National Institute for Health and Care Excellence

Draft

Hip Fracture (short update)

[B] Evidence reviews for total hip replacement vs hemiarthroplasty

NICE guideline CG124

Evidence reviews underpinning recommendation 1.6.3 and a research recommendation in the NICE guideline

October 2022

Draft for Consultation

These evidence reviews were developed by the Guideline Development Team

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN:

Contents

1 Total hip replacement vs hemiarthroplasty

1.1 Review question

In adults undergoing surgery for displaced intracapsular hip fracture what is the clinical and cost effectiveness of total hip replacement compared with hemiarthroplasty?

1.1.1 Introduction

Current NICE guidance recommends offering total hip replacement/total hip arthroplasty (THA) over hemiarthroplasty for people who are able to walk independently with no more than a stick, are not cognitively impaired and are medically fit for anaesthesia and the procedure. A recent NICE exceptional surveillance review indicates that there may be no significant clinically important benefit in THA compared to HA, therefore a full evidence review has been conducted to investigate if the recommendation to offer THA should be reconsidered.

1.1.2 Summary of the protocol

For full details see the review protocol in Appendix A.

Table 1: PICO characteristics of review question

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.
Intervention	• Hemiarthroplasty
Comparison	• Total hip replacement/Total hip arthroplasty
Outcomes	• All-cause mortality – early mortality, 1 year and any time point after • 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) overall and change from baseline at 6 weeks, 4 months (or early as defined by study), 1 year and any timepoint after • Length of stay in an acute trust • Return to original place of residence • 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)
Study design	• RCT

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 outlined in appendix M.

During development of the review question, a Cochrane systematic review (Lewis 2022, Arthroplasties for hip fracture in adults, Cochrane Database of Systematic Reviews 2022, Issue 2. Copyright © 2022 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.) was identified that included RCT comparisons relevant to this review question. An additional RCT search was performed by NICE to identify any RCTs published after the Cochrane review's final search date (6th July 2020). Analysis from the Cochrane review was used and is presented directly where possible. New data from the NICE sift beyond July 2020 was pooled with the Cochrane analysis to see if this altered the interpretation of effect. Where it didn't change the interpretation of the effect, data is presented as a separate analysis.

The protocol for this review included subgroups which were not a part of the Cochrane review. Where studies included information on relevant subgroups, data from the Cochrane review was analysed to see if there was a significant difference in results between these subgroups (P <0.05). Where the test for subgroup differences was significant, data from the Cochrane review was reanalysed to show these subgroups presented as new forest plots in this review.

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 evidence search from July 2020. • Systematic review risk of bias assessment (ROBIS). • RCT risk of bias assessment for studies not in Cochrane review • RCT meta-analysis and summary for studies not in Cochrane review. • RCT evidence tables for studies not in Cochrane review • GRADE assessment. • Subgroup analysis on people who are cognitively impaired.	• RCT evidence search to 6 July 2020. • RCT risk of bias assessments for studies in Cochrane review • RCT evidence tables for studies in Cochrane review. • RCT meta-analysis and summary of results from studies in Cochrane review.

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 17 RCTs that compare THA with HA were identified. A further search for RCTs published after the search dates for the Cochrane review was conducted. After de-duplication 304 references were screened, and 3 further studies met the inclusion criteria for this review. In total, 20 studies were included in the review.

See also the study selection flow chart in Appendix C, study evidence tables in Appendix D, forest plots in appendix F and GRADE tables in Appendix G.

1.1.4.2 Excluded studies

See the excluded studies list in Appendix K.

1.1.5 Summary of studies included in the effectiveness evidence

Table 3: Summary of studies included in the evidence review

Study	Longest Follow-up time	Population	Intervention	Comparator	Outcomes
Randomised controlled trials (from Lewis 2022 Cochrane review)					
Baker 2006	39 months	Displaced fracture, THA mean age: 74.2 HA mean age: 75.83	THA 8 mm femoral head articulating with an all-polyethylene Zimmer cemented acetabular cup	HA Endo Femoral Head (Zimmer); cemented; unipolar	Mortality
Blomfeldt 2007	48 months	Displaced fracture, THA mean age: 80.5, HA mean age: 80.7	THA . Modular Exeter femoral component; 28 mm head; OGEE cemented acetabular component	HA . Bipolar; modular Exeter, 28 mm head	ADL Delirium Function Mortality
Cadossi 2013	36 months	Displaced fracture, THA mean age: 82.3, HA mean age 84.2	THA Uncemented Conus stem and a large-diameter femoral head	HA Uncemented, bipolar	Mortality
Chammout 2019	24 months	Displaced fracture, THA mean age: 85, HA mean age 86	THA Cemented 32 mm cobalt chromium head; cemented highly cross-linked polyethylene acetabular component	HA Cemented, unipolar	ADL Delirium Function HRQoL Mortality Unplanned return to theatre

Study	Longest Follow-up time	Population	Intervention	Comparator	Outcomes
Dorr 1986	48 months	Displaced fracture, THA mean age: 72, HA cemented mean age:69 HA uncemented mean age: 66	THA 28 mm head size was used	HA cemented (n = 37) or uncemented (n = 13), bipolar	Unplanned return to theatre
HEALTH 2019	24 months	Displaced fracture, THA mean age: 79.1. HA mean age: 78.6	THA (surgeon's preference)	HA (surgeon's preference)	Function HRQoL Mobility Mortality Unplanned return to theatre
Iorio 2019	12 months	Displaced fracture, THA mean age 82 HA mean age: 83	THA mobility cup with cementless femoral stem	HA Cementless femoral stem with bipolar head	Mortality Unplanned return to theatre
Keating 2006	24 months	Displaced fracture THA mean age: 75.2 HA mean age: 75.4	THA (method/design not reported)	HA Bipolar, cemented	Delirium Function HRQoL Mortality Unplanned return to theatre
Macaulay 2008	24 months	Displaced fracture, THA mean age 82, HA mean age 77	THA (surgeon's preference)	HA (surgeon's preference)	Function HRQoL Mobility Mortality
Mouzouplos 2008	48 months	Displaced fracture, THA mean age: 73.07, HA mean age: 74.24	THA Plus (dePuy)	HA Merete	ADL Function Mortality Unplanned return to theatre
Parker 2019	12 months	Displaced fracture, THA mean age: 77.1, HA mean age: 77.1	THA CPCS stem (n=29), CPT Zimmer (n=23)	HA Monoblock Exeter Trauma Stem (n=22), CPT bipolar (n=4), CPT modular (n=27)	ADL Delirium Mobility Mortality Unplanned return to theatre
Ravikumar 2000	13 years	Displaced fracture, THA mean age 81.03, HA mean age: 82.06	THA cemented with Howse II	HA Uncemented Austin Moore	Mobility Mortality Unplanned return to theatre

Study	Longest Follow-up time	Population	Intervention	Comparator	Outcomes
Ren 2017	Not reported	Fracture type not reported THA mean age: 69.49 HA mean age: 69.73	THA (surgeon's preference)	HA cemented	Function
Sharma 2016	1 week	Displaced fracture, THA mean age: 78, HA mean age: 73	THA (method/design not reported)	HA (method/design not reported)	Mortality
Sonaje 2017	24 months	Displaced fracture, THA mean age: 66.4, HA mean age: 65.3	THA (method/design not reported)	HA (method/design not reported)	Function
Van De Bekerom 2010	60 months	Displaced fracture, THA mean age: 82.1, HA mean age: 80.3	THA cemented; 32 mm diameter modular head	HA Cemented, bipola	Mortality Unplanned return to theatre
Xu 2017	60 months	Fracture type not reported THA mean age 76.16, HA mean age: 75.45	THA Uncemented prosthesis	HA Bipolar; uncemented	Function Mortality
Randomised controlled trials from NICE search					
Li 2022	12 months	Traumatic femoral neck fracture, THA mean age 73.21 HA mean age: 73.161	THA (method/design not reported)	HA (method/design not reported)	Length of hospital stay Pain (VAS) Harris Hip Score Infection Periprosthetic fracture Pressure Ulcer DVT
Makeen 2021	24 months	Displaced fracture, THA mean age: 70.38, HA mean age: 71.12	THA dual mobility cup	HA Bipolar	Dislocation
Ukaj 2019	3 years	Displaced fracture, THA mean age: 78.11, HA mean age: 77.64	THA Cementless acetabular components: Dual Mobility Cup (HAP Quattro VPS	HA Bipolar cementless acetabular prosthesis UHL (GROUPE LE´PINE)	Harris Hip Score Functional independence Mortality Dislocation

Study	Longest Follow-up time	Population	Intervention	Comparator	Outcomes
			cup; Groupe Lepine, Genay, France)		

See Appendix D for full evidence tables.

1.1.6 Summary of the effectiveness evidence

Table 4 – Evidence from Cochrane Review

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
THA vs HA					
Early ADL (≤ 4 months, using categorical data)	2	225	RR 1.03 (95% CI 0.91, 1.18)	Low	Unable to differentiate
Early ADL (≤ 4 months; using social mobility scale[1] (>0 favours HA)	1	83	MD -0.10 (95% CI -0.46, 0.26)	Low	Unable to differentiate
ADL 12 months, using categorical data	2	217	RR 0.96 (95% CI 0.86, 1.07)	Low	Unable to differentiate
ADL (12 months; using Barthel Index) [2] (>0 favours THA)	1	63	MD -0.68 (95% CI -1.18, -0.17)	Low	Effect favouring HA, but less than the MID
ADL (12 months using social mobility scale)[3] (>0 favours THA)	1	78	MD 0.09 (95% CI -0.35, 0.53)	Moderate	Unable to differentiate
Late ADL (> 24 months; using Barthel Index[2] (>0 favours THA)	1	43	MD 5.70 (95% CI 0.21,11.19)	Very low	Effect favouring THA, but less than the MID
Early functional status ≤ 4 months (>0 favours THA)	3	395	Std MD 0.27 (95% CI 0.07,0.47)	Low	Effect favouring THA, but less than the MID
Functional status (12 months) (>0 favours THA)	8	1273	Std MD 0.29 (95% CI 0.14, 0.44)	Low	Effect favouring THA, but less than the MID
Functional status (HHS – good/excellent)	2	140	RR 1.07 (95% CI 0.98, 1.17)	Very low	Unable to differentiate

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Late functional status (>24 months using OHS or HHS[4] (>0 favours THA)	4	224	Std MD 0.65 (95% CI 0.23, 1.08	Very low	Favours THA
Early HRQoL (≤ 4 months) (>0 favours THA)	2	279	MD 0.03 (95% CI -0.06, 0.12)	Very low	Unable to differentiate
HRQoL (12 months) (>0 favours THA)	4	1158	Std MD 0.19 (95% CI 0.07,0.31)	Moderate	Effect favouring THA, but less than the MID
HRQoL (> 24 months. Using SF-36;[5]) (>0 favours THA)	1	34	5.90 (95% CI -1.99, 13.79)	low	Unable to differentiate
Early mobility (≤ 4 months)[1] (>0 favours HA)	1	83	MD -0.40 (95% CI -0.96, 0.16)	low	Unable to differentiate
Mobility (12 months, using TUG)[1] (>0 favours HA)	2	575	MD -2.74 (95% CI -6.82, 1.35)	Moderate	Unable to differentiate
Mobility (12 months, using 9-point mobility scale)[1] (>0 favours HA)	1	78	MD 0.40 (95% CI -0.32, 1.12)	low	Unable to differentiate
Mobility (12 months; able to ambulate independently)	2	175	RR 0.96 (0.71,1.31)	Very low	Unable to differentiate
Late mobility (> 24 months; able to ambulate independently)	1	32	RR 1.27 (0.71, 2.29)	Very low	Unable to differentiate
Early mortality (≤ 4 months)	6	725	RR 0.77 (95% CI 0.42, 1.42)	Very low	Unable to differentiate
Mortality (12 months)	11	2667	RR 1.00 (95% CI 0.83, 1.22)	Low	Unable to differentiate
Late mortality (> 24 months)	8	931	RR 1.00 (95% CI 0.81, 1.23)	Very low	Unable to differentiate
Unplanned return to theatre (end of follow up)	10	2594	RR 0.63 (95% CI 0.37, 1.07)	Very low	Unable to differentiate
Length of hospital stay (days) (>0 favours HA)	3	306	MD 0.80 (95% CI -1.12, 2.73)	Very low	Unable to differentiate
Pain (12 months)[8] (>0 favours HA	9	1435	Std MD -0.13 (95% CI -0.38, 0.12)	Very Low	Unable to differentiate

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Late pain (>24 months) uncemented THA, mixed cemented/uncemented HA[10] (>0 favours THA)	1	32	MD -3.50 (95% CI -7.19, 0.19)	Very low	Unable to differentiate
Late pain (>24 months) cemented THA and HA[9] (>0 favours THA)	1	83	MD 7.90 (95% CI 5.69, 10.11)	Low	Effect favouring THA, but less than the MID
'Pain (> 24 months)' *(late pain)* – categorical data – No Pain	1	135	RR 1.47 (95% CI 1.07, 2.00)	Very low	Favours THA
Early pain (≤ 4 months)[11](>0 favours THA)	5	572	Std MD 0.10 (95% CI -0.10, 0.30)	Low	Unable to differentiate
Discharge destination (own home)	2	1612	RR 0.97 (95% CI 0.87, 1.08)	Low	Unable to differentiate
Discharge destination (older persons ward)	1	120	RR 0.88 (95% CI 0.34, 2.26)	Very low	Unable to differentiate
Adverse events relating to implant, fracture or both					
Postoperative periprosthetic fracture	3	1557	RR 1.08 (95% CI 0.70, 1.66)	Very low	Unable to differentiate
Prosthetic loosening	4	1889	RR 0.64 (95% CI 0.17, 2.41)	Very low	Unable to differentiate
Deep infection	8	2343	RR 0.87 (95% CI 0.50, 1.54)	Very low	Unable to differentiate
Superficial infection	10	2495	RR 1.25 (95% CI 0.67, 2.30)	Very low	Unable to differentiate
Dislocation	12	2719	RR 1.96 (95% CI 1.17, 3.27)	Very low	Favours HA
Dislocation (non-cognitively impaired population)	11	2659	RR 2.22 (95% CI 1.52, 3.23)	Low	Favours HA
Dislocation (cognitively impaired only population)	1	60	RR 0.09 (95% CI 0.01, 1.57)	Very low	Unable to differentiate
Adverse events unrelated to implant, fracture or both					
Acute Kidney Injury	2	1561	RR 1.09 (95% CI 0.62,1.92)	Very low	Unable to differentiate
Blood transfusion	2	285	RR 2.14 (95% CI 1.27, 3.61)	Low	Favours HA

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Cerebrovascular accident	4	657	RR 1.63 (95% CI 0.63, 4.21)	Very low	Unable to differentiate
Pneumonia/chest infection (reported at > 4 months)	5	613	RR 0.87 (95% CI 0.38, 2.00)	Very low	Unable to differentiate
Myocardial infarction	4	460	RR 1.48 (95% CI 0.48, 4.58)	Very low	Unable to differentiate
Urinary Tract Infection	1	40	RR 0.19 (95% CI 0.01, 3.46)	Very low	Unable to differentiate
Venous thromboembolic phenomena (DVT)	4	486	RR 4.25 (95% CI 0.86, 21.06)	Very low	Unable to differentiate
Venous thromboembolic phenomena (pulmonary embolism)	5	673	RR 0.49 (95% CI 0.14, 1.63)	Very low	Unable to differentiate

1 Lower scores indicate better mobility
2 Higher scores indicate more independence
3 Lower scores indicate more independence
4 Higher scores indicate better function
5 Higher scores indicate better quality of life
6 THA: cement, stem, head (≥28 mm) and cup all at surgeons preference; HA: cement, stem and head all at surgeons preference; at 12 months
7 THA: cemented, Howse II stem, 32 mm head, semicaptive cup; HA: uncemented, Austin-Moore, unipolar; at 12 months
8 Lower scores indicate less pain
9 HHS (higher scores indicate less pain); THA: cemented, Exeter modular, OGEE (DePuy); HA: cemented, Exeter modular, 28mm bipolar; at 48 months
10 HHS (higher scores indicate less pain); THA: uncemented, Conus stem, large diameter head, polycarbonate-urethane cup; HA: mixed cemented and uncemented
11 Higher scores indicate less pain

Table 5 – Evidence from NICE search

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
THA vs HA					
Overall functional status - Harris hip score - 3 months[2]	1	86	MD 1.14 (95% CI -1.43, 3.71)	Very low	Unable to differentiate
Overall functional status Harris hip score - 1 year[2]	1	72	MD 3.83 (95% CI 0.22, 7.44)	Very low	Favours THA
Overall functional status Harris hip score – 3 years[2]	1	63	MD 4.16 (95% CI 0.71, 7.61)	Very low	Favours THA
Functional independence measure	1	94	MD 1.75 (95% CI -0.48, 3.98)	Very low	Unable to differentiate
Mortality - 3 months	1	94	RR 0.71 (95% CI 0.24, 2.09)	Very low	Unable to differentiate

Outcomes	No. studies	Sample size	Effect size (95% CI)	Quality	Interpretation of effect
Mortality – 1 year	1	94	RR 0.58 (95% CI 0.25, 1.35)	Very low	Unable to differentiate
Mortality – 3 years	1	94	RR 0.87 (95% CI 0.46, 1.62)	Very low	Unable to differentiate
Hospital length of stay (days)	1	132	MD 1.60 (95% CI 0.38, 2.82)	Very low	Favours HA
Pain (VAS) – 3 days[3]	1	132	MD 0.19 (95% CI 0.05, 0.33)	Very low	Effect favouring HA, but less than the MID
Pain (VAS) – 7 days[3]	1	132	MD 0.12 (95% CI 0.02, 0.22)	Very low	Effect favouring HA, but less than the MID
Harris Hip Score - 12 months (pain domain)[1]	1	132	MD 4.61 (95% CI 3.86, 5.36)	Very low	Favours THA
Adverse event related to implant or fracture					
Infection	1	132	RR 0.20 (95% CI 0.01, 4.09)	Very low	Unable to differentiate
Periprosthetic Fracture	1	132	RR 0.20 (95% CI 0.01, 4.09)	Very low	Unable to differentiate
Dislocation	2	127	RR 0.34 (95% CI 0.06, 2.08)	Very low	Unable to differentiate
Adverse event unrelated to implant or fracture					
Pressure ulcer	1	132	RR 0.50 (95% CI 0.05, 5.38)	Very low	Unable to differentiate
Deep vein thrombosis	1	132	RR 0.33 (95% CI 0.01, 8.04)	Very low	Unable to differentiate

1 Higher scores indicate less pain
2 Higher scores are better
3 Lower score indicates less pain

1

1.1.7 Economic evidence

1.1.7.1 Included studies

Four health economic studies with the relevant comparison were included in this review.{Axelrod 2020, Blythe 2020, Carroll 2011, Larranaga 2022} These are summarised in the health economic evidence profile below (Table 6) and the health economic evidence tables in Appendix I. A further health economic analysis was included in the review which was developed for the previous update of the guideline.

1.1.7.2 Excluded studies

One economic studies relating to this review question was identified but was excluded due to it not being a relevant study design.{Gao 2020 } These are listed in Appendix K, with reasons for exclusion given.

See also the health economic study selection flow chart in 0H.

1.1.8 Summary of included economic evidence

Table 6: Health economic evidence profile

Study	Applicability	Limitations	Other comments	Incremental			Uncertainty
				Cost[a] (£)	Effects (QALYs)	ICER[a] (£/QALY)	
Axelrod et al. 2020	Partially applicable	Potentially serious limitations[b]		£3,298	0.04	£91,045	**Deterministic:** Changing the discount rate 0% to 3% did not meaningfully change the ICER. Changing the cost of total hip replacement by 30% did not change the ICER significantly. **Probabilistic:** Probability of total hip replacement being cost effective was 12.8% and 32.8% for £30,020 and £60,040 willingness to pay threshold
Blythe et al. 2020	Partially applicable	Potentially serious limitations[c]		<75: £2,765,602 75-85: £3,430,353 >85:£6,952,647	<75:1,350 75-85: 3,193 >85: 4,615	<75: £2,049 75-85: £1,075 >85: £1,507	Scenario analysis assumed that all patients were equally suited to THR and HA. The analysis showed that the more patients receiving hybrid THA over cemented HA, the greater the costs and QALYs, with

Study	Applicability	Limitations	Other comments	Incremental			Uncertainty
				Cost[a] (£)	Effects (QALYs)	ICER[a] (£/QALY)	
							diminishing returns as patients aged.
Carroll et al. 2011	Directly applicable	Potentially serious limitations[d]		2 year time horizon: £4,837 3 year time horizon: £4,837 5 year time horizon: £4,837	2 year time horizon: 0.147 3 year time horizon: 0.285 5 year time horizon: 0.580	2 year time horizon: £32,769 3 year time horizon: £19,579 5 year time horizon: £9,643	Using data reported by Blomfeldt et al 2007 found the cost per QALY was £54,565, £36,999 and £22,932 at 2, 3 and 5 years, respectively
Larranaga et al 2022	Partially applicable	Potentially serious limitations[e]		£2,548	0.81	£3,162	Sensitivity analysis showed that partially hip replacement should be used in most patients with total hip replacement being reserved for younger patients
CG124 Model	Directly applicable	Potentially serious limitations[f]		£304	-0.54	Total hip replacement dominates	Sensitivity analysis showed that total hip replacement dominated in almost all of the scenarios including in the probabilistic sensitivity analysis

(a)Costs were adjusted for purchase price parities and inflated to 2020 British Pounds Sterling using Eppi-Centre Cost Converter. https://eppi.ioe.ac.uk/costconversion/default.aspx
(b) Time horizon is 2 years
(c) Time horizon is 5 years, focus is on cemented vs uncemented
(d) Time horizon is 2, 3 and 5 years, assumed there were no difference in costs after first year, no deterministic sensitivity analyses were reported
(e) Not all parameters were investigated, only age and anaesthesiology risk
(f) Swedish Hip Arthroplasty Register used, not sure if this applies to the UK populations, few expert assumptions and older data, costs uprated from 2000/01

1.1.9 Economic model

We updated the economic model that was developed for the 2017 update of the guideline. The findings from the updated economic model are summarised in the health economic evidence profile below (Table 7**Table**), with a full write up of the methods and results in Appendix J.

Table 7: Health economic evidence profile: Total hip replacement (THR) vs hemiarthroplasty (HA)

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
CG124 UK	Directly applicable	Minor limitations [b]	• Analysis type: Cost-utility analysis • Outcome: QALYs: ICER • Population: Adults presenting to the health service with a firm or provisional clinical diagnosis of fragility fracture of the hipComparators: Total hip replacement, hemiarthroplasty Model Type: Markov • Time horizon: Lifetime	£1,607[c]	0.33 QALYs	£4,819 per QALY gained	Probability total hip replacement cost effective (£20/£30K threshold): 95.6% A number of sensitivity analyses were completed but the parameter that affected the result was the long term utility. If the benefit of THR lasts for a lifetime then THR is the most cost effective option however, if the benefit only lasts for a year then HA is the most cost effective option.

Abbreviations: ICER= incremental cost-effectiveness ratio; QALYs= quality-adjusted life years

1.1.11 Evidence statements

Economic evidence

Four existing health economics studies were found for this review and a model that was built for the previous version of the guideline was updated. The evidence was contradictory with some studies showing total hip replacement to be cost effective whereas others showed hemiarthroplasty was cost effective.

1.1.12 The committee's discussion and interpretation of the evidence

1.1.12.1. The outcomes that matter most

The committee commented that for clinicians, their priority was to ensure someone who has had traumatic fracture could stand up and walk again, therefore functional status (Harris Hip Score), activities of daily living and measures of independence were all considered important outcomes. While pain was considered important, early pain when a patient is in hospital is highly variable and therefore not considered a reliable indicator of a successful procedure. Length of hospital stay was considered important from a patient and resource perspective, but not a reliable outcome for measuring the success of the procedure as other complications such as arranging social care or comorbidities could act as a confounder. Blood transfusion was also considered an important outcome directly related to the procedure and with greater medical and financial consequences than some of the other adverse events. Blood transfusion was considered more important to decision making than some of the other adverse events.

1.1.12.2 The quality of the evidence

The committee noted that most of the evidence was rated very low to low quality due to indirectness (many of the studies used uncemented and bipolar prosthesis which were not the direct population of interest), imprecision (due to a lot of outcomes from single study analysis and wide confidence intervals) and risk of bias (due to lack of information about blinding, allocation concealment, or not true randomisation). Much of the evidence also provided relatively short-term follow up data. The committee also commented that the studies had quite restrictive inclusion criteria, therefore patients included may have been likely to have better outcomes than those in the general population. However, it was noted that the HEALTH trial was more recent than most others in the analysis, more relevant to current UK practice and contributed more weight to the pooled total due to a larger sample size. For most outcomes the evidence was unable to differentiate between the two procedures, and for 5 of the 7 outcomes that favoured THA, the effect estimate did not meet the minimum clinical important difference threshold. This included functional status at 4 and 12 months and health related quality of life at 12 months which were considered important outcomes. Three of the 4 outcomes that favoured HA were above the minimum important difference threshold. The low-quality outcomes and the lack of a clear benefit of one type of arthroplasty over the other meant that the committee could not strongly recommend one of these procedures for all people with displaced intracapsular hip fracture.

The committee specified a number of subgroups in the protocol to represent people who they thought might have different outcomes for THA or HA. These included people with cognitive impairment, people of different age groups and people able to walk outdoors with no more than the use of a stick. Evidence was only found for one of these subgroups (people with cognitive impairment), and this was specific to a Dementia population. The committee felt that more evidence of the effect of THA vs HA in these subgroups, particularly between

different age groups, with more longer-term follow up data would have helped them to make a stronger and more focused recommendation.

1.1.12.3 Benefits and harms

It was noted that the clinical and health economic evidence suggested that in the longer term there may be functional benefits in offering people THA, but in the committee's clinical experience, some older people may not live for long enough to experience these benefits. People who were less mobile may also not be concerned or affected by some of the consequences of HA such as wear on the acetabulum. The committee also agreed that from clinical experience and as noted in the evidence, THA could result in higher dislocation rates and higher loss of blood, which can be directly linked to the risk of higher rates of blood transfusion. For this reason, they were confident that THA would not be the best treatment option for all people with displaced intracapsular fractures. Although there was a small increase in quality of life with THA, the committee agreed that this was not significant enough to warrant offering it as a treatment instead of HA for all people with displaced intracapsular hip fracture.

Overall, the committee thought that the evidence was unable to show a statistically or clinically significant benefit for recommending one treatment over the other. They recognised that THA may be more beneficial in the longer term making it a more cost-effective option, but that this would only be relevant for specific subgroups, specifically younger people who have no comorbidities that would otherwise make them unsuitable for the procedure. With no evidence to inform decisions on which subgroups would benefit the most from THA, the committee decided the recommendation should give clinicians discretion when deciding who will benefit from THA and who will see the same, or more, benefit from having HA. For this reason, they stated that THA should be considered if it is thought that someone will gain long-term benefits to their functional independence from the procedure. To provide clinicians with additional guidance when making these choices, the committee decided to specify that people should be able to walk independently out of doors with only a stick before being considered for THA. Although no evidence was identified for this subgroup, the committee felt that this was a good measure of whether someone was sufficiently mobile to be likely to benefit from THA. They also agreed that people who had cognitive impairment significant enough to increase their risk of dislocation would not be suitable for THA as they would be at an increased risk of falls and higher dislocation rates. The low-quality evidence with limited data on subgroups meant that the committee could not make more detailed recommendations on who should be offered THA or HA. Instead, they decided that it is important that future research provides data on the long-term benefits of each procedure in specific subgroups, such as younger age groups. They therefore made a research recommendation to reflect this.

1.1.12.4 Cost effectiveness and resource use

For this review question, we identified four published cost effectiveness studies and the economic analysis that was developed to support the previous update of this guideline that was published in 2017. Two of the five analyses were from the UK perspective. All studies had serious to very serious limitations, and therefore the committee considered that they were less important in their decision making. The model that was developed for the review question of this guideline in 2017 incorporated intervention costs from the 2000/2001 cost year, collected from trauma units in Scotland. The existing evidence for the cost effectiveness of THR compared with HA was contradictory, and largely depended on two factors: the time horizon that was taken for the analysis and extrapolation of benefits, and the age group in which the procedure was given. Some studies demonstrated that THR was not cost effective in the general hip fracture population but was more likely to be a cost effective

treatment for younger patients. Studies that presented results for shorter time horizons, such as Axelrod et al. (2020) and Carroll et al. (2011) showed that HA is the most cost-effective treatment for a 2-year time horizon. In contrast, Blythe et al. (2020), Larranaga et al. (2022), and the model from the previous version of the guideline all showed that THR is cost effective compared with HA.

Given the limitations with the existing cost effectiveness literature and the recent publication of a new large RCT comparing outcomes for THR compared with HA, we updated the model from the previous update of the guideline with new information from the clinical review and more recent costs. We also reviewed the assumptions around the extrapolation of benefits beyond the period of the trials in which they were measured, and we conducted some additional scenario analyses to test the robustness of these results to different sources of data and assumptions. To update the costs of the procedure, we obtained prosthesis costs from NHS Spend Comparison, and estimated the costs for each procedure associated with operating and with post-procedure recovery. We also updated the way in which the costs of unplanned return to theatre were estimated.

Our model found that THR is cost effective compared with HA, if we assume that the benefits in quality of life observed at 12 months are sustained for the patient lifetime. THR was associated with an additional cost of £1,607 and additional QALYs of 0.33, resulting in an ICER of £4,819. The majority of the additional cost of THR was due to increased prosthesis cost and a longer time spent in hospital after the procedure.

However, the committee felt that the parameters and assumptions were very uncertain and based on weak evidence due to low quality studies and a lack of long-term evidence for a trauma population. One of the assumptions that had the biggest impact on the results was regarding the long term QALY increase for THR relative to HA. As part of a scenario analysis, we assumed that the QALY improvement of THR at 12 months remains consistent for the rest of the model. The difference between the two procedures was very small, but if maintained for a number of years the cumulative impact of this difference means that THR may be cost effective. The committee were unsure of the validity of the assumption but there was no data to show how long the benefit would last. When the QALY benefit of THR was ceased after 12 months the ICER increased to £82,510, which is significantly over the £20,000 per QALY gained threshold. Similarly, when we considered the quality of life benefit of THR would last up to 24 months over a two year time horizon, the ICER was again above NICE's £20,000 per QALY gained threshold.

The committee discussed how long they thought the benefit of THR would have to last for it to be cost effective. Unfortunately, the model was not designed to find out the exact length of time that the THR benefit needed to last. However, a number of scenario analyses were conducted whereby we looked at the benefit of THR lasting two, three, four and five years over a lifetime horizon. It found that THR was cost effective if the benefit lasts somewhere between two and three years. Therefore, the committee made a recommendation that the patients should be given THR only if it is believed that they will gain the benefit of THR relative to HA beyond two years. It is worthy to note that the limited evidence meant that no subgroup analyses could have been conducted, for example age or other baseline characteristic, and therefore we were unable to fully investigate which group would benefit the most from a THR rather than a HA.

The committee compared the results of our study to those in Axelrod et al. (2020), a trial-based cost effectiveness analysis of the HEALTH study conducted over a two-year period. The committee felt that even though the HEALTH study was based in Canada and the costs are different to a UK population, this study is a relevant clinical study and provided similar cost effective conclusions to our model when we allowed the benefit of THR to last up to two years. This supported the committee's opinion that THR is more likely to be cost effective in those who have greater capacity to benefit for longer than two years.

Despite existing recommendations to use THR instead of HA in people who are medically fit, the committee acknowledged that around 25% of people that would qualify for a THR are currently offered it, due to perceptions in the medical community about the benefits of THR relative to HA. With the new recommendation suggesting that THR is considered for those who are likely to have a long-term benefit, it may be that people are more selectively offered the procedure. Therefore, the committee felt that it is unlikely that the number of THRs would increase and it is also unlikely that there would be a resource impact.

1.1.12.5 Other factors the committee took into account

The committee noted that one study from the Cochrane analysis and two from the NICE analysis used dual mobility cups for THA, and while this was not a subgroup of interest, it could produce a different effect from single articulation THA. However, the study included in the Cochrane analysis was also the study that reported on people with cognitive impairment. When the results of this study were analysed separately to the other studies, subgroup differences were only seen for the dislocations outcome. The results of this study are therefore not likely to have a major impact on the recommendations.

When making the research recommendation, the committee noted that there would be future government requirements for hospitals to record long term data on hip fractures in a national registry, and that this data would be useful for future guidance in this area.

1.1.13 Recommendations supported by this evidence review

This evidence review supports recommendation 1.6.3 and the research recommendation on the long-term effectiveness of total hip replacement.

1.1.14 References

(2019) Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. New England Journal of Medicine 381(23): 2199-2208

Baker, RP, Squires, B, Gargan, MF et al. (2006) Total hip arthroplasty and hemiarthroplasty in mobile, independent patients with a displaced intracapsular fracture of the femoral neck. A randomized, controlled trial. The Journal of bone and joint surgery. American volume 88(12): 2583-2589

Blomfeldt, R, Tornkvist, H, Eriksson, K et al. (2007) A randomised controlled trial comparing bipolar hemiarthroplasty with total hip replacement for displaced intracapsular fractures of the femoral neck in elderly patients. The Journal of bone and joint surgery. British volume 89(2): 160-165

Cadossi, M, Chiarello, E, Savarino, L et al. (2013) A comparison of hemiarthroplasty with a novel polycarbonate-urethane acetabular component for displaced intracapsular fractures of the femoral neck: a randomised controlled trial in elderly patients. The bone & joint journal: 609-615

Chammout, G, Kelly-Pettersson, P, Hedbeck, CJ et al. (2019) HOPE-Trial: Hemiarthroplasty Compared with Total Hip Arthroplasty for Displaced Femoral Neck Fractures in Octogenarians: A Randomized Controlled Trial. JB & JS open access 4(2): e0059

Dorr, LD, Glousman, R, Hoy, AL et al. (1986) Treatment of femoral neck fractures with total hip replacement versus cemented and noncemented hemiarthroplasty. The Journal of arthroplasty 1(1): 21-28

Iorio, R, Iannotti, F, Mazza, D et al. (2019) Is dual cup mobility better than hemiarthroplasty in patients with dementia and femoral neck fracture? A randomized controlled trial. SICOT-J 5: 38

Keating, JF, Grant, A, Masson, M et al. (2006) Randomized comparison of reduction and fixation, bipolar hemiarthroplasty, and total hip arthroplasty. Treatment of displaced intracapsular hip fractures in healthy older patients. The Journal of bone and joint surgery. American volume 88(2): 249-260

Li, X., Zhao, L., Chen, R. et al. (2022) Effects of total hip arthroplasty and hemiarthroplasty on hip function in patients with traumatic femoral neck fracture. Archives of Orthopaedic and Trauma Surgery

Macaulay, W, Nellans, KW, Garvin, KL et al. (2008) Prospective randomized clinical trial comparing hemiarthroplasty to total hip arthroplasty in the treatment of displaced femoral neck fractures: winner of the Dorr Award. The Journal of arthroplasty 23(6suppl1): 2-8

Makeen, T.M., Abdelazim Mohamed, H., Mohasseb, A.M. et al. (2021) Functional outcome after dual mobility cups total hip replacement versus bipolar hemiarthroplasty in femoral neck fractures in active elderly patients: A randomized controlled trial. Current Orthopaedic Practice 32(5): 468-473

Mouzopoulos, G, Stamatakos, M, Arabatzi, H et al. (2008) The four-year functional result after a displaced subcapital hip fracture treated with three different surgical options. International orthopaedics 32(3): 367-373

Parker, Martyn J and Cawley, Shirley (2019) Treatment of the displaced intracapsular fracture for the 'fitter' elderly patients: A randomised trial of total hip arthroplasty versus hemiarthroplasty for 105 patients. Injury 50(11): 2009-2013

Ravikumar, KJ and Marsh, G (2000) Internal fixation versus hemiarthroplasty versus total hip arthroplasty for displaced subcapital fractures of femur--13 year results of a prospective randomised study. Injury 31(10): 793-797

Ren CG, GY (2017) Comparison of total hip arthroplasty and hemiarthroplasty in elderly patients with femoral neck fracture. Biomedical Research (India) 2017;28(16):7127-30.

Sharma, V, Awasthi, B, Kumar, K et al. (2016) Outcome Analysis of Hemiarthroplasty vs. Total Hip Replacement in Displaced Femoral Neck Fractures in the Elderly. Journal of clinical and diagnostic research : JCDR 10(5): rc11

Sonaje, JC, Meena, PK, Bansiwal, RC et al. (2018) Comparison of functional outcome of bipolar hip arthroplasty and total hip replacement in displaced femoral neck fractures in elderly in a developing country: a 2-year prospective study. European journal of orthopaedic surgery & traumatology : orthopedie traumatologie 28(3): 493-498

Ukaj S, Zhuri O, Ukaj F et al. (2019) Dual Mobility Acetabular Cup Versus Hemiarthroplasty in Treatment of Displaced Femoral Neck Fractures in Elderly Patients: Comparative Study and Results at Minimum 3-Year Follow-up. Geriatric orthopaedic surgery & rehabilitation 10: 2151459319848610

van den Bekerom, MP, Hilverdink, EF, Sierevelt, IN et al. (2010) A comparison of hemiarthroplasty with total hip replacement for displaced intracapsular fracture of the femoral neck: a randomised controlled multicentre trial in patients aged 70 years and over. The Journal of bone and joint surgery. British volume 92(10): 1422-1428

Xu F, Ke R, Gu Y QW (2017) Bipolar hemiarthroplasty vs. total hip replacement in elderly. International Journal of Clinical and Experimental Medicine 2017;10(5):7911-20. [CORPUS ID: 53049671]

Appendices

Appendix A – Review protocols

Review protocol for total hip replacement vs hemiarthroplasty

ID	Field	Content
0.	PROSPERO registration number	CRD42022347384
1.	Review title	Clinical effectiveness and cost-effectiveness of total hip replacement compared with hemiarthroplasty in adults undergoing surgery for displaced intracapsular hip fracture.
2.	Review question	In adults undergoing surgery for displaced intracapsular hip fracture what is the clinical and cost effectiveness of total hip replacement compared with hemiarthroplasty?
3.	Objective	To establish which is more clinically and cost effective for displaced intracapsular hip fracture: total hip replacement or hemiarthroplasty
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 (searches for health economic evidence from June 2016) • English language • Human studies A Cochrane review is available that includes RCT evidence for this comparison up to July 2020. The Cochrane review is broader than the current review and so only a subset of studies are likely to be included in this review, however, all studies from the Cochrane review will be formally assessed for inclusion. A date limit for RCT searches will be set from July 2020, in order to identify RCTs that were published since the Cochrane review.

		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	Hemiarthroplasty (HA)
8.	Comparator/Reference standard/Confounding factors	Total hip replacement (THA)
9.	Types of study to be included	• RCTs
10.	Other exclusion criteria	• Other study types • RCTs with a crossover study design • Studies on non-isolated fracture
11.	Context	A NICE exceptional surveillance review indicates that there may be no significant clinically important benefit in THA compared to HA, therefore this 'strong' recommendation to offer THA should be updated.

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 – early mortality, 1 year and any time point after • 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) overall and change from baseline at 6 weeks, 4 months (or early as defined by study), 1 year and any timepoint after • Length of stay in an acute trust • Return to original place of residence • 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)	N/A
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	Evidence from the Cochrane systematic review will be presented as it appears in that review. Evidence from the >July 2020 sift will be presented as a separate analysis, unless the effect is considered to alter the results reported in the Cochrane review; in which circumstance they will be added to the Cochrane meta-analysis.

		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	• People with / without cognitive impairment • People able / not able to walk independently out of doors with no more than the use of a stick • Different age groups (as reported) • Level of independence as defined by the study
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	July 2022		
22.	Anticipated completion date	October 2022		
23.	Stage of review at time of this submission	**Review stage**	**Started**	**Completed**
		Preliminary searches	X	X
		Piloting of the study selection process	X	X

		Formal screening of search results against eligibility criteria	X	X
		Data extraction	X	X
		Risk of bias (quality) assessment	X	X
		Data analysis	X	X
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
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	Total hip replacement, total hip arthroplasty, hemiarthroplasty, adults, surgery, displaced, intracapsular
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

1
2

3

4

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 run on 30th June and 5th July 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 letters, editorials, news, and conferences in Embase were applied in adherence to standard NICE practice and the review protocol.

The searches were limited from June 2020 and September 2010 as defined in 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 and classifiers

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.

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.

Clinical searches

Main search – Databases

Database	Date searched	Database Platform	Database segment or version	No. of results downloaded
Cochrane Central Register of Controlled Trials (CENTRAL)	30/06/2022	Wiley	6 of 12 June 2022	97
Cochrane Database of Systematic Reviews (CDSR)	30/06/2022	Wiley	6 of 12 June 2022	6
Embase	30/06/2022	Ovid	1974 to 2022 June 29	228
MEDLINE	30/06/2022	Ovid	1946 to June 29 2022	141
MEDLINE-in-Process	30/06/2022	Ovid	1946 to June 29 2022	0
MEDLINE Epub Ahead-of-Print	30/06/2022	Ovid	June 29 2022	12

Search strategy history

Database name: Medline

1 exp Hip Fractures/ (27399)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (11194)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (41716)
4 or/1-3 (47341)
5 exp Arthroplasty, Replacement,/ (66558)
6 exp Joint Prosthesis/ (47647)
7 (arthroplast* or replace* or implant* or prosthe*).tw. (851829)
8 Hemiarthroplasty/ (1305)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (655965)
10 or/5-9 (1481752)
11 4 and 10 (12030)
12 randomized controlled trial.pt. (571534)
13 randomi?ed.mp. (921657)
14 placebo.mp. (217501)
15 or/12-14 (977557)
16 11 and 15 (916)
17 limit 16 to ed=20200601-20220630 (142)
18 animals/ not humans/ (4988972)
19 17 not 18 (142)
20 limit 19 to english language (141)

Database name: Medline in Process

1	exp Hip Fractures/	0	Advanced
	2	((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw.	0
	3	((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw.	5
	4	or/1-3	5
	5	exp Arthroplasty, Replacement,/	0
	6	exp Joint Prosthesis/	0
	7	(arthroplast* or replace* or implant* or prosthe*).tw.	88
	8	Hemiarthroplasty/	0
	9	(hemiarthroplast* or hemi-arthroplas* or partial*).tw.	66
	10	or/5-9	151
	11	4 and 10	2
	12	randomized controlled trial.pt.	0
	13	randomi?ed.mp.	99
	14	placebo.mp.	22
	15	or/12-14	106
	16	11 and 15	0
	17	limit 16 to dt=20200601-20220630	0
	18	animals/ not humans/	0
	19	17 not 18	0
	20	limit 19 to english language	0

Database name: Medline e pub ahead of print

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (139)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (644)
4 or/1-3 (646)
5 exp Arthroplasty, Replacement,/ (0)
6 exp Joint Prosthesis/ (0)

7 (arthroplast* or replace* or implant* or prosthe*).tw. (12347)
8 Hemiarthroplasty/ (0)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (7568)
10 or/5-9 (19391)
11 4 and 10 (180)
12 randomized controlled trial.pt. (1)
13 randomi?ed.mp. (13088)
14 placebo.mp. (2673)
15 or/12-14 (13947)
16 11 and 15 (12)
17 limit 16 to english language/ (12)

Database name: Embase

1 exp hip fracture/ (45229)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (15216)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (61740)
4 or/1-3 (74806)
5 exp replacement arthroplasty/ (38065)
6 exp joint prosthesis/ (73361)
7 (arthroplast* or replace* or implant* or prosthe*).tw. (1263185)
8 exp hemiarthroplasty/ (3143)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (933822)
10 or/5-9 (2159737)
11 4 and 10 (18822)
12 random:.tw. (1804479)
13 placebo:.mp. (496725)
14 double-blind:.tw. (231272)
15 or/12-14 (2072979)
16 11 and 15 (1609)
17 limit 16 to (books or chapter or conference abstract or conference paper or "conference review" or editorial or letter) (250)
18 16 not 17 (1359)
19 nonhuman/ not (human/ and nonhuman/) (5012277)
20 18 not 19 (1322)
21 limit 20 to dc=20200601-20220630 (242)
22 limit 21 to english language (228)

Database name: Cochrane

#1 MeSH descriptor: [Hip Fractures] explode all trees 1836
#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 (Word variations have been searched) 2129
#3 #1 or #2 3363
#4 MeSH descriptor: [Arthroplasty, Replacement] explode all trees 4973

#5 MeSH descriptor: [Joint Prosthesis] explode all trees 1997
#6 (arthroplast* or replace* or implant* or prosthe*):ti,ab,kw (Word variations have been searched) 86152
#7 MeSH descriptor: [Hemiarthroplasty] explode all trees 68
#8 (hemiarthroplast* or hemi-arthroplas* or partial*):ti,ab,kw (Word variations have been searched) 44409
#9 {OR #4-#8} 126841
#10 #3 and #9 1024
#11 conference:pt 199022
#12 #10 not #11 989
#13 (clinicaltrials or trialsearch):so 401307
#14 #12 not #13 with Cochrane Library publication date Between Jun 2020 and Jul 2022 103

Cost-effectiveness searches

Main search – Databases

Database	Date searched	Database Platform	Database segment or version	No. of results downloaded
EconLit	05/07/2022	OVID	1886 to June 23 2022	2
EED	05/07/2022	CRD		0
Embase	05/07/2022	Ovid	1974 to 2022 July 01	1730
HTA	05/07/2022	CRD		2
INAHTA	05/07/2022	INAHTA		14
MEDLINE	05/07/2022	Ovid	1946 to July 01 2022	715
MEDLINE-in-Process	05/07/2022	Ovid	1946 to July 01 2022	1
MEDLINE Epub Ahead-of-Print	05/07/2022	Ovid	July 01 2022	23

Search strategy history

Database name: Medline

1 exp Hip Fractures/ (27408)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (11197)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (41729)
4 or/1-3 (47354)
5 exp Arthroplasty, Replacement,/ (66585)
6 exp Joint Prosthesis/ (47658)
7 (arthroplast* or replace* or implant* or prosthe*).tw. (852149)
8 Hemiarthroplasty/ (1305)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (656159)
10 or/5-9 (1482262)
11 4 and 10 (12035)
12 Economics/ (27456)
13 exp "Costs and Cost Analysis"/ (258850)
14 Economics, Dental/ (1920)
15 exp Economics, Hospital/ (25592)
16 exp Economics, Medical/ (14343)
17 Economics, Nursing/ (4013)

18 Economics, Pharmaceutical/ (3070)
19 Budgets/ (11621)
20 exp Models, Economic/ (16124)
21 Markov Chains/ (15735)
22 Monte Carlo Method/ (31388)
23 Decision Trees/ (11979)
24 econom$.tw. (294025)
25 cba.tw. (10327)
26 cea.tw. (22826)
27 cua.tw. (1099)
28 markov$.tw. (21588)
29 (monte adj carlo).tw. (34503)
30 (decision adj3 (tree$ or analys$)).tw. (18511)
31 (cost or costs or costing$ or costly or costed).tw. (548814)
32 (price$ or pricing$).tw. (39636)
33 budget$.tw. (27114)
34 expenditure$.tw. (57175)
35 (value adj3 (money or monetary)).tw. (2551)
36 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (3800)
37 or/12-36 (1085329)
38 "Quality of Life"/ (245091)
39 quality of life.tw. (286084)
40 "Value of Life"/ (5792)
41 Quality-Adjusted Life Years/ (14915)
42 quality adjusted life.tw. (13842)
43 (qaly$ or qald$ or qale$ or qtime$).tw. (11345)
44 disability adjusted life.tw. (3825)
45 daly$.tw. (3389)
46 Health Status Indicators/ (24063)
47 (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. (25673)
48 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (1516)
49 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (6096)
50 (sf16 or sf 16 or short form 16 or shortform 16 or sf sixteen or sfsixteen or shortform sixteen or short form sixteen).tw. (33)
51 (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw. (408)
52 (euroqol or euro qol or eq5d or eq 5d).tw. (12414)
53 (qol or hql or hqol or hrqol).tw. (56005)
54 (hye or hyes).tw. (63)
55 health$ year$ equivalent$.tw. (38)
56 utilit$.tw. (207030)
57 (hui or hui1 or hui2 or hui3).tw. (1526)
58 disutili$.tw. (482)
59 rosser.tw. (100)
60 quality of wellbeing.tw. (26)
61 quality of well-being.tw. (422)
62 qwb.tw. (199)
63 willingness to pay.tw. (6224)

64 standard gamble$.tw. (829)
65 time trade off.tw. (1164)
66 time tradeoff.tw. (249)
67 tto.tw. (1074)
68 or/38-67 (593541)
69 37 or 68 (1595696)
70 11 and 69 (1188)
71 limit 70 to ed=20100901-20220705 (762)
72 limit 71 to english language (715)

Database name: Medline in Process

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (2)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (10)
4 or/1-3 (10)
5 exp Arthroplasty, Replacement,/ (0)
6 exp Joint Prosthesis/ (0)
7 (arthroplast* or replace* or implant* or prosthe*).tw. (149)
8 Hemiarthroplasty/ (0)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (85)
10 or/5-9 (230)
11 4 and 10 (3)
12 Economics/ (0)
13 exp "Costs and Cost Analysis"/ (0)
14 Economics, Dental/ (0)
15 exp Economics, Hospital/ (0)
16 exp Economics, Medical/ (0)
17 Economics, Nursing/ (0)
18 Economics, Pharmaceutical/ (0)
19 Budgets/ (0)
20 exp Models, Economic/ (0)
21 Markov Chains/ (0)
22 Monte Carlo Method/ (0)
23 Decision Trees/ (0)
24 econom$.tw. (92)
25 cba.tw. (1)
26 cea.tw. (5)
27 cua.tw. (0)
28 markov$.tw. (2)
29 (monte adj carlo).tw. (6)
30 (decision adj3 (tree$ or analys$)).tw. (11)
31 (cost or costs or costing$ or costly or costed).tw. (138)
32 (price$ or pricing$).tw. (7)
33 budget$.tw. (4)
34 expenditure$.tw. (13)
35 (value adj3 (money or monetary)).tw. (1)
36 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (0)

37 or/12-36 (246)
38 "Quality of Life"/ (0)
39 quality of life.tw. (86)
40 "Value of Life"/ (0)
41 Quality-Adjusted Life Years/ (0)
42 quality adjusted life.tw. (6)
43 (qaly$ or qald$ or qale$ or qtime$).tw. (8)
44 disability adjusted life.tw. (1)
45 daly$.tw. (1)
46 Health Status Indicators/ (0)
47 (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. (4)
48 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (1)
49 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (2)
50 (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)
51 (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)
52 (euroqol or euro qol or eq5d or eq 5d).tw. (7)
53 (qol or hql or hqol or hrqol).tw. (15)
54 (hye or hyes).tw. (0)
55 health$ year$ equivalent$.tw. (0)
56 utilit$.tw. (66)
57 (hui or hui1 or hui2 or hui3).tw. (1)
58 disutili$.tw. (0)
59 rosser.tw. (0)
60 quality of wellbeing.tw. (0)
61 quality of well-being.tw. (0)
62 qwb.tw. (0)
63 willingness to pay.tw. (4)
64 standard gamble$.tw. (0)
65 time trade off.tw. (0)
66 time tradeoff.tw. (0)
67 tto.tw. (0)
68 or/38-67 (154)
69 37 or 68 (372)
70 11 and 69 (1)
71 limit 70 to dt=20100901-20220705 (1)
72 limit 71 to english language (1)

Database name: Medline e pub ahead of print

1 exp Hip Fractures/ (0)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (143)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (649)
4 or/1-3 (651)

5 exp Arthroplasty, Replacement,/ (0)
6 exp Joint Prosthesis/ (0)
7 (arthroplast* or replace* or implant* or prosthe*).tw. (12349)
8 Hemiarthroplasty/ (0)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (7553)
10 or/5-9 (19372)
11 4 and 10 (184)
12 Economics/ (0)
13 exp "Costs and Cost Analysis"/ (0)
14 Economics, Dental/ (0)
15 exp Economics, Hospital/ (0)
16 exp Economics, Medical/ (0)
17 Economics, Nursing/ (0)
18 Economics, Pharmaceutical/ (0)
19 Budgets/ (0)
20 exp Models, Economic/ (0)
21 Markov Chains/ (0)
22 Monte Carlo Method/ (0)
23 Decision Trees/ (0)
24 econom$.tw. (7538)
25 cba.tw. (49)
26 cea.tw. (231)
27 cua.tw. (19)
28 markov$.tw. (602)
29 (monte adj carlo).tw. (840)
30 (decision adj3 (tree$ or analys$)).tw. (629)
31 (cost or costs or costing$ or costly or costed).tw. (12688)
32 (price$ or pricing$).tw. (1039)
33 budget$.tw. (549)
34 expenditure$.tw. (1043)
35 (value adj3 (money or monetary)).tw. (70)
36 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (45)
37 or/12-36 (21718)
38 "Quality of Life"/ (0)
39 quality of life.tw. (7715)
40 "Value of Life"/ (0)
41 Quality-Adjusted Life Years/ (0)
42 quality adjusted life.tw. (417)
43 (qaly$ or qald$ or qale$ or qtime$).tw. (338)
44 disability adjusted life.tw. (108)
45 daly$.tw. (99)
46 Health Status Indicators/ (0)
47 (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. (405)
48 (sf6 or sf 6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw. (47)
49 (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw. (163)
50 (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)

51 (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)
52 (euroqol or euro qol or eq5d or eq 5d).tw. (430)
53 (qol or hql or hqol or hrqol).tw. (1528)
54 (hye or hyes).tw. (1)
55 health$ year$ equivalent$.tw. (0)
56 utilit$.tw. (4354)
57 (hui or hui1 or hui2 or hui3).tw. (24)
58 disutili$.tw. (15)
59 rosser.tw. (0)
60 quality of wellbeing.tw. (2)
61 quality of well-being.tw. (8)
62 qwb.tw. (1)
63 willingness to pay.tw. (226)
64 standard gamble$.tw. (6)
65 time trade off.tw. (29)
66 time tradeoff.tw. (0)
67 tto.tw. (25)
68 or/38-67 (12424)
69 37 or 68 (32308)
70 11 and 69 (23)
71 limit 70 to english language (23)

Database name: Embase

1 exp hip fracture/ (45252)
2 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (15219)
3 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or intracapsular$) adj4 fracture$).tw. (61756)
4 or/1-3 (74841)
5 exp replacement arthroplasty/ (38130)
6 exp joint prosthesis/ (73410)
7 (arthroplast* or replace* or implant* or prosthe*).tw. (1263621)
8 exp hemiarthroplasty/ (3144)
9 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (934105)
10 or/5-9 (2160469)
11 4 and 10 (18841)
12 exp Health Economics/ (963353)
13 exp "Health Care Cost"/ (320680)
14 exp Pharmacoeconomics/ (219583)
15 Monte Carlo Method/ (46540)
16 Decision Tree/ (17698)
17 econom$.tw. (444813)
18 cba.tw. (13619)
19 cea.tw. (38742)
20 cua.tw. (1716)
21 markov$.tw. (36018)
22 (monte adj carlo).tw. (56001)
23 (decision adj3 (tree$ or analys$)).tw. (31445)
24 (cost or costs or costing$ or costly or costed).tw. (905090)
25 (price$ or pricing$).tw. (66623)

26 budget$.tw. (43836)
27 expenditure$.tw. (84544)
28 (value adj3 (money or monetary)).tw. (3952)
29 (pharmacoeconomic$ or (pharmaco adj economic$)).tw. (9277)
30 or/12-29 (2058272)
31 "Quality of Life"/ (560188)
32 Quality Adjusted Life Year/ (31765)
33 Quality of Life Index/ (3022)
34 Short Form 36/ (35161)
35 Health Status/ (142134)
36 quality of life.tw. (528833)
37 quality adjusted life.tw. (23764)
38 (qaly$ or qald$ or qale$ or qtime$).tw. (24105)
39 disability adjusted life.tw. (5353)
40 daly$.tw. (5151)
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. (46652)
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. (2753)
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. (11151)
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. (66)
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. (494)
46 (euroqol or euro qol or eq5d or eq 5d).tw. (26291)
47 (qol or hql or hqol or hrqol).tw. (117355)
48 (hye or hyes).tw. (151)
49 health$ year$ equivalent$.tw. (41)
50 utilit$.tw. (341637)
51 (hui or hui1 or hui2 or hui3).tw. (2781)
52 disutili$.tw. (1102)
53 rosser.tw. (135)
54 quality of wellbeing.tw. (62)
55 quality of well-being.tw. (542)
56 qwb.tw. (263)
57 willingness to pay.tw. (11248)
58 standard gamble$.tw. (1157)
59 time trade off.tw. (1910)
60 time tradeoff.tw. (308)
61 tto.tw. (1985)
62 or/31-61 (1171386)
63 30 or 62 (3042292)
64 11 and 63 (2450)
65 limit 64 to dc=20100901-20220705 (1807)
66 limit 65 to english language (1730)

Database name: Econlit

1 1 ((femur$ or femoral$) adj3 (head or neck or proximal) adj4 fracture$).tw. (1)
2 2 ((hip$ or femur$ or femoral$ or trochant$ or pertrochant$ or intertrochant$ or subtrochant$ or
3 intracapsular$) adj4 fracture$).tw. (49)
4 3 or/1-2 (49)
5 4 (arthroplast* or replace* or implant* or prosthe*).tw. (11307)
6 5 (hemiarthroplast* or hemi-arthroplas* or partial*).tw. (22999)
7 6 or/4-5 (33978)
8 7 3 and 6 (4)
9 8 limit 7 to yr="2010 -Current" (2)

10

11 **Database name: EED**
12

1	MeSH DESCRIPTOR Hip Fractures EXPLODE ALL TREES	252	Delete	
	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	Delete
	3	#1 OR #2	252	Delete
	4	MeSH DESCRIPTOR Arthroplasty, Replacement EXPLODE 1 2 3	79	Delete
	5	MeSH DESCRIPTOR Joint Prosthesis EXPLODE ALL TREES	261	Delete
	6	((arthroplast* or replace* or implant* or prosthe*))	5644	Delete
	7	MeSH DESCRIPTOR Hemiarthroplasty EXPLODE ALL TREES	13	Delete
	8	(hemiarthroplast* or hemi-arthroplas* or partial*)	2302	Delete
	9	#4 OR #5 OR #6 OR #7 OR #8	7598	Delete
	10	#3 AND #9	97	Delete
	11	* FROM 2010 TO 2022	43987	Delete
	12	#10 AND #11	44	Delete
	13	(*) and ((Economic evaluation:ZDT and Abstract:ZPS))	9541	Delete
	14	#12 AND #13	0	Delete

13

14 **Database name: HTA**

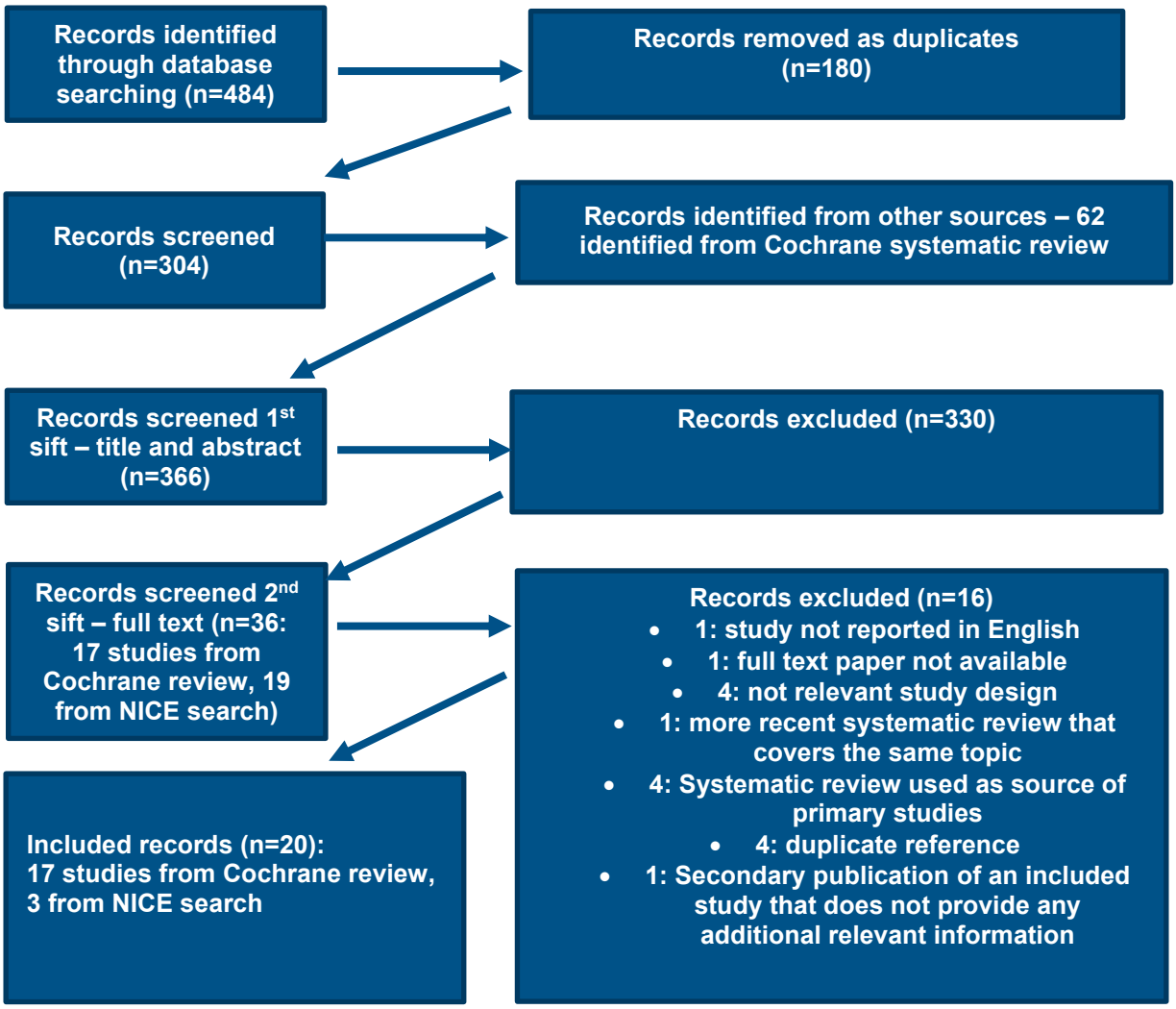
15

Line	Search	Hits	
1	MeSH DESCRIPTOR Hip Fractures EXPLODE ALL TREES	252	Delete
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	Delete
3	#1 OR #2	252	Delete
4	MeSH DESCRIPTOR Arthroplasty, Replacement EXPLODE 1 2 3	79	Delete
5	MeSH DESCRIPTOR Joint Prosthesis EXPLODE ALL TREES	261	Delete
6	((arthroplast* or replace* or implant* or prosthe*))	5644	Delete
7	MeSH DESCRIPTOR Hemiarthroplasty EXPLODE ALL TREES	13	Delete
8	(hemiarthroplast* or hemi-arthroplas* or partial*)	2302	Delete
9	#4 OR #5 OR #6 OR #7 OR #8	7598	Delete
10	#3 AND #9	97	Delete
11	* FROM 2010 TO 2022	43987	Delete
12	#10 AND #11	44	Delete
13	(*) and (Full publication record:ZDT)	15974	Delete
14	#12 AND #13	2	Delete

Database name: INAHTA

(* FROM 2010 TO 2022) AND ((((hemiarthroplast* or hemi-arthroplas* or partial*)[abs]) OR ((hemiarthroplast* or hemi-arthroplas* or partial*)[title]) OR ("Hemiarthroplasty"[mh]) OR ((arthroplast* or replace* or implant* or prosthe*)[abs]) OR ((arthroplast* or replace* or implant* or prosthe*)[title]) OR ("Arthroplasty, Replacement"[mhe])) AND (((hip* and fracture*)[abs]) OR ((hip* and fracture*)[title]) OR ((femur* or femoral* and fracture*)[abs]) OR ((femur* or femoral* and fracture*)[title]) OR ("Hip Fractures"[mhe])))

Appendix C – Effectiveness evidence study selection



Appendix D – Effectiveness evidence

Evidence table and risk of bias assessment for 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 Arthroplasties 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 (62 references)
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

1

2

Critical appraisal - GDT Crit App - ROBIS checklist

Section	Question	Answer
Study eligibility criteria	Concerns regarding specification of study eligibility criteria	Low *(Eligibility criteria reasonable for review question and protocol registered a priori)*
Identification and selection of studies	Concerns regarding methods used to identify and/or select studies	Low Appropriate use of sources/databases and restrictions
Data collection and study appraisal	Concerns regarding methods used to collect data and appraise studies	Low -no concerns
Synthesis and findings	Concerns regarding the synthesis and findings	Low -no concerns
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 not relevant to this review.)*

Evidence table and risk of bias assessment for RCTs identified in NICE search

Li, 2022

Bibliographic Reference	Li, X.; Zhao, L.; Chen, R.; Cao, H.; Wei, Y.; Wu, X.; Zhu, G.; Jiang, L.; Effects of total hip arthroplasty and hemiarthroplasty on hip function in patients with traumatic femoral neck fracture; Archives of Orthopaedic and Trauma Surgery; 2022

1 **Study details**

Trial registration number and/or trial name	Effects of total hip arthroplasty and hemiarthroplasty on hip function in patients with traumatic femoral neck fracture
Study type	Randomised controlled trial (RCT)
Study location	China
Study setting	Department of Trauma Orthopedics, Renmin Hospital, Hubei University of Medicine
Study dates	January 2019 to January 2021
Sources of funding	Unclear - likely to be funded from the University
Inclusion criteria	Inclusion Patient diagnosed with an initial unilateral traumatic femur fracture on X-ray and CT; patient had normal hip development and was a first-time hip arthroplasty; in addition, the patient had no contraindications to surgery and had good compliance
Exclusion criteria	Exclusion Patients who have a combination of multiple injuries; patients with severe cardiac, cerebral, hepatic and renal dysfunction or coagulation abnormalities; patients undergoing revision hip arthroplasty for failed internal fixation of femoral neck fractures; patients with femoral neck fractures who choose conservative treatment or those with poor compliance.
Intervention(s)	HA
Comparator	THA
Outcome measures	Operative time Blood loss Drainage volume Hospital stay

	Pain Range of joint motion Joint function Deformity Delayed union Infection Pressure ulcer Periprosthetic fracture Disarticulation DVT
Number of participants	n=132
Duration of follow-up	12 months
Loss to follow-up	No information
Methods of analysis	Unblinded randomised parallel trial. SPSS 21.0 software was used to analyze the data, mean ± SD was used to represent the measurement data of operation time, intraoperative blood loss, postoperative drainage volume and hospital stay, and t test was used. The counting data were expressed by rate (%) and chi-square $\chi 2$ test was used. $P < 0.05$ was considered statistically signifcant

1

Study arms

Total Hip Arthroplasty (THA) (N = 66)

Hemiarthroplasty (HA) (N = 66)

Characteristics

Arm-level characteristics

Characteristic	THA (N = 66)	HA (N = 66)
Age (years) Mean (SD)	73.21 (10.23)	73.16 (10.16)
Garden type (type of fracture)	n=35 : 3, n=31: 4	n=36: 3, n=30: 4
Osteoporosis	29	28
Diabetes mellitus	8	9
Hypertension	11	10

1 **Critical appraisal - GDT Crit App - Cochrane Risk of Bias tool (RoB 2.0) Normal RCT**

Section	Question	Answer
Domain 1: Bias arising from the randomisation process	Risk of bias judgement for the randomisation process	High *(Unclear if patients chose a different treatment following randomisation having been given the choice.)*
Domain 2a: Risk of bias due to deviations from the intended interventions (effect of assignment to intervention)	Risk of bias for deviations from the intended interventions (effect of assignment to intervention)	High *(Participants were given information about each procedure's predicted outcomes and asked to make a final choice. No information about whether deviations arose from the experimental context but likely that there could have been given the predicted risks associated with each treatment were explained to the patient and a choice then provided.)*
Domain 3. Bias due to missing outcome data	Risk-of-bias judgement for missing outcome data	Some concerns *(No information on loss to follow up or missing outcome data)*
Domain 4. Bias in measurement of the outcome	Risk-of-bias judgement for measurement of the outcome	High *(Outcome assessment could have been influenced by knowledge of intervention received for observational outcomes including HHS. (Author's state that 'patients are educated about the characteristics of THA as well as HA before choosing a surgical procedure. THA usually has greater walking distance, less residual pain, and less risk of reoperation after surgery. HA has a lower risk of infection due to the less invasive nature and shorter operative time, in addition to a lower risk of postoperative dislocation', so outcome assessors have a preconceived idea of likely outcomes for each arm.)*
Domain 5. Bias in selection of the reported result	Risk-of-bias judgement for selection of the reported result	Low
Overall bias and Directness	Risk of bias judgement	High *(Effect of randomisation unclear given participants were also given information on each treatment and then given a choice ; no information on whether or not participants switched or if ITT was carried out.)*

Section	Question	Answer
Overall bias and Directness	Overall Directness	Partially indirect – no information on type of implant – bipolar / uncemented – both exclusions from protocol

Makeen, 2021

Bibliographic Reference	Makeen, T.M.; Abdelazim Mohamed, H.; Mohasseb, A.M.; Elshabrawy, W.E.S.A.E.; Ashoub, M.M.; El Ganzoury, I.M.; Functional outcome after dual mobility cups total hip replacement versus bipolar hemiarthroplasty in femoral neck fractures in active elderly patients: A randomized controlled trial; Current Orthopaedic Practice; 2021; vol. 32 (no. 5); 468-473

Study details

Trial registration number and/or trial name	Functional outcome after dual mobility cups total hip replacement versus bipolar hemiarthroplasty in femoral neck fractures in active elderly patients: a randomized controlled trial FWA 000017585
Study type	Randomised controlled trial (RCT)
Study location	Egypt
Study setting	Ain Shams University Hospitals
Study dates	Beginning January 2018 - no further information
Sources of funding	No information
Inclusion criteria	60-80 years Displaced FNF
Exclusion criteria	Exclusion

	Patients with grade 3 hip osteoarthritis (Tönnis classification),7 acetabular dysplasia, or with previous ipsilateral hip surgeries
Intervention(s)	HA
Comparator	THA
Outcome measures	Harris Hip Score Mortality Dislocation Operative time Blood loss Pain
Number of participants	33
Duration of follow-up	2 years
Loss to follow-up	1 in each arm
Methods of analysis	Single blinded RCT Sample size was calculated using Software for Statistics and Data Science (STATA [StataCorp LLC, College Station, Texas]), setting alpha error at 5% and power at 80% Results from Rashed et al. 15 showed that the postoperative HHS at FIGURE 1. Consort flow diagram. Current Orthopaedic Practice www.c-orthopaedicpractice.com | 469 Copyright r 2021 Wolters Kluwer Health, Inc. All rights reserved. Supplied by the British Library 13 Jul 2022, 11:15 (BST) one year was 92.8 ± 11.1, and the preoperative HHS was 72.8± 22.1 for the traditional treatment group. Based on this information, the sample size that was needed was 30 patients (15 in each group). Statistical Package for Social Science (SPSS 15.0.1 for windows; SPSS Inc., Chicago, Illinois) was used. Data were presented as mean and standard deviation (±SD) for quantitative

	parametric data, median and interquartile range for quantitative nonparametric data. Frequency and percentage were used for presenting qualitative data. P less than 0.05 was considered significant.

1 **Study arms**

2 **HA (bipolar hemiarthroplasty) (N = 17)**

3

4 **Dual mobility THA (N = 16)**

5

6 **Characteristics**

7 **Arm-level characteristics**

Characteristic	**HA (bipolar hemiarthroplasty) (N = 17)**	**DM THA (N = 16)**
Age Mean (SD)	71.12 (6.28)	70.38 (5.73)
Female (%) Custom value	41.2%	56.3%

8

9

1 **Critical appraisal - GDT Crit App - Cochrane Risk of Bias tool (RoB 2.0) Normal RCT**

Section	Question	Answer
Domain 1: Bias arising from the randomisation process	Risk of bias judgement for the randomisation process	Some concerns No information about allocation concealment
Domain 2a: Risk of bias due to deviations from the intended interventions (effect of assignment to intervention)	Risk of bias for deviations from the intended interventions (effect of assignment to intervention)	Some concerns No information about blinding of participants or deviations from interventions.
Domain 3. Bias due to missing outcome data	Risk-of-bias judgement for missing outcome data?	Low Unlikely there was bias due to missing outcome data – 1 lost to follow up in each arm
Domain 4. Bias in measurement of the outcome	Risk-of-bias judgement for measurement of the outcome	Low Although outcome assessors aware of intervention this wouldn't have affected objective outcomes used in this review
Domain 5. Bias in selection of the reported result	Risk-of-bias judgement for selection of the reported result	Some concerns No information on trial protocol
Overall bias and Directness	Risk of bias judgement	Moderate *(No information about allocation concealment, blinding of participants or deviations from treatments.*
Overall bias and Directness	Overall Directness	Partially indirect - bipolar implant used which is an exclusion from protocol

2

1 **Ukaj, 2019**

Bibliographic Reference	Ukaj S; Zhuri O; Ukaj F; Podvorica V; Grezda K; Caton J; Prudhon JL; Krasniqi S; Dual Mobility Acetabular Cup Versus Hemiarthroplasty in Treatment of Displaced Femoral Neck Fractures in Elderly Patients: Comparative Study and Results at Minimum 3-Year Follow-up.; Geriatric orthopaedic surgery & rehabilitation; 2019; vol. 10

2

3 **Study details**

Trial registration number and/or trial name	Dual Mobility Acetabular Cup Versus Hemiarthroplasty in Treatment of Displaced Femoral Neck Fractures in Elderly Patients: Comparative Study and Results at Minimum 3-Year Follow-up
Study type	Quasi- randomised controlled trial 'In order to eliminate bias in patient selection for surgical procedures, treatment decisions were made in a random manner where even-numbered patients underwent DM and odd numbered patients underwent HA.'
Study location	University Clinical Centre of Kosovo
Study setting	a tertiary health-care institution
Study dates	January 2008 to January 2014.
Sources of funding	Likely to have come from the University. 'The author(s) received no financial support for the research, authorship, and/or publication of this article.'
Inclusion criteria	Inclusion (1) displaced femoral neck fractures; (2) patient aged 70 years or older; (3) informed consent obtained, (4) treated with HA or DM, and (5) followed-up for minimum 3 years.
Exclusion criteria	Exclusion

	(1) patients with pathological fractures; (2) patients with any type of neurological disorder that could affect (directly or indirectly) bone density or future recuperation (including paresis or hemiparesis, multiple sclerosis, Parkinson's disease, and other chronic neurodegenerative diseases); and (3) patients with preexisting coxarthrosis in the same hip.
Intervention(s)	Dual Mobility Acetabular Cup THA
Comparator	HA
Outcome measures	Harris Hip Score Functional Independence Mortality Dislocation
Number of participants	n=94
Duration of follow-up	3 years
Loss to follow-up	n=2
Methods of analysis	This was a prospective, comparative interventional study, single-blinded, performed in the University Clinical Center of Kosovo, a tertiary health-care institution.

Study arms

HA (N = 47)

Dual mobility HA (N = 47)

1 **Characteristics**

2 **Arm-level characteristics**

Characteristic	HA (N = 47)	DM HA (N = 47)
Age Mean (SD)	77.64 (4.7)	78.11 (5.41)
BMI (kg/m2) Mean (SD)	26.66 (3.36)	26.96 (3.32)
Sex (male/female) (%)	68/32	49/51

3

4

5 **Critical appraisal - GDT Crit App - Cochrane Risk of Bias tool (RoB 2.0) Normal RCT**

Section	Question	Answer
Domain 1: Bias arising from the randomisation process	Risk of bias judgement for the randomisation process	High *Patient selection for surgical procedures, treatment decisions were made in a random manner where even-numbered patients underwent DM and odd numbered patients underwent HA.)*
Domain 2a: Risk of bias due to deviations from the intended interventions (effect of assignment to intervention)	Risk of bias for deviations from the intended interventions (effect of assignment to intervention)	Low
Domain 3. Bias due to missing outcome data	Risk-of-bias judgement for missing outcome data	Some concerns

Section	Question	Answer
		2 lost to follow up in one arm and not included in overall analysis
Domain 4. Bias in measurement of the outcome	Risk-of-bias judgement for measurement of the outcome	Some concerns Outcome assessors knowledge of interventions could bias some subjective outcomes
Domain 5. Bias in selection of the reported result	Risk-of-bias judgement for selection of the reported result	Low
Overall bias and Directness	Risk of bias judgement	High *(Quasi randomised trial (even numbered patients to one arm, odd numbered to another), no allocation concealment and a lack of clarity over the use of intention to treat and final number of participants in the analysis)*
Overall bias and Directness	Overall Directness	Partially indirect - bipolar implant used which is an exclusion from protocol

1

2

3

1 # Appendix E – Economic Evaluation Checklist

2

Study identification Axelrod 2020		
Category	**Rating**	**Comments**
Applicability		
1.1 Is the study population appropriate for the review question?	Yes	
1.2 Are the interventions appropriate for the review question?	Yes	
1.3 Is the system in which the study was conducted sufficiently similar to the current UK context?	Yes	Canada
1.4 Is the perspective for costs appropriate for the review question?	Yes	
1.5 Is the perspective for outcomes appropriate for the review question?	Yes	
1.6 Are all future costs and outcomes discounted appropriately?	Partly	Used 1.5%
1.7 Are QALYs, derived using NICE's preferred methods, or an appropriate social care-related equivalent used as an outcome? If not, describe rationale and outcomes used in line with analytical perspectives taken (item 1.5 above).	Partly	EQ-5D-5L used
1.8 OVERALL JUDGEMENT	**PARTIALLY APPLICABLE**	
Limitations		
2.1 Does the model structure adequately reflect the nature of the topic under evaluation?	Yes	
2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?	No	The time horizon is 2 years
2.3 Are all important and relevant outcomes included?	Yes	
2.4 Are the estimates of baseline outcomes from the best available source?	Yes	
2.5 Are the estimates of relative intervention effects from the best available source?	Yes	

Study identification Axelrod 2020		
Category	**Rating**	**Comments**
2.6 Are all important and relevant costs included?	Yes	
2.7 Are the estimates of resource use from the best available source?	Partly	Based on a single clinical trial
2.8 Are the unit costs of resources from the best available source?	Yes	
2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?	Yes	
2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?	Yes	
2.11 Has no potential financial conflict of interest been declared?	Yes	
2.12 OVERALL ASSESSMENT	**POTENTIALLY SERIOUS LIMITATIONS**	

1

Study identification Blythe 2020		
Category	**Rating**	**Comments**
Applicability		
1.1 Is the study population appropriate for the review question?	Yes	
1.2 Are the interventions appropriate for the review question?	Yes	
1.3 Is the system in which the study was conducted sufficiently similar to the current UK context?	Yes	Australia
1.4 Is the perspective for costs appropriate for the review question?	Yes	
1.5 Is the perspective for outcomes appropriate for the review question?	Yes	
1.6 Are all future costs and outcomes discounted appropriately?	Partly	Used 3%
1.7 Are QALYs, derived using NICE's preferred methods, or an appropriate social care-related equivalent used as an outcome? If not, describe	Yes	

Study identification Blythe 2020		
Category	**Rating**	**Comments**
rationale and outcomes used in line with analytical perspectives taken (item 1.5 above).		
1.8 OVERALL JUDGEMENT	**PARTIALLY APPLICABLE**	
Limitations		
2.1 Does the model structure adequately reflect the nature of the topic under evaluation?	Yes	
2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?	No	The time horizon is 5 years
2.3 Are all important and relevant outcomes included?	Yes	
2.4 Are the estimates of baseline outcomes from the best available source?	Yes	
2.5 Are the estimates of relative intervention effects from the best available source?	Yes	
2.6 Are all important and relevant costs included?	Yes	
2.7 Are the estimates of resource use from the best available source?	Yes	
2.8 Are the unit costs of resources from the best available source?	Yes	
2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?	Yes	
2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?	Partly	There is also a focus on cemented vs uncemented
2.11 Has no potential financial conflict of interest been declared?	Yes	
2.12 OVERALL ASSESSMENT	**POTENTIALLY SERIOUS LIMITATIONS**	

1

2

Study identification Carroll et al 2011		
Category	**Rating**	**Comments**
Applicability		

Study identification Carroll et al 2011		
Category	**Rating**	**Comments**
1.1 Is the study population appropriate for the review question?	Yes	
1.2 Are the interventions appropriate for the review question?	Yes	
1.3 Is the system in which the study was conducted sufficiently similar to the current UK context?	Yes	
1.4 Is the perspective for costs appropriate for the review question?	Yes	
1.5 Is the perspective for outcomes appropriate for the review question?	Yes	
1.6 Are all future costs and outcomes discounted appropriately?	Partly	QALYs were discounted at 3.5%, Costs were not discounted as it was assumed that all the cost differences were in the first year
1.7 Are QALYs, derived using NICE's preferred methods, or an appropriate social care-related equivalent used as an outcome? If not, describe rationale and outcomes used in line with analytical perspectives taken (item 1.5 above).	Yes	
1.8 OVERALL JUDGEMENT	**DIRECTLY APPLICABLE**	
Limitations		
2.1 Does the model structure adequately reflect the nature of the topic under evaluation?	Yes	
2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?	No	Time horizons used were 2, 3 and 5 years
2.3 Are all important and relevant outcomes included?	Yes	
2.4 Are the estimates of baseline outcomes from the best available source?	Yes	
2.5 Are the estimates of relative intervention effects from the best available source?	Yes	
2.6 Are all important and relevant costs included?	Partly	It was assumed that there were no difference in costs after the first year
2.7 Are the estimates of resource use from the best available source?	Yes	

Study identification Carroll et al 2011		
Category	**Rating**	**Comments**
2.8 Are the unit costs of resources from the best available source?	Yes	
2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?	Yes	
2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?	Partly	No deterministic sensitivity analyses were reported
2.11 Has no potential financial conflict of interest been declared?	Yes	
2.12 OVERALL ASSESSMENT	**POTENTIALLY SERIOUS LIMITATIONS**	

1

2

Study identification Larranaga et al. 2022		
Category	**Rating**	**Comments**
Applicability		
1.1 Is the study population appropriate for the review question?	Yes	
1.2 Are the interventions appropriate for the review question?	Yes	
1.3 Is the system in which the study was conducted sufficiently similar to the current UK context?	Partly	Spain
1.4 Is the perspective for costs appropriate for the review question?	Yes	
1.5 Is the perspective for outcomes appropriate for the review question?	Yes	
1.6 Are all future costs and outcomes discounted appropriately?	No	No discounting was done
1.7 Are QALYs, derived using NICE's preferred methods, or an appropriate social care-related equivalent used as an outcome? If not, describe rationale and outcomes used in line with analytical perspectives taken (item 1.5 above).	Partly	Used EQ-5D-5L data in a Spanish population

Study identification Larranaga et al. 2022		
Category	**Rating**	**Comments**
1.8 OVERALL JUDGEMENT	**PARTIALLY APPLICABLE**	
Limitations		
2.1 Does the model structure adequately reflect the nature of the topic under evaluation?	Yes	
2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?	Yes	Lifetime
2.3 Are all important and relevant outcomes included?	Yes	
2.4 Are the estimates of baseline outcomes from the best available source?	Yes	
2.5 Are the estimates of relative intervention effects from the best available source?	Yes	
2.6 Are all important and relevant costs included?	Yes	
2.7 Are the estimates of resource use from the best available source?	Yes	
2.8 Are the unit costs of resources from the best available source?	Yes	
2.9 Is an appropriate incremental analysis presented or can it be calculated from the data?	Yes	
2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?	Partly	All parameters were not investigated but age and anaesthesiology risk were investigated
2.11 Has no potential financial conflict of interest been declared?	Yes	
2.12 OVERALL ASSESSMENT	**POTENTIALLY SERIOUS LIMITATIONS**	

1

Study identification CG124 model		
Category	**Rating**	**Comments**
Applicability		
1.1 Is the study population appropriate for the review question?	Yes	

Study identification CG124 model		
Category	**Rating**	**Comments**
1.2 Are the interventions appropriate for the review question?	Yes	
1.3 Is the system in which the study was conducted sufficiently similar to the current UK context?	Yes	UK
1.4 Is the perspective for costs appropriate for the review question?	Yes	
1.5 Is the perspective for outcomes appropriate for the review question?	Yes	
1.6 Are all future costs and outcomes discounted appropriately?	Yes	
1.7 Are QALYs, derived using NICE's preferred methods, or an appropriate social care-related equivalent used as an outcome? If not, describe rationale and outcomes used in line with analytical perspectives taken (item 1.5 above).	Yes	
1.8 OVERALL JUDGEMENT	**DIRECTLY APPLICABLE**	
Limitations		
2.1 Does the model structure adequately reflect the nature of the topic under evaluation?	Yes	
2.2 Is the time horizon sufficiently long to reflect all important differences in costs and outcomes?	Yes	
2.3 Are all important and relevant outcomes included?	Yes	
2.4 Are the estimates of baseline outcomes from the best available source?	Partly	The Swedish Hip Arthroplasty Register was used, not sure if this is applicable to a UK population
2.5 Are the estimates of relative intervention effects from the best available source?	Yes	
2.6 Are all important and relevant costs included?	Yes	
2.7 Are the estimates of resource use from the best available source?	No	There are a few expert assumptions and older data
2.8 Are the unit costs of resources from the best available source?	No	Costs from 2000/01 that were uprated
2.9 Is an appropriate incremental analysis presented	Yes	

Study identification CG124 model		
Category	**Rating**	**Comments**
or can it be calculated from the data?		
2.10 Are all important parameters whose values are uncertain subjected to appropriate sensitivity analysis?	Yes	
2.11 Has no potential financial conflict of interest been declared?	Yes	
2.12 OVERALL ASSESSMENT	**POTENTIALLY SERIOUS LIMITATIONS**	

1

2

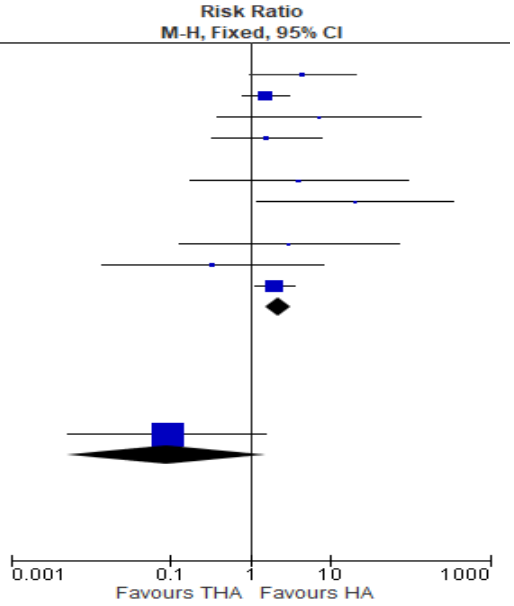
3

Appendix F - Forest plots

Forest plots from Cochrane studies can be found in Lewis 2022 comparison 8 – THA vs HA

Forest plot from subgroup analysis (Adverse events – dislocation) – cognitively impaired / not cognitively impaired

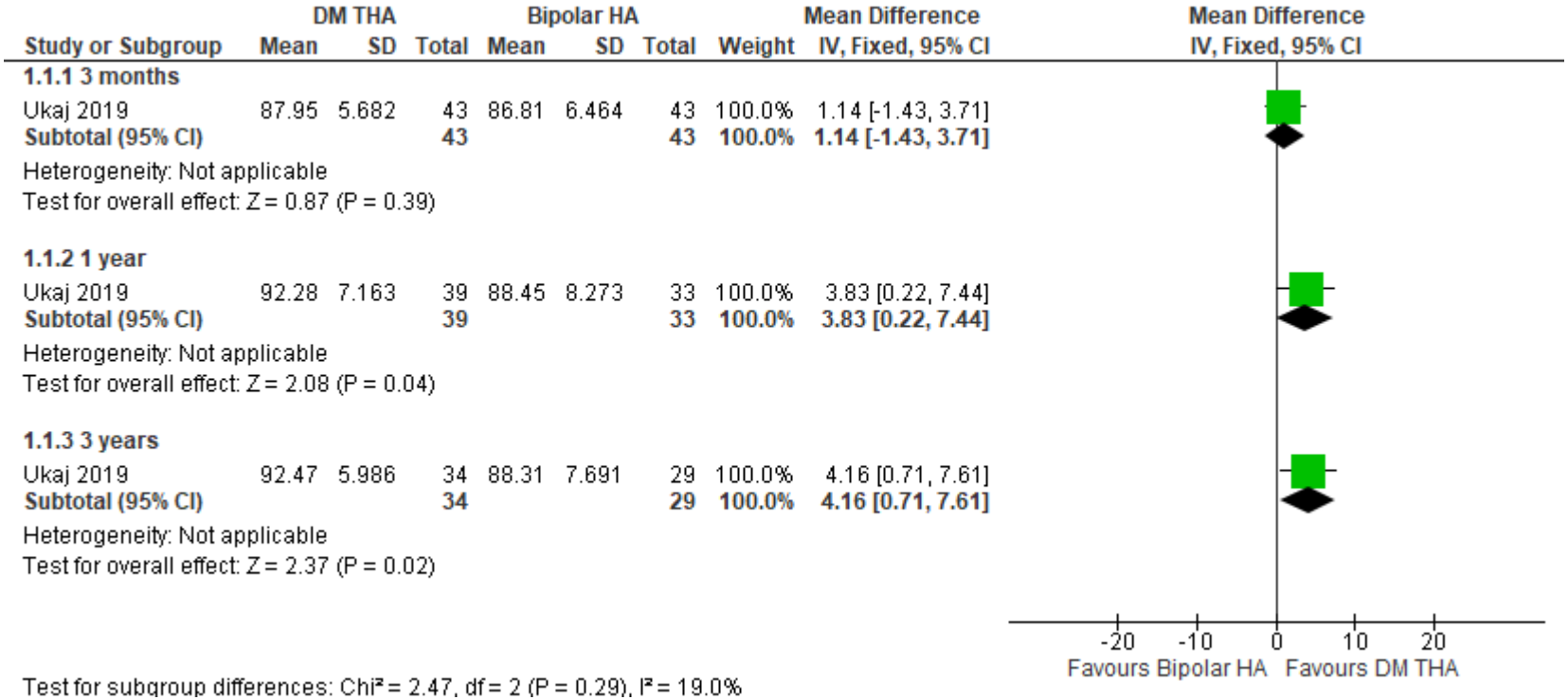
Study or Subgroup	THA Events	THA Total	HA Events	HA Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Year
8.30.5 Dislocation							
Dorr 1986 (1)	7	39	2	50	4.8%	4.49 [0.99, 20.41]	1986
Ravikumar 2000 (2)	18	89	12	91	32.7%	1.53 [0.79, 3.00]	2000
Baker 2006 (3)	3	40	0	41	1.4%	7.17 [0.38, 134.53]	2006
Keating 2006 (4)	3	69	3	111	6.3%	1.61 [0.33, 7.75]	2006
Blomfeldt 2007 (5)	0	60	0	60		Not estimable	2007
Macaulay 2008 (6)	1	17	0	23	1.2%	4.00 [0.17, 92.57]	2008
Van den Bekerom 2010 (7)	8	115	0	137	1.3%	20.22 [1.18, 346.66]	2010
Sharma 2016 (8)	0	40	0	40		Not estimable	2016
Xu 2017 (9)	1	38	0	38	1.4%	3.00 [0.13, 71.40]	2017
Chammout 2019 (10)	0	60	1	60	4.1%	0.33 [0.01, 8.02]	2019
HEALTH 2019 (11)	34	718	17	723	46.7%	2.01 [1.14, 3.57]	2019
Subtotal (95% CI)		**1285**		**1374**	**100.0%**	**2.22 [1.52, 3.23]**	
Total events	75		35				
Heterogeneity: Chi² = 6.74, df = 8 (P = 0.56); I² = 0%							
Test for overall effect: Z = 4.16 (P < 0.0001)							
8.30.6 Dislocation - cognitively imparied population							
Iorio 2019	0	30	5	30	100.0%	0.09 [0.01, 1.57]	2019
Subtotal (95% CI)		**30**		**30**	**100.0%**	**0.09 [0.01, 1.57]**	
Total events	0		5				
Heterogeneity: Not applicable							
Test for overall effect: Z = 1.65 (P = 0.10)							



Footnotes
(1) THA: cemented, but stem, head and cup not reported; HA: cemented and uncemented, bipolar; at 48 months
(2) THA: cemented, Howse II stem, 32mm head, semicaptive cup; HA: uncemented, Austin-Moore, unipolar; at 13 years
(3) THA: cemented, CPT stem, 28mm head, polyethylene cemented cup; HA: cemented, CPT stem, unipolar; at 30 days
(4) THA: all cemented, but stem, head and cup surgeons preference; HA: all cemented, stem surgeons preference, bipolar; at 24 months
(5) THA: cemented, Exeter modular, OGEE (DePuy); HA: cemented, Exeter modular, 28mm bipolar; at 12 months
(6) THA: cement, stem, head (≥28 mm) and cup all at surgeons preference; HA: cement, stem and head all at surgeons preference; at 6 months
(7) THA: cemented, 32mm head, no details for cup; HA: cemented, bipolar; at 60 months
(8) THA: details not reported; HA: details not reported; at 1 week
(9) THA: uncemented, no other details; HA: uncemented, bipolar; at 60 months
(10) THA: cemented, CPT stem, 32mm head, cross linked polyethylene cup; HA: cemented, CPT stem, unipolar; at 24 months
(11) THA: cement, stem, head and cup all at surgeons preference; HA: cement, stem and head all at surgeons preference; at 24 months

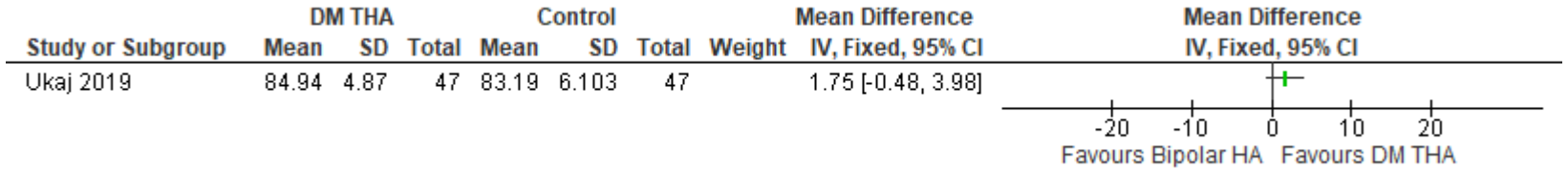
Forest plots from NICE search

Harris Hip Score – overall functional status



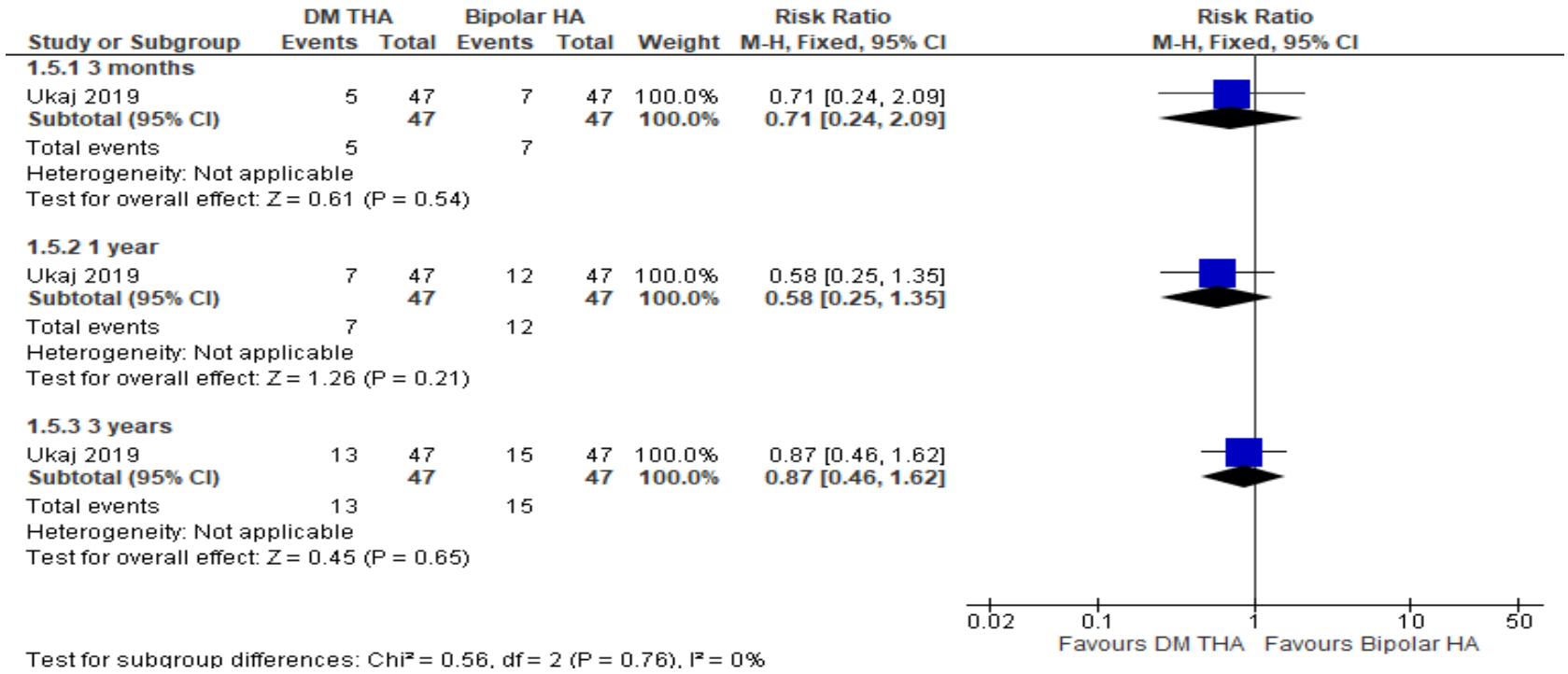
Test for subgroup differences: Chi² = 2.47, df = 2 (P = 0.29), I² = 19.0%

1 **Functional independence Measure**



2

3 **Mortality**



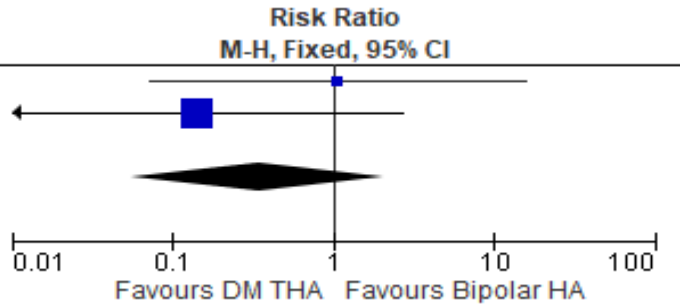
4 Test for subgroup differences: Chi² = 0.56, df = 2 (P = 0.76), I² = 0%

5

Dislocation

Study or Subgroup	DM THA Events	DM THA Total	Bipolar HA Events	Bipolar HA Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Makeen 2021	1	16	1	17	21.7%	1.06 [0.07, 15.60]
Ukaj 2019	0	47	3	47	78.3%	0.14 [0.01, 2.69]
Total (95% CI)		**63**		**64**	**100.0%**	**0.34 [0.06, 2.08]**
Total events	1		4			

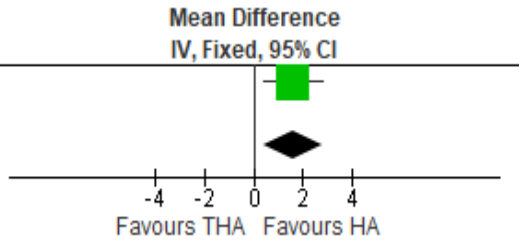
Heterogeneity: Chi² = 1.02, df = 1 (P = 0.31); I² = 2%
Test for overall effect: Z = 1.16 (P = 0.24)



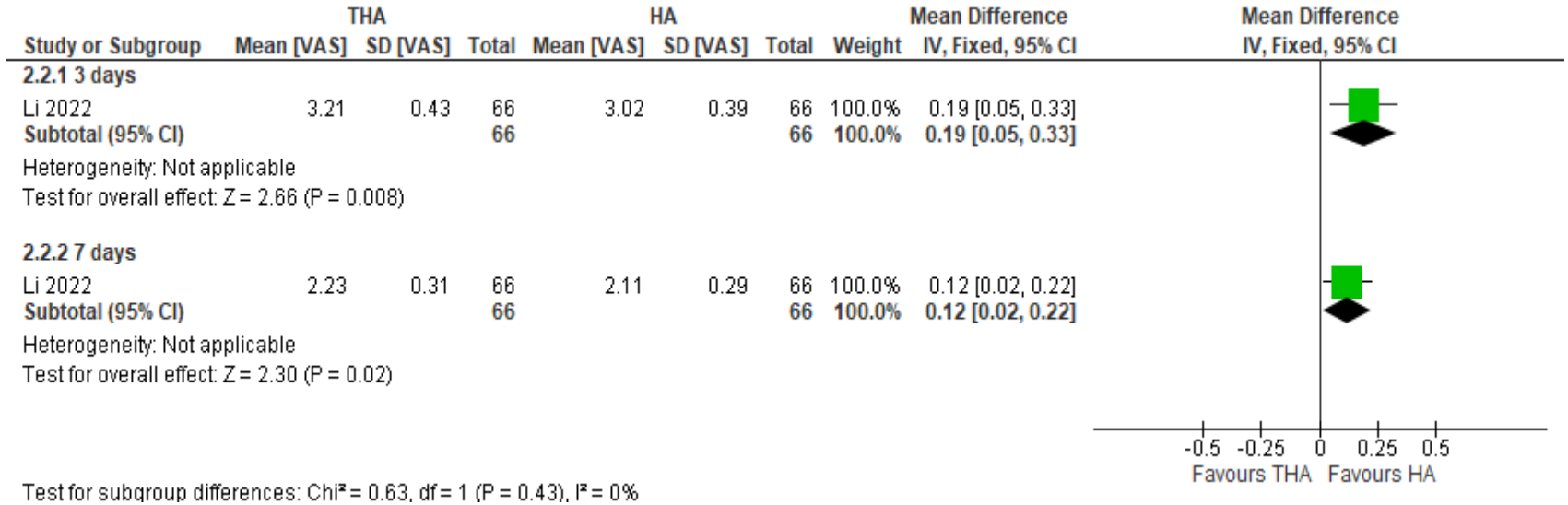
Hospital length of stay (days)

Study or Subgroup	THA Mean [Days]	THA SD [Days]	THA Total	HA Mean [Days]	HA SD [Days]	HA Total	Weight	Mean Difference IV, Fixed, 95% CI
Li 2022	19.92	3.61	66	18.32	3.52	66	100.0%	1.60 [0.38, 2.82]
Total (95% CI)			**66**			**66**	**100.0%**	**1.60 [0.38, 2.82]**

Heterogeneity: Not applicable
Test for overall effect: Z = 2.58 (P = 0.010)

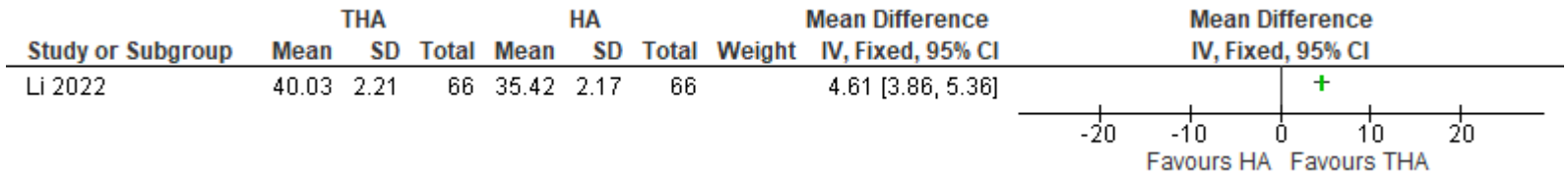


Pain (VAS)

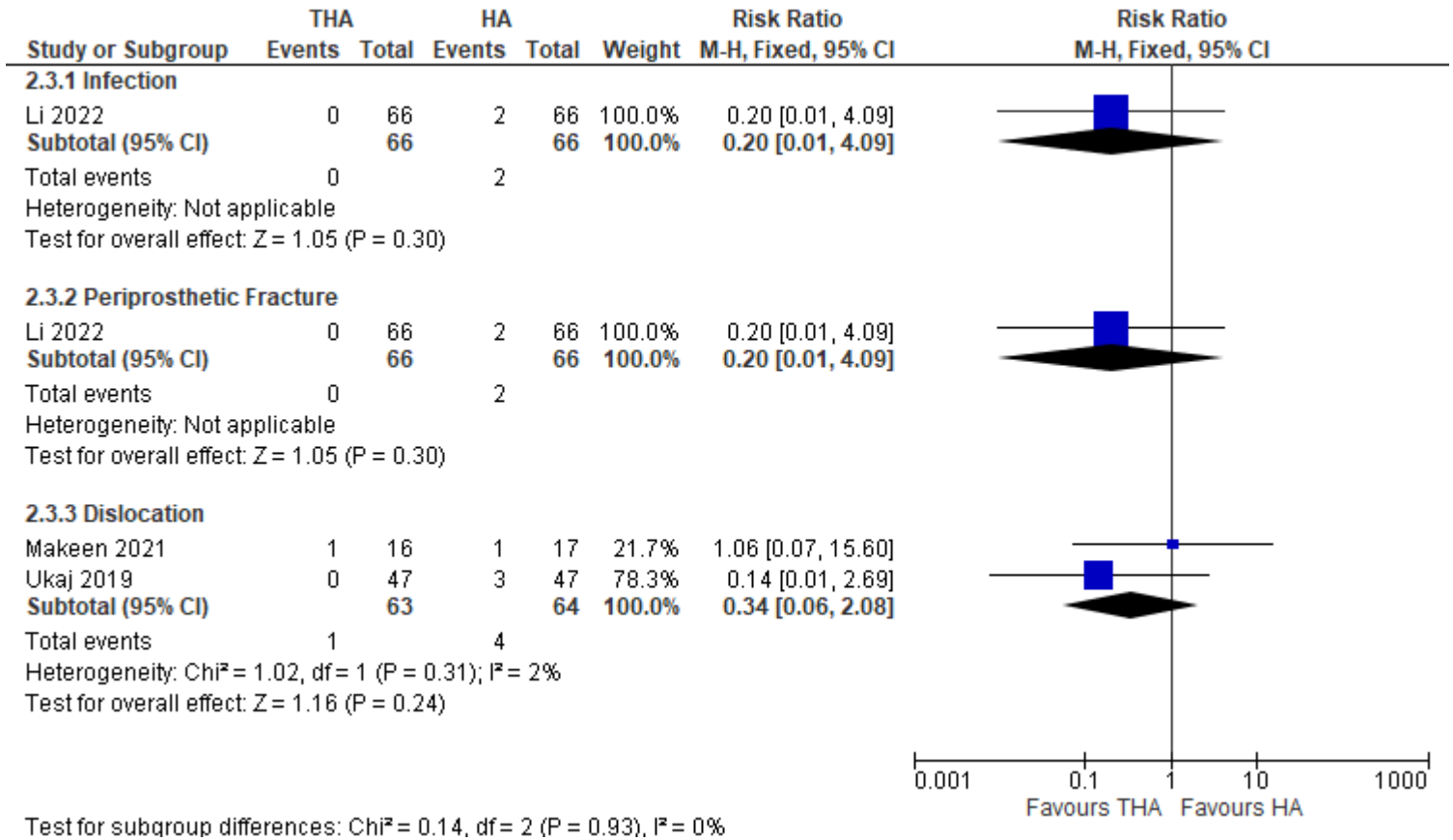


Test for subgroup differences: $Chi^2 = 0.63$, df = 1 (P = 0.43), $I^2 = 0\%$

Pain domain (Harris Hip Score) – at 12 months

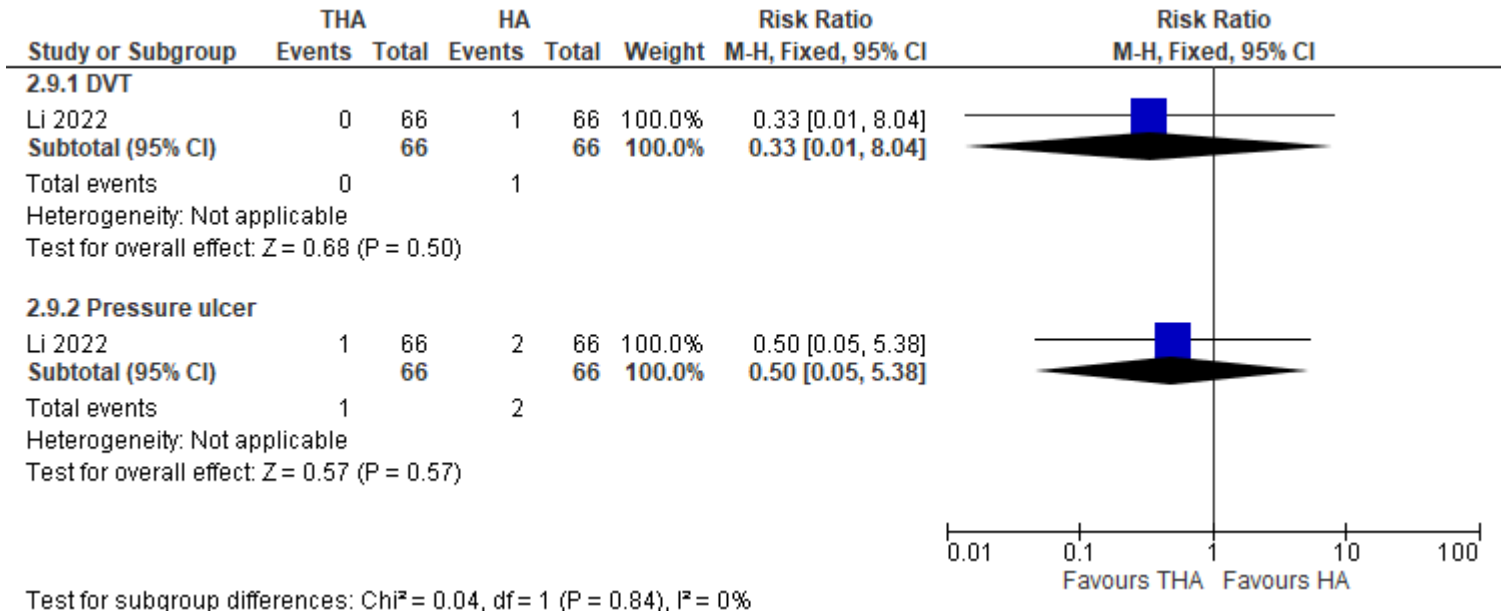


Adverse event related to implant or fracture



Test for subgroup differences: Chi² = 0.14, df = 2 (P = 0.93), I² = 0%

Adverse event unrelated to implant or fracture



Test for subgroup differences: $Chi^2 = 0.04$, df = 1 (P = 0.84), $I^2 = 0\%$

Appendix G - GRADE tables

Outcomes from Lewis 2022 Cochrane review evidence

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 ADL (≤ 4 months, using categorical data)											
2 [b, d]	RCT	225	RR 1.03 (95% CI 0.91, 1.18)	783 per 1000	806 per 1000	23 more per 1000 (70 fewer to 141 more)	Serious [19]	Not serious	Serious[20]	Not serious	Low
Early ADL (≤ 4 months; using social mobility scale[1]) >0 favours HA											
1 [k]	RCT	83	MD -0.10 (95% CI -0.46, 0.26)	-	-	-	Not serious	N/A	Serious[20]	Serious[24]	Low
ADL 12 months, using categorical data											
2 [b, d]	RCT	217	RR 0.96 (95% CI 0.86 – 1.07)	768 per 1000	737 per 1000	31 fewer per 1000 (108 fewer to 54 more)	Serious [19]	Not serious	Serious[20]	Not serious	Low
ADL (12 months; using Barthel Index)[2] > 0 favours THA											

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 [j]	RCT	63	MD -0.68 (95% CI -1.18, -0.17)	-	-	-	Serious [19]	N/A	Serious[20]	Not serious	Low
ADL (12 months using social mobility scale)[3] > 0 favours THA											
1 [k]	RCT	78	MD 0.09 (95% CI -0.35, 0.53)	-	-	-	Not serious	N/A	Serious[20]	Not serious	Moderate
Late ADL (> 24 months; using Barthel Index[2]) > 0 favours THA											
1 [j]	RCT	43	MD 5.70 (95% CI 0.21, 11.19)	-	-	-	Serious [19]	N/A	Serious[20]	Serious[25]	Very Low
Early functional status ≤ 4 months (HHS and Hip rating questionnaire) >0 favours THA											
3 [b, d, h]	RCT	395	Std MD 0.27 (95% CI 0.07, 0.47)	-	-	-	Serious [19]	Not serious	Serious[20]	Not serious	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
Functional status (12 months) (HHS, Johansen Hip Score, WOMAC), > 0 favours THA											
8 [b, d, f, h, l, j, o, q]	RCT	1273	Std MD 0.29 (95% CI 0.14, 0.44)	-	-	-	Serious [19]	Not serious	Serious[20]	Not serious	Low
Functional status (HHS – good/excellent)											
2 [m, o]	RCT	140	RR 1.07 (95% CI 0.98, 1.17)	957 per 1000 people	1024 per 1000 people	67 more per 1000 (19 fewer to 163 more)	Very serious	Not serious	Serious[20]	Not serious	Very low
Late functional status (>24 months using OHS or HHS)[4] > 0 favours THA											
4 [a, b, j ,q]	RCT	224	Std MD 0.65 (95% CI 0.23, 1.08	-	-	-	Serious [19]	Serious inconsistency[12]	Serious[20]	Serious[28]	Very low
Early HRQoL (≤ 4 months) EQ-5D, >0 favours THA											
2 [d, h]	RCT	279	MD 0.03 (95% CI -0.06, 0.12)	-	-	-	Serious [19]	Serious inconsistency[12]	Serious[20]	Not serious	Very 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
HRQoL (12 months) EQ-5D and SF-36, > 0 favours THA											
4 [d, f, h, l]	RCT	1158	Std MD 0.19 (95% CI 0.07, 0.31)	-	-	-	Not serious [19]	Not serious	Serious[20]	Not serious	Moderate
HRQoL (> 24 months. Using SF-36;[5]) > 0 favours THA											
1 [a]	RCT	34	MD 5.90 (95% CI -1.99, 13.79)	-	-	-	Serious [19]	N/A	Not serious	Serious[13]	low
Early mobility (≤ 4 months)[1] (using 10 point scoring system) >0 favours HA											
1 [k]	RCT	83	MD -0.40 (95% CI -0.96, 0.16)	-	-	-	Not serious	N/A	Serious[20]	Serious[26]	Low
Mobility (12 months, using TUG)[1] >0 favours HA											
2 [f, l]	RCT	575	MD -2.74 (95% CI -	-	-	-	Not serious	Not serious	Serious[20]	Not serious	Moderate

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
			6.82, 1.35)								
Mobility (12 months, using 9-point mobility scale)[1] >0 favours HA											
1 [k]	RCT	78	MD 0.40 (95% CI -0.32, 1.12)	-	-	-	Not serious	N/A	Serious[20]	Serious[27]	Low
Mobility (12 months; able to ambulate independently)											
2 [i, L]	RCT	175	RR 0.96 (95% CI 0.71,1.31)	709 per 1000 people	681 per 1000 people	28 fewer (206 fewer to 220 more)	Serious [19]	Serious inconsistency[12]	Serious[20]	Very serious[14]	Very low
Late mobility (> 24 months; able to ambulate independently)											
1 [L]	RCT	32	RR 1.27 (95% CI 0.71, 2.29)	684 people per 1000	869 people per 1000	185 per 1000 more (198 fewer to 882 more)	Serious [19]	N/A	Serious[20]	Very serious[14]	Very low
Early mortality (≤ 4 months)											
6 [b, g, h, k, l, n]	RCT	725	RR 0.77 (95% CI	62 per 1000 people	48 per 1000 people	14 fewer per 1000 (36 fewer	Serious [19]	Not serious	Serious[20]	Very serious[14]	Very 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
			0.42, 1.42)			to 26 more)					
Mortality (12 months)											
11 [b, c, d, f, g, h, I ,j ,k, l, p]	RCT	2667	RR 1.00 (95% CI 0.83, 1.22)	135 per 1000 people	135 per 1000 people	0 more (23 fewer to 30 more)	Serious [19]	Not serious	Serious[20]	Not serious	Low
Late mortality (> 24 months)											
8 [a b c I j l p q]	RCT	931	RR 1.00 (95% CI 0.81, 1.23)	450 per 1000 people	450 per 1000 people	0 more per 1000 (85 fewer to 104 more)	Serious [19]	Serious inconsistency[12]	Serious[20]	Not serious	Very low
Unplanned return to theatre (end of follow up)											
10 [a d e f g h j k l p]	RCT	2594	RR 0.63 (95% CI 0.37, 1.07)	85 per 1000 people	54 per 1000 people	31 fewer per 1000 (54 fewer to 6 more)	Serious [19]	Serious inconsistency[12]	Serious[20]	Serious[15]	Very low
Length of hospital stay (days)											

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
3 [h, l , j]	RCT	306	MD 0.80 (95% CI -1.12, 2.73)	-	-	-	Serious [19]	Serious inconsistency[12]	Serious[20]	Serious[18]	Very low
Pain (12 months)[8] (mixed scales)[22] >0 favours HA											
9 [b, c ,d ,f, h, l, k, o, q]	RCT	1435	Std MD -0.13 (95% CI -0.38, 0.12)	-	-	-	Serious [19]	Very serious[16]	Serious[20]	Not serious	Very Low
Late pain (>24 months)[10] (THA uncemented, HA mixed, HHS) >0 favours THA											
1 [c]	RCT	32	MD -3.50 (95% CI -7.19, 0.19)	-	-	-	Serious [19]	N/A	Serious[20]	Serious[17]	Very low
Late pain (>24 months)[9] (THA uncemented, HA cemented, bipolar, HHS) >0 favours THA											
1[b]	RCT	83	MD 7.90 (95% CI 5.69, 10.11)	-	-	-	Serious [19]	N/A	Serious[20]	Not serious	Low
Pain (> 24 months – categorical data – No Pain)											

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[L]	RCT	135	RR 1.47 (95% CI 1.07, 2.00)	667 people per 1000	980 people per 1000	313 more (47 more to 667 more)	Serious[19]	N/A	Serious[20]	Serious[15]	Very low
Early pain (≤ 4 months)[11] (mixed scales)[23] >0 favours THA											
5[b c d h k]	RCT	572	Std MD 0.10 (95% CI -0.10, 0.30)	-	-	-	Serious[19]	Not serious	Serious[20]	Not serious	Low
Discharge destination (own home)											
2[f h]	RCT	1612	RR 0.97 (95% CI 0.87, 1.08)	387 people per 1000	375 people per 1000	12 fewer per 1000 (50 fewer to 31 more)	Serious[19]	Not serious	Serious[20]	Not serious	Low
Discharge destination (older persons ward)											
1[d]	RCT	120	RR 0.88 (95% CI 0.34, 2.26)	133 people per 1000	117 people per 1000	16 fewer per 1000 (88 fewer to 168 more)	Serious[19]	N/A	Not serious	Very serious[14]	Very low
Adverse events relating to implant, fracture or both											

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
Postoperative periprosthetic fracture											
3 [f,o,q]	RCT	1557	RR 1.08 (95% CI 0.70, 1.66)	49 per 1000 people	53 per 1000 people	4 more (15 fewer to 32 more)	Not serious	Not serious	Serious[20]	Very serious[14]	Very low
Prosthetic loosening											
4 [b,f,p,q]	RCT	1889	RR 0.64 (95% CI 0.17, 2.41	10 per 1000 people	6 per 1000 people	4 fewer (8 fewer to 14 more)	Not serious	Not serious	Serious[20]	Very serious[14]	Very low
Deep infection											
8 [d, e, f,k,l,n,p,q]	RCT	2343	RR 0.87 (95% CI 0.50, 1.54)	23 per 1000 people	20 per 1000 people	3 fewer per 1000 (11 fewer to 12 more)	Not serious	Not serious	Serious[20]	Very serious[14]	Very low
Superficial infection											
10 [a, b, d, e, f,h,l,k,n,p	RCT	2495	RR 1.25 (95% CI 0.67, 2.30)	16 per 1000 people	20 per 1000 people	4 more per 1000 (5 fewer to 21 more)	Serious [19]	Not serious	Serious[20]	Very serious[14]	Very 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
Dislocation											
12 [a, b, d, e, f, g, h ,I, L, n, p, q]	RCT	2719	RR 1.96 (95% CI 1.17, 3.27)	28 per 1000 people	55 per 1000 people	27 more (5 more to 64 more)	Serious [19]	Not serious	Serious[20]	Serious[15]	Very Low
Dislocation – without cognitively impaired population											
11 [a, b, d, e, f, h ,I, L, n, p, q]	RCT	2659	RR 2.22 (95% CI 1.52, 3.23)	25 per 1000 people	56 per 1000 people	31 more per 1000 (13more to 56 more)	Serious [19]	Not serious	Serious[20]	Not serious	Low
Dislocation – cognitively impaired population only											
1[g]	RCT	60	RR 0.09 (95% CI 0.01, 1.57)	167 per 1000 people	15 per 1000 people	152 fewer per 1000 (165 fewer to 95 more)	Very serious [21]	Not serious	Serious[20]	Very serious[3]	Very 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
2[d,f]	RCT	1561	RR 1.09 (95% CI 0.62,1.92)	28 per 1000 people	31 per 1000 people	3 more per 1000 (11 fewer to 26 more)	Not serious	Not serious	Serious[20]	Very serious[14]	Very Low
Blood transfusion											
2 [h, k]	RCT	285	RR 2.14 (95% CI 1.27, 3.61)	116 per 1000 people	248 per 1000 people	132 more per 1000 (31 more to 303 more)	Serious [19]	Not serious	Serious[20]	Not serious	Low
Cerebrovascular accident											
4 [d, h ,k ,p]	RCT	657	RR 1.63 (95% CI 0.63, 4.21)	19 per 1000 people	31 per 1000 people	12 more per 1000 (7 fewer to 50 more)	Serious [19]	Not serious	Serious[20]	Very serious[14]	Very low
Pneumonia/chest infection (reported at > 4 months)											
5 [a, b, d, l, p]	RCT	613	RR 0.87 (95% CI 0.38, 2.00)	40 per 1000 people	35 per 1000 people	5 fewer (25 fewer to 40 more)	Serious [19]	Not serious	Not serious	Very serious[14]	Very low
Myocardial infarction											

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
4 [b, d ,h, i]	RCT	460	RR 1.48 (95% CI 0.48, 4.58)	24 per 1000 people	36 per 1000 people	12 more per 1000 (12 fewer to 24 more)	Serious [19]	Not serious	Serious[20]	Very serious[14]	Very low
Urinary Tract Infection											
1[i]	RCT	40	RR 0.19 (95% CI 0.01, 3.46	130 per 1000 people	25 per 1000 people	105 fewer per 1000 (129 fewer to 320 more)	Not serious	Not serious	Serious[20]	Very serious[14]	Very Low
Venous thromboembolic phenomena (DVT)											
4 [a, b, h ,k]	RCT	486	RR 4.25 (95% CI 0.86, 21.06)	4 per 1000 people	17 per 1000 people	13 more per 1000 (1 fewer to 80 more)	Serious [19]	Not serious	Serious[20]	Serious[15]	Very Low
Venous thromboembolic phenomena (pulmonary embolism)											
5 [a, d, h, l, p]	RCT	673	RR 0.49 (95% CI 0.14, 1.63)	30 per 1000 people	15 per 1000 people	15 fewer per 1000 (26 fewer to 19 more)	Serious [19]	Not serious	Serious[20]	Very serious[14]	Very low

1 Lower scores indicate better mobility
2 Higher scores indicate more independence

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

3 Lower scores indicate more independence
4 Higher scores indicate better function
5 Higher scores indicate better quality of life
6 THA: cement, stem, head (≥28 mm) and cup all at surgeons preference; HA: cement, stem and head all at surgeons preference; at 12 months
7 THA: cemented, Howse II stem, 32 mm head, semicaptive cup; HA: uncemented, Austin-Moore, unipolar; at 12 months
8 Lower scores indicate less pain
9 HHS (higher scores indicate less pain); THA: cemented, Exeter modular, OGEE (DePuy); HA: cemented, Exeter modular, 28mm bipolar; at 48 months
10 HHS (higher scores indicate less pain); THA: uncemented, Conus stem, large diameter head, polycarbonate-urethane cup; HA: mixed cemented and uncemented
11 Higher scores indicate less pain
12 I^2 between 33.3% and 66.6%
13 Confidence interval crosses MID at one end (MID 0.5 * median standard deviation of control group = 5.7)
14 Confidence interval crosses MID at both ends (0.8 – 1.25)
15 Confidence interval crosses MID at one end
16 I^2 above 66.6%
17 Confidence interval crosses MID at one end (MID 0.5 * median standard deviation of control group = 2.67
18 Confidence interval crosses MID at one end (MID 0.5 * median standard deviation of control group = 1.7
19 Greater than 33.3% of the weight in a meta-analysis came from studies at moderate or high risk of bias
20 Greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies
21 Greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias
22 Using Hip rating questionnaire, HHS, WOMAC, VAS, and 8-point pain scale
23 Using Hip rating questionnaire, HHS, 8-point plan scale, and VAS
24 Confidence interval crosses MID at one end (0.5* median SD of control group = 0.43)
25 Confidence interval crosses MID at one end (0.5* median SD of control group = 5.8)
26 Confidence interval crosses MID at one end (0.5* median SD of control group = 0.73)
27 Confidence interval crosses MID at one end (0.5* median SD of control group =0.77)
28 Confidence interval crosses MID at one end (for SMDs 0.5)

a) Baker 2006
b) Blomfeldt 2007

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
c) Cadossi 2013 d) Chammout 2019 e) Dorr 1986 f) HEALTH 2019 g) Iorio 2019 h) Keating 2006 I) Macaulay 2008 J) Mouzouplos 2008 K) Parker 2019) L) Ravikumar 2000 m) Ren 2017 n) Sharma 2016 o) Sonaje 2017 p) Van De Bekerom 2010 q) Xu 2017											

1

Outcomes from >July 2020 NICE search

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
THA vs HA											
Overall functional status - Harris Hip Score – 3 months[1] >0 favours THA											
1[a]	RCT	86	MD 1.14 (95% CI -1.43, 3.71)	-	-	-	Very Serious[6]	N/A	Serious[8]	Serious[9]	Very low
Overall functional status - Harris hip score - 1 year[1] >0 favours THA											
1[a]	RCT	72	MD 3.83 (95% CI 0.22, 7.44)	-	-	-	Very Serious[6]	N/A	Serious[8]	Serious[2]	Very low
Overall functional status - Harris hip score – 3 years[1] >0 favours THA											
1[a]	RCT	63	MD 4.16 (95% CI 0.71, 7.61)	-	-	-`	Very Serious[6]	N/A	Serious[8]	Serious[3]	Very Low
Functional independence measure[4] >0 favours THA											

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[a]	RCT	94	MD 1.75 (95% CI -0.48, 3.98)	-	-	-	Very Serious[6]	N/A	Serious[8]	Serious	Very low
Mortality - 3 months											
1[a]	RCT	94	RR 0.71 (95% CI 0.24, 2.09	149 per 1000 people	106 per 1000 people	43 fewer per 1000 (113 fewer to 162 more)	Very Serious[6]	N/A	Serious[8]	Very serious[5]	Very low
Mortality - 1 year											
1[a]	RCT	94	RR 0.58 (95% CI 0.25, 1.35	255 per 1000 people	148 per 1000 people	107 fewer per 1000 (191 fewer to 89 more)	Very Serious[6]	N/A	Serious[8]	Very serious[5]	Very low
Mortality - 3 years											
1[a]	RCT	94	RR 0.87 (95% CI 0.46, 1.62)	319 per 1000 people	278 per 1000 people	41 fewer per 1000 (172 fewer to 198 more)	Very Serious[6]	N/A	Serious[8]	Very serious[5]	Very low

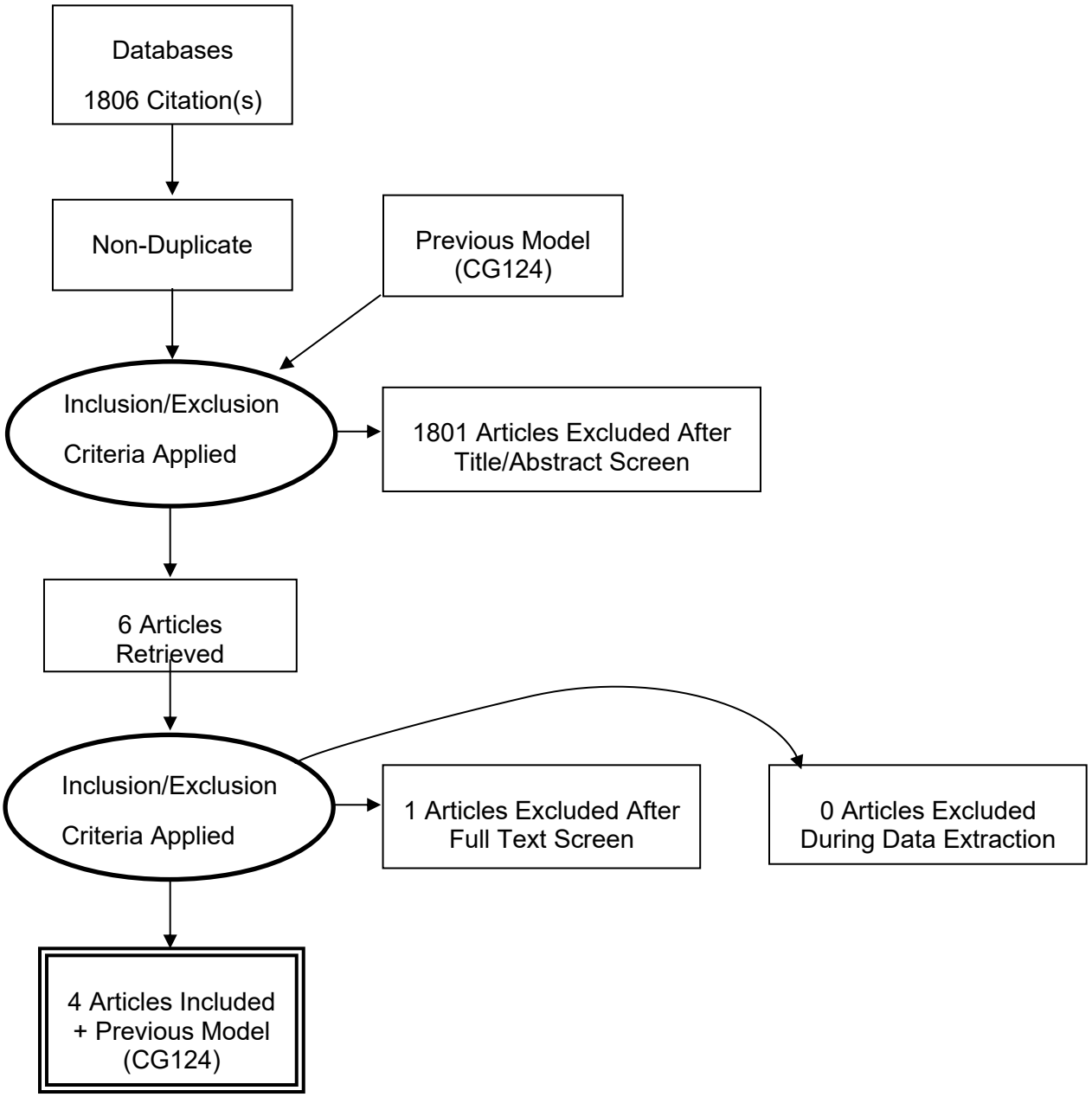
No. of studies	Study design	Sample size	Effect size (95% CI)	Absolute risk (control)	Absolute risk (intervention)	Absolute risk difference	Risk of bias	Inconsistency	Indirectness	Imprecision	Quality
Hospital length of stay (days) >0 favours HA											
1[c]	RCT	132	MD 1.60 (95% CI 0.38, 2.82)	-	-	-	Very Serious[6]	N/A	Serious[8]	Serious[11]	Very low
Pain (VAS) – 3 days[7] >0 favours HA											
1[c]	RCT	132	MD 0.19 (95% CI 0.05, 0.33)	-	-	-	Very Serious[6]	N/A	Serious[8]	Serious[12]	Very Low
Pain (VAS) – 7 days[7] >0 favours HA											
1[c]	RCT	132	MD 0.12 (95% CI 0.02, 0.22)	-	-	-	Very Serious[6]	N/A	Serious[8]	Serious[13]	Very Low
Harris Hip Score - 12 months (pain domain)[1] >0 favours THA											
1[c]	RCT	132	MD 4.61 (95% CI 3.86, 5.36)	-	-	-	Very Serious[6]	N/A	Serious[8]	Not serious	Very Low
Adverse event related to implant or 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
Infection											
1[c]	RCT	132	RR 0.20 (95% CI 0.01,4.09)	30 per 1000 people	6 per 1000 people	24 fewer per 1000 (30 fewer to 93 more)	Very Serious[6]	N/A	Serious[8]	Very Serious[5]	Very low
Periprosthetic Fracture											
1[c]	RCT	132	RR 0.20 (95% CI 0.01,4.09)	30 per 1000 people	6 per 1000 people	24 fewer per 1000 (30 fewer to 93 more)	Very Serious[6]	N/A	Serious[8]	Very Serious[5]	Very low
Dislocation											
2 [a, b]	RCT	127	RR 0.34 (95% CI 0.06, 2.08)	63 per 1000 people	21 per 1000 people	42 fewer per 1000 (59 fewer to 68 more)	Very Serious[6]	Not serious	Serious[8]	Very Serious[5]	Very low
Adverse event unrelated to implant or fracture											
Pressure ulcer											

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[c]	RCT	132	RR 0.50 (95% CI 0.05, 5.38)	30 per 1000 people	15 per 1000 people	15 fewer per 1000 (28 fewer to 131 more)	Very Serious [6]	N/A	Serious[8]	Very Serious[5]	Very low
DVT											
1[c]	RCT	132	RR 0.33 (95% CI 0.01, 8.04)	15 per 1000 people	5 per 1000 people	10 fewer per 1000 (15 fewer to 106 more)	Very Serious [6]	N/A	Serious[8]	Very Serious[5]	Very low

1) Higher scores are better
2) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 3.58)
3) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 3)
4) Higher scores indicate greater independence
5) Confidence interval crossed the MID at both ends (MID 0.8 - 1.25)
6) Greater than 33.3% of the weight in a meta-analysis came from studies at high risk of bias
7) Higher scores indicate greater pain
8) Greater than 33.3% of the weight in a meta-analysis came from partially indirect or indirect studies
9) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 2.84)
10) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 2.44)
11) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 1.76)
12) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 0.2)
13) Confidence interval crosses the MID at one end (MID 0.5* median SD of control group = 0.15)
    a) Ukaj 2019
    b) Makeen 2021
    c) Li 2022

Appendix H – Economic evidence study selection



Appendix I – Economic evidence tables

Study	Study type	Setting	Interventions	Population	Methods of analysis	Base-case results	Sensitivity analyses	Additional comments
Axelrod et al. 2020	Cost-utility study based on an RCT (HEALTH study)	Canada Single payer perspective	Total hip replacement (THR) vs hemiarthroplasty (HA)	THR (n=718) HA (n=723) Patients had to be 50 years or over and have a low energy displaced fracture of the femoral neck. Patients had to be able to ambulate without assistance before the fracture occurred. 80 participating sites in Canada, Spain, UK, the Netherlands, Norway, Finland, Australia, NZ, and South Africa.	Study collected data on health related quality of life, secondary procedures, serious adverse events, physiotherapy visits (assumed to be 7) and hospital readmissions. Unit costs were obtained from the Canadian Institute of Health Information. Implant costs were obtained from Hamilton General Hospital. EQ-5D-5L data was obtained from the HEALTH study. Time horizon: 2 years Discount rate: 1.5 %	THR: Cost CAD 32,851 QALY 1.4 HA: Cost CAD 27,358 QALY 1.36 Incremental: Cost CAD 5,493 QALY 0.04 ICER: CAD 151,640	Deterministic: Changing the discount rate 0% to 3% did not meaningfully change the ICER. Changing the cost of total hip replacement by 30% did not change the ICER significantly. Probabilistic: Probability of total hip replacement being cost effective was 12.8% and 32.8% for £30,020 to £60,040 willingness to pay threshold	Source of funding: Not reported Limitation: EQ-5D-5L data was not available for absent for some observations, multiple imputation was used to generate a compete data set. The time horizon was 2 years and more complications may occur after this time. Authors' conclusion: THR is not cost effective compared to HA. However, there may be benefit for younger patients.
Blythe et al 2020	Cost utility study	Australia Australian health care system	Total hip replacement (THR) vs hemiarthroplasty (HA)	Study was based on data from the Australian Orthopedic Association National Joint Replacement Registry collected in 2017	Costing data was obtained from the Metro North Clinical Costing and Reporting department. Further costing to differentiate between surgical types was obtained from the Metro North dataset, New South Wales Operating Theater Standard Costs Template, and the Australian Prostheses List 2017. QALY data was obtained from the literature	<75: THR: Cost $11,633,253 QALY 3528 HA: Cost $6,143,466 QALY 2178 Incremental: Cost $5,489,787 QALY 1350	Scenario analysis was completed that assumed that all patients were equally suited to THR and HA. The analysis showed that the more patients receiving hybrid THA over cemented HA, the greater the costs and QALYs, with diminishing returns as patients aged.	Source of funding: No grants from funding agencies in the public, commercial or not for profit sectors Limitation: Each age group was assumed to be homogeneous. Cycle length is one year and utilities were assumed to be constant for that entire year. Authors' conclusion: Cemented HA showed the greatest reduction in costs and increase in quality of life in older patients,

Study	Study type	Setting	Interventions	Population	Methods of analysis	Base-case results	Sensitivity analyses	Additional comments
						ICER: $4,067 75-85: THR: Cost $7,300,478 QALY 1942 HA: Cost $14,109,811 QALY 5135 Incremental: Cost $6,809,333 QALY 3193 ICER: $2,133 >85: THR: Cost $2,608,752 QALY 516 HA: Cost $16,409,925 QALY 5131 Incremental: Cost $13,801,173 QALY 4615		indicating that older patients may benefit more from this procedure.

Study	Study type	Setting	Interventions	Population	Methods of analysis	Base-case results	Sensitivity analyses	Additional comments
						ICER: €2,991		
Carroll et al. 2011	Cost utility study	UK NHS and PSS perspective	Total hip replacement (THR) vs hemiarthroplasty (HA)	Economic analysis was based on multiple RCT data. EQ-5D and cost data from Keating et al. for patients with displaced intracapsular hip fracture	Costs were obtained from Keating et al and included initial inpatient episode, hip-related admissions, non-hip related admissions, total hip related costs and total costs. EQ-5D data was obtained from Keating et al, providing values at 4, 12 and 24 months. Time horizon: 2, 3 and 5 years Discount rate: Utilities at 3.5%, it was assumed all differences in costs occurred in the first year therefore no discounting was done.	2 year Incremental: Cost £3,989 QALY 0.147 ICER: £27,023 3 year Incremental: Cost £3,989 QALY 0.285 ICER: £16,146 5 years Incremental: Cost £3,989 QALY 0.580 ICER: £7,952	Using data reported by Blomfeldt et al found the cost per QALY was £44,997, £30,511 and £18,932 at 2, 3 and 5 years, respectively	Source of funding: NIHR Limitation: Longer term consequences are not included in the analysis Authors' conclusion: THR is cost effective compared to HA. However, there is likely to be increase costs in the first 2 years.
Larranaga et al 2022	Cost utility study	Spain Spanish National Health System	Total hip replacement (THR) vs hemiarthroplasty (HA)	Data were collected from the corporative database, which contains administrative and clinical records of the Basque Health Service in an anonymized form including the variables: age, sex, socioeconomic status, hospital size, diagnoses required for calculating the Charlson comorbidity index,14 American Society	Costs were obtained from the Basque Heath Service. QALYs were obtained from a Spanish utility data set published in 2019	THR: Cost €13,704 QALY 3.01 HA: Cost €11,357 QALY 2.20 Incremental: Cost €2,346	Sensitivity analysis showed that partially hip replacement should be used in most patients with total hip replacement being reserved for younger patients	Source of funding: Basque Government Department of Health Limitation: Lack of utility information at a patient level, Authors' conclusion: HA in most patients and reserving THR for younger patients.

Study	Study type	Setting	Interventions	Population	Methods of analysis	Base-case results	Sensitivity analyses	Additional comments
				of Anesthesiologists (ASA) class, history of antithrombotic drug use, type of anaesthesia, type of prosthesis, time to surgery in days, surgical time in minutes, hospital stay, complications up to 1 year after surgery, life-long complications after, date of death and place of residence		QALY 0.81 ICER: €2,912		
CG124 Model	Cost utility study	UK NHS perspective	Total hip replacement (THR) vs hemiarthroplasty (HA)	Data were collected from previously published studies, particularly using Keating et al (2005)	Costs were obtained from Keating et al (2005) QALYs were obtained from Keating et al. (2005)	THR: Cost £11,083 QALY 4.05 HA: Cost £11,387 QALY 3.51 Incremental: Cost £304 QALY -0.54 ICER: THR dominates	Sensitivity analysis showed that total hip replacement dominated in almost all of the scenarios including in the probabilistic sensitivity analysis	Source of funding: NICE guideline Authors' conclusion: THR is cost effective

1 THR= Total hip replacement, HA=Hemiarthroplasty, QALY=Quality Adjusted Life Year

Appendix J – Health economic model

The model can be found in Health economic report for evidence review B

The references in the model were:

Curtis L, Burns A. (2021) Unit cost of health and social care 2021. University of Kent, UK

Garellick, G., Kärrholm, J., Lindahl, H., Malchau, H., Rogmark, C., Rolfson., 2014. The Swedish Hip Arthroplasty Register Annual Report 2014. Available at: http://www.shpr.se/Libraries/Documents/Annual_Report_2014_Eng.sflb.ashx [accessed 16th January 2017]

Getting it right first time (GRIFT) 2020 Available from: https://gettingitrightfirsttime.co.uk/wp-content/uploads/2020/02/GIRFT-orthopaedics-follow-up-report-February-2020.pdf

HEALTH Investigators, Bhandari M, Einhorn TA, Guyatt G, Schemitsch EH, Zura RD, Sprague S, Frihagen F, Guerra-Farfán E, Kleinlugtenbelt YV, Poolman RW, Rangan A, Bzovsky S, Heels-Ansdell D, Thabane L, Walter SD, Devereaux PJ. Total Hip Arthroplasty or Hemiarthroplasty for Hip Fracture. N Engl J Med. 2019 Dec 5;381(23):2199-2208. doi: 10.1056/NEJMoa1906190. Epub 2019 Sep 26. PMID: 31557429.

Janssen, M.F., Pickard, A.S. & Shaw, J.W. General population normative data for the EQ-5D-3L in the five largest European economies. Eur J Health Econ 22, 1467–1475 (2021). https://doi.org/10.1007/s10198-021-01326-9

Lewis SR; Macey R; Parker MJ; Cook JA; Griffin XL; Arthroplasties for hip fracture in adults.; The Cochrane database of systematic reviews 2022, Issue 2. Art. No.: CD013410. DOI: 10.1002/14651858.CD013410.pub2.National Institute for Health and Care Excellence (NICE). Developing NICE guidelines: the manual. 2018. Available from: www.nice.org.uk/process/pmg20.

NHS Improvement (2019) National schedule of reference costs 2019-20. Accessed at: https://www.england.nhs.uk/national-cost-collection/#ncc1920

Office for National Statistics (2020) National life tables: UK. Accessed at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables

Parsons N, Griffin XL, Achten J, Costa ML. Outcome assessment after hip fracture: is EQ-5D the answer? Bone Joint Res. 2014 Mar 19;3(3):69-75. doi: 10.1302/2046-3758.33.2000250. PMID: 24648420; PMCID: PMC3963508.

Appendix K – Excluded studies

Clinical evidence:

Study	Reason for exclusion
Blankstein, Michael, Schemitsch, Emil H, Bzovsky, Sofia et al. (2020) What Factors Increase Revision Surgery Risk When Treating Displaced Femoral Neck Fractures With Arthroplasty: A Secondary Analysis of the HEALTH Trial. Journal of orthopaedic trauma 34suppl3: 49-s54	- Full text paper not available *Secondary analysis of included primary study*

Study	Reason for exclusion
Chammout, G, Kelly-Pettersson, P, Hedbeck, C-J et al. (2019) HOPE-Trial: Hemiarthroplasty compared with total hip arthroplasty for displaced femoral neck fractures in octogenarians: a randomized controlled trial. Journal of Bone & Joint Surgery - American Volume 4(2e0059): 1-9	- Duplicate reference
Comeau-Gauthier, Marianne, Zura, Robert D, Bzovsky, Sofia et al. (2021) Heterotopic Ossification Following Arthroplasty for Femoral Neck Fracture. The Journal of bone and joint surgery. American volume 103(14): 1328-1334	- Not a relevant study design *Secondary analysis of included primary study*
DeAngelis, Ryan D, Minutillo, Gregory T, Stein, Matthew K et al. (2020) Who Did the Arthroplasty? Hip Fracture Surgery Reoperation Rates are Not Affected by Type of Training-An Analysis of the HEALTH Database. Journal of orthopaedic trauma 34suppl3: 64-s69	- Not a relevant study design *Secondary analysis of primary paper*
Hopley, C., Stengel, D., Ekkernkamp, A. et al. (2010) Primary total hip arthroplasty versus hemiarthroplasty for displaced intracapsular hip fractures in older patients: systematic review. BMJ (Clinical research ed.) 340: c2332	- More recent systematic review included that covers the same topic
Judge, Andrew, Metcalfe, David, Whitehouse, Michael R et al. (2020) Total hip arthroplasty versus hemiarthroplasty for intracapsular hip fracture. The bone & joint journal 102b(6): 658-660	- Not a relevant study design *Summary text of included primary study*
Ma, Hsuan-Hsiao, Chou, Te-Feng Arthur, Pai, Fu-Yuan et al. (2021) Outcomes of dual-mobility total hip arthroplasty versus bipolar hemiarthroplasty for patients with femoral neck fractures: a systematic review and meta-analysis. Journal of orthopaedic surgery and research 16(1): 152	- Systematic review used as source of primary studies
Macaulay, W, Nellans, K, Garvin, K et al. (2006) Prospective randomized clinical trial comparing hemiarthroplasty to total hip arthroplasty: functional outcomes in the treatment of displaced femoral neck fractures. Journal of arthroplasty 17: S238-9	- Duplicate reference
Migliorini, F., Maffulli, N., Trivellas, M. et al. (2022) Total hip arthroplasty compared to bipolar and unipolar hemiarthroplasty for displaced hip fractures in the elderly: a Bayesian	- Systematic review used as source of primary studies

Study	Reason for exclusion
network meta-analysis. European journal of trauma and emergency surgery : official publication of the European Trauma Society	
Parker, MJ and Cawley, S (2019) Treatment of the displaced intracapsular fracture for the 'fitter' elderly patients: A randomised trial of total hip arthroplasty versus hemiarthroplasty for 105 patients. Injury 50(11): 2009-2013	- Duplicate reference
Peng, Lin, Liu, Hongyu, Hu, Xiaoyi et al. (2020) Hemiarthroplasty versus total hip arthroplasty for displaced femoral neck fracture in patients older than 80 years: A randomized trial protocol. Medicine 99(50): e23530	- Not a relevant study design *Trial protocol only*
Peng, Wei, Bi, Na, Zheng, Jun et al. (2020) Does total hip arthroplasty provide better outcomes than hemiarthroplasty for the femoral neck fracture? A systematic review and meta-analysis. Chinese journal of traumatology = Zhonghua chuang shang za zhi 23(6): 356-362	- Systematic review used as source of primary studies
Sharma, V, Awasthi, B, Kumar, K et al. (2016) Outcome analysis of hemiarthroplasty vs. total hip replacement in displaced femoral neck fractures in the elderly. Journal of Clinical and Diagnostic Research 10(5): RC11-3	- Duplicate reference
Stengel, D; Mutschler, W; Renkawitz, T (2020) Surgical treatment of displaced hip fractures by total hip arthroplasty or hemiarthroplasty : Results of the multicentric international HEALTH trial. Der Unfallchirurg 123(8): 665-668	- Study not reported in English
Tang, Xiumei, Wang, Duan, Liu, Ying et al. (2020) The comparison between total hip arthroplasty and hemiarthroplasty in patients with femoral neck fractures: a systematic review and meta-analysis based on 25 randomized controlled trials. Journal of orthopaedic surgery and research 15(1): 596	- Systematic review used as source of primary studies
Tol, MC, van den Bekerom, MP, Sierevelt, IN et al. (2017) Hemiarthroplasty or total hip arthroplasty for the treatment of a displaced intracapsular fracture in active elderly patients: 12-year follow-up of randomised trial. The bone & joint journal: 250-254	- Secondary publication of an included study that does not provide any additional relevant information

1

2 Economic Studies:

Study	Code [Reason]
Gao, L.; Han, Z.; Xiong, A. (2020) Total hip arthroplasty or hemiarthroplasty for hip fracture. New England Journal of Medicine 382(11): 1072-1074	- Not a relevant study design *Does not contain costs, is not a cost utility study*

1

Appendix L - Research recommendation

What is the long-term clinical and cost-effectiveness for adults undergoing total hip replacement compared with hemiarthroplasty for displaced intracapsular hip fracture?

L.1 Why this is important

Evidence comparing total hip arthroplasty and hemiarthroplasty has focused mainly on short-term outcomes. Little is known about the long-term outcomes of each type of arthroplasty or whether some population groups will benefit more from a particular type of arthroplasty. Data on long-term effectiveness and cost-effectiveness of each type of arthroplasty is therefore needed to understand their relative benefits for a fragility fracture population and subgroup populations within that.

L.2 Rationale for research recommendation

Importance to 'patients' or the population	There is currently limited long-term evidence on outcomes for total hip arthroplasty and hemiarthroplasty in a fragility fracture population. By having a greater understanding of which types of arthroplasty have the best long-term outcomes, and how this varies for different subpopulations, people will be able to benefit from being given the most effective surgical option.
Relevance to NICE guidance	Total hip arthroplasty is currently recommended for people who are expected to have good functional long-term outcomes after surgery. However, there is limited understanding of whether there are specific population groups who would benefit the most from either total hip arthroplasty or hemiarthroplasty. Future research will help develop a more detailed understanding of the long-term benefits, harms and cost-effectiveness of both types of arthroplasty. This will enable future recommendations to be more specific about who should be offered total hip arthroplasty or hemiarthroplasty.
Relevance to the NHS	The outcomes of this research will help people to receive the most appropriate type of arthroplasty and avoid unnecessary complications. Reducing complications or long-term adverse events will also help to reduce costs to the NHS.
National priorities	Moderate
Current evidence base	Short-term data from 20 RCTs (5 UK-based RCTs), mostly with low or very low-quality outcomes
Equality considerations	There is currently limited knowledge about whether people from different population groups may benefit more from total hip arthroplasty or hemiarthroplasty.

L.1.3 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, such as different age groups and ethnic backgrounds.
Intervention	Total hip arthroplasty
Comparator	Hemiarthroplasty
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	RCT studies with follow up periods >2 years
Timeframe	Long term (>2 years)
Additional information	People from different population groups may have different long-term outcomes depending on whether total hip arthroplasty or hemiarthroplasty is used. It is important for research to provide subgroup analysis in these populations.

2

3

Appendix M– Methods

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

M.1.131 Incorporating published evidence syntheses

For all review questions where a literature search was undertaken looking for a particular study design, published evidence syntheses (quantitative systematic reviews or qualitative evidence syntheses) containing studies of that design were also included. All included studies from those syntheses were screened to identify any additional relevant primary studies not found as part of the initial search. Evidence syntheses that were used solely as a source of primary studies were not formally included in the evidence review (as they did not provide additional data) and were not quality assessed.

If published evidence syntheses were identified sufficiently early in the review process (for example, from the surveillance review or early in the database search), they were considered for use as the primary source of data, rather than extracting information from primary studies. Syntheses considered for inclusion in this way were quality assessed to assess their suitability using the appropriate checklist, as outlined in

Table 2. Note that this quality assessment was solely used to assess the quality of the synthesis in order to decide whether it could be used as a source of data, as outlined in Table 3, not the quality of evidence contained within it, which was assessed in the usual way as outlined in the section on 'Appraising the quality of evidence'.

Table 2: Checklists for published evidence syntheses

Type of synthesis	Checklist for quality appraisal
Systematic review of quantitative evidence	ROBIS
Network meta-analysis	Modified version of the PRISMA NMA tool (see appendix K of 'Developing NICE guidelines, the manual')
Qualitative evidence synthesis	ENTREQ reporting standard for published evidence synthesis (https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-181) is the generic reporting standard for QES, however specific reporting standards exist for meta-ethnography (eMERGe [https://emergeproject.org/]) and for realist synthesis (RAMESES II [https://www.ramesesproject.org/]). If these reporting standards are not appropriate to the QES then an adapted PRISMA framework is used (see Flemming K, Booth A, Hannes K, Cargo M, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses. Journal of Clinical Epidemiology 2018; 97: 79-85).
Individual patient data meta-analysis	Checklist based on Tierney, Jayne F., et al. "Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use." PLoS Med 12.7 (2015): e1001855.

Each published evidence synthesis was classified into one of the following three groups:

- High quality – It is unlikely that additional relevant and important data would be identified from primary studies compared to that reported in the review, and unlikely that any relevant and important studies have been missed by the review.

- Moderate quality – It is possible that additional relevant and important data would be identified from primary studies compared to that reported in the review, but unlikely that any relevant and important studies have been missed by the review.
- Low quality – It is possible that relevant and important studies have been missed by the review.

Each published evidence synthesis was also classified into one of three groups for its applicability as a source of data, based on how closely the review matches the specified review protocol in the guideline. Studies were rated as follows:

- Fully applicable – The identified review fully covers the review protocol in the guideline.
- Partially applicable – The identified review fully covers a discrete subsection of the review protocol in the guideline (for example, some of the factors in the protocol only).
- Not applicable – The identified review, despite including studies relevant to the review question, does not fully cover any discrete subsection of the review protocol in the guideline.

The way that a published evidence synthesis was used in the evidence review depended on its quality and applicability, as defined in Table 3. When published evidence syntheses were used as a source of primary data, data from these evidence syntheses were quality assessed and presented in GRADE/CERQual tables in the same way as if data had been extracted from primary studies. In questions where data was extracted from both systematic reviews and primary studies, these were checked to ensure none of the data had been double counted through this process.

Table 3: Criteria for using published evidence syntheses as a source of data

Quality	Applicability	Use of published evidence synthesis
High	Fully applicable	Data from the published evidence synthesis were used instead of undertaking a new literature search or data analysis. Searches were only done to cover the period of time since the search date of the review. If the review was considered up to date (following discussion with the guideline committee and NICE lead for quality assurance), no additional search was conducted.
High	Partially applicable	Data from the published evidence synthesis were used instead of undertaking a new literature search and data analysis for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. If the review was considered up to date (following discussion with the guideline committee and NICE lead for quality assurance), no additional search was conducted. For other sections not covered by the evidence synthesis, searches were undertaken as normal.
Moderate	Fully applicable	Details of included studies were used instead of undertaking a new literature search. Full-text papers of included studies were still retrieved for the purposes of data analysis. Searches were only done to cover the period of time since the search date of the review.
Moderate	Partially applicable	Details of included studies were used instead of undertaking a new literature search for the relevant subsection of the protocol. For this section, searches were only done to cover the period of time since the search date of the review. For other sections not covered by the evidence synthesis, searches were undertaken as normal.

M.1.122 Pairwise meta-analysis

Pairwise meta-analyses were performed in Cochrane Review Manager V5.4. A pooled relative risk was calculated for dichotomous outcomes (using the Mantel–Haenszel method) reporting numbers of people having an event.

A pooled mean difference was calculated for continuous outcomes (using the inverse variance method) when the same scale was used to measure an outcome across different studies.

For continuous outcomes analysed as mean differences, change from baseline values were used in the meta-analysis if they were accompanied by a measure of spread (for example standard deviation). Where change from baseline (accompanied by a measure of spread) were not reported, the corresponding values at the timepoint of interest were used.

Random effects models were fitted when there was significant between-study heterogeneity in methodology, population, intervention or comparator was identified by the reviewer in advance of data analysis. This decision was made and recorded before any data analysis was undertaken. For all other syntheses, fixed- and random-effects models were fitted, with the presented analysis dependent on the degree of heterogeneity in the assembled evidence. Fixed-effects models were the preferred choice to report, but in situations where the assumption of a shared mean for fixed-effects model were clearly not met, even after appropriate pre-specified subgroup analyses were conducted, random-effects results are presented. Fixed-effects models were deemed to be inappropriate if there was significant statistical heterogeneity in the meta-analysis, defined as $I^2 \geq 50\%$.

However, in cases where the results from individual pre-specified subgroup analyses were less heterogeneous (with $I^2 < 50\%$) the results from these subgroups were reported using fixed effects models. This may have led to situations where pooled results were reported from random-effects models and subgroup results were reported from fixed-effects models.

M.1.133 Intervention studies (relative effect estimates)

RCTs and quasi-randomised controlled trials were quality assessed using the Cochrane Risk of Bias 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.

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.

M.1.154 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.

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.

M.1.185 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 randomised controlled trials, non-randomised controlled trials and cohort studies (which were quality assessed using the Cochrane risk of bias tool or ROBINS-I) were initially rated as high quality while data from other study types were initially rated as low quality. The quality of the evidence for each outcome was downgraded or not from this initial point, based on the criteria given in Table 4.

Table 4: 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.

GRADE criteria	Reasons for downgrading quality
	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.
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.

1