 30 and 90 days • RCT evidence search from June 2020 • Systematic review risk of bias assessment (ROBIS) • Observational evidence search • Observational evidence summary of results • Observational evidence risk of bias assessment (ROBINS-I) • GRADE assessment	• RCT evidence search to 6 July 2020 • RCT risk of bias assessments • RCT meta-analysis and summary of results

Please refer to the Cochrane systematic review Lewis 2022 for methods used in RCT assessment and analysis.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.1.4 Effectiveness evidence

1.1.4.1 Included studies

A Cochrane systematic review (Lewis 2022) was identified which included comparisons relevant to this review question. The 62 references from this review were screened for inclusion and from this 2 RCTs that compare the Exeter Trauma Stem/Exeter Unitrax with the Thompson Stem were identified. A further search for RCTs was conducted. After de-duplication 330 references were screened, and no further RCTs met the inclusion criteria for this review.

Given the limited number of RCTs relevant to this review, a subsequent search for comparative observational studies was carried out. 4041 references were screened at title and abstract after de-duplication, with 22 articles ordered. After full-text screening, 2 observational studies that compared the Lubinus SP2 stem with either the Exeter stem or the Zimmer stem were identified for inclusion.

1.1.4.2 Excluded studies

See appendix I for excluded studies and reasons for exclusion.

1.1.5 Summary of studies included in the effectiveness evidence

Table 3: Summary of included clinical studies

Study	Follow-up time	Population	Intervention	Comparator	Outcomes
Randomised controlled trials (from Lewis 2022 Cochrane systematic review)					
Parker 2012 n=200 UK	1 year	Patients admitted with displaced intracapsular fracture Mean age (ETS): 84.9 Mean age (Thompson): 83.6 Thompson - Female: 89 ETS – Female: 86	Exeter Trauma Stem (ETS) (n=100) – Monoblock – 1 part	Thompson (n=100)	• mortality (120 days and 1 year) • length of hospital stay • blood transfusion • superficial infection • deep infection • dislocation • periprosthetic fracture *(operative fracture femur)* • complications *(pneumonia, DVT, pulmonary embolism, CVA, cardiac failure, delirium acute renal failure)* • pain (mean scores)

Study	Follow-up time	Population	Intervention	Comparator	Outcomes
					• mobility (change in mean scores at 1 year) • unplanned return to theatre
Sims 2018 n=964 UK	4 months	Patients over the age of 60 years, receiving a hemiarthroplasty for a type B3 fracture of the hip Mean age (ETS): 83.9 (SD 7.9) Mean age (Thompson): 83.7 (SD7.3) Female (ETS): 326 Female (Thompson): 326	Thompson (n=482)	Exeter Trauma Stem (n=482) Modular - 2 part	• EQ-5D-5L (4 months) • mobility • mortality (4 months) • length of stay
Observational evidence					
Mellner, 2019 n= 2528 Sweden	Median follow up 47 months	>60 years old admitted to participating hospitals that underwent primary hip arthroplasty for a displaced FNF with either a cemented Exeter stem or a cemented Lubinus SP2 stem Mean age (Exeter): 82 (SD 8), Mean age (SP2) 81 (SD 8) Female (Exeter): 69% Female (SP2): 68%	Tapered Exeter Stem n=1326	Lubinus SP2 n=1202	Periprosthetic fracture

Study	Follow-up time	Population	Intervention	Comparator	Outcomes
Mohammed 2019 n=1077 Sweden	2 years	Patients with a cemented hip arthroplasty Mean age (SP2): 82 (SD8.0), Mean age (Zimmer): 82 (SD 8.4) Female (SP2): 75%, Female (Zimmer) 71%	Lubinus SP2 (Anatomic stem)	Zimmer (Polished tapered stem)	Periprosthetic fracture

See appendix C for full evidence tables

1.1.6 Summary of the effectiveness evidence

Table 4: Results and quality assessment of clinical studies included in the evidence review: Randomised Controlled Trials (RCTs from Cochrane systematic review)

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Thompson vs Exeter Trauma Stem					
Early HRQoL ≤ 4 months (EQ-5D)[1] (>0 favours ETS)	1[2]	618	MD: 0.06 (0.00, 0.11)	Moderate	Cannot differentiate
Early mobility (freely mobile without aids, or able to walk outdoors with one aid) (>0 favours ETS)	1[2]	494	RR: 1.14 (0.83, 1.57)	Moderate	Cannot differentiate
Early mortality 30 days	1[4]	200	RR: 2.00 (0.62, 6.43)	Low	Cannot differentiate
Early mortality 90 days	1[4]	200	RR: 1.67 (0.86, 3.22)	Moderate	Cannot differentiate
Early mortality ≤ 4 months	2[2,4]	1164	RR: 1.20 (0.76, 1.88)	Very Low	Cannot differentiate
Mortality 12 months	1[4]	200	RR 1.44 (0.94, 2.21)	Moderate	Cannot differentiate
Unplanned return to theatre (end of follow up)	2[2,4]	1164	RR (0.46 (0.05, 3.89)	Very Low	Cannot differentiate
Adverse event related to implant, fracture, or both					
Intraoperative periprosthetic fracture	1[4]	200	RR: 1.00 (0.21, 4.84)	Low	Cannot differentiate
Superficial infection	1[4]	200	RR: 3.00 (0.32, 28.35)	Low	Cannot differentiate

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Dislocation	1[4]	200	RR: 0.20 (0.01, 4.11)	Low	Cannot differentiate
Adverse event unrelated to implant or fracture					
Acute Kidney Injury	1[4]	200	RR: 1.00 (0.06, 15.77)	Low	Cannot differentiate
Blood transfusion	1[4]	200	RR: 1.00 (0.54, 1.84)	Low	Cannot differentiate
Cerebrovascular accident	1[4]	200	RR: 2.00 (0.18, 21.71)	Low	Cannot differentiate
Chest infection/pneumonia	1[4]	200	RR: 1.67 (0.41, 6.79)	Low	Cannot differentiate
Myocardial infarction	1[4]	200	RR: 5.00 (0.24, 102.85)	Low	Cannot differentiate
Venous thromboembolic phenomena (DVT)	1[4]	200	RR: 1.00 (0.21, 4.84)	Low	Cannot differentiate

1 EQ-5D (higher scores indicate better QoL)
2 Simms 2018
3 Minimum clinically important difference 0.08 taken from Simms 2018
4 Parker 2012

Table 5: Results and quality assessment of clinical studies included in the evidence review: Observational studies

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Tapered Exeter Stem vs Lubinus SP2					
Periprosthetic fracture (>1 favours Lubinus SP2)	1[1]	2528	HR 5.40 (2.37, 12.32)	Low	Favours Lubinus SP2
Lubinus SP2 vs Zimmer					
Periprosthetic fracture (>1 favours Zimmer)	1[2]	1077	HR 0.10 (0.02, 0.50)	Low	Favours Lubinus SP2

1 Mellner 2019
2 Mohammed 2019

See appendix E for full GRADE tables

1.1.7 Economic evidence

1.1.7.1 Included studies

A search was performed to identify published economic evaluations of relevance, this search retrieved 660 studies. Based on title and abstract screening 660 studies were excluded and therefore no economic studies were included for this review question.

1.1.7.2 Excluded studies

There were no excluded studies for this review question.

1.1.8 Summary of included economic evidence

There are no existing economic studies for this review question.

1.1.9 Economic model

No economic modelling was completed for this review question.

1.1.10 Unit costs

Resource	Mean Cost	Lower Bound	Upper Bound	Source
Thompsons	£179.15	£133.57	£210.95	Royal College of Physicians, 2019
Exeter Trauma stem	£240.63	£160.07	£489.74	Royal College of Physicians, 2019
Exeter V40 stem + Unitrax head + sleeve	£640.27	£420.84	£1,258.27	Royal College of Physicians, 2019
Exeter V40 stem + head + bipolar head	£791.85	£534.26	£1,510.56	Royal College of Physicians, 2019
C stem + unipolar head	£444.16	£255.47	£848.04	Royal College of Physicians, 2019
C stem + head + bipolar head	£526.84	£322.25	£1,004.92	Royal College of Physicians, 2019

Costs were adjusted for purchase price parities and inflated to 2022 British Pounds Sterling using Eppi-Centre Cost Converter.

CCEMG - EPPI-Centre Cost Converter v.1.4 (ioe.ac.uk)

1.1.11 The committee's discussion and interpretation of the evidence

1.1.11.1. The outcomes that matter most

The committee agreed that while long term adverse events such as periprosthetic fractures were important from a surgical point of view, these would always be a secondary outcome, following primary patient reported outcomes such as health related quality of life (HRQOL). They agreed that HRQOL could be measured at 4 months (as per one of the RCTs included) and this would be representative of future measurements, but periprosthetic fractures would require longer term follow up and more data on this for a fragility fracture population (as opposed to those who have elective surgery) is required. The committee agreed that other outcomes reported including early mobility, mortality, unplanned return to theatre and adverse events relating to the implant were all important outcomes for measuring the effectiveness of femoral components.

1.1.11.2 The quality of the evidence

The committee noted that the Cochrane review which was used as a source for RCT evidence had a low risk of bias and was partially applicable due to it being on a much broader review question. However, the section of the review that related to this question was fully applicable and included evidence from 2 UK-based RCTs. The quality of outcomes from RCT evidence ranged from very low to moderate, with most outcomes reported from single studies (Parker 2012, Simms 2018) and this was why they were rated down for imprecision. Although the interventions differed slightly in the two papers (Parker – Exeter Trauma Stem (monoblock); Simms – Exeter with Unitrax Head), the committee agreed that they could be combined in a meta-analysis for two outcomes (early mortality and unplanned return to theatre) as they are similar enough not to have a differential effect on the outcome. The

committee noted that for all outcomes, the evidence was unable to differentiate between Thompson or Exeter Trauma Stem/Exeter Unitrax components.

The committee also noted that observational studies were rated down in quality due to confounding and that the types of femoral components used in these were not directly relevant to UK practice. However, although these femoral components were not used in the UK and therefore could not be used to recommend specific types of femoral component, they did show the benefit that registry data in Sweden had in providing long-term follow up data on adverse events in people who have had hemiarthroplasty.

1.1.11.3 Benefits and harms

The committee discussed that while there is a National Joint Registry (NJR) in the UK, it only records data on total hip arthroplasty patients. Long term real-world follow-up data on hemiarthroplasties in a UK population would be valuable for making decisions regarding femoral components in future, in a similar way to the data used in the Swedish observational studies included in this review. Therefore, they made a recommendation about reporting patient outcomes for hemiarthroplasties, in addition to total hip arthroplasties, in a national joint registry.

The committee also discussed whether they knew of any populations at risk of less favourable outcomes dependent on the femoral component used. There is currently limited knowledge about inequalities for people who are given hemiarthroplasties, and they thought it was important that future research considers groups who are potentially at risk of less favourable outcomes . By identifying whether any groups have better or worse outcomes from a particular femoral component, clinicians will be able to ensure that patients receive the most effective treatment options in future. For this reason, they specified that population subgroups should be considered in the research recommendation (see Appendix J). They also commented that this data on subgroups could also be collected in a national joint registry.

The committee commented that the original recommendation for using a proven femoral component was based on an elective surgery population, where ODEP (Orthopaedic Device Evaluation Panel) ratings are used to evaluate the longevity of femoral components, and this was extrapolated to a fragility fracture population. They questioned the appropriateness of this, as there are differences in the two populations that could affect outcomes, such as bone strength, and commented that more long-term follow-up evidence on both adverse events and patient reported outcomes was required on femoral components in a fragility fracture population. For this reason, they decided to make a research recommendation aimed at establishing the long-term outcomes of different types of femoral component for people with fragility fractures.

While patient outcomes were considered important, the committee also wanted to highlight the importance of standardised treatment in a recommendation and choosing the most cost-effective option. They agreed hospitals should choose one femoral component that medical teams are trained to use and are familiar with implanting, rather than have different types. This may help improve outcomes for patients as medical teams will build up greater knowledge and expertise in performing the operation with one specific type of femoral component. They also thought that hospitals may want to consider a femoral component that is suitable for both hemiarthroplasty and total hip arthroplasty to allow consistency in practice, and this should also be considered when deciding which component was the most cost-effective. Hospitals should consider the cost of training needs as well as component costs when switching to a different component, alongside any future costs relating to adverse outcomes.

1.1.11.4 Cost effectiveness and resource use

The committee had no published cost-effectiveness evidence to support their decision making. However, they considered evidence on the mean costs for different implanted components that was most recently published by the National Hip Fracture Database (NHFD) in 2019. The committee acknowledged that there was no long-term comparative effectiveness data and that the available short-term evidence showed that there was no statistical or meaningful difference between the Thompson and the Exeter/Unitrax stems. Therefore, the committee felt that an economic model based on short-term outcomes only would not provide any useful information, but a cost comparison analysis of the cost of femoral stem components would be useful for decision making. The committee decided to include a recommendation for long-term data on the long-term effectiveness of different stem designs for hemiarthroplasty to be collected through the National Joint Registry. Should the long-term data suggest that there are other differences between stems that this review was unable to capture (such as periprosthetic fractures for a fragility fracture population), economic modelling may be warranted in the future.

The committee acknowledged that the main differences in the cost of a hemiarthroplasty using different femoral stems is the costs of the stem itself, because the rest of the surgery takes a similar amount of time and needs the same resources, and that revision rates were not found to differ. The committee noted that the report from NHFD suggested that there is a large variation in the types of stems used in different trusts across England. There are differences in mean cost between the different types of stem, and national differences in the cost for a given stem. For example, an Exeter V40 stem with a Unitrax head and sleeve has a mean cost of £640, and ranges from £421 to £1,258. This is due to the trusts having different agreements with the suppliers of the stems.

The committee felt that introducing recommendations for hospitals to use a single type of cemented femoral component for hemiarthroplasties and to consider the cost of the stem within this context of training requirements, team familiarity and overall costs would reduce the variation in practice and in costs across the country, and would reduce the cost to the National Health Service.

1.1.11.5 Other factors the committee took into account

The committee questioned why the evidence search had excluded cemented vs uncemented hemiarthroplasties and commented that this is an area that is an important comparison, particularly given that there is new evidence from the WHITE 5 trial on this. The NICE team confirmed that there is an existing recommendation to use cemented implants in patients undergoing arthroplasties and this had recently been reviewed by the NICE surveillance team. Evidence from studies which compared outcomes from cemented and uncemented implants, including the WHITE 5 trial, was included in a Cochrane review which was considered by NICE in an exceptional surveillance review. The outcomes from this review showed no difference between cemented and uncemented implants for some outcomes, while others such as mortality at 12 months, functional status, health-related quality of life and mobility were improved with the use of cemented implants. The outcomes favouring the use of cemented over uncemented implants supported the existing recommendation to use cemented implants and so it was decided that the recommendation did not need updating, and that this review should only consider cemented implants. The committee commented that clinicians reading the guidance might assume this recommendation had not been reviewed so the NICE team confirmed they would take this into consideration when discussing the implementation of the guideline.

The committee also discussed that a new regulatory environment will require companies to produce data in relation to implants, and this would have an impact on femoral components used. As a result of this it was likely that older components will no longer be used in future

and therefore, they did not want to recommend specific types of femoral stems for use in hemiarthroplasties.

1.1.12 Recommendations supported by this evidence review

This evidence review supports recommendations 1.6.5 – 1.6.7 and the research recommendation on long-term outcomes from different femoral component designs.

1.1.13 References – included studies

1.1.13.1 Effectiveness

Lewis SR, Macey R, Parker MJ et al. Arthroplasties for hip fracture in adults. The Cochrane database of systematic reviews 2: CD013410

Parker MJ (2012) Cemented Thompson hemiarthroplasty versus cemented Exeter Trauma Stem (ETS) hemiarthroplasty for intracapsular hip fractures: a randomised trial of 200 patients. Injury 43(6): 807-810

Sims AL, Parsons N, Achten J et al. A randomized controlled trial comparing the Thompson hemiarthroplasty with the Exeter polished tapered stem and Unitrax modular head in the treatment of displaced intracapsular fractures of the hip: the WHiTE 3: HEMI Trial. The bone & joint journal: 352-360

Mellner, Carl, Mohammed, Jabbar, Larsson, Magnus et al. (2021) Increased risk for postoperative periprosthetic fracture in hip fracture patients with the Exeter stem than the anatomic SP2 Lubinus stem. European journal of trauma and emergency surgery : official publication of the European Trauma Society 47(3): 803-809

Mohammed, Jabbar, Mukka, Sebastian, Hedbeck, Carl-Johan et al. (2019) Reduced periprosthetic fracture rate when changing from a tapered polished stem to an anatomical stem for cemented hip arthroplasty: an observational prospective cohort study with a follow-up of 2 years. Acta orthopaedica 90(5): 427-432

1.1.13.2 Economic

Royal College of Physicians. National Hip Fracture Database annual report 2019. London: RCP, 2019

Appendix A – Review protocols

Review protocol for Hip Fracture

ID	Field	Content
0.	PROSPERO registration number	CRD42022324242
1.	Review title	Clinical effectiveness and cost-effectiveness of different femoral component designs for hemiarthroplasties in adults undergoing surgery for displaced intracapsular hip fracture.
2.	Review question	In adults undergoing surgery for displaced intracapsular hip fracture what is the most clinically effective and cost-effective femoral component design used for hemiarthroplasties?
3.	Objective	To establish which femoral component designs for hemiarthroplasties should be used in surgery for displaced intracapsular hip fracture.
4.	Searches	The following databases will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR)

		• Embase • MEDLINE Searches will be restricted by: • July 2020 onwards • English language • Human studies A Cochrane review is available that includes RCTs up to July 2020. The Cochrane review is broader than the current review and so only a subset of the included studies from the Cochrane review will be included in this review. All included studies from the Cochrane review will be assessed for inclusion. The search dates for RCTs will be set to begin from July 2020 to identify any RCTs that were published after this date and therefore not included in the review. If observational studies are required, there will be no search date restrictions for these. The full search strategies for MEDLINE database will be published in the final review.
5.	Condition or domain being studied	Management of displaced intracapsular fracture in adult patients.
6.	Population	Inclusion: • Adults presenting to the health service with a firm or provisional clinical diagnosis of fragility fracture of the hip.

		• Adults with displaced intracapsular hip fracture.
		
		Exclusion:
		• People with fractures caused by specific pathologies other than osteoporosis or osteopaenia (because these would require more condition-specific guidance).
		
		• Adults with the following types of hip fracture:
		  ○ undisplaced intracapsular
		  ○ extracapsular (trochanteric and subtrochanteric)

7.	Intervention/Exposure/Test	Femoral component designs used for cemented hemiarthroplasties, for example: • Exeter Trauma stem (ETS) monoblock • Exeter V40 stem • Unitrax stem • Austin Moore stem • Thompson stem • C stem • Corail stem

8.	Comparator/Reference standard/Confounding factors	Femoral component designs used for cemented hemiarthroplasties, for example: • Exeter Trauma stem (ETS) monoblock • Exeter V40 stem • Unitrax stem • Austin Moore stem • Thompson stem • C stem • Corail stem
9.	Types of study to be included	• RCTs • Non-randomised controlled or comparative observational studies with a concurrent control group and adjustment for confounding factors including a minimum of age and sex. Observational evidence will only be used if insufficient RCT evidence is available and in the committee's view, observational studies could reasonably be expected to provide robust information on an outcome to inform decision making. Adjustment must use one of the methods specified in NICE TSD 17: The use of observational data to inform estimates of treatment effectiveness in technology appraisal.

10.	Other exclusion criteria	• Other study types • RCTs with a crossover study design • Studies on femoral component designs used only with uncemented hemiarthroplasty • Studies on bipolar implants • Studies on non-isolated fracture
11.	Context	The NICE surveillance review identified new evidence from a trial conducted in the UK as part of the World Hip Trauma Evaluation Study that indicates that Thomson stems may provide similar clinical outcomes to the Exeter/Unitrax stem design. During the development of the original guideline no randomised studies were found that compared older stem designs with modern stem designs in patients with hip fractures, thus evidence was extrapolated from studies looking at the selection of prosthesis for primary total hip replacement and expert opinion. Furthermore, surveillance conducted in 2013 and 2015 did not find any studies that addressed this recommendation.
12.	Primary outcomes (critical outcomes)	Except where stated, outcomes will be reported at 30 days, 90 days, 1 year and >1 year • All-cause mortality • Unplanned return to theatre (including number of reoperations or surgical revisions)

		• Functional status (using any validated measure such as the Barthel Index, mobility component of the EQ5D, Nottingham Extended Activities of Daily Living, WOMAC score, Harris hip score) • Pain (measured by any validated scale) • Health-related quality of life (measured by any validated scale) • Length of stay in an acute trust • Place of residence at 120 days • Periprosthetic fracture • Surgical site infection (grouped by SSIs up to 30 days and 1 year) • Number of adverse events (if data is available this will be grouped by those related to the femoral component (e.g. loosening of prosthesis, dislocation, leg length discrepancy, etc.) and those unrelated to the femoral component (e.g. thrombosis, embolism, neurological adverse events)
13.	Secondary outcomes (important outcomes)	None

14.	Data extraction (selection and coding)	All references identified by the searches and from other sources will be uploaded into EPPI reviewer and de-duplicated. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. The full text of potentially eligible studies will be retrieved and will be assessed in line with the criteria outlined above. A standardised form will be used to extract data from studies (see Developing NICE guidelines: the manual section 6.2).
15.	Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual. If studies from the Cochrane review are included, we will refer to the published Cochrane review for risk of bias judgments, as outlined in the GSD
16.	Strategy for data synthesis	Pairwise meta-analyses will be performed in Cochrane Review Manager V5.3. A pooled relative risk will be calculated for dichotomous outcomes (using the Mantel–Haenszel method) reporting numbers of people having an event. A pooled mean difference will be calculated for continuous outcomes (using the inverse variance method) when the same scale will be used to measure an outcome across different studies. Where different studies presented continuous data measuring the same outcome but using different numerical scales these outcomes will be all converted to the same scale before meta-analysis is

		conducted on the mean differences. Where outcomes measured the same underlying construct but used different instruments/metrics, data will be analysed using standardised mean differences (SMDs, Hedges' g). Fixed effects models will be fitted unless there is significant statistical heterogeneity in the meta-analysis, defined as $I^2 \geq 50\%$, when random effects models will be used instead. Where 10 or more studies are included as part of a single meta-analysis, a funnel plot will be produced to graphically assess the potential for publication bias. • GRADE will be used to assess the quality of any pair-wise analysis of outcomes. Outcomes using evidence from RCTs will be rated as high quality initially and downgraded from this point. Reasons for upgrading the certainty of the evidence will also be considered.
17.	Analysis of sub-groups	The committee did not think that there were any subgroups of patient characteristics in this cohort of people that are likely to affect outcomes.
18.	Type and method of review	☒ Intervention ☐ Diagnostic

		☐ Prognostic ☐ Qualitative ☐ Epidemiologic ☐ Service Delivery ☐ Other (please specify)		
19.	Language	English		
20.	Country	England		
21.	Anticipated or actual start date	March 2022		
22.	Anticipated completion date	October 2022		
23.	Stage of review at time of this submission	**Review stage**	**Started**	**Completed**

		Preliminary searches	☐	☐
		Piloting of the study selection process	☐	☐
		Formal screening of search results against eligibility criteria	☐	☐
		Data extraction	☐	☐

		Risk of bias (quality) assessment	☐	☐
		Data analysis	☐	☐
24.	Named contact	**5a. Named contact** Guideline Development Team **5b Named contact e-mail** hipupdate@nice.org.uk **5e Organisational affiliation of the review** National Institute for Health and Care Excellence (NICE)		
25.	Review team members	From the Guideline Development Team: • Technical Lead: Clare Dadswell • Technical Analyst: Anthony Gildea • Health Economics Lead: Lindsay Claxon • Health Economics Analyst: Steph Armstrong		

		• Information Specialist: Elizabeth Barrett • Project Manager: Jon Littler
26.	Funding sources/sponsor	This systematic review is being completed by the Guideline Development Team which receives funding from NICE.
27.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
28.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: Project information | Hip fracture: management (update) | Guidance | NICE

29.	Other registration details	None
30.	Reference/URL for published protocol	None
31.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.
32.	Keywords	Femoral stem design, adults, surgery, hip fracture, arthroplasty, hemiarthroplasty, Austin Moore stem, Thompson stem, Exeter stem, Unitrax stem.
33.	Details of existing review of same topic by same authors	This is a new review question that will update the surgical procedures section in the NICE Guideline: Hip fracture: management (2017) NICE guideline CG124.
34.	Current review status	☒ Ongoing

		☐ Completed but not published ☐ Completed and published ☐ Completed, published and being updated ☐ Discontinued
35..	Additional information	None
36.	Details of final publication	www.nice.org.uk

Appendix B – Literature search strategies

Background and development

Search design and peer review

A NICE information specialist conducted the literature searches for the evidence review. The searches were originally run on the 23rd and 29th March 2022. This search report is compliant with the requirements of PRISMA-S.

The MEDLINE strategy below was quality assured (QA) by a trained NICE information specialist. All translated search strategies were peer reviewed to ensure their accuracy. Both procedures were adapted from the 2016 PRESS Checklist.

The principal search strategy was developed in MEDLINE (Ovid interface) and adapted, as appropriate, for use in the other sources listed in the protocol, taking into account their size, search functionality and subject coverage.

Review management

The search results were managed in EPPI-Reviewer v5. Duplicates were removed in EPPI-R5 using a two-step process. First, automated deduplication is performed using a high-value algorithm. Second, manual deduplication is used to assess 'low-probability' matches. All decisions made for the review can be accessed via the deduplication history.

Prior work

The search strategy was based on the terms used for the CG124 NICE guideline (2011). Modifications were made to these original search strategies for the specifications in the review protocol.

Limits and restrictions

English language limits were applied in adherence to standard NICE practice and the review protocol.

Limits to exclude conferences in Embase were applied in adherence to standard NICE practice and the review protocol.

The limit to remove animal studies in the searches was the standard NICE practice, which has been adapted from: Dickersin, K., Scherer, R., & Lefebvre, C. (1994). Systematic Reviews: Identifying relevant studies for systematic reviews. *BMJ*, 309(6964), 1286.

Search filters

Clinical searches

- RCT filters:
    - McMaster Therapy – Medline - "best balance of sensitivity and specificity" version.

Haynes RB et al. (2005) Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ*, 330, 1179-1183.

- McMaster Therapy – Embase "best balance of sensitivity and specificity" version.

Wong SSL et al. (2006) Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. Journal of the Medical Library Association, 94(1), 41-47.

- Observational filter:
    - The terms used for observational studies are standard NICE practice that have been developed in house.

Cost effectiveness searches

The following search filters were applied to the search strategies in MEDLINE and Embase to identify cost-effectiveness studies:

- Glanville J et al. (2009) Development and Testing of Search Filters to Identify Economic Evaluations in MEDLINE and EMBASE. Alberta: Canadian Agency for Drugs and Technologies in Health (CADTH)

Several modifications have been made to these filters over the years that are standard NICE practice.

Key decisions

The search strategy was developed to find evidence on for the specified population and intervention in the review protocol.

Clinical searches

Database	Date searched	Database Platform	Database segment or version	No. of results downloaded
Cochrane Central Register of Controlled Trials (CENTRAL)	23/03/2022	Wiley	2 of 12 February 2022	211
Cochrane Database of Systematic Reviews (CDSR)	23/03/2022	Wiley	3 of 12 March 2022	34
Embase	23/03/2022	Ovid	1974 to 2022 March 22	RCTs 111 Observational 1961
MEDLINE	23/03/2022	Ovid	1946 to March 22 2022	RCTs 96 Observational 3308
MEDLINE-in-Process	23/03/2022	Ovid	1946 to March 22 2022	RCTs 1 Observational 7
MEDLINE Epub Ahead-of-Print	23/03/2022	Ovid	March 22 2022	RCTs 4 Observational 26

Search strategy history

Database name: MEDLINE

1 exp Hip Fractures/ (26866)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (10984)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (40913)
4 or/1-3 (46481)
5 Femur Neck/su or Femoral Neck Fractures/su or Prosthesis Design/ (64287)
6 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (2214)
7 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (20244)
8 ("c-stem" or "c stem").tw. (96)
9 or/5-8 (80732)
10 4 and 9 (8564)
11 randomized controlled trial.pt. (561907)
12 randomi?ed.mp. (903093)
13 placebo.mp. (214108)
14 or/11-13 (958282)
15 10 and 14 (613)
16 limit 15 to ed=20200601-20220323 (96)
17 animals/ not humans/ (4943119)
18 16 not 17 (96)

19 limit 18 to english language/ (96)
20 Observational Studies as Topic/ (7616)
21 Observational Study/ (123441)
22 Epidemiologic Studies/ (9039)
23 exp Case-Control Studies/ (1298466)
24 exp Cohort Studies/ (2315891)
25 Cross-Sectional Studies/ (416516)
26 Controlled Before-After Studies/ (690)
27 Historically Controlled Study/ (220)
28 Interrupted Time Series Analysis/ (1551)
29 Comparative Study.pt. (1910621)
30 case control$.tw. (128412)
31 case series.tw. (73262)
32 (cohort adj (study or studies)).tw. (228100)
33 cohort analy$.tw. (8704)
34 (follow up adj (study or studies)).tw. (48983)
35 (observational adj (study or studies)).tw. (113562)
36 longitudinal.tw. (245209)
37 prospective.tw. (573852)
38 retrospective.tw. (551362)
39 cross sectional.tw. (362073)
40 or/20-39 (4853891)
41 10 and 40 (4099)
42 animals/ not humans/ (4943119)
43 41 not 42 (4062)
44 limit 43 to english language/ (3308)

Database name: MEDLINE in Process

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (55)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (206)
4 or/1-3 (206)
5 Femur Neck/su or Femoral Neck Fractures/su or Prosthesis Design/ (0)
6 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (11)
7 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (169)
8 ("c-stem" or "c stem").tw. (0)
9 or/5-8 (178)
10 4 and 9 (18)
11 randomized controlled trial.pt. (0)
12 randomi?ed.mp. (4404)
13 placebo.mp. (922)
14 or/11-13 (4647)
15 10 and 14 (1)
16 limit 15 to dt=20200601-20220323 (1)
17 animals/ not humans/ (0)
18 16 not 17 (1)
19 limit 18 to english language/ (1)

20 Observational Studies as Topic/ (0)
21 Observational Study/ (0)
22 Epidemiologic Studies/ (0)
23 exp Case-Control Studies/ (0)
24 exp Cohort Studies/ (0)
25 Cross-Sectional Studies/ (0)
26 Controlled Before-After Studies/ (0)
27 Historically Controlled Study/ (0)
28 Interrupted Time Series Analysis/ (0)
29 Comparative Study.pt. (0)
30 case control$.tw. (726)
31 case series.tw. (548)
32 (cohort adj (study or studies)).tw. (3113)
33 cohort analy$.tw. (116)
34 (follow up adj (study or studies)).tw. (194)
35 (observational adj (study or studies)).tw. (1297)
36 longitudinal.tw. (2309)
37 prospective.tw. (3908)
38 retrospective.tw. (4976)
39 cross sectional.tw. (3502)
40 or/20-39 (15559)
41 10 and 40 (7)
42 animals/ not humans/ (0)
43 41 not 42 (7)
44 limit 43 to english language/ (7)

Database name: MEDLINE ePubs

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (165)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (688)
4 or/1-3 (690)
5 Femur Neck/su or Femoral Neck Fractures/su or Prosthesis Design/ (0)
6 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (49)
7 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (335)
8 ("c-stem" or "c stem").tw. (1)
9 or/5-8 (381)
10 4 and 9 (66)
11 randomized controlled trial.pt. (1)
12 randomi?ed.mp. (13522)
13 placebo.mp. (2818)
14 or/11-13 (14415)
15 10 and 14 (4)
16 animals/ not humans/ (0)
17 15 not 16 (4)
18 limit 17 to english language/ (4)
19 Observational Studies as Topic/ (0)
20 Observational Study/ (1)

21 Epidemiologic Studies/ (0)
22 exp Case-Control Studies/ (0)
23 exp Cohort Studies/ (0)
24 Cross-Sectional Studies/ (0)
25 Controlled Before-After Studies/ (0)
26 Historically Controlled Study/ (0)
27 Interrupted Time Series Analysis/ (0)
28 Comparative Study.pt. (0)
29 case control$.tw. (2383)
30 case series.tw. (2517)
31 (cohort adj (study or studies)).tw. (9112)
32 cohort analy$.tw. (337)
33 (follow up adj (study or studies)).tw. (602)
34 (observational adj (study or studies)).tw. (4223)
35 longitudinal.tw. (6693)
36 prospective.tw. (12049)
37 retrospective.tw. (18221)
38 cross sectional.tw. (10948)
39 or/19-38 (50908)
40 10 and 39 (26)
41 animals/ not humans/ (0)
42 40 not 41 (26)
43 limit 42 to english language/ (26)

Database name: Embase

1 exp hip fracture/ (44297)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (14975)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (60680)
4 or/1-3 (73484)
5 femoral neck fracture/su (1616)
6 prosthesis design/ (6868)
7 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (3186)
8 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (28648)
9 ("c-stem" or "c stem").tw. (101)
10 or/5-9 (39163)
11 4 and 10 (5956)
12 limit 11 to (books or chapter or conference abstract or conference paper or "conference review" or editorial or letter) (774)
13 11 not 12 (5182)
14 nonhuman/ not (human/ and nonhuman/) (4950513)
15 13 not 14 (5129)
16 limit 15 to english language/ (4332)
17 random:.tw. (1767901)
18 placebo:.mp. (491338)
19 double-blind:.tw. (228598)
20 or/17-19 (2034439)
21 16 and 20 (419)

22 limit 21 to dc=20200601-20220323 (111)
23 Clinical study/ (157579)
24 Case control study/ (185675)
25 Family study/ (25397)
26 Longitudinal study/ (169653)
27 Retrospective study/ (1218722)
28 comparative study/ (941801)
29 Prospective study/ (753983)
30 Randomized controlled trials/ (222820)
31 29 not 30 (745255)
32 Cohort analysis/ (820627)
33 cohort analy$.tw. (16210)
34 (Cohort adj (study or studies)).tw. (384226)
35 (Case control$ adj (study or studies)).tw. (155828)
36 (follow up adj (study or studies)).tw. (68742)
37 (observational adj (study or studies)).tw. (212417)
38 (epidemiologic$ adj (study or studies)).tw. (115039)
39 (cross sectional adj (study or studies)).tw. (281701)
40 case series.tw. (127289)
41 prospective.tw. (986280)
42 retrospective.tw. (1076577)
43 or/23-28,31-42 (4756505)
44 16 and 43 (1961)

Database name: Cochrane Library

#1 MeSH descriptor: [Hip Fractures] explode all trees 1807
#2 (((hip* or pertrochant* or intertrochant* or trochant* or subtrochant* or intracapsular*) or (femur* or femoral*)) NEAR/3 (neck or proximal) NEAR/4 fracture*):ti,ab,kw 2137
#3 #1 or #2 3345
#4 MeSH descriptor: [Femur Neck] this term only and with qualifier(s): [surgery - SU] 29
#5 MeSH descriptor: [Femoral Neck Fractures] this term only and with qualifier(s): [surgery - SU] 298
#6 MeSH descriptor: [Prosthesis Design] this term only 2485
#7 ((hemiarthroplast* or partial*) near/4 (hip* or femor* or femur*)):ti,ab,kw 368
#8 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) near/4 (stem* or implant* or prosthe*)):ti,ab,kw 3119
#9 ("c-stem" or "c stem"):ti,ab,kw 8
#10 {OR #4 - #9} 1300138
#11 #3 and #10 2423
#12 conference:pt 195680
#13 #11 not #12 2176
#14 (clinicaltrials or trialsearch):so 391193
#15 #13 not #14 with Publication Year from 2020 to 2022, in Trials 211
#16 #13 not #14 1679

Cost-effectiveness searches

Database	Date searched	Database Platform	Database segment or version	No. of results downloaded
EconLit	29/03/2022	OVID	1886 to March 17 2022	0
EED	29/03/2022	CRD		0
Embase	29/03/2022	Ovid	1974 to 2022 March 28	455
HTA	29/03/2022	CRD		1
INAHTA	29/03/2022	INAHTA		73
MEDLINE	29/03/2022	Ovid	1946 to March 28 2022	385
MEDLINE-in-Process	29/03/2022	Ovid	1946 to March 28 2022	1
MEDLINE Epub Ahead-of-Print	29/03/2022	Ovid	March 28 2022	7

Search strategy history

Database name: MEDLINE

1 exp Hip Fractures/ (26900)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (10998)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (40956)
4 or/1-3 (46530)
5 Femur Neck/su or Femoral Neck Fractures/su or Prosthesis Design/ (64316)
6 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (2215)
7 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (20286)
8 ("c-stem" or "c stem").tw. (96)
9 or/5-8 (80802)
10 4 and 9 (8571)
11 Economics/ (27435)
12 exp "Costs and Cost Analysis"/ (256370)
13 Economics, Dental/ (1920)
14 exp Economics, Hospital/ (25534)
15 exp Economics, Medical/ (14332)
16 Economics, Nursing/ (4013)
17 Economics, Pharmaceutical/ (3059)

18 Budgets/ (11582)
19 exp Models, Economic/ (16092)
20 Markov Chains/ (15651)
21 Monte Carlo Method/ (30998)
22 Decision Trees/ (11917)
23 econom$.tw. (285059)
24 cba.tw. (10234)
25 cea.tw. (22473)
26 cua.tw. (1086)
27 markov$.tw. (20959)
28 (monte adj carlo).tw. (33757)
29 (decision adj3 (tree$ or analys$)).tw. (17592)
30 (cost or costs or costing$ or costly or costed).tw. (535238)
31 (price$ or pricing$).tw. (38586)
32 budget$.tw. (26569)
33 expenditure$.tw. (55972)
34 (value adj3 (money or monetary)).tw. (2486)
35 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (3752)
36 or/11-35 (1060708)
37 "Quality of Life"/ (236395)
38 quality of life.tw. (277385)
39 "Value of Life"/ (5782)
40 Quality-Adjusted Life Years/ (14531)
41 quality adjusted life.tw. (13388)
42 (qaly$ or qald$ or qale$ or qtime$).tw. (10958)
43 disability adjusted life.tw. (3643)
44 daly$.tw. (3227)
45 Health Status Indicators/ (24055)
46 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw. (25252)
47 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (1481)
48 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (5920)
49 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw. (31)
50 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw. (406)
51 (euroqol or euro qol or eq5d or eq 5d).tw. (11885)
52 (qol or hql or hqol or hrqol).tw. (54176)
53 (hye or hyes).tw. (63)
54 health$ year$ equivalent$.tw. (38)
55 utilit$.tw. (201486)
56 (hui or hui1 or hui2 or hui3).tw. (1491)
57 disutili$.tw. (467)
58 rosser.tw. (99)
59 quality of wellbeing.tw. (24)
60 quality of well-being.tw. (414)
61 qwb.tw. (196)
62 willingness to pay.tw. (5941)
63 standard gamble$.tw. (826)

64 time trade off.tw. (1148)
65 time tradeoff.tw. (248)
66 tto.tw. (1046)
67 or/37-66 (578329)
68 36 or 67 (1558035)
69 10 and 68 (593)
70 animals/ not humans/ (4945537)
71 69 not 70 (591)
72 limit 71 to ed=20100901-20220329 (410)
73 limit 72 to english language/ (385)

Database name: MEDLINE in Process

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (48)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (182)
4 or/1-3 (182)
5 Femur Neck/su or Femoral Neck Fractures/su or Prosthesis Design/ (0)
6 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (9)
7 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (146)
8 ("c-stem" or "c stem").tw. (0)
9 or/5-8 (154)
10 4 and 9 (15)
11 Economics/ (0)
12 exp "Costs and Cost Analysis"/ (0)
13 Economics, Dental/ (0)
14 exp Economics, Hospital/ (0)
15 exp Economics, Medical/ (0)
16 Economics, Nursing/ (0)
17 Economics, Pharmaceutical/ (0)
18 Budgets/ (0)
19 exp Models, Economic/ (0)
20 Markov Chains/ (0)
21 Monte Carlo Method/ (0)
22 Decision Trees/ (0)
23 econom$.tw. (1911)
24 cba.tw. (29)
25 cea.tw. (90)
26 cua.tw. (4)
27 markov$.tw. (168)
28 (monte adj carlo).tw. (197)
29 (decision adj3 (tree$ or analys$)).tw. (226)
30 (cost or costs or costing$ or costly or costed).tw. (3387)
31 (price$ or pricing$).tw. (263)
32 budget$.tw. (130)
33 expenditure$.tw. (347)
34 (value adj3 (money or monetary)).tw. (16)
35 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (8)
36 or/11-35 (5821)

37 "Quality of Life"/ (0)
38 quality of life.tw. (2211)
39 "Value of Life"/ (0)
40 Quality-Adjusted Life Years/ (0)
41 quality adjusted life.tw. (106)
42 (qaly$ or qald$ or qale$ or qtime$).tw. (82)
43 disability adjusted life.tw. (46)
44 daly$.tw. (39)
45 Health Status Indicators/ (0)
46 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw. (90)
47 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (6)
48 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (40)
49 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw. (1)
50 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw. (0)
51 (euroqol or euro qol or eq5d or eq 5d).tw. (116)
52 (qol or hql or hqol or hrqol).tw. (481)
53 (hye or hyes).tw. (0)
54 health$ year$ equivalent$.tw. (0)
55 utilit$.tw. (1600)
56 (hui or hui1 or hui2 or hui3).tw. (9)
57 disutili$.tw. (1)
58 rosser.tw. (1)
59 quality of wellbeing.tw. (0)
60 quality of well-being.tw. (4)
61 qwb.tw. (1)
62 willingness to pay.tw. (62)
63 standard gamble$.tw. (0)
64 time trade off.tw. (2)
65 time tradeoff.tw. (1)
66 tto.tw. (9)
67 or/37-66 (3907)
68 36 or 67 (9267)
69 10 and 68 (1)
70 animals/ not humans/ (0)
71 69 not 70 (1)
72 limit 71 to dt=20100901-20220329 (1)
73 limit 72 to english language/ (1)

Database name: MEDLINE ePubs

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (165)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (696)
4 or/1-3 (698)
5 Femur Neck/su or Femoral Neck Fractures/su or Prosthesis Design/ (0)

6 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (49)
7 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (346)
8 ("c-stem" or "c stem").tw. (1)
9 or/5-8 (392)
10 4 and 9 (66)
11 Economics/ (0)
12 exp "Costs and Cost Analysis"/ (0)
13 Economics, Dental/ (0)
14 exp Economics, Hospital/ (0)
15 exp Economics, Medical/ (0)
16 Economics, Nursing/ (0)
17 Economics, Pharmaceutical/ (0)
18 Budgets/ (0)
19 exp Models, Economic/ (0)
20 Markov Chains/ (0)
21 Monte Carlo Method/ (0)
22 Decision Trees/ (0)
23 econom$.tw. (8130)
24 cba.tw. (55)
25 cea.tw. (254)
26 cua.tw. (16)
27 markov$.tw. (637)
28 (monte adj carlo).tw. (836)
29 (decision adj3 (tree$ or analys$)).tw. (604)
30 (cost or costs or costing$ or costly or costed).tw. (13035)
31 (price$ or pricing$).tw. (1106)
32 budget$.tw. (585)
33 expenditure$.tw. (1115)
34 (value adj3 (money or monetary)).tw. (73)
35 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (58)
36 or/11-35 (22587)
37 "Quality of Life"/ (0)
38 quality of life.tw. (7971)
39 "Value of Life"/ (0)
40 Quality-Adjusted Life Years/ (0)
41 quality adjusted life.tw. (442)
42 (qaly$ or qald$ or qale$ or qtime$).tw. (372)
43 disability adjusted life.tw. (112)
44 daly$.tw. (101)
45 Health Status Indicators/ (0)
46 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw. (419)
47 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (43)
48 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (152)
49 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw. (0)
50 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw. (3)

51 (euroqol or euro qol or eq5d or eq 5d).tw. (482)
52 (qol or hql or hqol or hrqol).tw. (1598)
53 (hye or hyes).tw. (1)
54 health$ year$ equivalent$.tw. (0)
55 utilit$.tw. (4502)
56 (hui or hui1 or hui2 or hui3).tw. (25)
57 disutili$.tw. (19)
58 rosser.tw. (0)
59 quality of wellbeing.tw. (2)
60 quality of well-being.tw. (6)
61 qwb.tw. (3)
62 willingness to pay.tw. (235)
63 standard gamble$.tw. (4)
64 time trade off.tw. (23)
65 time tradeoff.tw. (1)
66 tto.tw. (26)
67 or/37-66 (12844)
68 36 or 67 (33532)
69 10 and 68 (7)
70 animals/ not humans/ (0)
71 69 not 70 (7)
72 limit 71 to english language/ (7)

Database name: Embase

1 exp hip fracture/ (44319)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (14986)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (60701)
4 or/1-3 (73537)
5 femoral neck fracture/su (1619)
6 prosthesis design/ (6884)
7 ((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw. (3184)
8 ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw. (28709)
9 ("c-stem" or "c stem").tw. (101)
10 or/5-9 (39242)
11 4 and 10 (5961)
12 exp Health Economics/ (950321)
13 exp "Health Care Cost"/ (316116)
14 exp Pharmacoeconomics/ (217382)
15 Monte Carlo Method/ (45772)
16 Decision Tree/ (17022)
17 econom$.tw. (434359)
18 cba.tw. (13479)
19 cea.tw. (38197)
20 cua.tw. (1692)
21 markov$.tw. (35274)
22 (monte adj carlo).tw. (55085)
23 (decision adj3 (tree$ or analys$)).tw. (30235)

24 (cost or costs or costing$ or costly or costed).tw. (887377)
25 (price$ or pricing$).tw. (65394)
26 budget$.tw. (43163)
27 expenditure$.tw. (83177)
28 (value adj3 (money or monetary)).tw. (3891)
29 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (9196)
30 or/12-29 (2020599)
31 "Quality of Life"/ (547615)
32 Quality Adjusted Life Year/ (31127)
33 Quality of Life Index/ (2991)
34 Short Form 36/ (34400)
35 Health Status/ (140112)
36 quality of life.tw. (517279)
37 quality adjusted life.tw. (23278)
38 (qaly$ or qald$ or qale$ or qtime$).tw. (23612)
39 disability adjusted life.tw. (5149)
40 daly$.tw. (4958)
41 (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shortform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw. (46052)
42 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (2692)
43 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (10914)
44 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw. (65)
45 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw. (487)
46 (euroqol or euro qol or eq5d or eq 5d).tw. (25647)
47 (qol or hql or hqol or hrqol).tw. (114991)
48 (hye or hyes).tw. (149)
49 health$ year$ equivalent$.tw. (41)
50 utilit$.tw. (334600)
51 (hui or hui1 or hui2 or hui3).tw. (2729)
52 disutili$.tw. (1076)
53 rosser.tw. (135)
54 quality of wellbeing.tw. (59)
55 quality of well-being.tw. (534)
56 qwb.tw. (263)
57 willingness to pay.tw. (10926)
58 standard gamble$.tw. (1152)
59 time trade off.tw. (1885)
60 time tradeoff.tw. (307)
61 tto.tw. (1957)
62 or/31-61 (1146823)
63 30 or 62 (2983810)
64 11 and 63 (684)
65 limit 64 to (books or chapter or conference abstract or conference paper or "conference review" or editorial or letter) (108)
66 64 not 65 (576)
67 nonhuman/ not (human/ and nonhuman/) (4955125)
68 66 not 67 (575)

69 limit 68 to dc=20100901-20220329 (479)
70 limit 69 to english language/ (455)

Database name: Econlit

1	((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw.	1
2	((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw.	48
3	1 or 2	48
4	((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*)).tw.	0
5	((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*)).tw.	2
6	("c-stem" or "c stem").tw.	0
7	4 or 5 or 6	2
8	3 and 7	0

Database name: CRD databases

EED

1	MeSH DESCRIPTOR Hip Fractures EXPLODE ALL TREES	252
2	((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture) OR ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$)	0
3	#1 OR #2	252
4	MeSH DESCRIPTOR femur neck EXPLODE ALL TREES WITH QUALIFIER SU	1
5	MeSH DESCRIPTOR femoral neck fractures EXPLODE ALL TREES WITH QUALIFIER SU	30
6	MeSH DESCRIPTOR Prosthesis Design EXPLODE ALL TREES	274
7	((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*))	35
8	((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*))	182
9	("c-stem" or "c stem")	0

10	#4 OR #5 OR #6 OR #7 OR #8 OR #9	456
11	#3 AND #10	46
12	(*) and ((Economic evaluation:ZDT and Abstract:ZPS)) FROM 2010 TO 2022	1370
13	#11 AND #12	0

HTA

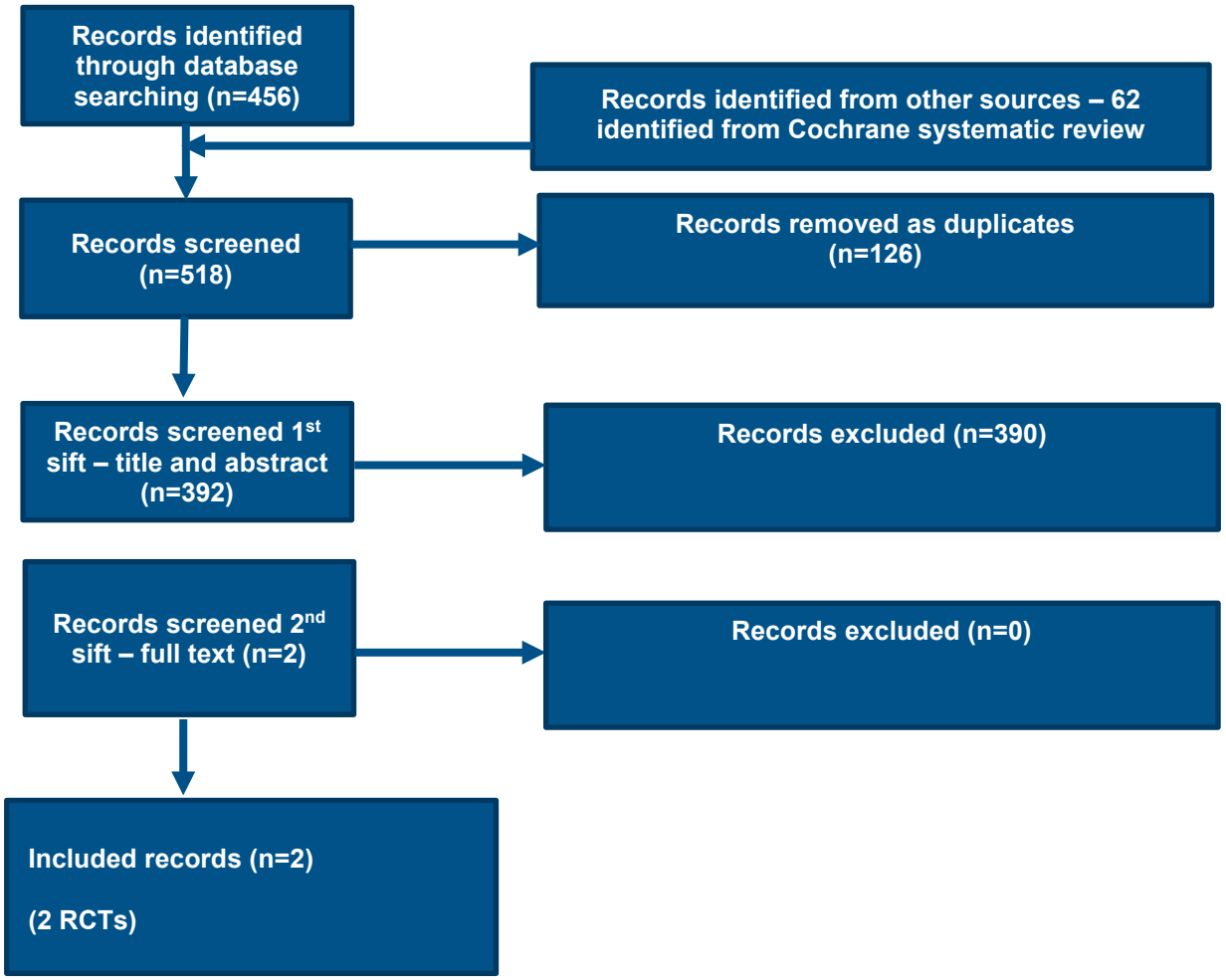
1	MeSH DESCRIPTOR Hip Fractures EXPLODE ALL TREES	252
2	((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture) OR ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$)	0
3	#1 OR #2	252
4	MeSH DESCRIPTOR femur neck EXPLODE ALL TREES WITH QUALIFIER SU	1
5	MeSH DESCRIPTOR femoral neck fractures EXPLODE ALL TREES WITH QUALIFIER SU	30
6	MeSH DESCRIPTOR Prosthesis Design EXPLODE ALL TREES	274
7	((hemiarthroplast* or partial*) adj4 (hip* or femor* or femur*))	35
8	((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) adj4 (stem* or implant* or prosthe*))	182
9	("c-stem" or "c stem")	0
10	#4 OR #5 OR #6 OR #7 OR #8 OR #9	456
11	#3 AND #10	46
12	(*) and (Full publication record:ZDT) FROM 2010 TO 2022	7429
13	#11 AND #12	1

Database name: INAHTA (International HTA Database)

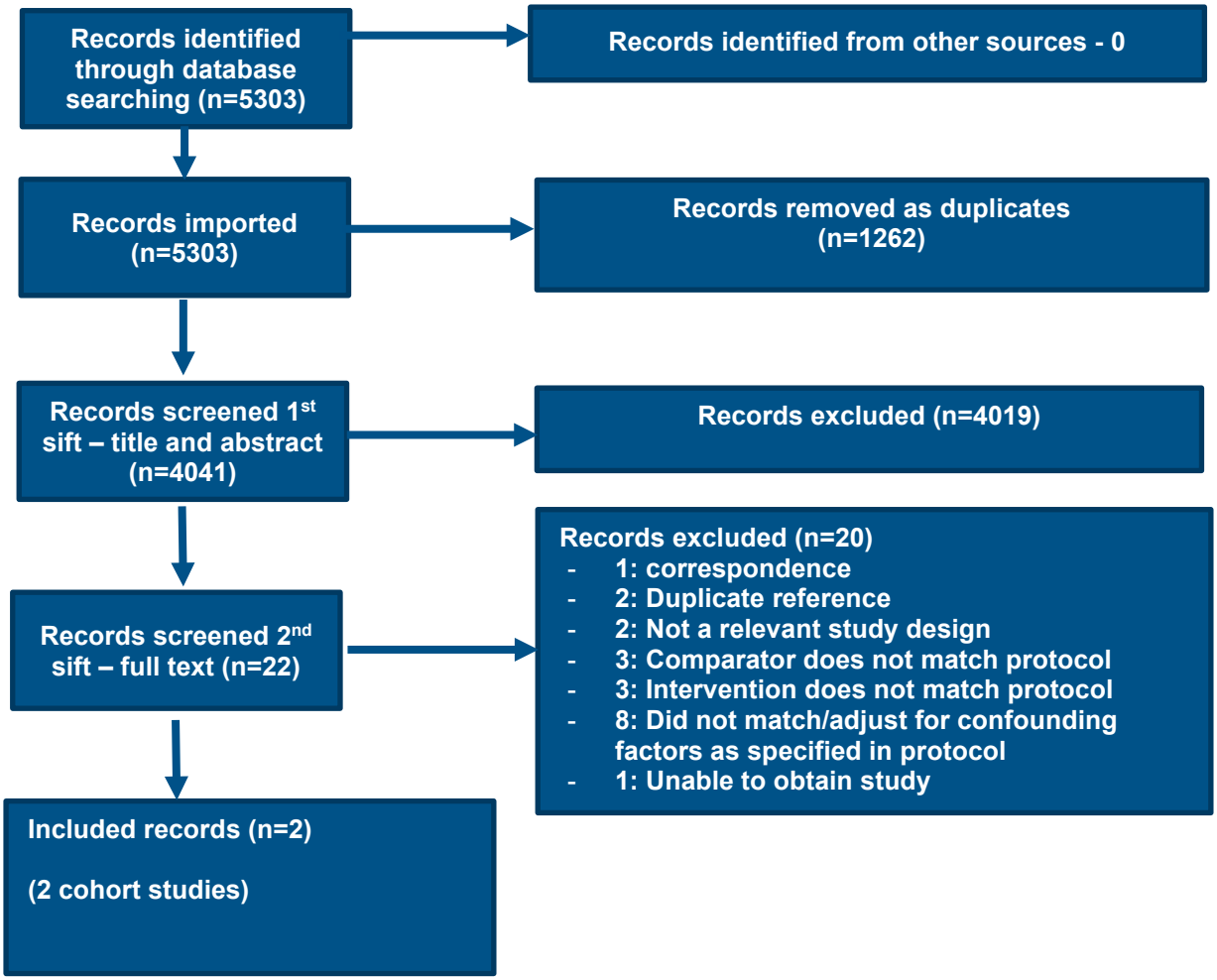
(("c-stem" or "c stem")[Title] OR ("c-stem" or "c stem")[abs]) OR (((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) AND (stem* or implant* or prosthe*))[Title] OR ((femor* or femur* or hip* or exeter or unitrax or thompson* or "austin moore" or corail or furlong) AND (stem* or implant* or prosthe*))[abs]) OR (((hemiarthroplast* or partial*) AND (hip* or femor* or femur*))[Title] OR ((hemiarthroplast* or partial*) AND (hip* or femor* or femur*))[abs]) OR ("Prosthesis Design"[mh]) OR ("Femoral Neck Fractures"[mh]) OR ("Femur Neck"[mh])) AND (((hip* or Femur* or femoral* or trochant*or pertrichant* or intertrochant* or subtrochant* or intracapsular* AND fracture*)[Title] OR (hip* or Femur* or femoral* or trochant*or pertrichant* or intertrochant* or subtrochant* or intracapsular* AND fracture*)[abs]) OR ((femur* or femoral* AND (head or neck or proximal) AND Fracture*)[Title] OR (femur* or femoral* AND (head or neck or proximal) AND Fracture*)[abs]) OR ("Hip Fractures"[mhe]))

– Effectiveness evidence study selection

B.1.1.1 RCT search



B.1.1.2 Observational search



Appendix C – Effectiveness evidence

Cochrane Systematic Review
Lewis et al. 2022

Bibliographic Reference	Lewis SR; Macey R; Parker MJ; Cook JA; Griffin XL; Arthroplasties for hip fracture in adults.; The Cochrane database of systematic reviews; vol. 2

Study Characteristics

Study design	Systematic review
Study details	Dates searched Up to July 2020
Inclusion criteria	Randomised controlled trials (RCTs) and quasi-RCTs comparing different arthroplasties for treating fragility intracapsular hip fractures in older adults. THAs and HAs inserted with or without cement, and comparisons between different articulations, sizes, and types of prostheses
Exclusion criteria	Excluded studies of people with specific pathologies other than osteoporosis and with hip fractures resulting from high-energy trauma.
Intervention(s)	Different Arthorplasties THAs and HAs inserted with or without cement, and comparisons between different articulations, sizes, and types of prostheses.
Outcome(s)	• Activities of daily living (e.g. Barthel Index (BI), Functional Independence Measure (FIM)) • Delirium using recognised assessment scores, such as Mini mental test score or 4AT • Functional status (region specific) (e.g. hip rating questionnaire, Harris Hip Score, Oxford Hip Score) • Health-related Quality-of-Life (HRQoL) (e.g. SF36, EQ-5D)

	• Mobility (e.g. indoor/outdoor walking status, Cumulated Ambulation Score, Elderly Mobility Scale Score, Timed up and go, Short Physical Performance Battery, self-reported walking scores (e.g. Mobility Assessment Tool - short form)) • Mortality • Unplanned return to theatre: secondary procedure required for a complication resulting directly or indirectly from the index operation/primary procedure
Number of studies included in the systematic review	58 studies
Studies from the systematic review that are relevant for use in the current review	Parker 2012 Sims 2018
Studies from the systematic review that are not relevant for use in the current review	Abdelkhalek 2011 Baker 2006 Blomfeldt 2007 Brandfoot 2000 Cadossi 2013 Calder 1995 Calder 1996 Cao 2017

	Chammout 2017
	Chammout 2019
	Cornell 1998
	Davison 2001
	DeAngelis 2012
	Dorr 1986
	Emery 1991
	Fernandez 2022
	Figved 2009
	Figved 2018
	Griffin 2016
	Harper 1994
	HEALTH 2019
	Hedbeck 2011
	Inngul 2015
	Iorio 2019

	Jeffcote 2010
	Kanto 2014
	Keating 2006
	Kim 2012
	Lim 2020
	Livesley 1993
	Macaulay 2008
	Malhotra 1995
	Moerman 2017
	Moroni 2002
	Mouzopoulos 2008
	Movrin 2020
	Parker 2010c
	Parker 2019
	Parker 2020
	Patel 2008

	Raia 2003
	Rashed_2020
	Ravikumar 2000
	Rehman 2014
	Ren 2017
	Sadr 1977
	Santini 2005
	Sharma 2016
	Sonaje 2017
	Sonne-Holm 1982
	Stoffel 2013
	Talsnes 2013
	Taylor 2012
	Van den_Bekerom 2010
	Vidovic 2013
	Xu 2017

Additional comments	Summary details of included RCTs available in summary table 3 and full evidence tables and risk of bias assessments can be found in Lewis 2022

Critical appraisal - GDT Crit App - ROBIS checklist

Section	Question	Answer
Study eligibility criteria	Did the review adhere to pre-defined objectives and eligibility criteria?	Yes *(Protocol registered with PROSPERO CRD42019149095)*
Study eligibility criteria	Were the eligibility criteria appropriate for the review question?	Yes
Study eligibility criteria	Were eligibility criteria unambiguous?	Yes
Study eligibility criteria	Were all restrictions in eligibility criteria based on study characteristics appropriate (e.g. date, sample size, study quality, outcomes measured)?	Yes
Study eligibility criteria	Were any restrictions in eligibility criteria based on sources of information appropriate (e.g. publication status or format, language, availability of data)?	Yes
Study eligibility criteria	Concerns regarding specification of study eligibility criteria	Low *(Eligibility criteria reasonable for review question and protocol registered a priori)*

Section	Question	Answer
Identification and selection of studies	Did the search include an appropriate range of databases/electronic sources for published and unpublished reports?	Yes
Identification and selection of studies	Were methods additional to database searching used to identify relevant reports?	Yes *(hand searched these conference abstracts from 2016 to November 2018:• Fragility Fractures Network Congress;• British Orthopaedic Association Congress;• Orthopaedic World Congress (SICOT);• Orthopaedic Trauma Association Annual Meeting;• Bone and Joint Journal Orthopaedic Proceedings;• American Academy of Orthopaedic Surgeons Annual Meeting.)*
Identification and selection of studies	Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Yes
Identification and selection of studies	Were restrictions based on date, publication format, or language appropriate?	Yes *(no restrictions on language, date, or publication status - no date restriction)*
Identification and selection of studies	Were efforts made to minimise error in selection of studies?	Yes
Identification and selection of studies	Concerns regarding methods used to identify and/or select studies	Low
Data collection and study appraisal	Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	Yes
Data collection and study appraisal	Were all relevant study results collected for use in the synthesis?	Yes

Section	Question	Answer
Data collection and study appraisal	Was risk of bias (or methodological quality) formally assessed using appropriate criteria?	Yes
Data collection and study appraisal	Were efforts made to minimise error in risk of bias assessment?	Yes *(For each domain, two review authors judged whether study authors made sufficient attempts to minimise bias in their design.)*
Data collection and study appraisal	Concerns regarding methods used to collect data and appraise studies	Low
Synthesis and findings	Did the synthesis include all studies that it should?	Probably yes *(Authors 'planned to investigate the potential for publication bias and explore possible small-study biases using funnel plots. However, we had insufficient studies (fewer than 10 studies) for most outcomes (Sterne 2017). For outcomes with 10 or more studies, we constructed a funnel plot and interpreted the plot using a visual inspection and the Harbord modified test in Stata; for the critical review outcomes, we reported P values for the Harbord modified test. We incorporated this judgement into the assessment of publication bias within the GRADE assessment. To assess outcome reporting bias, we screened clinical trials registers for protocols and registration documents of included studies that were prospectively published, and we sourced all clinical trials register documents that were reported in the study reports of included studies. We used evidence of prospective registration to judge whether studies were at risk of selective reporting bias.')*
Synthesis and findings	Were all pre-defined analyses reported or departures explained?	Yes *(Protocol published and deviations from this explained)*
Synthesis and findings	Was the synthesis appropriate given the nature and similarity in the research questions, study designs and outcomes across included studies?	Yes *(Synthesis appropriate within the relevant comparisons - THA vs HA, cemtented vs uncemented etc. Authors 'conducted meta-analyses only when meaningful; that is, when the treatments, participants, and the underlying clinical question were similar enough for pooling to make sense. We pooled results of*

Section	Question	Answer
		comparable groups of trials using random-effects models. We chose this model after careful consideration of the extent to which any underlying effect could truly be thought to be fixed, given the complexity of the interventions included in this review. We presented 95% CIs throughout. We found that some studies reported outcome data at more than one time point and we reported the data within three time point windows for the studies. Early data included data up to four months, with priority given to data closest to four months; 12-month data included a window from later than four months up to 24 months, but with priority given to data at 12 months; and late data, which included data reported after 24 months at the latest time point reported by study authors. For studies that reported outcome data using more than one measurement tool, we selected the tool that was used most commonly by other studies in the comparison group, or which reported data for the largest number of participants.')
Synthesis and findings	Was between-study variation (heterogeneity) minimal or addressed in the synthesis?	Yes *'(We used the I^2 statistic, automatically calculated in Review Manager 2014 software, to quantify the possible degree of heterogeneity of treatment effects between trials. We assumed moderate heterogeneity when the I^2 was between 30% and 60%; substantial heterogeneity when it was between 50% and 90%; and considerable heterogeneity when it was between 75% and 100%. We noted the importance of I^2 depending on: 1) magnitude and direction of effects; and 2) strength of evidence for heterogeneity. We pooled results of comparable groups of trials using random-effects models. We chose this model after careful consideration of the extent to which any underlying effect could truly be thought to be fixed, given the complexity of the interventions included in this review)'*
Synthesis and findings	Were the findings robust, e.g. as demonstrated through funnel plot or sensitivity analyses?	Yes
Synthesis and findings	Were biases in primary studies minimal or addressed in the synthesis?	Yes

Section	Question	Answer
Synthesis and findings	Concerns regarding the synthesis and findings	Low
Overall study ratings	Overall risk of bias	Low *(No concerns with study eligibility criteria, search strategy, data collection or data synthesis)*
Overall study ratings	Applicability as a source of data	Partially applicable *(Some comparisons (THA vs HA, cemented vs uncemented) not relevant to this review.)*

RCTs

Details of included RCTs available in summary table 3 and full evidence tables and risk of bias assessments can be found in Lewis 2022

Observational Studies
Mellner, 2021

Bibliographic Reference	Mellner, Carl; Mohammed, Jabbar; Larsson, Magnus; Esberg, Sandra; Szymanski, Maciej; Hellstrom, Nils; Chang, Cecilia; Berg, Hans E; Skoldenberg, Olof; Knutsson, Bjorn; Morberg, Per; Mukka, Sebastian; Increased risk for postoperative periprosthetic fracture in hip fracture patients with the Exeter stem than the anatomic SP2 Lubinus stem.; European journal of trauma and emergency surgery : official publication of the European Trauma Society; 2021; vol. 47 (no. 3); 803-809

Study details

Trial registration number and/or trial name	NCT03326271
Study type	Retrospective cohort study

Study location	Sweden
Study setting	Three Swedish hospitals: the orthopedic department of Sundsvall Hospital, the orthopedic department of Sunderby Hospital, and the orthopedic department of Karolinska University Hospital Huddinge, Stockholm
Study dates	2006-2014
Sources of funding	funded by grants from the regional agreement on medical training and clinical research (ALF) between Västerbotten County Council and Umeå University and the Research and Development Centre (FoU) for Västernorrland, Norrbotten, Sörmland County Councils and the Visare Norr Fund, Northern County Councils.
Inclusion criteria	Age >60 Underwent primary hip arthroplasty for a displaced FNF with either a cemented Exeter stem or a cemented Lubinus SP2 stem.
Exclusion criteria	Patients with pathological fractures were excluded.
Intervention(s)	collarless, polished, tapered Exeter stem
Comparator	anatomic Lubinus SP2 stem
Number of participants	Periprosthetic Fractures
Duration of follow-up	Median - 47 months
Loss to follow-up	Retrospective design
Methods of analysis	Retrospective comparative observational study. Cox proportional hazards for regression modelling with follow-up time as time to death, PPF, or end of follow-up (min 2 years after surgery). The selection of variables for the analyses was an a priori hypothesis based on the literature search for known predictors of the outcome of interest. Our main outcome variable was the presence of a PPF during the study period and we adjusted for exposure variable (type of stem), age, sex, surgical approach (direct lateral or posterior), and type of arthroplasty (hemi- or total hip arthroplasty) achieving 8–10 events per predictor variable.

Study arms
Tapered Exeter Stem (N = 1326)

Lubinus SP2 (N = 1202)

Characteristics

Arm-level characteristics

Characteristic	Tapered Exeter Stem (N = 1326)	Lubinus SP2 (N = 1202)
% Female Custom value	69%	68%
Mean age (SD) Standardised Mean (SD)	82 (8)	81 (8)
Type of Arthroplasty	Hemiarthroplasty: 84%	Hemiarthroplasty: 83%

Critical appraisal - GDT Crit App - ROBINS-I: a tool for non-randomised studies of interventions

Section	Question	Answer
1. Bias due to confounding	1.1 Is there potential for confounding of the effect of intervention in this study?	Yes *(Observational study - regression analysis was carried out but unable to address all confounding factors)*
1. Bias due to confounding	1.2. Was the analysis based on splitting participants' follow up time according to intervention received?	No
1. Bias due to confounding	1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?	Not applicable
1. Bias due to confounding	1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Probably yes
1. Bias due to confounding	1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Yes
1. Bias due to confounding	1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	No
1. Bias due to confounding	1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable

Section	Question	Answer
1. Bias due to confounding	1.8. If Y/PY to 1.7: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable
1. Bias due to confounding	Risk of bias judgement for confounding	Moderate *(Appropriate regression analysis used but residual unknown confounding expected)*
2. Bias in selection of participants into the study	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If N/PN to 2.1: go to 2.4	No
2. Bias in selection of participants into the study	2.2. If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention?	Not applicable
2. Bias in selection of participants into the study	2.3 If Y/PY to 2.2: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	Not applicable
2. Bias in selection of participants into the study	2.4. Do start of follow-up and start of intervention coincide for most participants?	Probably yes
2. Bias in selection of participants into the study	2.5. If Y/PY to 2.2 and 2.3, or N/PN to 2.4: Were adjustment techniques used that are	Not applicable

Section	Question	Answer
	likely to correct for the presence of selection biases?	
2. Bias in selection of participants into the study	Risk of bias judgement for selection of participants into the study	Low
3. Bias in classification of interventions	3.1 Were intervention groups clearly defined?	No *(Although stem choice was clearly defined, a 'unipolar or bipolar' head was used for HAs, with no indication as to what proportion in each arm)*
3. Bias in classification of interventions	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Yes
3. Bias in classification of interventions	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	No
3. Bias in classification of interventions	Risk of bias judgement for classification of interventions	Moderate *(both unipolar and biploar heads used)*
4. Bias due to deviations from intended interventions	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	No
4. Bias due to deviations from intended interventions	4.2. If Y/PY to 4.1: Were these deviations from intended intervention unbalanced between groups and likely to have affected the outcome?	Not applicable

Section	Question	Answer
4. Bias due to deviations from intended interventions	4.3. Were important co-interventions balanced across intervention groups?	Yes *(Antibiotic-loaded bone cement was used for all patients.Prophylactic antibiotics were administered 30 min preoperativelyand two more times over 24 h postoperatively.Low molecular weight heparin was administered for14–30 days postoperatively.Patients were mobilized according to a standard physiotherapeuticprogram and full weight bearing with theuse of crutches was encouraged)*
4. Bias due to deviations from intended interventions	4.4. Was the intervention implemented successfully for most participants?	Probably yes
4. Bias due to deviations from intended interventions	4.5. Did study participants adhere to the assigned intervention regimen?	Not applicable
4. Bias due to deviations from intended interventions	4.6. If N/PN to 4.3, 4.4 or 4.5: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable
4. Bias due to deviations from intended interventions	Risk of bias judgement for deviations from intended interventions	Low
5. Bias due to missing data	5.1 Were outcome data available for all, or nearly all, participants?	Yes
5. Bias due to missing data	5.2 Were participants excluded due to missing data on intervention status?	Yes *(One patient was excluded due to insufficient documentation but not included in this retrospective analysis)*

Section	Question	Answer
5. Bias due to missing data	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	No
5. Bias due to missing data	5.4 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?	Not applicable *(excluded before analysis)*
5. Bias due to missing data	5.5 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Is there evidence that results were robust to the presence of missing data?	Not applicable
5. Bias due to missing data	Risk of bias judgement for missing data	Low
6. Bias in measurement of outcomes	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	No
6. Bias in measurement of outcomes	6.2 Were outcome assessors aware of the intervention received by study participants?	Probably yes
6. Bias in measurement of outcomes	6.3 Were the methods of outcome assessment comparable across intervention groups?	Yes
6. Bias in measurement of outcomes	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	No

Section	Question	Answer
6. Bias in measurement of outcomes	Risk of bias judgement for measurement of outcomes	Low
7. Bias in selection of the reported result	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain?	No
7. Bias in selection of the reported result	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple analyses of the intervention-outcome relationship?	Probably no *(Results presented as adjusted hazard ratio)*
7. Bias in selection of the reported result	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different subgroups?	No
7. Bias in selection of the reported result	Risk of bias judgement for selection of the reported result	Low
Overall bias	Risk of bias judgement	Moderate *(Although study adjusted for confounding factors as specified in protocol, residual confounding expected with observational evidence)*
Overall bias	Directness	Partially Applicable *(17% of procedures were THA and not HA as specified in the protocol. Authors also state that either unipolar or bipolar heads were used for HA's but do not provide respective data on how many of each.)*

Mohammed, 2019

Bibliographic Reference	Mohammed, Jabbar; Mukka, Sebastian; Hedbeck, Carl-Johan; Chammout, Ghazi; Gordon, Max; Skoldenberg, Olof; Reduced periprosthetic fracture rate when changing from a tapered polished stem to an anatomical stem for cemented hip arthroplasty: an observational prospective cohort study with a follow-up of 2 years.; Acta orthopaedica; 2019; vol. 90 (no. 5); 427-432

Study details

Study type	Prospective cohort study
Study location	Sweden
Study setting	Orthopedic Department of Danderyd Hospital in Stockholm, Sweden
Study dates	between 2012 and the beginning of 2018 (inclusion period 2012–2015)
Sources of funding	Funded by the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet and by a research grant from LINK.
Inclusion criteria	All patients operated between 2012 and 2015 with a cemented hip arthroplasty.
Exclusion criteria	Uncemented stems and bilateral cases of cemented stems
Intervention(s)	Anatomic stem (AS group) (Lubinus SP2, Waldemar Link, Hamburg, Germany).
Comparator	Polished tapered stem (PTS group) (CPT, Zimmer Inc., Warsaw, IN, USA)
Outcome measures	Periprosthetic fracture
Number of participants	n=1077
Duration of follow-up	2 years
Loss to follow-up	none

Methods of analysis	Prospective cohort study Cox proportional hazards with follow-up time defined as time to death, reoperation, or end of follow-up (max. 2 years after surgery). Our main outcome variable was the occurrence of a PPF during the study period and we adjusted for exposure variable (PTS/AS), age, sex, ASA category, cognitive impairment, BMI, whether the indication was fracture or not, and surgical approach. Results are presented as hazard ratios (HRs) with 95% confidence intervals (CI). The statistical analysis is based on the assumption that the studied observations are independent; therefore, no bilateral fractures were included. In patients with 2 fractures during the study period, only the 1st fracture was included. All continuous variables were left as continuous but checked for non-linearity using ANOVA. We investigated the proportional hazards assumption using Grambsch and Therneau analysis of Schoenfeld residuals. All analyses were performed using R 3.5.2 (R Project for Statistical Computing, Vienna, Austria), using the rms package (v. 5.1-3) for survival modelling, knitr (v. 1.21) for reproducible research, ggplot2 for plots (v. 3.1.0) and Gmisc (v. 1.8) with Greg (v. 1.3) for table output.

Study arms

Lubinus SP2 (N = 534)

Zimmer (N = 543)

Characteristics

Arm-level characteristics

Characteristic	**Lubinus SP2 (N = 534)**	**Zimmer (N = 543)**
% Female Custom value	75%	71%

Characteristic	Lubinus SP2 (N = 534)	Zimmer (N = 543)
Mean age (SD) Mean (SD)	82 (8)	82 (8.4)
BMI (kg/m2) Mean (SD)	24 (4.5)	24 (4.1)
Type of Arthroplasty	54% HA	61% HA

Critical appraisal - GDT Crit App - ROBINS-I: a tool for non-randomised studies of interventions

Section	Question	Answer
1. Bias due to confounding	1.1 Is there potential for confounding of the effect of intervention in this study?	Probably yes *(Regression analysis taking into account appropriate confounding factors, but residual confounding still expected)*
1. Bias due to confounding	1.2. Was the analysis based on splitting participants' follow up time according to intervention received?	No
1. Bias due to confounding	1.3. Were intervention discontinuations or switches likely to be related to factors that are prognostic for the outcome?	Not applicable

Section	Question	Answer
1. Bias due to confounding	1.4. Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Probably yes *(Cox proportional hazards with follow-up time defined as time to death, reoperation, or end of follow-up (max. 2 years after surgery). Adjusted for exposure variable (PTS/AS), age, sex, ASA category, cognitive impairment,BMI, whether the indication was fracture or not, and surgicalapproach.)*
1. Bias due to confounding	1.5. If Y/PY to 1.4: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Probably yes
1. Bias due to confounding	1.6. Did the authors control for any post-intervention variables that could have been affected by the intervention?	No
1. Bias due to confounding	1.7. Did the authors use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding?	Not applicable
1. Bias due to confounding	1.8. If Y/PY to 1.7: Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Not applicable
1. Bias due to confounding	Risk of bias judgement for confounding	Moderate *(Appropriate adjustments for variables made but residual confounding expected with observational evidence)*
2. Bias in selection of participants into the study	2.1. Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention? If N/PN to 2.1: go to 2.4	No

Section	Question	Answer
2. Bias in selection of participants into the study	2.2. If Y/PY to 2.1: Were the post-intervention variables that influenced selection likely to be associated with intervention?	Not applicable
2. Bias in selection of participants into the study	2.3 If Y/PY to 2.2: Were the post-intervention variables that influenced selection likely to be influenced by the outcome or a cause of the outcome?	Not applicable
2. Bias in selection of participants into the study	2.4. Do start of follow-up and start of intervention coincide for most participants?	Yes
2. Bias in selection of participants into the study	2.5. If Y/PY to 2.2 and 2.3, or N/PN to 2.4: Were adjustment techniques used that are likely to correct for the presence of selection biases?	Not applicable
2. Bias in selection of participants into the study	Risk of bias judgement for selection of participants into the study	Low
3. Bias in classification of interventions	3.1 Were intervention groups clearly defined?	Yes *(Although HA and THA done, the percentage of these in each group is given)*
3. Bias in classification of interventions	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Yes
3. Bias in classification of interventions	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	No

Section	Question	Answer
3. Bias in classification of interventions	Risk of bias judgement for classification of interventions	Low
4. Bias due to deviations from intended interventions	4.1. Were there deviations from the intended intervention beyond what would be expected in usual practice?	No
4. Bias due to deviations from intended interventions	4.2. If Y/PY to 4.1: Were these deviations from intended intervention unbalanced between groups and likely to have affected the outcome?	Not applicable
4. Bias due to deviations from intended interventions	4.3. Were important co-interventions balanced across intervention groups?	Yes *(Antibiotic-loaded bone cement was used for all patients. Prophylactic antibiotics were administered 30minutes preoperatively and twice more over 24 h postoperatively. Low-molecular-weight heparin was administered for 10–30 days postoperatively)*
4. Bias due to deviations from intended interventions	4.4. Was the intervention implemented successfully for most participants?	Yes
4. Bias due to deviations from intended interventions	4.5. Did study participants adhere to the assigned intervention regimen?	Yes
4. Bias due to deviations from intended interventions	4.6. If N/PN to 4.3, 4.4 or 4.5: Was an appropriate analysis used to estimate the effect of starting and adhering to the intervention?	Not applicable

Section	Question	Answer
4. Bias due to deviations from intended interventions	Risk of bias judgement for deviations from intended interventions	Low
5. Bias due to missing data	5.1 Were outcome data available for all, or nearly all, participants?	Probably no *(No loss to follow up but authors admit that there is a risk of under-reporting reoperations to the Swedish hip arthroplasty registry of those PPFs treated with open reduction and internal fixation without change of implant)*
5. Bias due to missing data	5.2 Were participants excluded due to missing data on intervention status?	No
5. Bias due to missing data	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	No
5. Bias due to missing data	5.4 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Are the proportion of participants and reasons for missing data similar across interventions?	No information
5. Bias due to missing data	5.5 If PN/N to 5.1, or Y/PY to 5.2 or 5.3: Is there evidence that results were robust to the presence of missing data?	Yes
5. Bias due to missing data	Risk of bias judgement for missing data	Low
6. Bias in measurement of outcomes	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	No

Section	Question	Answer
6. Bias in measurement of outcomes	6.2 Were outcome assessors aware of the intervention received by study participants?	Probably yes
6. Bias in measurement of outcomes	6.3 Were the methods of outcome assessment comparable across intervention groups?	Probably yes
6. Bias in measurement of outcomes	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	Probably no
6. Bias in measurement of outcomes	Risk of bias judgement for measurement of outcomes	Low
7. Bias in selection of the reported result	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain?	No
7. Bias in selection of the reported result	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple analyses of the intervention-outcome relationship?	No
7. Bias in selection of the reported result	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different subgroups?	No
7. Bias in selection of the reported result	Risk of bias judgement for selection of the reported result	Low

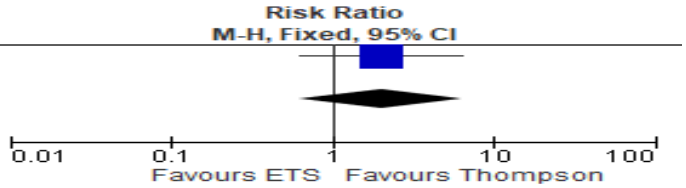
Section	Question	Answer
Overall bias	Risk of bias judgement	Moderate *(Residual confounding expected in observational evidence and also a risk that some PPF were not reported)*
Overall bias	Directness	Partially Applicable *(Intervention contained both HA and THA procedures)*

Appendix D – Forest plots

RCT evidence
Early mortality at 30 days

Study or Subgroup	ETS Events	ETS Total	Thompson Events	Thompson Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Parker 2012	8	100	4	100	100.0%	2.00 [0.62, 6.43]
Total (95% CI)		**100**		**100**	**100.0%**	**2.00 [0.62, 6.43]**
Total events	8		4			

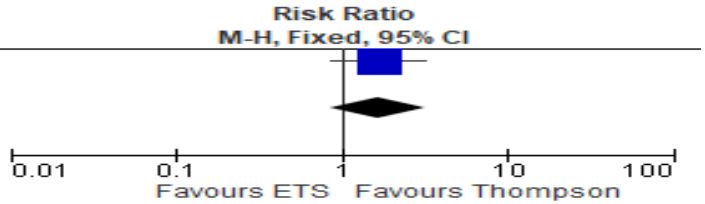
Heterogeneity: Not applicable
Test for overall effect: Z = 1.16 (P = 0.24)



Early mortality at 90 days

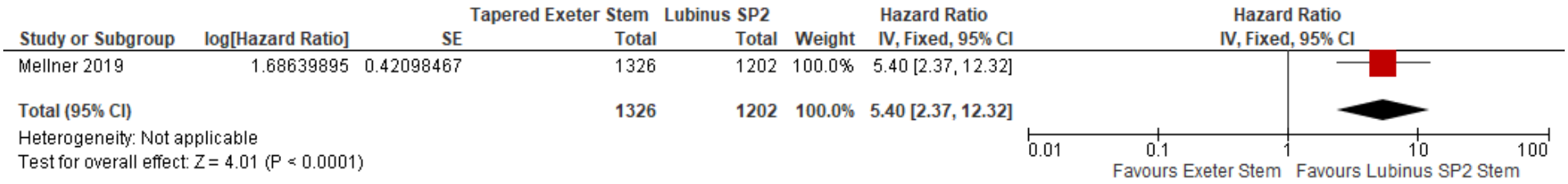
Study or Subgroup	ETS Events	ETS Total	Thompson Events	Thompson Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Parker 2012	20	100	12	100	100.0%	1.67 [0.86, 3.22]
Total (95% CI)		**100**		**100**	**100.0%**	**1.67 [0.86, 3.22]**
Total events	20		12			

Heterogeneity: Not applicable
Test for overall effect: Z = 1.52 (P = 0.13)

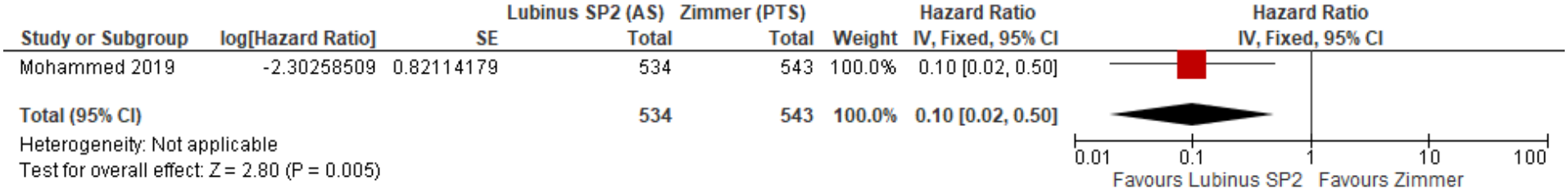


For all other forest plots of RCT evidence, please see systematic review Lewis 2022 – Analysis 6.2 - 6.8, Pg 235-238

Observational studies
Periprosthetic fracture: Exeter Stem Vs Lubinus SP2



Periprosthetic Fracture: Lubinus SP2 vs Zimmer



Appendix E – GRADE tables

Thompson (intervention) Vs Exeter Trauma Stem (comparator) – RCT evidence from Cochrane review

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Absolute risk difference	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Early health related quality of life ≤ 4 months (>0 favours ETS)											
1[1]	RCT	618	MD 0.06 (0.00, 0.11)	-	-	-	No serious risk of bias	N/A[2]	No serious indirectness	Serious imprecision[4]	Moderate
Early mobility (freely mobile without aids, or able to walk outdoors with one aid) (>0 favours ETS)											
1[1]	RCT	494	RR 1.14 (0.83, 1.57)	250 per 1000.[9]	285 per 1000.[10]	35 more per 1000 (42 fewer to 143 more)	No serious risk of bias	N/A[2]	No serious indirectness	Serious imprecision[5]	Moderate
Early mortality – 30 days *(evidence from primary study – outcome not included in Cochrane review)*											
1[8]	RCT	200	RR 2.00 (0.62, 6.43)	40 per 1000[10]	80 per 1000[9]	40 per 1000 (15 fewer to 217 more)	No serious risk of bias	N/A[2]	No serious indirectness	Very serious imprecision[7]	Low
Early mortality – 90 days *(evidence from primary study – outcome not included in Cochrane review)*											

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Absolute risk difference	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1[8]	RCT	200	RR: 1.67 (0.86, 3.22)	120 per 1000	200 per 1000	80 more per 1000 (17 fewer to 266 more)	No serious risk of bias	N/A[2]	No serious indirectness	Serious Imprecision[5]	Moderate
Early mortality ≤ 4 months											
2[2]	RCT	1164	RR 1.20 (0.76, 1.88)	149 per 1000.[10]	179 per 1000.[9]	30 more per 1000 (36 fewer to 131 more)	No serious risk of bias	Serious inconsistency[6]	No serious indirectness	Very serious imprecision[7]	Very Low
Mortality 12 months											
1[8]	RCT	200	RR 1.44 (0.94, 2.21)	250 per 1000.[10]	360 per 1000.[9]	110 more per 1000 (15 fewer to 303 more)	No serious risk of bias	N/A[2]	No serious indirectness	Serious Imprecision[5]	Moderate
Unplanned return to theatre (end of follow up)											
2[2]	RCT	1164	RR 0.46 (0.05, 3.89)	12 per 1000.[10]	6 per 1000.[9]	6 fewer per 1000 (11 fewer to 35 more)	No serious risk of bias	Serious inconsistency[6]	No serious indirectness	Very serious imprecision[7]	Very low
Adverse event related to implant, fracture or both											
Intraoperative periprosthetic fracture											

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Absolute risk difference	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1[8]	RCT	200	RR 1.00 (0.21, 4.84)	30 per 1000.[10]	30 per 1000. [9]	0 more per 1000 (2 fewer to 12 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Deep infection											
1[8]	RCT	200	Not estimable	-	-	-	-	-	-	-	-
Superficial infection											
1[8]	RCT	200	RR 3.00 (0.32, 28.35)	10 per 1000.[10]	30 per 1000.[9]	20 more per 1000 (7 fewer to 274 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Dislocation											
1[8]	RCT	200	RR 0.20 (0.01, 4.11)	20 per 1000. [10]	4 per 1000.[9]	16 fewer per 1000 (20 fewer to 62 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Adverse events unrelated to implant, fracture or both											
Acute Kidney Injury											

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Absolute risk difference	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
1[8]	RCT	200	RR 1.00 (0.06, 15.77)	10 per 1000.[10]	10 per 1000.[9]	0 more per 1000 (9 fewer to 148 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Blood Transfusion											
1[8]	RCT	200	RR 1.00 (0.54, 1.84)	170 per 1000. [10]	170 per 1000. [9]	0 more per 1000 (78 fewer to 143 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Cerebrovascular event											
1[8]	RCT	200	RR 2.00 (0.18, 21.71)	10 per 1000.[10]	20 per 1000.[9]	10 more per 1000 (8 fewer to 207 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Chest infection/pneumonia											
1[8]	RCT	200	RR 1.67 (0.41, 6.79)	30 per 1000.[10]	50 per 1000.[9]	20 more per 1000 (18 fewer to 174 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Myocardial Infarction											
1[8]	RCT	200	RR 5.00	0 per 1000	0 per 1000	0 more per 1000	No serious	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Absolute risk difference	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
			(0.24, 102.85)				risk of bias				
Venous thromboembolic phenomena (DVT)											
1[8]	RCT	200	RR 1.00 (0.21, 4.84)	30 per 1000	30 per 1000	0 more per 1000 (24 fewer to 115 more)	No serious risk of bias	N/A[3]	No serious indirectness	Very serious imprecision[7]	Low
Venous thromboembolic phenomena (pulmonary embolism)											
1[8]	RCT	200	Not estimable	-	-	-	-	-	-	-	-

1. Simms 2018
2. Simms 2018, Parker 2012
3. Single study. Inconsistency not applicable
4. Confidence interval crosses the minimum clinically important difference threshold 0.08 - taken from Simms 2018 – quality downgraded one level
5. Confidence interval crosses one end of the minimum important difference threshold (0.8 – 1.25) – quality downgraded one level
6. I^2 between 33.3% and 66.7%. Quality downgraded 1 level
7. Confidence interval crosses both ends of the minimum important difference threshold (0.8 – 1.25) – quality downgraded two levels
8. Parker 2012
9. Exeter Stem
10. Thompson Stem

Exeter Stem (intervention) Vs Lubinus SP2 (control) – observational evidence from NICE review

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Periprosthetic fracture (>1 favours Lubinus SP2)										
1[1]	Cohort	2528	HR 5.40 (2.37, 12.32)	23 per 1000.[6]	7 per 1000.[5]	Serious risk of bias[2]	N/A[3]	Serious indirectness[4]	No serious imprecision	Low

1. Mellner 2019
2. Moderate risk of bias rating using ROBINS-I – rated down once
3. Single study. Inconsistency not applicable
4. Partially applicable rating using ROBINS-I – rated down once
5. Lubinus SP2
6. Exeter

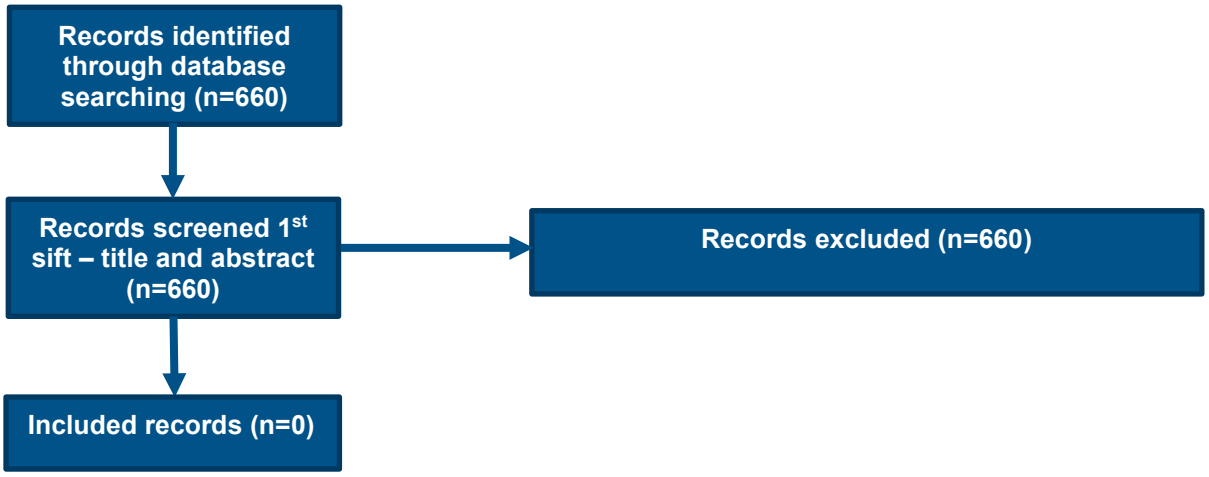
Lubinus SP2 (intervention) Vs Zimmer (control) – observational evidence from NICE review

No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Periprosthetic Fracture (>1 favours Zimmer)										
1[1]	Cohort	1077	HR 0.10 (0.02, 0.50)	33 per 1000.[5]	4 per 1000. [6]	Serious risk of bias[2]	N/A[3]	Serious indirectness[4]	No serious imprecision	Low

1. Mohammed 2019
2. Moderate risk of bias rating using ROBINS-I – rated down once
3. Single study. Inconsistency not applicable
4. Partially applicable rating using ROBINS-I – rated down once

5. Zimmer
6. Lubinus SP2

Appendix F – Economic evidence study selection



Appendix G – Economic evidence tables

No economic evidence was found for this review question.

Appendix H – Health economic model

No original health economic modelling was done for this review question.

Appendix I – Excluded studies

Studies excluded from the observational search

Study	Reason for exclusion
Akinola, Bola, Collins, Ruaraidh, Sim, Francis C J et al. (2018) Does a fixed offset hemiarthroplasty implant have any effect on pain and function in patients with a femoral neck fracture?. Injury 49(8): 1577-1580	- Not a relevant study design *Does not contain a comparator*
Bidwai, Amit S C and Willett, Keith M (2012) Comparison of the Exeter Trauma Stem and the Thompson hemiarthroplasty for intracapsular hip fractures. Hip international : the journal of clinical and experimental research on hip pathology and therapy 22(6): 655-60	- Did not match or adjust for confounding factors as specified in protocol
Chan, Gareth K, Aladwan, Rahmeh, Hook, Samantha E et al. (2020) Thompson Hemiarthroplasty for Femoral Neck Fracture Is Associated With Increased Risk of Dislocation. The Journal of arthroplasty 35(6): 1606-1613	- Did not match or adjust for confounding factors as specified in protocol
Dawe, E.J.C., Lindisfarne, E.A.O., Nicol, S. et al. (2014) Does using a modular variable offset hemiarthroplasty reduce length of stay after hip fracture? Early experience with the Exeter Unipolar hemiarthroplasty. European Orthopaedics and Traumatology 5(1): 49-55	- Did not match or adjust for confounding factors as specified in protocol
Garellick, Goran, Karrholm, Johan, Lindahl, Hans et al. (2016) Substantially higher prevalence of postoperative periprosthetic fractures in octogenarians with hip fractures operated with a cemented, polished tapered stem rather than an anatomic stem: A prospective cohort study involving 979 hips. Acta orthopaedica 87(6): 653	- Correspondence *Correspondence to authors, not full study*
Hsu, A.Y.-C. (2018) Changes in rehabilitation outcomes by new guidelines of Hong Kong Hospital Authority in implant choice for femoral neck fractures-Austin Moore versus cemented Exeter hemiarthroplasty. Journal of Orthopaedics, Trauma and Rehabilitation 25: 37-48	- Comparator in study does not match that specified in protocol *Uncemented designs excluded from protocol*
Joanroy, Rajzan, Stork-Hansen, Jesper, Rotwitt, Lars et al. (2021) Cemented hemiarthroplasty for femoral neck fracture patients: collarless, polished tapered stem (CPT) versus anatomic matte stem (Lubinus SP2). European journal of	- Did not match or adjust for confounding factors as specified in protocol

Study	Reason for exclusion
orthopaedic surgery & traumatology : orthopedie traumatologie 31(5): 855-860	
Kaltsas, D S and Klugman, D J (1986) Acetabular erosion: a comparison between the Austin Moore and Monk hard top prostheses. Injury 17(4): 230-6	- Did not match or adjust for confounding factors as specified in protocol
Kennedy, John W, Ng, Nigel Y B, Young, David et al. (2021) Cement-in-cement femoral component revision : a comparison of two different taper-slip designs with medium-term follow up. The bone & joint journal 103b(7): 1215-1221	- Study does not contain a relevant intervention
Kwok, D C and Cruess, R L (1982) A retrospective study of Moore and Thompson hemiarthroplasty. A review of 599 surgical cases and an analysis of the technical complications. Clinical orthopaedics and related research: 179-85	- Unable to obtain study
Laflamme, Melissa, Angers, Michele, Vachon, Jessica et al. (2020) High Incidence of Intraoperative Fractures With a Specific Cemented Stem Following Intracapsular Displaced Hip Fracture. The Journal of arthroplasty 35(2): 485-489	- Did not match or adjust for confounding factors as specified in protocol
Lin, X., Yang, K., Tan, H. et al. (2021) Comparison of the Curative Effects of Hip Arthroplasty with Bio-Type Femoral Stem and Cemented Femoral Stem in Elderly Patients with Unstable Osteoporotic Intertrochanteric Femur Fractures. Journal of Medical and Biological Engineering 41(4): 523-533	- Comparator in study does not match that specified in protocol *Uncemented comparator*
Meyer, S (1981) Prosthetic replacement in hip fractures: a comparison between the Moore and Christiansen endoprostheses. Clinical orthopaedics and related research: 57-62	- Comparator in study does not match that specified in protocol *Uncemented comparator*
Mukka, Sebastian, Mellner, Carl, Knutsson, Bjorn et al. (2016) Substantially higher prevalence of postoperative peri-prosthetic fractures in octogenarians with hip fractures operated with a cemented, polished tapered stem rather than an anatomic stem. Acta orthopaedica 87(3): 257-61	- Did not match or adjust for confounding factors as specified in protocol

Study	Reason for exclusion
Parker, Martyn J (2012) Cemented Thompson hemiarthroplasty versus cemented Exeter Trauma Stem (ETS) hemiarthroplasty for intracapsular hip fractures: a randomised trial of 200 patients. Injury 43(6): 807-10	- Duplicate reference *Included in RCT sift*
Pongkunakorn, Anuwat; Thisayukta, Phornphinit; Palawong, Pattanapong (2009) Invention technique and clinical results of Lampang cement injection gun used in hip hemiarthroplasty. Journal of the Medical Association of Thailand = Chotmaihet thangphaet 92suppl6: 232-8	- Did not match or adjust for confounding factors as specified in protocol
Pritchett, J.W. (2008) Curved-stem hip resurfacing: Minimum 20-year followup. Clinical Orthopaedics and Related Research 466(5): 1177-1185	- Study does not contain a relevant intervention
Schweizer, A., Luem, M., Riede, U. et al. (2005) Five-year results of two cemented hip stem models each made of two different alloys. Archives of Orthopaedic and Trauma Surgery 125(2): 80-86	- Study does not contain a relevant intervention *THA not HA*
Sims, A L, Parsons, N, Achten, J et al. (2018) A randomized controlled trial comparing the Thompson hemiarthroplasty with the Exeter polished tapered stem and Unitrax modular head in the treatment of displaced intracapsular fractures of the hip: the WHiTE 3: HEMI Trial. The bone & joint journal 100b(3): 352-360	- Duplicate reference *Study already included in RCT evidence*
Siow, J.W.X. and Kwek, E.B.K. (2021) Mismatch between conventional femoral arthroplasty stems and hip morphology in the elderly chinese hip fracture population. Malaysian Orthopaedic Journal 15(2): 101-106	- Not a relevant study design *Not a direct comparison of femoral component designs*

Appendix J – Research recommendations – full details

J.1.1 Research recommendation

In adults undergoing hemiarthroplasty for displaced intracapsular hip fracture (including in different subgroups), which femoral component design has the best long-term outcomes?

J.1.2 Why this is important

Recommendations for femoral components in a fragility fracture population have previously been based on evidence from an elective surgery population. There is also no evidence for specific subpopulations and if they are at risk of less favourable outcomes from different femoral component types. It will be important to have long-term data on patient reported and adverse event outcomes for different femoral components used in hemiarthroplasty and to understand their relative benefit for a fragility fracture population and subgroup populations within that.

J.1.3 Rationale for research recommendation

Importance to 'patients' or the population	There is not enough long-term evidence on outcomes for different femoral components in a fragility fracture population. This population could benefit in future from more effective treatment by understanding which components have the best long-term outcomes.
Relevance to NICE guidance	Cemented femoral components are recommended for hemiarthroplasties but there are only a small number of studies which have compared the effectiveness of different types of components. These are unable to differentiate between interventions and not enough long-term evidence exists.
Relevance to the NHS	The outcome could affect which femoral component is offered by the NHS to this population in the future and could reduce variation in practice between hospitals and trusts.
National priorities	Moderate
Current evidence base	Minimal short-term data (2 UK-based RCTs, 2 Swedish observational studies)
Equality considerations	It is unknown whether people from different population groups may have less favourable outcomes depending on which femoral component is used .

J.1.4 Modified PICO table

Population	Adults presenting to the health service with a firm or provisional clinical diagnosis of fragility fracture of the hip. Adults with displaced intracapsular hip fracture. Subgroups of people from different populations and ethnic backgrounds.
Intervention	Femoral component designs for use in cemented hemiarthroplasty procedures
Comparator	Femoral component designs for use in cemented hemiarthroplasty procedures
Outcome	• All-cause mortality • Unplanned return to theatre (including number of reoperations or surgical revisions) • Functional status (using any validated measure such as the Barthel Index, mobility component of the EQ5D, Nottingham Extended Activities of Daily Living, WOMAC score, Harris hip score) • Pain (measured by any validated scale) • Health-related quality of life (measured by any validated scale) • Length of stay in an acute trust • Place of residence at 120 days • Periprosthetic fracture • Surgical site infection • Number of adverse events (grouped by those related to the femoral component (e.g. loosening of prosthesis, dislocation, leg length discrepancy, etc.) and those unrelated to the femoral component (e.g. thrombosis, embolism, neurological adverse events)
Study design	Comparative observational studies that adjust or match for a minimum of age and sex and with follow up periods >2 years RCT studies with follow up periods >2 years
Timeframe	Long term
Additional information	People from different population groups may have less favourable outcomes depending on

	which femoral component is used. It will be important for research to provide subgroup analysis in these populations.

Appendix K – Methods

Please see Cochrane systematic review Lewis 2022 for methods used in the RCT analysis.

K.1.1.1 Pairwise meta-analysis

There was only one study for each of the comparisons that came from observational data and so pairwise meta-analyses could not be performed with pooled relative risks. The observational studies reported hazard ratios and so these were presented in forest plots and GRADE tables using Cochrane Review Manager V5.3. Both relative and absolute risks were presented, with absolute risks calculated by applying the relative risk to the risk in the comparator arm of the meta-analysis (calculated as the total number events in the comparator arms of studies in the meta-analysis divided by the total number of participants in the comparator arms of studies in the meta-analysis).

K.1.1.2 Intervention studies (relative effect estimates)

Non-randomised controlled trials and cohort studies were quality assessed using the ROBINS-I tool. Evidence on each outcome for each individual study was classified into one of the following groups:

- Low risk of bias – The true effect size for the study is likely to be close to the estimated effect size.
- Moderate risk of bias – There is a possibility the true effect size for the study is substantially different to the estimated effect size.
- High risk of bias – It is likely the true effect size for the study is substantially different to the estimated effect size.
- Critical risk of bias (ROBINS-I only) - It is very likely the true effect size for the study is substantially different to the estimated effect size.

Each individual study was also classified into one of three groups for directness, based on if there were concerns about the population, intervention, comparator and/or outcomes in the study and how directly these variables could address the specified review question. Studies were rated as follows:

- Direct – No important deviations from the protocol in population, intervention, comparator and/or outcomes.
- Partially indirect – Important deviations from the protocol in one of the following areas: population, intervention, comparator and/or outcomes.
- Indirect – Important deviations from the protocol in at least two of the following areas: population, intervention, comparator and/or outcomes.

K.1.1.3 Minimally important differences (MIDs) and clinical decision thresholds

The Core Outcome Measures in Effectiveness Trials (COMET) database was searched to identify published minimal clinically important difference thresholds relevant to this guideline that might aid the committee in identifying clinical decision thresholds for the purpose of GRADE. Identified MIDs were assessed to ensure they had been developed and validated in a methodologically rigorous way, and were applicable to the populations, interventions and outcomes specified in this guideline. In addition, the Guideline Committee were asked to prospectively specify any outcomes where they felt a consensus clinical decision threshold could be defined from their experience. In particular, any questions looking to evaluate non-

inferiority (that one treatment is not meaningfully worse than another) required a clinical decision threshold to be defined to act as a non-inferiority margin.

Clinical decision thresholds were used to assess imprecision using GRADE and aid interpretation of the size of effects for different outcomes. Clinical decision threshold that were used in the guideline are given in Table and also reported in the relevant evidence reviews.

Table 4: Identified Clinical decision thresholds

Outcome	Clinical decision threshold	Source
Health related quality of life	0.08	Simms 2008
All other dichotomous outcomes	0.8 - 1.25	Default

For continuous outcomes expressed as a mean difference where no other clinical decision threshold was available, a clinical decision threshold of 0.5 of the median standard deviations of the comparison group arms was used (Norman et al. 2003). For continuous outcomes expressed as a standardised mean difference where no other clinical decision threshold was available, a clinical decision threshold of 0.5 standard deviations was used. For SMDs that were back converted to one of the original scales to aid interpretation, rating of imprecision was carried out before back calculation. For relative risks and hazard ratios, where no other clinical decision threshold was available, a default clinical decision threshold for dichotomous outcomes of 0.8 to 1.25 was used. Odds ratios were converted to risk ratios before presentation to the committee to aid interpretation.

K.1.1.4 GRADE for intervention studies analysed using pairwise analysis

GRADE was used to assess the quality of evidence for the outcomes specified in the review protocol. Data from cohort studies (which were quality assessed ROBINS-I) were initially rated as high quality. The quality of the evidence for each outcome was downgraded or not from this initial point, based on the criteria given in Table .

Table 5: Rationale for downgrading quality of evidence for intervention studies

GRADE criteria	Reasons for downgrading quality
Risk of bias	Not serious: If less than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias, the outcome was downgraded two levels. Extremely serious: If greater than 33.3% of the weight in a meta-analysis came from studies at critical risk of bias, the outcome was downgraded three levels
Indirectness	Not serious: If less than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the overall outcome was not downgraded. Serious: If greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies, the outcome was downgraded one level. Very serious: If greater than 33.3% of the weight in a meta-analysis came from indirect studies, the outcome was downgraded two levels.

GRADE criteria	Reasons for downgrading quality
Inconsistency	Concerns about inconsistency of effects across studies, occurring when there is unexplained variability in the treatment effect demonstrated across studies (heterogeneity), after appropriate pre-specified subgroup analyses have been conducted. This was assessed using the I^2 statistic. N/A: Inconsistency was marked as not applicable if data on the outcome was only available from one study. Not serious: If the I^2 was less than 33.3%, the outcome was not downgraded. Serious: If the I^2 was between 33.3% and 66.7%, the outcome was downgraded one level. Very serious: If the I^2 was greater than 66.7%, the outcome was downgraded two levels.
Imprecision	If an MID other than the line of no effect was defined for the outcome, the outcome was downgraded once if the 95% confidence interval for the effect size crossed one line of the MID, and twice if it crosses both lines of the MID. If the line of no effect was defined as an MID for the outcome, it was downgraded once if the 95% confidence interval for the effect size crossed the line of no effect (i.e. the outcome was not statistically significant), and twice if the sample size of the study was sufficiently small that it is not plausible any realistic effect size could have been detected. Outcomes meeting the criteria for downgrading above were not downgraded if the confidence interval was sufficiently narrow that the upper and lower bounds would correspond to clinically equivalent scenarios.
Publication bias	Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias. When a funnel plot showed convincing evidence of publication bias, or the review team became aware of other evidence of publication bias (for example, evidence of unpublished trials where there was evidence that the effect estimate differed in published and unpublished data), the outcome was downgraded once. If no evidence of publication bias was found for any outcomes in a review (as was often the case), this domain was excluded from GRADE profiles to improve readability.

For outcomes that were originally assigned a quality rating of 'low' (when the data was from observational studies that were not appraised using the ROBINS-I checklist), the quality of evidence for each outcome was upgraded if any of the following three conditions were met and the risk of bias for the outcome was rated as 'no serious':

- Data from studies showed an effect size sufficiently large that it could not be explained by confounding alone.
- Data showed a dose-response gradient.
- Data where all plausible residual confounding was likely to increase our confidence in the effect estimate.