CADTH Health Technology Review

Robotic-Assisted Surgical Systems for Hip Arthroplasty

Rapid Review

Table of Contents

List of Tables

List of Figures

Abbreviations

CI	confidence interval
cTHA	conventional total hip arthroplasty
DDH	developmental dysplasia of the hip
FJS	Forgotten Joint Score
HA	hip arthroplasty
HHS	Harris Hip Score
MDA	Merle d'Aubigné
mTHA	manual total hip arthroplasty
NRS	non-randomized study
OR	odds ratio
RCT	randomized controlled trial
rHA	robotic-assisted hip arthroplasty
rTHA	robotic-assisted total hip arthroplasty
SD	standard deviation
SR	systematic review
THA	total hip arthroplasty
UCLA	University of California Los Angeles
WOMAC	Western Ontario and McMaster University's Osteoarthritis Index

Key Messages

- Studies describing the clinical effectiveness of robotic-assisted surgical systems for hip arthroplasty reported variable results, with some findings indicating a benefit of robotic-assisted surgical systems, few findings indicating a benefit of conventional or manual surgical procedures, and most findings describing no difference between interventions.
- Complications of robotic-assisted surgical systems for hip arthroplasty as compared to conventional or manual surgical procedures were generally found to have few differences found between treatment groups.
- Cost-effectiveness evidence describing robotic-assisted surgical systems for hip arthroplasty was scarce, with 1 study identified by this review that bore limited relevance to the Canadian context.
- The quality of currently available evidence describing the clinical effectiveness of robotic-assisted surgical systems for hip arthroplasty is low, indicating the importance of more rigorous research (including randomized controlled trials) addressing this topic.
- Cost-effectiveness evidence relevant to the Canadian and/or public-payer context is needed.

Context and Policy Issues

For patients with osteoarthritis and other conditions affecting the hip joint, the impacts on quality of life and overall health can be significant, from symptoms including pain, limited mobility, and deleterious effects on the activities of daily living, to negative effects on mental health.[1] There are several treatments available to address these conditions and their symptoms, including pharmacological and non-pharmacological therapies, which are minimally invasive.[2] However, when these minimally invasive treatments fail, surgical interventions are also available; these include hemiarthroplasty, hip resurfacing, hip osteotomy, and total hip arthroplasty (THA; also known colloquially as hip replacement).[2]

Developed in the mid-20th century, THA has been characterized as one of the most successful technological advances in the history of modern medicine.[3] In Canada, there were more than 63,000 hip replacement procedures conducted between 2019 and 2020, representing a 2.4% increase from the previous year,[4] and an even larger increase than noted in a report from 2015, when the number of hip replacement procedures per annum in Canada was estimated to be closer to 24,000.[5] Nonetheless, the procedure is not without risk,[6] with some of the most common complications among Canadians including infection, aseptic loosening, and aseptic instability, all of which may necessitate revision surgery.[4] The impact of risks and complications can be significant for both patients and health care systems; patients who experience complications may have pain and discomfort and encounter the risk of undergoing revision surgery,[1] while the inpatient costs of primary THA in Canada are estimated at more than $10,000, with costs of revision surgery exceeding $16,000 (excluding postoperative costs, such as rehabilitation and other expenses).[4]

While there have been few developments in the field of hip arthroplasty (HA) over the past several decades, there have been significant improvements in metallurgy and components that proffer important benefits to patients undergoing the procedure.[7] More recently, the advent of robotic-assisted technology has been investigated in the context of HA.[8] Whereas

robotic-assisted procedures for indications other than joint arthroplasty have been in use for 20 years or more, robotic-assisted HA and THA (rHA and rTHA) are relatively new. Nevertheless, rHA and rTHA have been described as gaining rapid acceptance among surgical providers,[9] with a growing sense of acceptance in the scientific community as a valid intervention.[10] This upsurge in interest and acceptance puts pressure on health care decision-makers, administrators, and policy-makers to consider the adoption and implementation of these technologies. However, barriers to these technologies have been described as a limited base of robust evidence and challenges with effective implementation.[11-13]

Importantly, uncertainty about the clinical effectiveness of rHA has been highlighted in the literature (i.e., whereas there is some evidence showing a benefit of rTHA in terms of the accuracy of joint placement, improvements in leg-length discrepancy, and other non-clinical parameters, it remains unclear whether these advantages translate into clinical benefits for patients).[14,15] Clinical evidence to-date has been characterized as consisting of low-quality, non-randomized studies and therefore lacking clarity as to the clinical and cost-effectiveness of rHA for patients requiring HA,[8,10,16] precipitating calls in the literature for high-quality randomized controlled trials (RCTs).[15] As jurisdictions around the world consider the value of incorporating robotic technology into hospitals and the surgeries they perform, an emphasis on the importance of high-quality evidence describing clinical and cost-effectiveness and patient-relevant outcomes is clear.[11] Simultaneously, the importance of a robust implementation plan is critical to ensure that the new technology is adopted and integrated in a way that maximizes any of its potential clinical or cost-effectiveness benefits.[12,13]

Given the uncertainty concerning the clinical and cost-effectiveness of rHA,[8] alongside the potential for risk and/or harm to patients,[16] there is a need to summarize and assess the strengths and limitations of the currently available evidence on the topic to inform Canadian health care providers and decision-makers as to the value, benefits, and harms of rHA. Therefore, this review aims to summarize the currently available evidence describing the clinical and cost-effectiveness of robotic-assisted surgical systems for patients requiring HA.

Research Questions

1. What is the clinical effectiveness of robotic-assisted surgical systems used for patients requiring hip arthroplasty?
2. What is the cost-effectiveness of robotic-assisted surgical systems used for patients requiring hip arthroplasty?

Methods

Literature Search Methods

A limited literature search was conducted by an information specialist on key resources, including MEDLINE, the Cochrane Database of Systematic Reviews, the International HTA Database, the websites of Canadian and major international health technology agencies, as well as a focused internet search. The search strategy comprised both controlled vocabulary,

such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were robotic-assisted surgical systems and HA. No filters were applied to limit the retrieval by study type. Comments, newspaper articles, editorials, and letters were excluded. Where possible, retrieval was limited to the human population. The search was also limited to English-language documents published between January 1, 2017, and January 21, 2022.

Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1, were duplicate publications, or were published before 2017. Systematic reviews (SRs) in which all relevant studies were captured in other more recent or more comprehensive SRs were excluded. Primary studies retrieved by the search were excluded if they were captured in 1 or more included SRs.

Critical Appraisal of Individual Studies

The included publications were critically appraised by 1 reviewer using the following tools as a guide: A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2)[17] for SRs, the Downs and Black checklist[18] for randomized and non-randomized studies, and the Drummond checklist[19] for economic evaluations. Summary scores were not calculated for the included studies; rather, the strengths and limitations of each included publication were described narratively.

Table 1: Selection Criteria

Criteria	Description
Population	Patients (any age) requiring total or partial (i.e., hemiarthroplasty) hip arthroplasty due to any cause (e.g., osteoarthritis, rheumatoid arthritis, psoriatic arthritis, trauma, dysplasia)
Intervention	Arthroplasty performed using robotic-assisted surgical systems (e.g., Zimmer Biomet's ROSA, Stryker's Mako)
Comparator	Conventional or traditional surgery performed without robotic-assisted surgical systems
Outcomes	Q1: Clinical effectiveness (e.g., postoperative functional status and mobility, range of motion, pain reduction, length of hospital stay, health-related quality of life, patient satisfaction, blood loss, operative time) and safety (e.g., rates of adverse events, hospital readmission, surgical revisions, complications [e.g., infections, femoral fractures]) Q2: Cost-effectiveness (e.g., cost per quality-adjusted life-year, cost per health benefit gained)
Study designs	Health technology assessments, systematic reviews, randomized controlled trials, non-randomized studies, economic evaluations

Summary of Evidence

Quantity of Research Available

A total of 222 citations were identified in the literature search. Following screening of titles and abstracts, 187 citations were excluded and 35 potentially relevant reports from the electronic search were retrieved for full-text review. Three potentially relevant publications were retrieved from the grey literature search for full-text review. Of these potentially relevant articles, 30 publications were excluded for various reasons, and 8 publications met the inclusion criteria and were included in this report. These comprised 3 SRs, 4 non-randomized studies (NRSs), and 1 economic evaluation. Appendix 1 presents the PRISMA[20] flow chart of the study selection.

Summary of Study Characteristics

Study Design

The 3 SRs included in this review were published in 2021[21,22] and 2019,[23] and included literature searches that ranged from as early as 1994 to as recent as 2021.[21-23] The SRs identified and summarized a preponderance of NRS,[21-23] but 2 SRs identified and summarized RCTs, as well.[22,23] Most of the primary studies included in the 3 SRs were relevant to the research questions posed by this review; however, 2 of the SRs were broader in their scope of eligible outcomes as compared to the current review (i.e., 1 SR contained 11 of 17 included studies that were eligible for inclusion in this review[21]; another SR contained 18 studies, all of which were eligible for summary in this review[22]; and the third SR contained 11 of 14 studies that were eligible for and summarized in this report[23]). Methods for the synthesis of included studies were both meta-analyses and narrative syntheses for all 3 SRs.[21-23] There was considerable overlap in the primary studies included across the 3 SRs identified in this review, which is detailed in Appendix 5.

There were 4 NRSs included in this review, all of which were published in 2021.[24-27] All 4 studies reported the use of a retrospective, cohort design using matching on a variety of patient characteristics and other variables (e.g., year of surgery) to pair patients who received the intervention of interest with those receiving the control intervention.[24-27] Two of the retrospective matched cohort studies drew data from the same large, regional inpatient discharge abstracts database;[24,25] another study drew data from a large, longitudinal, national health care claims database,[26] whereas the fourth study used data from a single centre.[27]

One eligible economic evaluation was identified in this review, published in 2021 and designed using a payer's perspective.[28] The study employed a cost-effectiveness analysis using a Markov model and spanning a 5-year time horizon.[28] Model inputs included clinical effectiveness and utility data from literature sources, as well as primary data from a single-centre prospective cohort, and cost data from a Medicare source.[28] The possible patient outcomes in the model following the primary surgical procedure were infection, dislocation, revision surgery (as well as re-revision in some cases) or no major complications.[28]

Country of Origin

All authors of 1 SR were affiliated with a single site in the UK;[21] another SR listed all authors from a single site in the US;[22] and all but 1 author were listed as being based in China in the third SR.[23] One of the retrospective matched cohort studies was conducted in China,[27] and the

other 3 retrospective matched cohort studies,[24-26] as well as the economic evaluation,[28] were conducted in the US.

Patient Population

One SR sought and included patients using a similar criterion as was applied to the population in the current review (i.e., patients undergoing THA regardless of indication or condition).[22] Another SR sought and included patients with arthritis of the hip undergoing THA,[21] and the other SR sought patients with a variety of severe hip conditions undergoing THA[23] — rendering the populations from all 3 SRs relevant to this report.

Eligible patient populations identified in the retrospective matched cohort studies included in this review were patients who underwent primary THA between 2008 and 2018 in 1 study,[24] and between 2010 and 2014 in another study,[25] using data from the same discharge abstracts database. Similarly, another retrospective matched cohort study examined records from a different database describing patients who underwent THA between 2010 and 2018.[26] The fourth retrospective matched cohort study examined records from patients with developmental dysplasia of the hip (DDH) who underwent primary THA at a single centre between 2019 and 2020.[27] Because matching algorithms were used for all 4 retrospective matched cohort studies, patient characteristics and other variables (e.g., year of surgery) were similar between treatment groups in all of the retrospective matched cohort studies summarized in this review.[24-27] However, differences in patient characteristics between these studies were observed; for instance, whereas the 3 studies examining patients with any or multiple conditions tended toward older groups of patients (e.g., mean ages of 64 years in 2 studies[25,26] and more than 88% of patients ≥ 50 years of age in another[24]), the study examining patients with DDH constituted a younger group of patients with a mean age of 50 years.[27] Likewise, a larger proportion of patients who were female were represented in the study of patients with DDH (i.e., 75% female),[27] as compared to the 3 retrospective matched cohort studies that examined patients with a variety of conditions (i.e., ranging from 55.1% to 58.3% female).[24-26]

The cohort study informing clinical inputs for the Markov model used in the economic evaluation examined patients undergoing primary THA in the intervention group between 2011 and 2017.[28] The group of 555 patients in the intervention group was 53.7% female and had a mean age of 59.2 years (standard deviation [SD] = 9.7 years).[28] Data for the comparison group were reported as having been drawn from literature sources and were not further described.[28]

Interventions and Comparators

All 8 publications included in this report examined rTHA as the intervention of interest, compared against manual (also known as conventional) THA (mTHA or cTHA).[21-28] One of the retrospective matched cohort studies also examined a comparison group of patients treated with computer-assisted THA, with no description provided,[24] but data for these patients were ineligible for this review and therefore not included or summarized in this report. Five of the 8 reports described the brand name of the rTHA system(s) used in the procedure[21-23,27,28]; these included Mako,[21,27,28] ROBODOC,[22,23] Orthodoc,[23] and CASPAR.[23]

Outcomes

A broad range of outcomes and measures describing clinical effectiveness were used across the 7 publications included in this report[21-27]; only those eligible for inclusion in this review are described herein.

Functional outcomes were reported in all 3 of the SRs and described several measures, none of which were described in detail as to which features of functionality were measured, or how the measures were scored or interpreted.[21-23] Therefore, the direction of effect and interpretation of functional outcomes were necessarily inferred in this review, where possible, by examining forest plots or associating qualitative descriptions in the reports of findings (e.g., significantly improved) with quantitative scores reported numerically. For some measures, there was insufficient information reported to interpret the finding; data from these measures were neither included nor summarized in this report. Among the measures with sufficient information for interpreting the findings, several were reported in more than 1 SR: the Forgotten Joint Score (FJS), which renders a lower score for a more favourable outcome;[21,22] the Harris Hip Score (HHS), which renders a higher score for a more favourable outcome;[21-23] the Merle d'Aubigné (MDA) Hip Score, which renders a lower score for a more favourable outcome;[22,23] the University of California Los Angeles (UCLA) Physical Activity Scale, which renders a higher score for a more favourable outcome,[21,22] and the Western Ontario and McMaster University's Osteoarthritis Index (WOMAC), which renders a lower score for a more favourable outcome.[21-23] Other measures of functionality were ambulation, measured using walking speed and distance in miles per hour and metres, as well as gait asymmetry and range of motion, measured in degrees.[22] Follow-up timing was often not reported; where it was, it ranged between 3 months and 14 years.[21-23]

Two SRs reported on measures of quality of life, using Short Form surveys (SF-12 and SF-36), with follow-up timing ranging between 10 months and 5 years,[21,22] and 2 EQ-5D quality of life scales at 10 to 12 months of follow-up.[22] Two SRs described pain using either the number of patients experiencing postoperative knee or thigh pain (with follow-up timing not reported),[23] or various measures (with follow-up timing ranging from < 3 months to 14 years),[22] neither of which were described in terms of scoring or interpretation of findings. One SR and all 4 retrospective matched cohort studies reported on measures of health care utilization,[23-27] including duration of the surgical procedure, using an unreported measurement of time in 1 SR[23] and minutes in 1 retrospective matched cohort study,[27] hospital length of stay (in days),[24-26] discharge disposition (i.e., routine or to a care setting),[24,26] and the hospital readmission rates for patients at 90 days and 1 year of follow-up.[26] Two retrospective matched cohort studies described the number of patients who died; follow-up timing for this outcome was not clearly reported.[24,25] One SR described findings from several additional measures that were reported only as "patient-reported outcomes" (Table 5 in the SR[22]) and provided little to no information concerning the features of the outcomes or their scoring (i.e., the statistical significance of the difference between groups was characterized by P values, but not otherwise interpreted as to relative benefit, lack thereof, or harm between the groups).[22] Where the information was insufficient for interpreting the findings, data were neither included nor summarized in this report.

Clinical effectiveness was also characterized in terms of the comparative safety of rTHA versus cTHA — described as complications by number of patients or number of events in all 7 publications.[21-27] These complications included descriptions of any and/or all complications observed in a study (follow-up timing either not reported or ranging from 90 days to a minimum of 5 years);[21,23-25] intraoperative complications;[21,23] cardiovascular or blood complications reported either at 90 days or an unspecified time frame;[24,26,27] dislocations at either a range of 90 days to 14 years of follow-up, or an unreported time frame;[21-24,26,27] fractures at either a range of 90 days to 14 years of follow-up, or an unreported time frame;[22,24,26,27] heterotopic ossification at either a range of 24 to 67 months of follow-up, or an unreported time frame;[22,23] infections at either a range of 90 days to 5 years of follow-up, or an

unreported time frame;[21,22,24,26] mechanical complications (no follow-up time frame provided);[24] nerve injury (no follow-up timing reported); and[22,23] revision surgery at either a range of 90 days to 14 years of follow-up, or an unreported time frame.[21-23,26]

Cost-effectiveness outcomes of relevance to this report as described in the 1 eligible and included economic evaluation were quality-adjusted life-years and incremental cost-effectiveness ratio.[28]

Additional details regarding the characteristics of included publications are provided in Appendix 2.

Summary of Critical Appraisal

Systematic Reviews

The 3 SRs demonstrated both strengths and limitations. All 3 papers clearly reported inclusion criteria for the review and described the use of duplicate study selection.[21-23] Two of the SRs described risk of bias assessments, provided a clear description of an appropriately designed meta-analysis, and identified the source(s) of funding supporting the work on the SR,[21,23] whereas the third SR did not describe these important features.[22] In particular, an assessment of risk of bias has been described as a "crucial" feature of the SR method,[29] because it informs the reader as to the confidence that can be placed in the contribution of individual studies to the findings of the SR. Similarly, a description of the methods used for the meta-analysis is important for ensuring transparency around how the quantitative synthesis was performed to ensure methodological rigour and reproducibility. Further, acknowledgement of funding is important for all research studies because it helps the reader to assess the extent to which any potential external sources of bias may have impacted the methods and findings. The impact of these missing features from the 1 SR included in this review is a reduced level of confidence in the results reported.[22] Other strengths included a protocol registered with the PROSPERO database for 1 of the SRs,[21] and an assessment of publication bias performed for another SR.[23] Both of these features are important in the SR methodological approach: first, to ensure that the methods are carried out as planned and not subject to any post hoc changes based on the findings as the review progresses, and second, to assure the readers of the SR that publication bias has not unduly affected the findings of the review (e.g., by potentially inflating the estimates of effectiveness). Whereas 1 of the SRs did not report either of these features,[22] the others failed to report on either the development of a protocol,[23] or an assessment of publication bias.[21]

An important limitation demonstrated by all 3 SRs was a lack of details describing some features of the included studies (i.e., whereas an adequate description of some features of the included studies was provided in a readable, tabular format, other important details were notably missing).[21-23] For instance, 1 SR failed to report follow-up time for any of the included studies, preventing the reader from considering the extent to which the outcomes reported were affected by the variable of time.[23] Importantly, none of the SRs described in detail the particular features of the outcomes that were measured, or the scoring mechanisms for outcome measures summarized in their reports, making it a challenge to interpret the findings.[21-23] Other limitations of the included SRs were search strategies that were missing information; excluded studies that were not listed and did not include rationales for their ineligibility reported; lack of clarity around whether the meta-analyses accounted for the various lengths of patient follow-up; and sources of funding for included studies that were not reported.[21-23] While these limitations reduce the confidence that can be placed in the validity

and reliability of the SRs, they must also be considered alongside the strengths demonstrated to establish an informed assessment of the contribution these SRs make toward answering the research questions posed by this report.

Non-Randomized Studies

The 4 retrospective matched cohort studies included in this report also demonstrated both strengths and limitations. All 4 studies clearly reported the study aim, objectives, patient characteristics, potential confounders, and estimates of random variability.[24-27] Adverse events (often referred to as complications) were also described in all 4 of the papers,[24-27] although 1 study combined complications into 2 categories (i.e., minor complications and major complications), which reduced the detail and clarity describing adverse events.[25] Three of the studies relied on large regional patient databases, which contributes to the confidence in the studies' external validity.[24-26] However, while 2 of the studies used all of the available and eligible patient data,[24,27] the other 2 used matching algorithms that produced subsamples of patients,[25,26] potentially reducing the extent to which the findings may be applicable to the broader population. All 4 studies used matching techniques to pair study patients bearing similar characteristics,[24-27] which can mitigate the extent to which confounding poses a risk to the internal validity of the studies. In addition, the findings were generated from planned analyses, with adjustments made to account for potentially confounding factors — including an algorithm to account for patients who underwent surgery at different points in time.[24-27] Further, the use of standardized data available in large clinical or health administrative data can contribute to consistency across outcome measures; however, because the data were not collected for research purposes, the level of detail available, and relevance to the research question(s), may be limited[30] and can affect the extent to which authors are able to generate meaningful insights in answer to the research question(s) their study pose(s). One of the most important limitations was a lack of randomization of patients to treatment groups, by virtue of the studies' designs being observational; further, patients, clinicians, and those assessing outcome data were not blinded to the interventions received,[24-27] which might have affected the internal validity of the findings. Because 3 of the studies used very large databases, they had relatively large sample sizes, which contribute to the power of studies to detect statistically significant differences between treatment groups; however, none of the 4 studies explicitly addressed, described, or provided a power calculation.[24-27]

Economic Evaluation

The reporting of the study design was clear in the paper describing the 1 economic evaluation included in this report (i.e., the research aim and its economic importance, the viewpoint of the analysis, and the type of economic evaluation were clearly described).[28] These are important features of economic evaluations that justify the value of undertaking the study and increase relevance in the results that are generated.[31] The methods informing data collection demonstrated more strengths than limitations, with a sufficient description of the sources for effectiveness estimates, primary outcomes, methods to value benefits and estimate costs and quantities, currency and price information, and the model used, as well as a justification for its use.[28] However, an assessment of indirect benefits (or productivity costs) was not reported,[28] which can be a valuable feature of economic evaluations to inform the reader about a broader range of costs relevant to a health intervention, and the potential impact on cost-effectiveness.[31] The analyses and interpretation of the economic evaluation demonstrated several strengths and few limitations (i.e., the study's time horizon and discount rate were clearly reported, sensitivity and incremental analyses were included, and the conclusions clearly addressed the research aim).[28] These features are important to ensure maximum methodological transparency and to support the assessment of an economic

evaluation's relevance and applicability.[31] The few limitations identified included a lack of information describing the comparator group (i.e., details reported for the intervention group only), and confidence intervals that were missing in the reporting of results, which impacts the degree to which the reader can assess uncertainty in the findings.[31]

Additional details regarding the strengths and limitations of the included publications are provided in Appendix 3.

Summary of Findings

Clinical Effectiveness of Robotic-Assisted Surgical Systems for Hip Arthroplasty

Functional Status

Functional status was described in 3 SRs using 5 standardized outcome measures, 4 of which had sufficient data to inform meta-analyses.

Harris Hip Score

Scores from the HHS scale were reported in 3 SRs using meta-analyses — 1 of which combined results from 4 studies representing data from 669 patients across a range of follow-up between 3 months and 5 years;[21] another SR pooling data from 10 studies reporting on 1,330 patients across a range of follow-up from 3 months to 5 years;[22] and another SR combining data across 4 studies describing 328 patients (follow-up timing not reported).[23] Because of the overlap in primary studies across the 3 SRs, a total of 12 individual studies were represented across the 3 meta-analyses. Two of the 3 meta-analyses found no statistically significant difference in HHS scores between patients who underwent rTHA as compared to patients who underwent mTHA (i.e., P = 0.26 and P = 0.98). [22,23] On the other hand, the meta-analysis from 1 SR identified a statistically significant difference in the pooled mean post-operative HHS scores that favoured patients who underwent rTHA as compared to mTHA (i.e., 3.05; 95% confidence interval [CI], 0.46 to 5.64; P = 0.02).[21]

Forgotten Joint Score

Findings generated from the FJS scale by 2 of the included SRs were described using narrative synthesis of 1 primary study in 1 SR[21] and meta-analysis of 4 primary studies in another.[22] Notably, the 1 primary study in the first SR[21] was also included in the second SR and combined in the pooled results reported in the latter.[22] The first SR reported a statistically significant difference in FJS between patients who underwent rTHA as compared to mTHA at a minimum of 5 years of follow-up in the 1 included study with data describing 132 patients; however, the direction of reported effect was not described.[21] However, the meta-analysis of 4 primary studies representing 966 patients across a range of follow-up between 10 months and 5 years from another included SR identified no statistically significant difference in FJS between the treatment groups (i.e., pooled mean difference in post-operative scores of 8.73; 95% CI, −4.79 to 22.25; P = 0.21).[22]

Merle d'Aubigné Hip Score

The 2 SRs presenting meta-analyses of post-operative scores from the MDA scale combined the results from the same 3 primary studies,[22,23] with 1 SR including a fourth primary study in its analysis.[23] Both SRs reported no statistically significant difference between the treatment groups.[22,23] The more recent SR pooled data from 348 patients across 3 studies and a range of follow-up between 3 months and 2 years, producing a pooled mean difference in MDA between groups of 0.24 (95% CI, −0.05 to 0.52), P = 0.10.[22] The 2019 SR combined data from

4 studies including 426 patients (with no follow-up timing reported) and producing a pooled mean difference in MDA scores of 0.06 (95% CI, −0.56 to 0.68; P = 0.86).[23]

University of California Los Angeles Physical Activity Scale

Two SRs reported findings from the UCLA scale with no meta-analyses (i.e., narrative syntheses only); both SRs included 1 common primary study that overlapped between the SRs,[21,22] and 1 SR included a second primary study, as well.[22] The first SR reported data from the 1 common primary study investigating 200 patients and found a statistically significant improvement in the UCLA scale score at 1 year of follow-up in patients who underwent rTHA as compared to those who underwent mTHA (reported narratively only).[21] The other SR identified the same primary study and a second primary study, and reported mixed results: whereas the first primary study of 200 patients overlapped by both SRs was reported to have found a statistically significant benefit of rTHA compared to mTHA (i.e., P = 0.033), the other primary study of 67 patients at 14 years of follow-up found no statistically significant difference between the treatment groups (i.e., P = 0.417).[22]

Western Ontario and McMaster University's Osteoarthritis Index

All 3 SRs reported data from measurement with the WOMAC index;[21-23] 1 SR described findings narratively from 2 primary studies representing 307 patients with a range of follow-up between 1 and 2 years, and no meta-analysis;[21] another SR pooled data from 4 studies representing 505 patients across a range of follow-up from 1 to 14 years;[22] and the third SR combined data from 2 studies representing 116 patients (no follow-up timing reported).[23] Both of the primary studies from the first SR found no statistically significant difference between the treatment groups in post-operative WOMAC scores at 1 year and 2 years of follow-up.[21] On the other hand, meta-analysis from the second SR indicated a statistically significant improvement in mean post-operative WOMAC scores among patients who received rTHA as compared to patients who received mTHA (i.e., −3.57; 95% CI, −5.62 to −1.52; P = 0.0006).[22] Of note, subgroup meta-analyses by type of robotic system indicated that the 2 studies with 389 patients investigating the MAKO system similarly generated a statistically significant benefit for those who received rTHA versus mTHA (i.e., −4.39; 95% CI, −6.86 to −1.93; P = 0.0005), whereas the 2 studies with 116 patients investigating the ROBODOC system showed no statistically significant difference between treatment groups (−1.71; 95% CI, −5.62 to 1.99; P = 0.36).[22] The third SR reported findings from the same 2 studies meta-analyzed in the 2021 SR by Samuel et al. that investigated the ROBODOC system, and generated the same finding of no statistically significant difference between treatment groups (i.e., −1.71; 95% CI, −5.62 to 1.99; P = 0.36).[23]

Ambulation and Range of Motion

Ambulation was described in 1 SR as reported in 2 of its included studies, and no statistically significant difference was reported in patients having undergone rTHA as compared to those having undergone mTHA in terms of walking speed or gait asymmetry (follow-up time not reported), as well as the number of days post-surgery that patients could walk at least 500 m independently.[22] On the other hand, the proportion of patients who could walk at least 6 blocks independently at 13 days post-surgery was statistically significantly higher in the rTHA as compared to the mTHA group (i.e., 53% versus 39%; P < 0.05).[22] The same SR contained 1 study describing range of motion by comparing the surgically treated hip with the contralateral hip in each patient from both treatment groups at 1 year of follow-up, and reporting no statistically significant difference in the frontal and axial plane for either patients treated with rTHA or mTHA, as well as in the net sagittal plane for patients treated

with mTHA; however, patients treated with rTHA had statistically significantly better range of motion in their contralateral hip as compared to their surgically treated hip (P = 0.043).[22]

Overall Health, Mental Health, Physical Health, and Health-Related Quality of Life

Global measures of health and health-related quality of life were reported in 2 SRs included in this review using a variety of measures and scales.[21,22] In 1 primary study of 120 patients from 1 SR, no statistically significant differences were found in post-operative scores at 10 months to 12 months of follow-up between treatment groups in either the EQ-5D questionnaire or the EQ visual analogue scale (i.e., P = 0.562 and P = 0.355, respectively).[22] Both SRs included in this review described data generated from the use of SF surveys in 2 studies[21] and 3 studies,[22] respectively, with narrative synthesis reported. However, due to the overlap across the SRs in the primary studies included, both studies in the paper reported by Ng et al.[21] were also summarized in the paper by Samuel et al.;[22] thus, only the findings for the latter SR[22] are summarized narratively here (details for both studies are tabulated in the appendices of this report). Two of the 3 primary studies in the SR reported by Samuel et al. used SF-12 scales and found no statistically significant difference between the treatment groups in either the global measure at 1 year of follow-up, or the mental health scale at a minimum of 5 years of follow-up.[22] Similarly, the third primary study of 93 patients found no statistically significant difference between the treatment groups at 2 years of follow-up using the SF-36 scale.[22] However, 1 of the 2 primary studies, which included 132 patients, did report a statistically significant benefit for the rTHA group in physical health at 5 years of follow-up (i.e., P = 0.002) as compared to the mTHA group.[22]

Pain

Measures of pain were reported in 2 SRs included in this review using a variety of measures and scales.[22,23] The more recent SR conducted a meta-analysis of post-operative scores from the visual analogue scale including 4 studies representing 477 patients across a range of 10 months to 14 years of follow-up, and reported no statistically significant difference between the treatment groups (i.e., pooled mean difference was −0.19; 95% CI, −0.65 to 0.27; P = 0.41).[22] Both SRs described knee pain and thigh pain narratively from the same 2 primary studies;[22,23] as the earlier publication included data from an additional, third primary study and reported the findings of quantitative meta-analyses,[23] the results from those analyses are preferentially summarized here. Han and colleagues combined findings from 2 studies comprising 264 patients (follow-up timing not reported) and reported no statistically significant finding between patients who underwent rTHA versus mTHA when comparing the odds of experiencing knee pain (i.e., odds ratio [OR] = 1.40; 95% CI, 0.25 to 7.78; P = 0.70).[23] On the other hand, a similar meta-analysis of 3 studies (including the 2 reporting on knee pain, as above) indicated that there were statistically significant findings when comparing the odds of having the occurrence of thigh pain in favour of rTHA versus mTHA (i.e., OR = 0.32; 95% CI, 0.12 to 0.87; P = 0.03); however, it is important to note that follow-up timing was not reported for these studies and outcomes.[23] Nonetheless, another primary study of 67 patients included in the SR by Samuel and colleagues described 2 measures of pain, and found a statistically significant benefit at 14 years of follow-up that favoured mTHA over rTHA when pain was assessed using the Health Status Questionnaire (i.e., postoperative difference in mean scores was 11.10; 95% CI, not reported; P = 0.019) and the modified HHS (i.e., postoperative difference in mean scores was 2.72; 95% CI, not reported; P = 0.025).[22]

Health Care Utilization

Data describing the use of health care was reported in 1 SR and 4 retrospective matched cohort studies.[23-27] The SR included 8 studies representing 1,077 hips (and published between

2003 to 2017) in a meta-analysis to compare the duration of surgical procedures with rTHA as compared to mTHA.[23] Authors reported their findings in mean units of time that were not specified (all other studies in this review describing this outcome used minutes), reporting a statistically significant mean difference in favour of a shorter duration for mTHA procedures compared to the rTHA procedures (i.e., 20.72; 95% CI, 7.50 to 33.94; P = 0.002).[23] Procedure duration was reported in 1 other study in this review: a retrospective matched cohort study of 158 patients followed between 2019 and 2020 reported no statistically significant difference in the mean number of minutes between the treatment groups (i.e., P = 0.602).[27] The length of stay in hospital was measured in days by 3 retrospective matched cohort studies in this review (collectively representing more than 4 million patients followed between 2008 and 2018), which reported a statistically significantly shorter stay among patients who underwent rTHA versus mTHA (i.e., P < 0.001) for all 3 studies.[24-26] The proportion of patients who underwent rTHA or cTHA and experienced hospital readmission was investigated in 1 retrospective matched cohort study following 9,260 patients from 2010 to 2018, which reported no statistically significant difference between the treatment groups at 90 days of follow-up, but a statistically significant difference with more patients in the rTHA group having been readmitted at 1 year of follow-up (i.e., P = 0.037), favouring cTHA.[26] Finally, an assessment of discharge disposition (i.e., discharged to home or a non-home setting) found no statistically significant difference between the treatment groups (i.e., P = 0.205).[24]

Mortality

Two retrospective matched cohort studies included in this review reported on inpatient mortality, comparing the proportion of patients who died between the treatment groups, and neither reported a statistically significant difference between the treatment groups.[25,32]

Safety and Complications

All of the 7 publications included in this review reporting on clinical effectiveness outcomes contained findings describing the comparative safety (referred to as complications in the studies) of rTHA versus mTHA or cTHA in patients.[21-27]

All or Various Complications (Combined)

Two SRs and 2 retrospective matched cohort studies combined complications into broad, non-specific categories, describing the numbers and/or proportions of patients experiencing any complication.[21,23-25] Both SRs presented meta-analyses describing complications; 1 pooled results from 8 studies representing 3,208 patients, which was published between 2016 and 2020,[21] and the other combined data from 6 studies representing 907 patients, which was published between 1998 and 2017.[23] Notably, 1 of the primary studies appeared in both SRs, resulting in a total of 13 individual primary studies represented across the 2 SRs.[21,23] Neither SR identified a statistically significant difference in the odds of patients experiencing any complication (i.e., OR = 0.61; 95% CI, 0.30 to 1.24; P = 0.81;[21] and 0.83; 95% CI, 0.49 to 1.40; P = 0.48).[23] This finding was corroborated by both of the retrospective matched cohort studies, 1 of which combined perioperative complications into 2 categories, describing them as major or minor; meaning no statistically significant difference was found between the adjusted odds of experiencing a major (i.e., OR = 0.732; 95% CI, 0.320 to 1.672) or minor complication (i.e., OR = 1.294; 95% CI, 0.938 to 1.785) between the treatment groups.[25] Similarly, another retrospective matched cohort study identified no statistically significant difference between the treatment groups in the occurrence of non-mechanical complications in their assessment of more than 150,000 procedures from 2008 to 2018 (i.e., P = 0.296).[24]

Cardiovascular and/or Blood Complications

Three retrospective matched cohort studies reported on outcomes relevant to the cardiovascular system or describing conditions of the blood and/or blood loss.[24,26,27] Two retrospective matched cohort studies described the occurrence of pulmonary embolism.[24,26] One of the 2 studies described the pulmonary embolism data in combination with infarction across more than 150,000 procedures occurring between 2008 to 2010 and found a statistically significant difference in favour of the mTHA group compared to the rTHA group (i.e., P = 0.015).[24] However, the other study described pulmonary embolus individually in more than 9,000 patients at 90 days of follow-up between 2010 and 2018 and reported no statistically significant difference between the treatment groups (i.e., P = 0.159).[26]

The same 2 studies also reported on the occurrence of blood transfusions between the treatment groups and both found a statistically significant difference between the treatment groups (i.e., P < 0.001 in both studies).[24,26] While 1 study reported a smaller proportion of blood transfusions in the rTHA group,[24] the other found a smaller proportion among the mTHA group.[26] A third retrospective matched cohort study reported on blood loss in millilitres and found no statistically significant difference between the rTHA and mTHA groups (P = 0.098).[27]

Two of the studies also reported on the occurrence of deep-vein thrombosis; neither of them identified a statistically significant difference between the treatment groups at either the inpatient phase of care,[24] or at 90 days of follow-up.[26] Similarly, Emara and colleagues found no statistically significant differences between the treatment groups in the occurrence of cardiac or peripheral vascular complications, hematoma or seroma, or acute post-hemorrhagic anemia in inpatients.[24] Likewise, Remily et al. also assessed cerebrovascular accidents at 90 days of follow-up and found no statistically significant differences between the rTHA and mTHA groups.[26]

Dislocation

Three SRs and 3 retrospective matched cohort studies reported on the number of dislocations that occurred in the rTHA and mTHA groups.[21-24,26,27] One SR pooled data from across 7 primary studies representing 1,147 patients and ranging from 90 days to 5 years of follow-up (with the exception of 1 primary study for which follow-up was not reported), reported no statistically significant difference in the odds of dislocation between the treatment groups (i.e., OR = 0.66; 95% CI 0.15 to 3.02; P = 0.60).[21] Similarly, Samuel and colleagues conducted a meta-analysis of data describing dislocation, combining findings from 8 studies representing 1,190 patients who were followed between a range of 3 months and 14 years, and reported no statistically significant difference between rTHA and mTHA in the odds of experiencing a dislocation (i.e., OR = 1.81; 95% CI, 0.71 to 4.58; P = 0.11).[22] These findings were corroborated by 1 retrospective matched cohort study, which found no statistically significant difference between the treatment groups in the number of dislocations at 90 days, 1 year, and 5 years of follow-up.[26] Likewise, Zhou and colleagues reported a similar number of dislocations in their assessment of 118 patients with DDH, with 1 in the mTHA group and none in the rTHA group (the difference in terms of statistical significance was not reported).[27]

On the other hand, another SR pooled data describing dislocations from 6 studies representing 899 patients (whose follow-up timing was not reported) found a statistically significant benefit in favour of mTHA over rTHA (i.e., OR = 2.28; 95% CI, 1.12 to 4.67; P = 0.02).[23] Conversely, 1 retrospective matched cohort study examining more than 100,000

procedures across a 10-year time span found statistically significantly fewer dislocations in the rTHA group compared to the mTHA group (i.e., P < 0.001).

Fracture

One SR and 3 retrospective matched cohort studies reported comparative findings describing either intraoperative or periprosthetic fractures in patients who underwent rTHA versus mTHA.[22,24,26,27] Samuel and colleagues reported findings narratively from 6 primary studies, 5 of which described intraoperative fractures; whereas 2 primary studies found statistically significantly more patients in the mTHA group than in the rTHA group who experienced intraoperative fracture, the other 2 primary studies reported no significant difference between the treatment groups (and the fifth primary study did not describe the statistical difference between the groups, reporting only the number of events per group).[22] Of note, across the 5 primary studies reporting on intraoperative fracture included in this SR, all of the patient groups who received rTHA were reported as experiencing 0 intraoperative fractures, whereas all of the patient groups who received mTHA were reported as experiencing 2 or more intraoperative fractures.[22] This pattern was observed again in 1 of the retrospective matched cohort study included in the current review examining 118 patients, which reported 1 intraoperative fracture in the mTHA group and none in the rTHA group (however, the statistical difference between the groups was not described, with only the number of events per group reported).[27]

Periprosthetic fracture was reported in 1 SR and 2 retrospective matched cohort studies.[22,24,26] Findings from 2 primary studies were reported in the SR, with no statistical characterization of the difference between the treatment groups in either study, but 1 periprosthetic fracture was reported in each of the mTHA groups from both studies, and 1 periprosthetic fracture was reported in 1 of the rTHA groups (with no periprosthetic fractures in the other primary study's rTHA group).[22] In the large inpatient database follow-up of THA patients, a statistically significantly greater number of periprosthetic fractures were reported in the mTHA group compared to the rTHA group (i.e., P < 0.001).[24] In another retrospective matched cohort study, periprosthetic fracture was assessed at 90 days, 1 year, and 5 years, in almost 10,000 patients, with the authors reporting no statistically significant difference between patients who underwent rTHA as compared to cTHA.[26]

Heterotopic Ossification

Two SRs reported findings describing heterotopic ossification; the first included results from 2 primary studies[22] and the second reported on a meta-analysis of findings from 4 primary studies representing 416 patients (follow-up timing not reported).[23] One of the primary studies appeared in both SRs, resulting in a total of 5 primary studies represented across the 2 SRs.[22,23] None of the findings summarized from either of the SRs indicated a statistically significant difference in the incidence of heterotopic ossification between patients who underwent rTHA as compared to mTHA.[22,23]

Infection

Outcome data for multiple types of infections were reported in 2 SRs (describing 5 primary studies each) and 2 retrospective matched cohort studies in this review.[22-24,26] In the 2 SRs, 3 of the same primary studies were included in both reviews. Across all of the relevant data identified in this review, none indicated any statistically significant difference between rTHA and mTHA for any type of infection.[22-24,26]

Intraoperative Complications

Two SRs pooled findings from primary studies reporting data on intraoperative complications;[21,23] 1 SR published in 2021 combined data from 2 primary studies representing 386 patients,[21] and another SR published in 2019 combined data from 9 studies representing 1,121 patients (follow-up durations not reported).[23] Whereas the 2021 SR reported no statistically significant difference between patients who underwent rTHA as compared to mTHA (i.e., OR = 0.30; 95% CI, 0.05 to 1.88; P = 0.20), the 2019 SR reported statistically significantly fewer intraoperative complications in the rTHA group compared to the mTHA group (i.e., OR = 0.32; 95% CI, 0.14 to 0.72; P = 0.006).[23]

Mechanical Complications

One retrospective matched cohort study describing inpatient discharge abstracts reported on several types of mechanical complications that could impact clinical effectiveness and patient-reported outcomes (i.e., prosthetic joint implant failure, periprosthetic osteolysis, mechanical loosening, and unspecified [as well as "all"] complications).[24] All of the outcomes measured (with the exception of periprosthetic osteolysis, which did not characterize the difference between treatment groups statistically) indicated statistically significantly fewer mechanical complications in the rTHA group as compared to the mTHA group (i.e., P < 0.001).[24]

Nerve Injury

Two SRs reported on nerve injuries — both included 3 primary studies each in their analyses,[22,23] with 1 describing findings narratively,[22] and 1 reporting on the findings of a meta-analysis.[23] The findings from both SRs identified no statistically significant differences between patients having undergone rTHA as compared to mTHA, with the meta-analysis generating OR = 4.47; 95% CI, 0.72 to 27.90; P = 0.11.[22,23]

Revision Surgery

All 3 SRs and 1 retrospective matched cohort study compared the odds of requiring revision surgery between patients who underwent rTHA versus mTHA.[21-23,26] All of the findings reported indicated no statistically significant difference between the rTHA and mTHA groups, with the exception of the meta-analysis from 1 SR, which reported statistically significantly fewer revisions in the mTHA group across 3 studies representing 508 patients (with follow-up timing not reported) (i.e., OR = 2.88; 95% CI, 1.11 to 7.50; P = 0.03).[23]

Cost-Effectiveness of Robotic-Assisted Surgical Systems for Hip Arthroplasty

The cost-effectiveness evaluation included in this review concluded that rTHA was cost-effective when compared to mTHA, producing an incremental cost-effectiveness ratio of −$23,625 per quality-adjusted life-year at a willingness-to-pay threshold of $50,000.[28] Sensitivity analyses indicated that the utilities of both rTHA and mTHA were the only model parameters sensitive to the preferred intervention.[28]

Appendix 4 presents the main study findings.

Limitations

Whereas this review identified a considerable body of recently published evidence describing the clinical effectiveness of rTHA as compared to mTHA (i.e., 27 unique and relevant primary studies summarized across 3 SRs, plus 4 retrospective matched cohort studies), less evidence was found describing cost-effectiveness (i.e., 1 study). Many additional eligible SRs and NRSs were also identified but were overlapped by the SRs already included in this review (see Appendix 5), suggesting that the number of unique studies describing the clinical effectiveness of rHA may be more limited than it appears by the number of SRs on this topic. In addition, all publications included in this review described THA; no studies describing other types of HA were identified.

Arguably, one of the most important limitations of the evidence identified and summarized in this review is the low methodological quality of the studies included: no RCTs were identified and few RCTs were included in the SRs identified by this review, resulting in most of the evidence being summarized from NRSs. While many outcomes and outcome measures were reported across multiple studies — resulting in large sample sizes assembled from meta-analyses[21-23] or large, regional health administrative databases[24-26] — there was notable inconsistency in many of the findings, with disagreement in the magnitude and/or direction of effect between treatment groups observed across the studies. These discrepant findings, perhaps associated with the low quality of evidence available, make it challenging to draw conclusions as to the comparative clinical effectiveness of rTHA versus mTHA. In addition, while some statistically significant findings were identified, the extent to which these findings have clinical significance was not consistently made clear by the authors of the included publications, further complicating their utility in a real-world context.

Another considerable limitation of some included publications was the lack of clarity regarding outcomes (i.e., important details describing outcome measures and follow-up time were either absent from and/or unclear in all of the included publications).[21-27] For instance, 2 of the SRs did not contain a description of the scoring and/or interpretation of the measures summarized from the primary studies they included, resulting in a critical lack of clarity as to what and how the outcome was measured;[21,22] this necessitated the exclusion of some data that could not be interpreted but might have been relevant. Further, whereas some studies reported on the timing of follow-up for some or all outcomes, some did not report length of follow-up at all.[23,27] This is particularly problematic when considered in the context of meta-analyses, where variable lengths of follow-up across studies can introduce bias into the findings.[33,34] This, and other sources of possible heterogeneity, represent threats to the validity of the findings presented by the SRs included in this review.[21-23] Finally, the quality and clarity of reporting in the SRs was generally poor, with inconsistencies and errors identified in the papers,[21-23] which reduce the extent to which the findings can be clearly interpreted and applied to a decision-making context.

In addition, there were few details reported in several of the included studies describing the patient populations and procedures, often preventing the reader from being able to consider the extent to which, for instance, patients with different characteristics (beyond age and sex) may respond differently to the interventions under study. Whereas many of the studies did match patients on several characteristics, there remains uncertainty as to whether other patient variables may impact patient responses and could thereby impact the results of the studies. Alternatively, different features of the surgical procedures, surgical device(s), and/or hospital settings may have had an impact on the findings. In particular, several meta-analyses

combined results across multiple rTHA devices without characterizing the potential impact that the introduction of this interventional heterogeneity may have had on the findings of the studies.[22,23] Similarly, the particular rTHA device that was used was not always specified,[24,25] preventing the reader from considering how this may have affected the outcomes in the intervention group.

Lastly, in addition to a limited amount of evidence identified describing cost-effectiveness, the relevance of the 1 study included in this review to the Canadian context is likely limited by its focus on a US payer perspective, which necessarily drew information and provided insights from a mostly privatized health care system,[28] bearing little relevance to that of Canada's single-payer, provincial public systems.

Conclusions and Implications for Decision- or Policy-Making

This review identified 8 publications addressing the clinical and cost-effectiveness of rTHA compared to mTHA for patients undergoing THA (i.e., 3 SRs, 4 retrospective matched cohort studies, and 1 CEA).[21-28] Seven of these publications reported findings that described the clinical effectiveness of rTHA compared with mTHA,[21-27] and 1 study described the cost-effectiveness of rTHA versus mTHA.[28]

The strengths of this review include the identification of many unique studies describing clinical effectiveness — most of which were summarized in the 3 included SRs — and 4 retrospective matched cohort studies that were additionally identified by this review. The number of studies reporting findings from the same standardized outcome measures allowed for multiple meta-analyses, which combined patient data across studies and increased the number of patients from which data were available.[21-23] Likewise, 3 of the retrospective matched cohort studies drew from large, regional databases, which produced large sample sizes,[24-26] increasing the confidence in the findings generated. Notably however, there was a lack of consistency in the outcomes reported across studies, despite some degree of overlap represented in many of the meta-analyses and narrative syntheses (i.e., there was rarely consensus across 2 or more SRs as to a statistically significant difference between the treatment groups in a given outcome measure).[21-23] This was broadly observed across the retrospective matched cohort studies, as well (i.e., findings for similar outcomes were often discordant across studies).[24-27]

One finding that appeared consistently across measures of functionality, quality of life, pain, mortality, and complications was no statistically significant difference between rTHA and mTHA.[21-27] Nonetheless, when a statistically significant difference was reported, it rarely favoured mTHA, with most of these results indicating a benefit of rTHA. The underlying reasons for these inconsistencies are not entirely clear; however, heterogeneity in the intervention groups and/or the possible discrepancy in lengths of follow-up time may have had an impact on the validity of the findings. Other factors that could impact the study results could be variability across surgeons in experience, skill level, or other individual features,[35] as well as differences between patients that were not sufficiently accounted for by the matching algorithms or other statistical methods.

Importantly, there were some outcomes that did produce consistent and/or robust findings across studies (i.e., the length of stay in hospital was found to be significantly shorter in patients who underwent rTHA as compared to mTHA across 3 retrospective matched cohort studies).[24-26] And while 1 SR and 1 retrospective matched cohort study reporting on duration of surgical procedure did not agree,[22,27] the SR identified a statistically significant difference favouring mTHA in that rTHA was associated with a significantly longer duration of surgery;[22] however, the retrospective matched cohort study found no difference between the groups.[27] Because the SR used a meta-analysis of 8 studies representing more than 1,000 patients, and the retrospective matched cohort study reported a small analysis of 118 patients, the discrepancy might have been a result of different levels of statistical power.[22,27] However, it should also be noted that some of the 8 studies from the pooled analysis were conducted long ago (i.e., 5 of 8 were published in 2010 or earlier).[22] On the other hand, the small retrospective matched cohort study was conducted between 2019 and 2020.[27] These and other factors can introduce variability in the findings across NRSs and create a challenge when trying to draw conclusions about clinical effectiveness.

The question regarding cost-effectiveness of rTHA — as posed in a Canadian context — could not sufficiently be answered by the 1 study included in this review[28] due to the dissimilarity between the Canadian and American health systems. Nonetheless, it is worth considering that the conclusions drawn by Maldonado and colleagues were unequivocal with regard to the cost-effectiveness of rTHA in that the cost-effectiveness demonstrated in their model was robust;[28] however, the question of whether cost-effectiveness would be demonstrated in a Canadian context would require a more comparable dataset, and/or an analysis particular to the Canadian health system.

Previous CADTH work on the subject of robot-assisted surgery reflects some of the findings related to clinical effectiveness that are reported here; for example, a report comparing robotic surgical systems with conventional procedures for gynecological and urological procedures found inconsistent results across studies with regard to functional outcomes and surgical duration.[36] However, the report also described shorter length of stay in hospital for patients who underwent robot-assisted surgery, and no differences between treatment groups in complications caused by the procedures.[36] Another CADTH report summarized an analysis of qualitative studies describing the experiences and expectations of health care workers and patients concerning robotic-assisted surgery, highlighting the steep learning curve for surgical teams,[37] which may bear relevance to the inconsistency in findings across studies on the topic.

Given the uncertainty in the evidence identified in this review concerning the clinical effectiveness of rHA, more research is needed; particularly high-quality, RCTs.[15] Several such RCTs are being initiated or are under way,[38,39] and are poised to contribute to the emergence of higher-quality evidence describing the clinical and cost-effectiveness of rHA.
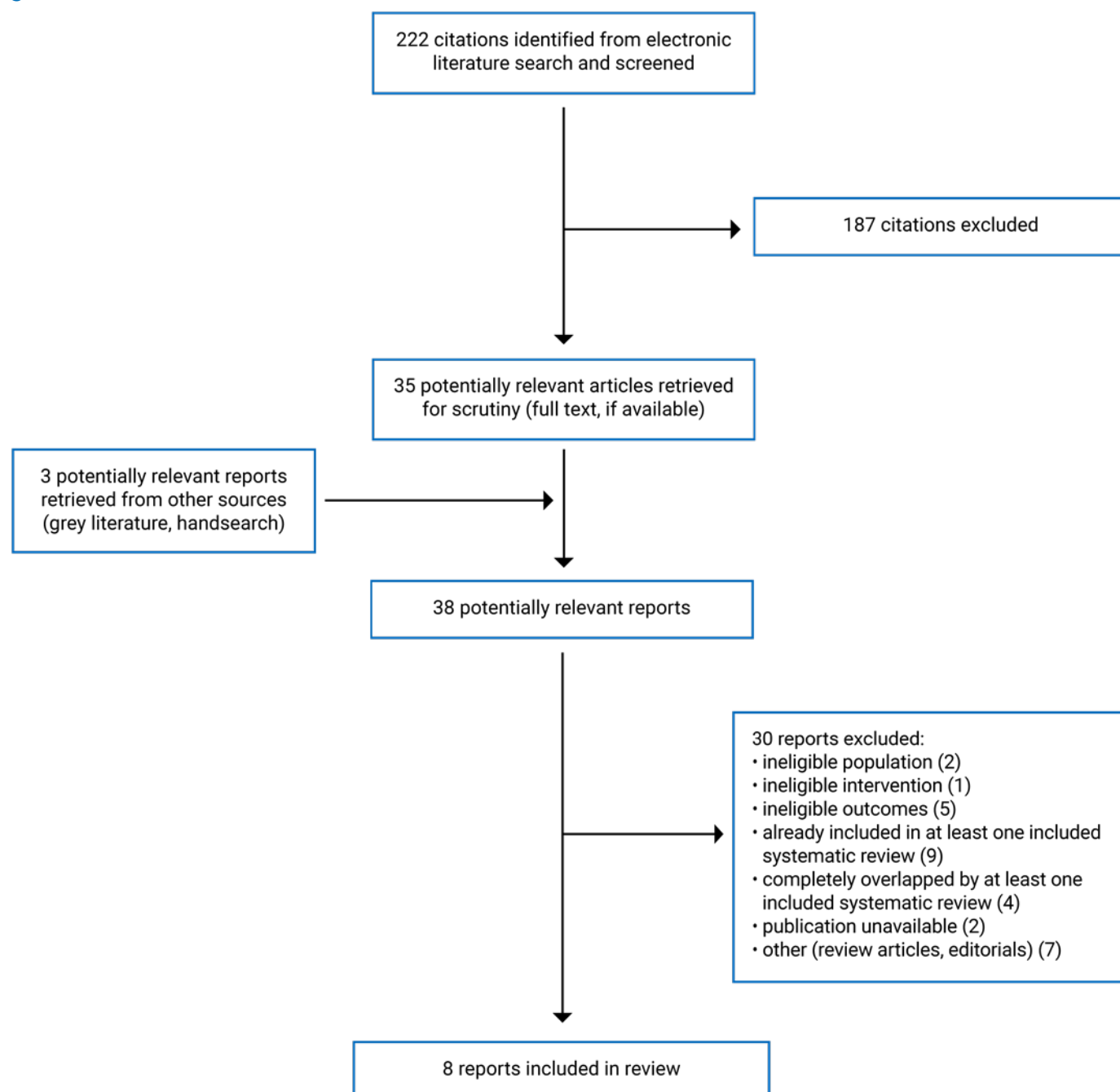
References

1. Ray GS, Ekelund P, Nemes S, Rolfson O, Mohaddes M. Changes in health-related quality of life are associated with patient satisfaction following total hip replacement: an analysis of 69,083 patients in the Swedish Hip Arthroplasty Register. *Acta Orthop.* 2020;91(1):48-52. PubMed

2. Mandl LA, Martin GM. Overview of surgical therapy of knee and hip osteoarthritis. In: T.W. P, ed. *UpToDate* Waltham (MA): UpToDate; 2021: http://www.uptodate.com. Accessed 2022 Mar 4.

3. Erens GA, Walter B, Crowley M. Total hip arthroplasty. In: Post TW, ed. *UpToDate*. Waltham (MA): UpToDate; 2021: http://www.uptodate.com. Accessed 2022 Mar 4.

4. CIHI. CJRR annual report: Hip and knee replacements in Canada. 2021 Jun 15; https://www.cihi.ca/en/cjrr-annual-report-hip-and-knee-replacements-in-canada. Accessed 2022 Mar 4.

5. Petis S, Howard JL, Lanting BL, Vasarhelyi EM. Surgical approach in primary total hip arthroplasty: anatomy, technique and clinical outcomes. *Can J Surg.* 2015;58(2):128-139. PubMed

6. Perets I, Mu BH, Mont MA, Rivkin G, Kandel L, Domb BG. Current topics in robotic-assisted total hip arthroplasty: a review. *Hip Int.* 2020;30(2):118-124. PubMed

7. Matar HE, Platt SR, Board TN, Porter ML. Overview of randomized controlled trials in primary total hip arthroplasty (34,020 patients): What have we learnt? *J Am Acad Orthop Surg Glob Res Rev.* 2020;4(8):e20.00120.

8. Robotic-assisted orthopedic surgical platforms for hip arthroplasty. Plymouth Meeting (PA): ECRI Institute; 2021 Dec: www.ecri.org. Accessed 2022 Mar 4.

9. Sherman WF, Wu VJ. Robotic surgery in total joint arthroplasty: A survey of the AAHKS membership to understand the utilization, motivations, and perceptions of total joint surgeons. *J Arthroplasty.* 2020;35(12):3474-3481.e3472. PubMed

10. Misso D, Zhen E, Kelly J, Collopy D, Clark G. A progressive scholarly acceptance analysis of robot-assisted arthroplasty: a review of the literature and prediction of future research trends. *J Robot Surg.* 2021;15(5):813-819. PubMed

11. Flynn A, Verhoeven A. Measuring value in new health technology assessments: a focus on robotic surgery in public hospitals. Deakin (AU): Deeble Institute for Health Policy Research, Australian Healthcare and Hospitals Association; 2020: https://ahha.asn.au/system/files/docs/publications/deeble_brief_no._37_-_measuring_value_in_new_health_technology_assessments.pdf. Accessed 2022 Mar 4.

12. Randell R, Honey S, Hindmarsh J, Alvarado N, Greenhalgh J, Pearman A. A realist process evaluation of robot-assisted surgery: integration into routine practice and impacts on communication, collaboration and decision-making. *Health Serv Deliv Res.* 2017;5(10). https://www.journalslibrary.nihr.ac.uk/hsdr/hsdr05200#/full-report. Accessed 2022 Mar 4. PubMed

13. National Institute for Health Research. A real-world, in-situ, evaluation of the introduction and scale-up of robot-assisted surgical services in the NHS: Evaluating its impact on clinical and service delivery, effectiveness and cost. 2022 Jan; https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/NIHR131537/#/. Accessed 2022 Mar 4.

14. Banerjee S, Cherian JJ, Elmallah RK, Pierce TP, Jauregui JJ, Mont MA. Robot-assisted total hip arthroplasty. *Expert Rev Med Devices.* 2016;13(1):47-56. PubMed

15. Kouyoumdjian P, Mansour J, Assi C, Caton J, Lustig S, Coulomb R. Current concepts in robotic total hip arthroplasty. *Sicot-J.* 2020;6:45. PubMed

16. ECRI Institute. Unproven surgical robotic procedures may put patients at risk. Hazard #5—2020 top 10 health technology hazards. *Health Devices.* 2019 Sep 26.

17. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008. PubMed

18. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health.* 1998;52(6):377-384. PubMed

19. Higgins JPT, Green S, editors. Figure 15.5.a: Drummond checklist (Drummond 1996). *Cochrane handbook for systematic reviews of interventions*. London (GB): The Cochrane Collaboration; 2011: http://handbook-5-1.cochrane.org/chapter_15/figure_15_5_a_drummond_checklist_drummond_1996.htm. Accessed 2022 Mar 4.

20. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34. PubMed

21. Ng N, Gaston P, Simpson PM, Macpherson GJ, Patton JT, Clement ND. Robotic arm-assisted versus manual total hip arthroplasty: a systematic review and meta-analysis. *Bone Joint J.* 2021;103-B(6):1009-1020. PubMed

22. Samuel LT, Acuna AJ, Mahmood B, Emara AK, Kamath AF. Comparing early and mid-term outcomes between robotic-arm assisted and manual total hip arthroplasty: a systematic review. *J Robot Surg.* 2021;30:30. PubMed

23. Han PF, Chen CL, Zhang ZL, et al. Robotics-assisted versus conventional manual approaches for total hip arthroplasty: A systematic review and meta-analysis of comparative studies. *Int J Med Robot.* 2019;15(3):e1990. PubMed

24. Emara AK, Zhou G, Klika AK, et al. Is there increased value in robotic arm-assisted total hip arthroplasty?: a nationwide outcomes, trends, and projections analysis of 4,699,894 cases. *Bone Joint J.* 2021;103-B(9):1488-1496. PubMed

25. Kirchner GJ, Lieber AM, Haislup B, Kerbel YE, Moretti VM. The cost of robot-assisted total hip arthroplasty: comparing safety and hospital charges to conventional total hip arthroplasty. *J Am Acad Orthop Surg.* 2021;29(14):609-615. PubMed

26. Remily EA, Nabet A, Sax OC, Douglas SJ, Pervaiz SS, Delanois RE. Impact of robotic assisted surgery on outcomes in total hip arthroplasty. *Arthroplast Today.* 2021;9:46-49. PubMed

27. Zhou Y, Shao H, Huang Y, Deng W, Yang D, Bian T. Does robotic assisted technology improve the accuracy of acetabular component positioning in patients with DDH? *J Orthop Surg (Hong Kong).* 2021;29(2):23094990211025325. PubMed

28. Maldonado DR, Go CC, Kyin C, et al. Robotic arm-assisted total hip arthroplasty is more cost-effective than manual total hip arthroplasty: A Markov model analysis. *J Am Acad Orthop Surg.* 2021;29(4):e168-e177. PubMed

29. Online Appendix 1: AMSTAR 2 guidance document. Data supplement to: Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.: https://www.bmj.com/content/bmj/suppl/2017/09/21/bmj.j4008.DC1/sheb036104.ww1.pdf. Accessed 2022 Feb 25.

30. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol.* 2012;65(2):126-131. PubMed

31. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ.* 1996;313(7052):275-283. PubMed

32. Emara AK, Samuel LT, Acuna AJ, Kuo A, Khlopas A, Kamath AF. Robotic-arm assisted versus manual total hip arthroplasty: Systematic review and meta-analysis of radiographic accuracy. *Int J Med Robot.* 2021;17(6):e2332. PubMed

33. Guevara JP, Berlin JA, Wolf FM. Meta-analytic methods for pooling rates when follow-up duration varies: a case study. *BMC Med Res Methodol.* 2004;4:17-17. PubMed

34. Buccheri S, Sodeck GH, Capodanno D. Statistical primer: methodology and reporting of meta-analyses†. *Eur J Cardiothorac Surg.* 2018;53(4):708-713. PubMed

35. Kong X, Yang M, Li X, et al. Impact of surgeon handedness in manual and robot-assisted total hip arthroplasty. *J.* 2020;15(1):159.

36. Li Y, Spry C. Robotic surgical systems for use in gynecologic oncology or urologic surgery: clinical effectiveness, cost-effectiveness, and guidelines. Ottawa (ON): CADTH; 2020 Jan: https://www.cadth.ca/sites/default/files/pdf/htis/2020/RB1430%20Robotic%20Surgical%20Systems%20Final.pdf. Accessed 2022 Mar 3.

37. Martinello N, Loshak H. Experiences with and expectations of robotic surgical systems: a rapid qualitative review. Ottawa (ON): CADTH; 2020 Mar: https://www.cadth.ca/sites/default/files/pdf/htis/2020/RC1251%20RSS%20for%20Gyno%20Uro%20Surgery%20Final.pdf. Accessed 2022 Mar 4.

38. National Institute for Health Research. Robotic arthroplasty: a clinical and cost effectiveness randomised controlled trial for hips (RACER-HIP). 2021; https://www.journalslibrary.nihr.ac.uk/programmes/hta/NIHR131407#/documentation. Accessed 2022 Mar 4.

39. Kayani B, Konan S, Tahmassebi J, Ayuob A, Haddad FS. Computerised tomography-based planning with conventional total hip arthroplasty versus robotic-arm assisted total hip arthroplasty: study protocol for a prospective randomised controlled trial. *Trials.* 2020;21(1):776. PubMed

Appendix 1: Selection of Included Studies

Figure 1: Selection of Included Studies

```
┌─────────────────────────────────────┐
│ 222 citations identified from        │
│ electronic literature search and     │
│ screened                             │
└─────────────────────────────────────┘
                  │
                  │────────────────────────▶ ┌──────────────────────────┐
                  │                           │ 187 citations excluded   │
                  ▼                           └──────────────────────────┘
┌─────────────────────────────────────┐
│ 35 potentially relevant articles     │
│ retrieved for scrutiny (full text,   │
│ if available)                        │
└─────────────────────────────────────┘
┌──────────────────────────────┐      │
│ 3 potentially relevant reports│─────▶│
│ retrieved from other sources  │      │
│ (grey literature, handsearch) │      ▼
└──────────────────────────────┘  ┌─────────────────────────────────┐
                                   │ 38 potentially relevant reports │
                                   └─────────────────────────────────┘
                  │
                  │────▶ ┌─────────────────────────────────────────────┐
                  │      │ 30 reports excluded:                        │
                  │      │ • ineligible population (2)                 │
                  │      │ • ineligible intervention (1)               │
                  │      │ • ineligible outcomes (5)                   │
                  │      │ • already included in at least one included │
                  │      │   systematic review (9)                     │
                  │      │ • completely overlapped by at least one     │
                  │      │   included systematic review (4)            │
                  │      │ • publication unavailable (2)               │
                  │      │ • other (review articles, editorials) (7)   │
                  ▼      └─────────────────────────────────────────────┘
┌─────────────────────────────────────┐
│ 8 reports included in review         │
└─────────────────────────────────────┘
```

Appendix 2: Characteristics of Included Publications

Note that this appendix has not been copy-edited.

Table 2: Characteristics of Included Systematic Reviews

Study citation, country, funding source	Study designs and numbers of primary studies included	Population characteristics	Intervention and comparator(s)	Clinical outcomes, length of follow-up
Ng 2021[21] Country: UK Funding: Reported as none	Studies sought and included: Comparative prospective and retrospective studies, N=17 Included studies relevant to this report: Comparative prospective and retrospective NRS, N=11	Studies sought and included: Patients with arthritis of the hip undergoing/ who underwent THA, N=4,201 Included studies of relevance to this report: Patients with arthritis of the hip undergoing/ who underwent THA, N=1,798	Studies sought, included and of relevance to this report: • Intervention: ○ Semi-active rTHA ○ N.B. All eligible studies identified and included investigated the Stryker MAKO device • Comparator: mTHA	Outcomes in the studies sought and included: Learning curve, implant accuracy, functional outcomes, complications, survival Outcomes in the included studies of relevance to this review: Functional outcomes, complications Synthesis methods: Quantitative meta-analyses and narrative syntheses Follow-up in the studies of relevance to this review, range: Between 6wk to 5yr (minimum)
Samuel 2021[22] Country: US Funding: NR	Studies sought included and of relevance to this report: Controlled prospective RCTs, prospective NRSs, and retrospective NRSs, N=18	Studies sought, included and of relevance to this report: • Patients undergoing/ who underwent THA, N=2,811 • Additional detail re. patient condition/ indication NR	Studies sought, included and of relevant to this report: • Intervention: ○ rTHA ○ N.B. All eligible studies identified and included investigated either the ROBODOC or MAKO device • Comparator: mTHA	Outcomes in the studies sought, included and of relevance to this report: Functional outcomes, complications Synthesis methods: Quantitative meta-analyses and narrative syntheses Follow-up: Between 90d to 14yr

Study citation, country, funding source	Study designs and numbers of primary studies included	Population characteristics	Intervention and comparator(s)	Clinical outcomes, length of follow-up

Han 2019[23]
 	
Country: China (1st author)
 	
Funding: National Natural Science Foundation of China, Grant/Award Number: 81601949; International Science and Technology Cooperation Program of China, Grant/Award Number: 2015DFA33050
	Studies sought and included: Controlled prospective and retrospective studies, N=14
	
	Included studies relevant to this report: Controlled prospective RCTs, prospective NRSs and retrospective NRSs, N=11
		Studies sought and included: Adult patients with severe hip disease (e.g., OA, DDH, avascular osteonecrosis, RA, Paget's disease, etc.) undergoing/who underwent THA, N=2,324
		
		Included studies of relevance to this report: Patients undergoing/who underwent THA, N=NR (i.e., authors report the total N of pts for all studies included in the SR and not per study)
			Studies sought, included and of relevant to this report:
			• Intervention:
			  ○ rTHA
			  ○ N.B. All eligible studies identified and included investigated the ROBODOC, MAKO, Orthodoc or CASPAR systems
			• Comparator: mTHA
				Outcomes in the studies sought and included: Functional scores, complications, cup angle, safe zone of cup, stem alignment, leg length discrepancy, duration of surgical procedure
				
				Outcomes in the studies of relevance to this review: Functional scores, complications, duration of surgical procedure
				
				Synthesis methods: Quantitative meta-analyses and narrative syntheses
				
				Follow-up: NR

cTHA = conventional total hip arthroplasty; caTHA = computer-assisted total hip arthroplasty; mTHA = manual total hip arthroplasty; N.B. = nota bene; NR = not reported; OA = osteoarthritis; RA = rheumatoid arthritis; RCT = randomized controlled trial; rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; THA = total hip arthroplasty; wk = week(s); yr = year(s)

Table 3: Characteristics of Included Primary Clinical Studies

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Clinical outcomes, length of follow-up

Emara 2021[24]
 	
Country: US
 	
Funding: Reported as none
	Retrospective PS-matched* cohort using data from the NIS
	
	*Matched on age, sex, race, insurance status, elective nature, year of surgery, hospital region, hospital size, income quartile, and Elixhauser[†] comorbidity category
		Patients included and of relevance to this report:
		• Inpatients in the NIS who underwent primary THA (with robot or manual technique) between 2008 and 2018, N=4,634,242
		  ○ Intervention: 32,898 (0.7%)
		  ○ Comparator: 4,601,344 (99.3%)
		Age ≤50yr, n (%)
		• Intervention: 3,690 (11.2)
		• Comparator: 401,121 (8.8)
		Age >50yr, n (%)
		• Intervention: 29,208 (88.8)
		• Comparator: 4,200,222 (91.3)
		Female sex, n %
		• Intervention: 18,112 (55.1)
		• Comparator: 2,680,312 (58.3)
		Non-elective/ emergency procedures, n (%)
		• Intervention: 1,475 (4.5)
		• Comparator: 1,036,627 (22.5)
			Interventions included:
			• Interventions:
			  ○ rTHA
			  ○ caTHA
			• Comparator: mTHA
			Interventions of relevance to this report:
			• Interventions: rTHA
			• Comparator: mTHA
				Outcomes of relevance to this report:
				• Hospital LOS
				• Complications
				• Mortality
				Follow-up: NR

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Clinical outcomes, length of follow-up
(continued)		Elixhauser[†] high risk category, n pts (%) • Intervention: 200 (0.6) • Comparator: 162,717 (3.5) Elixhauser[†] medium risk category, n pts (%) • Intervention: 2,485 (7.6) • Comparator: 650,035 (14.1) Elixhauser[†] low risk category, n pts (%) • Intervention: 30,213 (91.8) • Comparator: 3,788,592 (82.3)		

Kirchner 2021[25]
Country: US
Funding: Reported as none
	Retrospective, matched* cohort using data from the NIS
	*Matched on age, sex, race, comorbidities, hospital type, and calendar year
		Patients included and of relevance to this report:
		• Inpatients in the NIS who underwent primary THA between 2010 and 2014, N=1,516
		  ○ Intervention, n: 758
		  ○ Comparator, n: 758
		Age, mean yr
		• Intervention: 64.5
		• Comparator: 64.8
		Sex, % female: 57.1
		Diabetes, % pts: 9.4
		Obesity (not otherwise defined), % pts: 16.0
		COPD, % pts: 8.7
		Hypertension, % pts: 54.0
			Interventions included and of relevance to this report:
			• Interventions: rTHA
			• Comparator: cTHA
				Outcomes:
				• Hospital LOS
				• Complications
				• Mortality
				Follow-up: NR

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Clinical outcomes, length of follow-up
Remily 2021[26] Country: US Funding: NR	Retrospective, PS-matched* cohort using data from Mariner database *Matched on age, sex, CCI, smoking status, and BMI	Patients included and of relevance to this report: • Patients in the Mariner database who underwent THA between 2010 and 2018, N=9,260 • Intervention, n: 4,630 • Comparator, n: 4,630 Age, mean yr (SD) • Intervention: 64 (10.5) • Comparator: 64 (10.4) Sex, % female: 55.5 Obesity (as measured using BMI), % pts: 14.6 Tobacco use, % pts: 28.3 Hypertension, % pts: 76.5 CCI, mean (SD): 0.8 (1.6)	Interventions included and of relevance to this report: • Intervention: rTHA • Comparator: cTHA	Outcomes: • Hospital LOS • Readmission rate • Complications Follow-up: 90d, 1yr, 5yr
Zhou 2021[27] Country: China Funding: Beijing Municipal Administration of Hospitals Incubating Program, pX2020019	Retrospective matched* cohort using data from a single centre *Matched on age, sex, BMI, Crowe type, and date of surgery	Patients included and of relevance to this report: • Patients with DDH who underwent primary THA between 2019 and 2020, N=118 • Intervention, n: 59 • Comparator, n: 59	Interventions included and of relevance to this report: • Intervention: rTHA using the MAKO system • Comparator: cTHA	Outcomes: • Complications • Duration of surgical procedure Follow-up: NR

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Clinical outcomes, length of follow-up
(continued)		Age, years (type of summary statistics NR; mean and (SD) have been assumed) • Intervention: 49.9 (11.2) • Comparator: 49.7 (11.5) Sex, % female: 74.6 CROWE status$^{\Omega}$, % pts • Type I: 61 • Type II or III: 22 • Type IV: 17 BMI, kg/m^2 (type of summary statistic NR; mean and (SD) have been assumed) • Intervention: 24.5 (3.3) • Comparator: 24.7 (2.8)		

caTHA = computer-assisted total hip arthroplasty; CCI = Charlson Comorbidity Index; COPD = chronic obstructive pulmonary disease ; cTHA = conventional total hip arthroplasty; caTHA = computer-assisted total hip arthroplasty; d = day(s); DDH = developmental dysplasia of the hip; LOS = length of stay; mTHA = manual total hip arthroplasty; N.B. = nota bene; NIS = National Inpatient Sample; NR = not reported; NRS = non-randomized study; PS = propensity score; rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; THA = total hip arthroplasty; wk = week(s); yr = year(s)

†The Elixhauser Comorbidity Index is a measure of patient risk associated with comorbidities.

Ω CROWE status is a measure of the severity of hip dysplasia.

Table 4: Characteristics of Included Economic Evaluation

Study citation country, funding source	Type of analysis, time horizon, perspective	Population characteristics	Intervention and comparator(s)	Approach	Source of clinical, cost, and utility data used in analysis	Main assumptions
Maldonado 2021[28] Country: US Funding: NR (but COI statements suggested authors have received multiple sources of funding, including from industry)	Type of analysis: Cost-effectiveness Time horizon: 5yr Perspective: Payer	Adult patients undergoing primary THA	• Intervention: rTHA using the MAKO system • Comparator: mTHA	Markov model	Clinical data: Literature sources and consecutive surgical cases from 1 centre Cost data: Inpatient Medicare 100% Standard Analytic File and using a discount rate of 3% Utility data: Literature sources and consecutive surgical cases from 1 centre, and using a discount rate of 3%	• Diagnosis was accurate • Treatment was appropriate • Rates of revision were identical for both intervention and comparator after infection/dislocation • Rates of re-revision were identical for both intervention and comparator after infection/dislocation • Terminal state was re-revision

COI = conflict of interest; mTHA = manual total hip arthroplasty; rTHA = robotic-assisted total hip arthroplasty; THA = total hip arthroplasty; wk = week(s); yr = year(s)

Appendix 3: Critical Appraisal of Included Publications

Note that this appendix has not been copy-edited.

Table 5: Strengths and Limitations of Systematic Reviews Using AMSTAR 2[17]

Strengths	Limitations
Ng 2021[21]	
• PICOS described within the research aim and methods section • Authors registered the review with PROSPERO (i.e., published the methods *a priori*) • Study selection was performed in duplicate • Included studies described in sufficient detail • Satisfactory methods used to conduct risk of bias assessment • Appropriate methods used for meta-analysis • Funding for the review work was described	• Literature search was somewhat comprehensive but missing some information (i.e., no mention of language restrictions or grey literature search and no justification or rationale provided for the exclusion of some study designs) • There was no mention of data abstraction being performed in duplicate • Excluded studies were not listed and the rationales for exclusion were not reported for individual studies • It was unclear whether the meta-analyses accounted for various lengths of follow up time • Outcome measures were poorly reported (i.e., limited detail and no information concerning scoring) • Sources of funding for included studies were not reported • The potential impact of risk of bias or heterogeneity was not discussed in the interpretation of the results of the meta-analyses or the discussion/interpretation of the findings • There was no mention of an assessment of publication bias
Samuel 2021[22]	
• PICOS described within the methods section • Study selection was performed in duplicate • Included studies described in sufficient detail	• An explicit description of a review protocol or *a priori* method was not reported • Literature search was somewhat comprehensive but missing some information (i.e., no rationale provided for language restrictions or limitation on the inclusion of some study designs, and no mention of a grey literature search) • There was no mention of data abstraction being performed in duplicate • Excluded studies were not listed and the rationales for exclusion were not reported for individual studies • No assessment of risk of bias was described • Outcome measures were poorly reported (i.e., limited detail and no information concerning scoring) • The methods for meta-analyses were not described in sufficient detail so as to assess their appropriateness • It was unclear whether the meta-analyses accounted for various lengths of follow up time • Multiple typos and errors (e.g., incorrect citations, incorrectly reported findings) were observed in the study report

Strengths	Limitations
(continued)	• Sources of funding for included studies were not reported • The potential impact of risk of bias or heterogeneity was not discussed in the interpretation of the results of the meta-analyses or the discussion/interpretation of the findings • There was no mention of an assessment of publication bias • COI was described but no funding was acknowledged
Han 2019[23]	
• PICOS described within the methods section • Study selection and data abstraction performed in duplicate • Included studies described in sufficient detail • Satisfactory methods used to conduct risk of bias assessment • Appropriate methods used for meta-analysis • Publication bias was adequately investigated • Funding for the review work was described	• An explicit description of a review protocol or *a priori* method was not reported • Literature search was somewhat comprehensive but missing some information (i.e., no rationale provided for language restrictions or limitation on the inclusion of some study designs, and no mention of a grey literature search) • Excluded studies were not listed and the rationales for exclusion were not reported for individual studies • Follow-up time was not reported for any of the included studies/outcomes • Outcome measures were poorly reported (i.e., limited detail and no information concerning scoring) • It was unclear whether the meta-analyses accounted for various lengths of follow up time • Risk of bias and heterogeneity were acknowledged in the interpretation of findings, but their possible implications were not described • Sources of funding for included studies were not reported

AMSTAR 2 = A MeaSurement Tool to Assess systematic Reviews 2; COI = conflict of interest; PICOS = population, intervention, comparator, outcome, study design; PROSPERO = prospectively registered systematic reviews in health and social care

Table 6: Strengths and Limitations of Clinical Studies Using the Downs and Black Checklist[18]

Strengths	Limitations
Emara 2021[24]	
Reporting • Aim, objectives, patient characteristics, potential confounders and estimates of random variability all clearly reported • Numerators, denominators and actual probability values reported for outcome data • Adverse events reported **External Validity:** Study used a large regional discharge abstracts with a large sample size, contributing to confidence concerning external validity **Internal Validity** • Data were standardized (i.e., ICD-10 codes and hospital form data), contributing to consistency in measurement of outcomes • Statistical tests appeared appropriate to the data • No data dredging was apparent • While patients were treated at various points across time, a matching algorithm was used to compare patients treated within the same timeframe • Propensity score matching between patients in the treatment groups was performed to reduce the risk of confounding **Study Power:** While no power calculation was provided, the sample size was very large, contributing to confidence in the validity of the findings	**Reporting** • Details describing the interventions were not included • There was no mention of losses to follow up **Internal Validity** • Data were not collected for research purposes (i.e., health administrative data) and therefore may be missing important detail or information necessary to ensure an accurate and comprehensive analysis • Patients were not randomized to the interventions • Patients/clinicians/research staff were not blinded to the intervention received by the patients • It was unclear whether study patients had consistent duration of follow up **Study Power:** The paper did not describe a power calculation demonstrating sufficient power to detect a statistically significant difference between treatment groups
Kirchner 2021[25]	
Reporting • Aim, objectives, patient characteristics, potential confounders and estimates of random variability all clearly reported • Actual probability values were reported for outcome data **External Validity:** Study used a large regional discharge abstracts with a large sample size, contributing to confidence concerning external validity **Internal Validity** • Data were standardized (i.e., ICD-10 codes and hospital form data), contributing to consistency in measurement of outcomes • Statistical tests appeared appropriate to the data • No data dredging was apparent • While patients were treated at various points across time, a matching algorithm was used to compare patients treated within the same timeframe	**Reporting** • Details describing the interventions were not included • Numerators and denominators were reported for some, but not all, outcome data • There was no mention of losses to follow up • Adverse events were described dichotomously (i.e., as major and minor), with insufficient detail describing which events occurred **External Validity** • The method used for matching patients between treatment groups eliminated a sizable proportion from the source population • Authors demonstrated that distribution of patient characteristics was same in both the sample and population, but only for the intervention group (and not the control group)

Strengths	Limitations
• Matching between patients in the treatment groups was performed to reduce the risk of confounding	**Internal Validity** • Data were not collected for research purposes (i.e., health administrative data) and so, may be missing important detail or information necessary to ensure an accurate and comprehensive analysis • Patients were not randomized to the interventions • Patients/clinicians/research staff were not blinded to the intervention received by the patients • It was unclear whether study patients had consistent duration of follow up **Study Power:** The paper did not describe a power calculation demonstrating sufficient power to detect a statistically significant difference between treatment groups
Remily 2021[26]

Strengths	Limitations
Reporting • Aim, objectives, patient characteristics, potential confounders and estimates of random variability all clearly reported • Numerators, denominators and actual probability values reported for outcome data • Adverse events reported **External Validity:** Study used a large, regional and longitudinal health payer database with a large sample size, contributing to confidence concerning external validity **Internal Validity** • Data were standardized (i.e., ICD-10 codes and hospital form data), contributing to consistency in measurement of outcomes • Statistical tests appeared appropriate to the data • No data dredging was apparent • While patients were treated at various points across time, a matching algorithm was used to compare patients treated within the same timeframe • Propensity score matching between patients in the treatment groups was performed to reduce the risk of confounding **Study Power:** While no power calculation was provided, the sample size was very large, contributing to confidence in the validity of the findings	**Reporting** • Details describing the interventions were not included • There was no mention of losses to follow up **External Validity** • The method used for matching patients between treatment groups eliminated a sizable proportion from the source population • Authors did not describe whether the distribution of patient characteristics was the same in both the sample and population groups **Internal Validity** • Data were not collected for research purposes (i.e., health administrative data) and therefore may be missing important detail or information necessary to ensure an accurate and comprehensive analysis • Patients were not randomized to the interventions • Patients/clinicians/research staff were not blinded to the intervention received by the patients **Study Power:** The paper did not describe a power calculation demonstrating sufficient power to detect a statistically significant difference between treatment groups

Strengths	Limitations
Zhou 2021[27]	
Reporting • Aim, objectives, patient characteristics, potential confounders and estimates of random variability all clearly reported • Numerators, denominators and actual probability values reported for outcome data • Adverse events reported **External Validity:** Study used all eligible patients from a single, representative centre, contributing to confidence concerning external validity **Internal Validity** • Data were standardized (i.e., ICD-10 codes and hospital form data), contributing to consistency in measurement of outcomes • Statistical tests appeared appropriate to the data • No data dredging was apparent • While patients were treated at various points across time, a matching algorithm was used to compare patients treated within the same timeframe • Matching between patients in the treatment groups was performed to reduce the risk of confounding	**Reporting** • Details describing the interventions were not included • There was no mention of losses to follow up **Internal Validity** • Data were not collected for research purposes (i.e., health administrative data) and therefore may be missing important detail or information necessary to ensure an accurate and comprehensive analysis • Patients were not randomized to the interventions • Patients/clinicians/research staff were not blinded to the intervention received by the patients • It was unclear whether study patients had consistent duration of follow up **Study Power:** The paper did not describe a power calculation demonstrating sufficient power to detect a statistically significant difference between treatment groups

Table 7: Strengths and Limitations of Economic Evaluation Using the Drummond Checklist[19]

Strengths	Limitations
Maldonado 2021[28]	
Study design • The research hypothesis and its economic importance are clear • The viewpoint of the analysis is clearly stated and justified • The type of economic evaluation and the alternatives being compared are clearly reported and the rationale for the comparison is justified **Data collection** • Sources for estimates for effectiveness are clearly provided • Details describing patients from whom data were collected are clearly reported for the intervention arm • Primary outcomes and measures are stated • Methods to value benefits are stated • Methods for the estimation of quantities and costs are stated • Currency and price information are included • Details describing the model used are provided, including justification for the key parameters selected • Details describing currency of price adjustments for inflation are given **Analysis and interpretation** • The time horizon is included and justified • The discount rate is reported and a literature source is cited in support of its use • Sensitivity analyses, including variables and ranges selected, are described • Incremental analyses are reported and relevant alternatives are compared • The research hypothesis is addressed and conclusions reflect the data reported, including a description of study limitations	**Data collection** • Details describing the design and results of the studies from which effectiveness estimates were sourced are not provided • Details describing patients from whom data were collected are not reported for the comparator arm • Productivity changes were not assessed **Analysis and interpretation:** Confidence intervals are not reported

Appendix 4: Main Study Findings

Note that this appendix has not been copy-edited.

Table 8: Summary of Findings for Clinical Effectiveness — Functional Outcomes

Study Citation and Design	Outcomes
Ambulation	

Samuel 2021
 SR (2 studies)
	Peng 2019 (N=28 pts)
	 • Walking speed, mph (f/u timing NR)
	    ○ rTHA (n=20 pts), mean (SD); range: 1.91 (0.49); 0.5 to 2.9
	    ○ mTHA (n=8 pts), mean (SD); range: 1.76 (0.34); 1.0 to 2.6
	    ○ Difference between groups, NS: P = 0.262
	 • Gait asymmetry, degrees (f/u timing NR)
	    ○ rTHA (n=20 pts): NR
	    ○ mTHA (n=8 pts): NR
	    ○ Difference between groups (reported narratively only): NS

(continued)	**Nishihara 2006 (N=149 pts)**
	 • Number of days post-surgery to walk ≥500m independently, mean (range)
	    ○ rTHA (n=77 pts): 14 (7 to 31)
	    ○ mTHA (n=72 pts): 16 (7 to 46)
	    ○ Difference between groups (reported narratively only): NS
	 • Patients able to walk ≥6 blocks independently 13d post-surgery
	    ○ rTHA (n=77 pts), n: 41
	    ○ mTHA (n=72 pts), n: 28
	    ○ Difference between groups: P < 0.05

Forgotten Joint Score (FJS)	

Ng 2021
 SR (1 study)
	Domb 2020 (N=132 pts)
	 • Post-operative scores (min. 5yr f/u)
	    ○ rTHA (n=66 pts), mean (SD): 82.7 (NR)
	    ○ mTHA (n=66 pts), mean (SD): 70.6 (NR)
	    ○ Mean difference between groups (95% CI)
	      ■ 12.10 (NR)
	      ■ Difference between groups reported only as "a significant difference" (p. 1015) with no information on direction of the effect

Samuel 2021
 SR (4 studies w/MA)
	Clement 2021 (N=120 pts)
	 • Post-operative scores (10 to 12mos f/u)
	    ○ rTHA (n=40 pts), mean (SD): 78.0 (24.2)
	    ○ mTHA (n=80 pts), mean (SD): 56.9 (28.0)
	    ○ Mean difference between groups (95% CI)
	      ■ 21.10 (11.41 to 30.79), favours mTHA
	      ■ P < 0.001

Study Citation and Design	Outcomes

(continued)	**Singh 2021 (N=556 pts)**
	• Post-operative scores (3mo f/u)
	  ○ rTHA (n=NR pts), mean (SD): 53.93 (29.07)
	  ○ mTHA (n=NR pts), mean (SD): 54.59 (27.54)
	  ○ Mean difference between groups (95% CI)
	    ▪ -0.66 (NR), NS
	    ▪ *P* = 0.064
	• Post-operative scores (1yr f/u)
	  ○ rTHA (n=NR pts), mean (SD): 66.95 (27.51)
	  ○ mTHA (n=NR pts), mean (SD): 67.50 (27.52)
	  ○ Mean difference between groups (95% CI)
	    ▪ -0.55 (NR), NS
	    ▪ *P* = 0.230
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=132 pts), mean (SD): 64.72 (31.38)
	  ○ mTHA (n=424 pts), mean (SD): 73.35 (25.33)
	  ○ Mean difference between groups (95% CI)
	    ▪ -8.63 (-14.50 to -2.76), favours rTHA
	    ▪ *P* = 0.004

(continued)	**Domb 2020 (N=132 pts)**
	• Post-operative scores (min. 5yr f/u)
	  ○ rTHA (n=66 pts), mean (SD): 82.69 (21.53)
	  ○ mTHA (n=66 pts), mean (SD): 70.61 (26.74)
	  ○ Mean difference between groups (95% CI)
	    ▪ 12.08 (3.80 to 20.36), favours mTHA
	    ▪ *P* = 0.002

(continued)	**Perets 2020 (N=158 pts), MAKO system**
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=80 pts), mean (SD): 80.2 (21.3)
	  ○ mTHA (n=78 pts), mean (SD): 68.6 (27.3)
	  ○ Mean difference between groups (95% CI)
	    ▪ 11.60 (3.95 to 19.25), favours mTHA
	    ▪ *P* = 0.003

Study Citation and Design	Outcomes

(continued)	**Quantitative synthesis across 4 studies, MA using random effects (N=966 pts)**
	• Post-operative scores
	 ○ rTHA (n=318 pts), mean (SD): NR
	 ○ mTHA (n=648 pts), mean (SD): NR
	 ○ Pooled mean difference between treatment groups (95% CI)
	  ▪ 8.73 (-4.79 to 22.25), NS
	  ▪ *P* = 0.21

Harris Hip Score (HHS)
Ng 2021 SR (5 NRS w/MA)	**Domb 2020 (N=132 pts)** • Post-operative scores (min. 5yr f/u) ○ rTHA (n=66 pts), mean (SD): 90.6 (13.5) ○ mTHA (n=66 pts), mean (SD): 84.6 (14.5) ○ Mean difference between groups (95% CI): 6.00 (1.22 to 10.78), favours rTHA

(continued)	**Kong 2020a (N=186 pts)**
	• Post-operative scores (92.7d to 93.8d f/u)
	 ○ rTHA (n=86 pts), mean (SD): 83.4 (NR)
	 ○ mTHA (n=100 pts), mean (SD): 85.8 (NR)
	 ○ Mean difference between groups (95% CI): 2.40 (NR), NS (reported narratively only)
	• Pre-operative to post-operative change (reported as 'improvement') in scores (92.7d to 93.8d f/u)
	 ○ rTHA (n=86 pts), mean (SD): 43.6 (NR)
	 ○ mTHA (n=100 pts), mean (SD): 47.6 (NR)
	 ○ Mean difference between groups (95% CI)
	  ▪ 4.00 (NR)
	  ▪ Statistical significance of the difference NR

(continued)	**Kong 2020b (N=230 pts)** • Post-operative scores (3mo f/u) ○ rTHA (n=106 pts), mean (SD): 84.1 (6.3) ○ mTHA (n=124 pts), mean (SD): 82.4 (9.1) ○ Mean difference between groups (95% CI): 1.70 (-0.30 to 3.70), NS

Study Citation and Design	Outcomes

(continued)	**Banchetti 2018 (N=107 pts)**
	• Post-operative scores (24mo f/u)
	  ○ rTHA (n=56 pts), mean (SD): 85.6 (8.1)
	  ○ mTHA (n=51 pts), mean (SD): 85.2 (7.7)
	  ○ Mean difference between groups (95% CI): 0.40 (-2.59 to 3.39), NS
	• Pre-operative to post-operative change (reported as 'improvement') in scores (24mo f/u)
	  ○ rTHA (n=56 pts), mean (SD): 41.3 (NR)
	  ○ mTHA (n=51 pts), mean (SD): 39.2 (NR)
	  ○ Mean difference between groups (95% CI)
	    ▪ 2.10 (NR)
	    ▪ Statistical significance of the difference NR

(continued)	**Bukowski 2016 (N=200 pts)**
	• Post-operative scores (374 days [IQR 359 to 500] for rTHA pts; 367 days [IQR 353 to 387] for mTHA pts f/u)
	  ○ rTHA (n=100pts), mean (SD): 92.1 (10.5)
	  ○ mTHA (n=100 pts), mean (SD): 86.1 (16.2)
	  ○ Mean difference between groups (95% CI): 6.00 (2.2 to 9.78), favours rTHA
	• Pre-operative to post-operative change (reported as 'improvement') in scores (f/u timing NR)
	  ○ rTHA (n=100 pts), mean (SD): 43.0 (NR)
	  ○ mTHA (n=100 pts), mean (SD): 37.4 (NR)
	  ○ Mean difference between groups (95% CI): 5.6 (NR), favours rTHA (reported narratively only)

(continued)	**Quantitative synthesis across 4 studies, MA using random effects (N=669 pts)**
	• Post-operative scores
	  ○ rTHA (n=328 pts), mean (SD): NR
	  ○ mTHA (n=341 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ▪ 3.05 (0.46 to 5.64), favours rTHA
	    ▪ P = 0.02
	Narrative synthesis across 5 studies (N pts NR)
	• Pre-operative to post-operative change (reported as 'improvement') in scores (f/u timing NR)
	  ○ rTHA, range: 41.3 to 44.1
	  ○ mTHA, range: 37.4 to 47.6
	  ○ Difference between groups: NR

Study Citation and Design	Outcomes

Samuel 2021
SR (10 studies w/MA)
	Domb 2020 (N=132 pts), MAKO system
	• Post-operative scores (5yr f/u)
	  ○ rTHA (n=66 pts), mean (SD): 90.57 (13.46)
	  ○ mTHA (n=66 pts), mean (SD): 84.62 (14.45)
	  ○ Mean difference between groups (95% CI)
	    ■ 5.95 (1.19 to 10.71), favours mTHA*
	    ■ *P* < 0.001
	This finding is presented as reported in the review paper; however, the data supporting the finding indicate that the benefit favours the rTHA group as opposed to the mTHA group.

(continued)	**Hadley 2020 (N=191 pts), MAKO system**
	• Post-operative scores (min 16mo f/u)
	  ○ rTHA (n=94 pts), mean (SD): 86.7 (10.2)
	  ○ mTHA (n=95 pts), mean (SD): 83.6 (10.1)
	  ○ Mean difference between groups (95% CI)
	    ■ 2.40 (-0.49 to 5.29), NS*
	    ■ *P* < 0.05*
	These findings are presented as reported in the review paper; however, they are inconsistent.

(continued)	**Kong 2020a (N=115 pts), MAKO system**
	• Post-operative scores (3mo f/u)
	  ○ rTHA (n=53 pts), mean (SD): 84.14 (6.29)
	  ○ mTHA (n=62 pts), mean (SD): 62.0 (9.11)
	  ○ Mean difference between groups (95% CI)
	    ■ 22.14 (19.31 to 24.97)*, favours rTHA
	    ■ *P* = 0.231*
	These findings are presented as reported in the review paper; however, they are inconsistent.

(continued)	**Kong 2020b (N=186 pts), MAKO system**
	• Post-operative scores (3mo f/u)
	  ○ rTHA (n=53 pts), mean (SD): 83.35 (7.43)
	  ○ mTHA (n=62 pts), mean (SD): 85.77 (9.22)
	  ○ Mean difference between groups (95% CI)
	    ■ -2.42 (-4.81 to -0.03)*, favours mTHA
	    ■ *P* = 0.993*
	These findings are presented as reported in the review paper; however, they are inconsistent.

(continued)	**Perets 2020 (N=158 pts), MAKO system**
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=80 pts), mean (SD): 91.0 (12.4)
	  ○ mTHA (n=78 pts), mean (SD): 84.4 (14.9)
	  ○ Mean difference between groups (95% CI)
	    ■ 6.60 (2.32 to 10.88), favours rTHA
	    ■ *P* < 0.001

Study Citation and Design	Outcomes

(continued)	**Bargar 2018 (N=67 pts), ROBODOC system**
	• Post-operative scores (14yr f/u)
	  ○ rTHA (n=45 pts), mean (SD): 93.5 (8.77)
	  ○ mTHA (n=22 pts), mean (SD): 89.5 (12.03)
	  ○ Mean difference between groups (95% CI)
	    ■ 4.40 (-1.29 to 4.75), NS
	    ■ *P* > 0.05

(continued)	**Bukowski 2016 (N=200 pts), MAKO system**
	• Post-operative scores (min. 1yr f/u)
	  ○ rTHA (n=100 pts), mean (SD): 92.0 (10.5)
	  ○ mTHA (n=100 pts), mean (SD): 86.1 (16.2)
	  ○ Mean difference between groups (95% CI)
	    ■ 5.90 (2.12 to 9.68), favours rTHA
	    ■ *P* = 0.002

(continued)	**Lim 2015 (N=49 pts), ROBODOC system**
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=24 pts), mean (SD): 93.0 (12.2)
	  ○ mTHA (n=25 pts), mean (SD): 95.0 (8.9)
	  ○ Mean difference between groups (95% CI)
	    ■ -2.00 (-8.00 to 4.00), NS
	    ■ *P* = 0.512

(continued)	**Honl 2003 (N=141 pts), ROBODOC system**
	• Post-operative scores (3mo f/u)
	  ○ rTHA (n=61 pts), mean (SD): 52.6 (12.3)
	  ○ mTHA (n=80 pts), mean (SD): 51.7 (10.6)
	  ○ Mean difference between groups (95% CI)
	    ■ 0.90 (NR), NS
	    ■ = 0.08
	• Post-operative scores (6mo f/u)
	  ○ rTHA (n=61 pts), mean (SD): 74.4 (16.4)
	  ○ mTHA (n=80 pts), mean (SD): 68.3 (18.7)
	  ○ Mean difference between groups (95% CI)
	    ■ 6.10 (NR), NS
	    ■ = 0.06

Study Citation and Design	Outcomes
(continued)	• Post-operative scores (1yr f/u) 　○ rTHA (n=61 pts), mean (SD): 85.9 (12.0) 　○ mTHA (n=80 pts), mean (SD): 73.2 (16.9) 　○ Mean difference between groups (95% CI) 　　■ 12.70 (NR), favours rTHA 　　■ < 0.001 • Post-operative scores (2yr f/u) 　○ rTHA (n=61 pts), mean (SD): 85.9 (12.0) 　○ mTHA (n=80 pts), mean (SD): 83.6 (11.9) 　○ Mean difference between groups (95% CI) 　　■ 2.30 (-1.68 to 6.28), NS 　　■ = 0.06
(continued)	**Bargar 1998 (N=93 pts), ROBODOC system** • Post-operative scores (3mo f/u) 　○ rTHA (n=48 pts), mean (SD): 79.9 (NR) 　○ mTHA (n=45 pts), mean (SD): 80.3 (NR) 　○ Mean difference between groups (SD) 　　■ -0.4 (NR) 　　■ $P > 0.05$ • Post-operative scores (1yr f/u) 　○ rTHA (n=48 pts), mean (SD): 88.5 (0) 　○ mTHA (n=45 pts), mean (SD): 91.1 (0) 　○ Difference between groups 　　■ NE 　　■ $P > 0.05$ • Post-operative scores (2yr f/u) 　○ rTHA (n=48 pts), mean (SD): 89.4 (NR) 　○ mTHA (n=45 pts), mean (SD): 86.6 (NR) 　○ Difference between groups 　　■ NR 　　■ $P > 0.05$

Study Citation and Design	Outcomes

(continued)	**Quantitative synthesis across 10 studies, MA using random effects (N=1,330 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=657 pts), mean (SD): NR
	  ○ mTHA (n=673 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ■ 5.05 (-0.88 to 10.98), NS
	    ■ *P* = 0.25
	Quantitative synthesis across 6 studies investigating MAKO systems, MA using random effects (N=980 pts)
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=479 pts), mean (SD): NR
	  ○ mTHA (n=501 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ■ 6.76 (-1.23 to 14.75), NS
	    ■ *P* = 0.10

(continued)	**Quantitative synthesis across 4 studies investigating ROBODOC systems, MA using random effects (N=350 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=178 pts), mean (SD): NR
	  ○ mTHA (n=172 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ■ 1.73 (-1.29 to 4.75), NS
	    ■ *P* = 0.26

Han 2019
SR (4 studies w/MA)
	Bargar 2018 (N=67 pts), ROBODOC system
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=45 pts), mean (SD): 93.49 (8.77)
	  ○ mTHA (n=22 pts), mean (SD): 89.5 (12.03)
	  ○ Mean difference between groups (95% CI): 3.99 (-1.65 to 9.63), NS

(continued)	**Lim 2015 (N=49 pts), ROBODOC system**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=24 pts), mean (SD): 93.0 (12.2)
	  ○ mTHA (n=25 pts), mean (SD): 95.0 (8.9)
	  ○ Mean difference between groups (95% CI): -2.00 (-8.00 to 4.00), NS

(continued)	**Siebel 2005 (N=71 pts), f/u timing NR**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=36 pts), mean (SD): 88.5 (8.3)
	  ○ mTHA (n=35 pts), mean (SD): 92.7 (11.4)
	  ○ Mean difference between groups (95% CI): -4.20 (-8.85 to 0.45), NS

Study Citation and Design	Outcomes

(continued)	**Honl 2003 (N=141 pts), ROBODOC system**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=61 pts), mean (SD): 85.9 (12)
	  ○ mTHA (n=80 pts), mean (SD): 83.6 (11.9)
	  ○ Mean difference between groups (95% CI): 2.30 (-1.68 to 6.28), NS

(continued)	**Quantitative synthesis across 4 studies, MA using random effects (N=328 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=166 pts), mean (SD): NR
	  ○ mTHA (n=162 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ▪ 0.04 (-3.70 to 3.78), NS
	    ▪ *P* = 0.98

	Merle d'Aubigné (MDA) Hip Score

Samuel 2021
SR (3 studies w/MA)
	Hananouchi 2007 (N=58 pts), ROBODOC system
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=31 pts), mean (SD): 17.8 (0.6)
	  ○ mTHA (n=27 pts), mean (SD): 17.7 (0.7)
	  ○ Mean difference between groups (95% CI)
	    ▪ 0.10 (-0.24 to 0.44), NS
	    ▪ *P* > 0.05

(continued)	**Nishihara 2006 (N=149 pts), ROBODOC system**
	• Post-operative scores (3mo f/u)
	  ○ rTHA (n=77 pts), mean (SD): 15.8 (NR)
	  ○ mTHA (n=79 pts), mean (SD): 15.3 (NR)
	  ○ Mean difference between groups (95% CI)
	    ▪ 0.50 (NR), NS
	    ▪ *P* > 0.05
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=77 pts), mean (SD): 17.4 (1.94)
	  ○ mTHA (n=79 pts), mean (SD): 17.1 (2.78)
	  ○ Mean difference between groups (95% CI)
	    ▪ 0.50 (NR), NS
	    ▪ *P* = 0.05

Study Citation and Design	Outcomes

(continued)	**Honl 2003 (N=141 pts), ROBODOC system**
	• Post-operative scores (3mo f/u)
	  ○ rTHA (n=61 pts), mean (SD): 9.7 (1.9)
	  ○ mTHA (n=80 pts), mean (SD): 10.1 (1.8)
	  ○ Mean difference between groups (95% CI)
	    ■ -0.40 (NR), NS
	    ■ $P > 0.05$
	• Post-operative scores (6mo f/u)
	  ○ rTHA (n=61 pts), mean (SD): 13.0 (2.8)
	  ○ mTHA (n=80 pts), mean (SD): 12.6 (3.4)
	  ○ Mean difference between groups (95% CI)
	    ■ 1.6 (NR), favours mTHA
	    ■ $P = 0.04$

(continued)	• Post-operative scores (1yr f/u)
	  ○ rTHA (n=61 pts), mean (SD): 15.7 (2.2)
	  ○ mTHA (n=80 pts), mean (SD): 14.4 (2.6)
	  ○ Mean difference between groups (95% CI)
	    ■ 1.30 (NR), NS
	    ■ $P > 0.05$
	• Post-operative scores (2yr f/u)
	  ○ rTHA (n=61 pts), mean (SD): 15.7 (2.2)
	  ○ mTHA (n=80 pts), mean (SD): 14.9 (2.1)
	  ○ Mean difference between groups (95% CI)
	    ■ 0.8 (0.08 to 1.52)*, favours mTHA
	    ■ $P > 0.05$*
	These findings are presented as reported in the review paper; however, they are inconsistent.

(continued)	**Quantitative synthesis across 3 studies, MA using fixed effects (N=348 pts)**
	• Post-operative scores
	  ○ rTHA (n=169 pts), mean (SD): NR
	  ○ mTHA (n=179 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ■ 0.24 (-0.05 to 0.52), NS
	    ■ $P = 0.10$

Han 2019
SR (4 studies w/MA)
	Hananouchi 2007 (N=58 pts)
	• Post-operative scores, f/u timing NR
	  ○ rTHA (n=31 pts), mean (SD): 17.8 (0.6)
	  ○ mTHA (n=27 pts), mean (SD): 17.7 (0.7)
	  ○ Mean difference between groups (95% CI): 0.10 (-0.24 to 0.44), NS

Study Citation and Design	Outcomes
(continued)	**Nishihara 2006 (N=156 pts)** • Post-operative scores, f/u timing NR ○ rTHA (n=78 pts), mean (SD): 17.4 (1.94) ○ mTHA (n=78 pts), mean (SD): 17.1 (2.78) ○ Mean difference between groups (95% CI): 0.30 (-0.45 to 1.05), NS
(continued)	**Siebel 2005 (N=71 pts)** • Post-operative scores, f/u timing NR ○ rTHA (n=36 pts), mean (SD): 16 (1.6) ○ mTHA (n=35 pts), mean (SD): 17 (1.6) ○ Mean difference between groups (95% CI): -1.00 (-1.74 to -0.26), NS
(continued)	**Honl 2003 (N=141 pts), ROBODOC system** • Post-operative scores, f/u timing NR ○ rTHA (n=61 pts), mean (SD): 15.7 (2.2) ○ mTHA (n=80 pts), mean (SD): 14.9 (2.1) ○ Mean difference between groups (95% CI): 0.80 (0.08 to 1.52), NS
(continued)	**Quantitative synthesis of 4 studies, MA using fixed effects (N=426 pts)** • rTHA (n=206 pts), n events: NR • mTHA (n=220 pts), n events: NR • Pooled mean difference between treatment groups (95% CI) ○ 0.06 (-0.56 to 0.68), favours rTHA ○ *P* = 0.86
Range of Motion	

Samuel 2021
SR (1 study)
	Peng 2019 (N=28 pts)
	• Peak range of motion compared to contralateral hip, frontal/axial planes, 1yr f/u
	 ○ rTHA (n=20 pts), degrees of difference (reported narratively only): NS
	 ○ mTHA (n=8 pts), degrees of difference (reported narratively only): NS
	• Peak range of motion compared to contralateral hip, net sagittal plane, 1yr f/u
	 ○ rTHA (n=20 pts), mean (SD) degrees of difference
	  ▪ -3 (4.9), favours contralateral hip
	  ▪ *P* = 0.043
	 ○ mTHA (n=8 pts), degrees of difference: *P* > 0.05, no difference between hips

Study Citation and Design	Outcomes
University of California Los Angeles (UCLA) Physical Activity Scale	

Ng 2021
SR (1 NRS)
	Bukowski 2016 (N=200 pts)
	• Post-operative scores, 374 days f/u (IQR 359 to 500) for rTHA pts; 367 days f/u (IQR 353 to 387) for mTHA pts
	  ○ rTHA, mean (SD): 6.3 (NR)
	  ○ mTHA, mean (SD): 5.8 (NR)
	  ○ Difference between groups (reported narratively only): Favours rTHA
	• Pre-operative to post-operative change (reported as 'improvement') in scores, 374 days f/u (IQR 359 to 500) for rTHA pts; 367 days f/u (IQR 353 to 387) for mTHA pts
	  ○ rTHA (n=100 pts), mean (SD): 1.2 (NR)
	  ○ mTHA (n=100 pts), mean (SD): 1.0 (NR)
	  ○ Mean difference between groups (95% CI)
	    ▪ 0.2 (NR)
	    ▪ Statistical difference NR

Samuel 2021
SR (2 studies w/o MA)
	Bargar 2018 (N=67 pts)
	• Post-operative scores, 14 yr f/u
	  ○ rTHA (n=45 pts), mean (SD): 6.09 (1.89)
	  ○ mTHA (n=22 pts), mean (SD): 5.71 (1.45)
	  ○ Difference between groups, NS: *P* = 0.417

(continued)	**Bukowski 2016 (N=200 pts)**
	• Post-operative scores, min. 1yr f/u
	  ○ rTHA, mean (SD): 6.3 (1.8)
	  ○ mTHA, mean (SD): 5.8 (1.7)
	  ○ Difference between groups, favours rTHA: *P* = 0.033
	• Pre-operative to post-operative change (reported as 'improvement') in scores, min. 1yr f/u
	  ○ rTHA (n=100 pts), mean (SD): 1.2 (NR)
	  ○ mTHA (n=100 pts), mean (SD): 1.0 (NR)
	  ○ Mean difference between groups (95% CI): 0.2 (NR), NR

Study Citation and Design	Outcomes
Western Ontario and McMaster University's Osteoarthritis Index (WOMAC)	

Ng 2021
SR (2 NRS w/o MA)
	Banchetti 2018 (N=107 pts)
	• Post-operative scores (24mo f/u)
	  ○ rTHA, mean (SD): 6.8 (NR)
	  ○ mTHA, mean (SD): 6.9 (NR)
	  ○ Mean difference between groups (95% CI)
	    ■ 0.1 (NR), NS
	    ■ Statistical difference between groups reported narratively only
	• Pre-operative to post-operative change (reported as 'improvement') in scores
	  ○ rTHA (n=100 pts), mean (SD): 66.3 (NR)
	  ○ mTHA (n=100 pts), mean (SD): 62.0 (NR)
	  ○ Mean difference between groups (95% CI)
	    ■ 1.3 (NR), NS
	    ■ Statistical difference between groups reported narratively only

(continued)	**Bukowski 2016 (N=200 pts)**
	• Post-operative scores, 374 days f/u (IQR 359 to 500) for rTHA pts; 367 days f/u (IQR 353 to 387) for mTHA pts
	  ○ rTHA: NR
	  ○ mTHA: NR
	  ○ Difference between groups (reported narratively only)
	    ■ NS
	    ■ Statistical difference between groups reported narratively only

(continued)	• Pre-operative to post-operative change (reported as 'improvement') in scores, 374 days f/u (IQR 359 to 500) for rTHA pts; 367 days f/u (IQR 353 to 387) for mTHA pts
	  ○ rTHA (n=100 pts), mean (SD): 29.6 (NR)
	  ○ mTHA (n=100 pts), mean (SD): 28.5 (NR)
	  ○ Mean difference between groups (95% CI)
	    ■ 1.1 (NR)
	    ■ Statistical difference between groups NR

(continued)	**Narrative synthesis across 2 studies**
	• Pre-operative to post-operative change (reported as 'improvement') in scores, 1-2yr f/u
	  ○ rTHA: 29.6 to 66.3
	  ○ mTHA: 28.5 to 62.0
	  ○ Difference between groups: NR

Samuel 2021
SR (4 studies w/MA)
	Bukowski 2016 (N=200 pts), MAKO system
	• Post-operative scores, 1yr f/u
	  ○ rTHA (n=100 pts), mean (SD): 16.0 (14.9)
	  ○ mTHA (n=100 pts), mean (SD): 17.3 (15.5)
	  ○ Mean difference between groups (95% CI): -1.30 (-5.51 to 2.91), NS

Study Citation and Design	Outcomes

(continued)	**Hadley 2020 (N=191 pts), MAKO system**
	• Post-operative scores, min. 16mo f/u
	  ○ rTHA (n=94 pts), mean (SD): 11.7 (10.9)
	  ○ mTHA (n=95 pts), mean (SD): 17.7 (10.4)
	  ○ Mean difference between groups (95% CI): -6.00 (-9.04 to -2.96), favours rTHA

(continued)	**Bargar 2018 (N=67 pts), ROBODOC system**
	• Post-operative scores, 14yr f/u
	  ○ rTHA (n=45 pts), mean (SD): 8.44 (11.48)
	  ○ mTHA (n=22 pts), mean (SD): 11.32 (11.92)
	  ○ Mean difference between groups (95% CI): -2.88 (-8.89 to 3.13), NS

(continued)	**Lim 2015 (N=49 pts), ROBODOC system**
	• Post-operative scores, 2yr f/u
	  ○ rTHA (n=24 pts), mean (SD): 11.0 (10.5)
	  ○ mTHA (n=25 pts), mean (SD): 12.0 (5.4)
	  ○ Mean difference between groups (95% CI): -1.00 (-5.70 to 3.70), NS

(continued)	**Quantitative synthesis across 4 studies, MA using random effects (N=505 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=263 pts), mean (SD): NR
	  ○ mTHA (n=242 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ▪ -3.57 (-5.62 to -1.52), favours rTHA
	    ▪ *P* = 0.0006

(continued)	**Quantitative synthesis across 2 studies investigating MAKO systems, MA using random effects (N=389 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=194 pts), mean (SD): NR
	  ○ mTHA (n=195 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ▪ -4.39 (-6.86 to -1.93), favours rTHA
	    ▪ *P* = 0.0005

(continued)	**Quantitative synthesis across 2 studies investigating ROBODOC systems, MA using random effects (N=116 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=69 pts), mean (SD): NR
	  ○ mTHA (n=47 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ▪ -1.71 (-5.62 to 1.99), NS
	    ▪ *P* = 0.36

Study Citation and Design	Outcomes

Han 2019
SR (2 studies w/MA)
	Bargar 2018 (N=67 pts), ROBODOC system
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=45 pts), mean (SD): 8.44 (11.48)
	  ○ mTHA (n=22 pts), mean (SD): 11.32 (11.92)
	  ○ Mean difference between groups (95% CI): -2.88 (-8.89 to 3.13), NS

(continued)	**Lim 2015 (N=49 pts), ROBODOC system**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=24 pts), mean (SD): 11 (10.5)
	  ○ mTHA (n=25 pts), mean (SD): 12 (5.4)
	  ○ Mean difference between groups (95% CI): -1.00 (-5.70 to 3.70), NS

(continued)	**Quantitative synthesis across 2 studies, MA using random effects (N=116 pts)**
	• Post-operative scores (f/u timing NR)
	  ○ rTHA (n=69 pts), mean (SD): NR
	  ○ mTHA (n=47 pts), mean (SD): NR
	  ○ Pooled mean difference between treatment groups (95% CI)
	    ■ -1.71 (-5.42 to 1.99), NS
	    ■ *P* = 0.36

CI = confidence interval; d = day(s); f/u = follow up; FJS = Forgotten Joint Score; HHS = Harris Hip Score; IQR = interquartile range; m = metre(s); MA = meta-analysis; MDA = Merle d'Aubigné Hip Score; mo = month(s); mph = ; mTHA = manual total hip arthroplasty; NR = not reported; NRS = non-randomized study; NS = not significant; P = probability value; pt/pts = patient(s); rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; SR = systematic review; wk = week(s); w/o = without; WOMAC = Western Ontario and McMaster University's Osteoarthritis Index; yr = year(s)

Table 9: Summary of Findings for Clinical Effectiveness — Overall Health, Mental Health, Physical Health, and Health-Related Qualify of Life

Study Citation and Design	Outcomes
Euro-Quality of Life Five Dimension (EQ-5D) questionnaire

Samuel 2021
SR (1 study)
	Clement 2020 (N=120 pts)
	• Post-operative Role Physical scores (10-12mo f/u)
	  ○ rTHA (n=40 pts), mean (SD): 0.883 (0.150)
	  ○ mTHA (n=80 pts), mean (SD): 0.866 (0.157)
	  ○ Mean difference between groups (95% CI)
	    ■ 0.017 (NR), NS
	    ■ *P* = 0.562

Study Citation and Design	Outcomes
Euro-Quality of Life Visual Analog Scale (EQ-VAS)	

Samuel 2021
 	
SR (1 study)
	Clement 2020 (N=120 pts)
	• Post-operative Role Physical scores (10-12mo f/u)
	  ○ rTHA (n=40 pts), mean (SD): 88.6 (9.5)
	  ○ mTHA (n=80 pts), mean (SD): 86.1 (15.0)
	  ○ Mean difference between groups (95% CI)
	    ▪ 2.5 (NR), NS
	    ▪ *P* = 0.355

Short-Form (SF) surveys	

Ng 2021
 	
SR (2 NRS w/o MA)
	Domb 2020 (N=132 pts)
	• Post-operative scores, SF-12, min. 5yr f/u
	  ○ rTHA: NR
	  ○ mTHA: NR
	  ○ Difference between groups (reported narratively only): Favours rTHA

(continued)	**Bukowski 2016 (N=200 pts)**
	• Post-operative scores, SF-12 , 374 days f/u (IQR 359 to 500) for rTHA pts; 367 days f/u (IQR 353 to 387) for mTHA pts
	  ○ rTHA: NR
	  ○ mTHA: NR
	  ○ Difference between groups (reported narratively only): NS
	• Pre-operative to post-operative change (reported as 'improvement') in scores, SF-12 , 374 days f/u (IQR 359 to 500) for rTHA pts; 367 days f/u (IQR 353 to 387) for mTHA pts
	  ○ rTHA (n=100 pts), mean (SD): 0.4 (NR)
	  ○ mTHA (n=100 pts), mean (SD): 0.5 (NR)
	  ○ Mean difference between groups (95% CI): 0.1 (NR), NS

Samuel 2021
 	
(3 studies w/o MA)
	Domb 2020 (N=132 pts)
	• Post-operative SF-12 physical health scores, 5yr f/u
	  ○ rTHA, mean (SD): 50.30 (8.83)
	  ○ mTHA, mean (SD): 45.92 (9.44)
	  ○ Difference between groups, favours rTHA: P = 0.002
	• Post-operative SF-12 mental health scores, 5yr f/u
	  ○ rTHA: NR
	  ○ mTHA: NR
	  ○ Difference between groups, NS: P = 0.17

(continued)	**Bukowski 2016 (N=200 pts)**
	• Post-operative SF-12 scores, 1yr f/u
	  ○ rTHA: NR
	  ○ mTHA: NR
	  ○ Difference between groups (reported narratively only): NS

Study Citation and Design	Outcomes

(continued)	**Bargar 1998 (N=93)**
	• Post-operative SF-36 mental health scores, 2yr f/u
	  ○ rTHA (n=48 pts): NR
	  ○ mTHA: NR
	  ○ Difference between groups (reported narratively only): NS

CI = confidence interval; d = day(s); EQ-5D = Euro-Quality of Life Five Dimension questionnaire; EQ-VAS = Euro-Quality of Life Visual Analog Scale; f/u = follow up; IQR = interquartile range; MA = meta-analysis; mo = month(s); mTHA = manual total hip arthroplasty; NR = not reported; NRS = non-randomized study; NS = not significant; P = probability value; pt/pts = patient(s); rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; SF = Short-Form ; SR = systematic review; wk = week(s); w/o = without; yr = year(s)

Table 10: Summary of Findings for Clinical Effectiveness — Pain

Study Citation and Design	Outcomes
Health Status Questionnaire (HSQ)	

Samuel 2021
SR (1 study)
	Bargar 2018 (N=67 pts), ROBODOC system
	• Post-operative scores, 14yr f/u
	  ○ rTHA (n=45 pts), mean (SD): 83.75 (20.40)
	  ○ mTHA (n=22 pts), mean (SD): 72.65 (16.31)
	  ○ Mean difference between groups (95% CI)
	    ■ 11.10 (NR), favours mTHA
	    ■ P = 0.019

Postoperative Knee Pain	
Han 2019 SR (2 studies w/MA)	**Nakamura 2010 (N=146 pts), f/u timing NR** • rTHA (n=75 pts), n events: 0 • mTHA (n=71 pts), n events: 2 • Odds ratio (95% CI): 4.86 (0.23 to 103.09), NS
(continued)	**Nakamura 2009 (N=118 pts), f/u timing NR** • rTHA (n=40 pts), n events: 2 • mTHA (n=78 pts), n events: 0 • Odds ratio (95% CI): 0.38 (0.02 to 8.06), NS

(continued)	**Quantitative synthesis of 2 studies, MA using fixed effects (N=264 pts)**
	• rTHA (n=115 pts), n events: 2
	• mTHA (n=149 pts), n events: 2
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 1.40 (0.25 to 7.78), NS
	  ○ P = 0.70

Study Citation and Design	Outcomes
modified Harris Hip Score (mHHS)	

Samuel 2021
SR (1 study)
	Bargar 2018 (N=67 pts), ROBODOC system
	• Post-operative scores, 14yr f/u
	  ○ rTHA (n=45 pts), mean (SD): 41.81 (5.05)
	  ○ mTHA (n=22 pts), mean (SD): 39.09 (7.37)
	  ○ Mean difference between groups (95% CI)
	    ▪ 2.72 (NR), favours mTHA
	    ▪ *P* = 0.025

Persistent Pain	
Samuel 2021 SR (2 studies)	**Nakamura 2010 (N=NR pts), ROBODOC system at <1yr f/u** • rTHA (n=NR pts), n pts: 3 • mTHA (n=NR pts), n pts: 4 • Statistical difference between treatment groups: NR
(continued)	**Nishihara 2006 (N=NR pts), ROBODOC system at <3mo f/u** • rTHA (n=NR pts), n pts: 4 • mTHA (n=NR pts), n pts: 11 • Statistical difference between treatment groups: *P* = 0.6494
Postoperative Thigh Pain	
Han 2019 SR (3 studies w/MA)	**Nakamura 2010 (N=146 pts), f/u timing NR** • rTHA (n=75 pts), n events: 1 • mTHA (n=71 pts), n events: 4 • Odds ratio (95% CI): 0.23 (0.02 to 2.08), NS
(continued)	**Nakamura 2009 (N=118 pts), f/u timing NR** • rTHA (n=40 pts), n events: 0 • mTHA (n=78 pts), n events: 1 • Odds ratio (95% CI): 0.64 (0.03 to 16.01), NS
(continued)	**Nishihara 2006 (N=156 pts), f/u timing NR** • rTHA (n=78 pts), n events: 4 • mTHA (n=78 pts), n events: 11 • Odds ratio (95% CI): 0.33 (0.10 to 1.08), NS

(continued)	**Quantitative synthesis of 3 studies, MA using fixed effects (N=420 pts)**
	• rTHA (n=193 pts), n events: 5
	• mTHA (n=227 pts), n events: 16
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.32 (0.12 to 0.87), favours rTHA
	  ○ *P* = 0.03

Study Citation and Design	Outcomes
Visual Analog Scale (VAS)

Samuel 2021
SR (4 studies)
	Clement 2021 (N=120 pts)
	• Post-operative scores, 10-12 mo f/u:
	  ○ rTHA (n=40 pts), mean (SD): 88.9 (16.1)*
	  ○ mTHA (n=80 pts), mean (SD): 85.5 (21.4)
	  ○ Mean difference between groups (95% CI): 3.40 (-3.45 to 10.25), NS
	
	These findings are presented as reported in the study paper; however, the scoring is not described and appears inconsistent with other reports from the same measure.

(continued)	**Domb 2020 (N=132 pts)**
	• Post-operative scores, min. 5yr f/u
	  ○ rTHA (n=66 pts), mean (SD): 1.27 (2.2)
	  ○ mTHA (n=66 pts), mean (SD): 1.07 (1.87)
	  ○ Mean difference between groups (95% CI): 0.20 (-0.50 to 0.90), NS

(continued)	**Perets 2020 (N=158 pts), MAKO system**
	• Post-operative scores, 2 yr f/u
	  ○ rTHA (n=80 pts), mean (SD): 0.90 (1.80)
	  ○ mTHA (n=78 pts), mean (SD): 1.40 (2.10)
	  ○ Mean difference between groups (95% CI): -0.50 (-1.11 to 0.11), NS

(continued)	**Bargar 2018 (N=67 pts), ROBODOC system**
	• Post-operative scores, 14 yr f/u
	  ○ rTHA (n=45 pts), mean (SD): 4.69 (10.15)
	  ○ mTHA (n=22 pts), mean (SD): 6.42 (10.89)
	  ○ Mean difference between groups (95% CI): -1.73 (-7.16 to 3.70), NS

(continued)	**Quantitative synthesis of 4 studies, MA using fixed effects (N=477 pts)**
	• rTHA (n=231 pts), mean (SD): NR
	• mTHA (n=246 pts), mean (SD): NR
	• Mean difference between treatment groups (95% CI)
	  ○ -0.19 (-0.65 to 0.27), NS
	  ○ *P* = 0.41
	
	Quantitative synthesis of 3 studies investigating MAKO systems, MA using fixed effects (N=410 pts)
	• rTHA (n=186 pts), mean (SD): NR
	• mTHA (n=224 pts), mean (SD): NR
	• Mean difference between treatment groups (95% CI)
	  ○ -0.18 (-0.64 to 0.28), NS
	  ○ *P* = 0.44

Study Citation and Design	Outcomes

(continued)	**1 study investigating ROBODOC system, incorporated into MA using fixed effects (N=67 pts)**
	• rTHA (n=45 pts), mean (SD): NR
	• mTHA (n=22 pts), mean (SD): NR
	• Mean difference between treatment groups (95% CI)
	  ○ -1.73 (-7.16 to 3.70), NS
	  ○ *P* = 0.53

CI = confidence interval; f/u = follow up; HSQ = Health Status Questionnaire; MA = meta-analysis; mHHS = modified Harris Hip Score; mo = month(s); mTHA = manual total hip arthroplasty; NR = not reported; NRS = non-randomized study; NS = not significant; P = probability value; pt/pts = patient(s); rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; SR = systematic review; VAS = Visual Analog Scale; wk = week(s); w/ = with; yr = year(s)

Table 11: Summary of Findings for Clinical Effectiveness — Health Care Utilization

Study Citation and Design	Outcomes
	Duration of Surgical Procedure
Han 2019 SR (8 studies w/MA)	**Kamara 2017, MAKO system (N=296 pts), units of time NR** • rTHA (n=98 pts), mean (SD): 114 (57.5) • mTHA (n=198 pts), mean (SD): 113 (58.1) • Mean difference between treatment groups (95% CI): 1.00 (-12.97 to 14.97), NS
(continued)	**Domb 2014, MAKO system (N=100 pts), units of time NR, (min. 5yr f/u)** • rTHA (n=50 pts), mean (SD): 109.8 (25.5) • mTHA (n=50 pts), mean (SD): 101.5 (21.6) • Mean difference between treatment groups (95% CI): 8.30 (-0.96 to 17.56), NS
(continued)	**Lim 2015, ROBODOC system (N=49 pts), units of time NR** • rTHA (n=24 pts), mean (SD): 103 (71.1) • mTHA (n=25 pts), mean (SD): 78 (39.3) • Mean difference between treatment groups (95% CI): 25.00 (-7.35 to 57.35), NS
(continued)	**Nakamura 2010, MAKO system (N=146 pts), units of time NR** • rTHA (n=75 pts), mean (SD): 120 (27) • mTHA (n=71 pts), mean (SD): 108 (38) • Mean difference between treatment groups (95% CI): 12.00 (1.25 to 22.75), favours mTHA
(continued)	**Nakamura 2009, MAKO system (N=118 pts), units of time NR** • rTHA (n=40 pts), mean (SD): 146 (33) • mTHA (n=78 pts), mean (SD): 121 (26) • Mean difference between treatment groups (95% CI): 25.00 (13.26 to 36.74), favours mTHA
(continued)	**Nishihara 2006, Orthodoc system (N=156 pts), units of time NR** • rTHA (n=78 pts), mean (SD): 122 (268) • mTHA (n=78 pts), mean (SD): 102 (280.8) • Mean difference between treatment groups (95% CI): 20.00 (-66.14 to 106.14), NS

Study Citation and Design	Outcomes
(continued)	**Siebel 2005, CASPAR system (N=71 pts), units of time NR** • rTHA (n=36 pts), mean (SD): 99 (22.4) • mTHA (n=35 pts), mean (SD): 50 (13.1) • Mean difference between treatment groups (95% CI): 49.00 (40.49 to 57.51), favours mTHA
(continued)	**Honl 2003, ROBODOC system (N=141 pts), units of time NR** • rTHA (n=61 pts), mean (SD): 107.1 (29.1) • mTHA (n=80 pts), mean (SD): 82.4 (23.4) • Mean difference between treatment groups (95% CI): 24.70 (15.78 to 33.62), favours mTHA
(continued)	**Quantitative synthesis of 8 studies, MA using random effects (N=1,077 pts), units of time NR** • rTHA (n=462 pts), mean (SD): NR • mTHA (n=615 pts), mean (SD): NR • Mean difference between treatment groups (95% CI) 　○ 20.72 (7.50 to 33.94), favours mTHA 　○ *P* = 0.002
Zhou 2021 Retrospective matched cohort study	• **Surgical time in min (N=158), f/u timing NR** 　○ rTHA (n= 59 pts), mean (SD): 85.3 (25.7) 　○ cTHA (n= 59 pts), mean (SD): 88.6 (41.2) 　○ Statistical difference between treatment groups, NS: *P* = 0.602
Hospital Length-of-Stay
Emara 2021 Retrospective matched cohort study	**Days in hospital** • rTHA (n= 32,898 pts), mean (SD): 2.2 (1.4) • mTHA (n= 4,601,344 pts), mean (SD): 2.3 (1.8) • Mean difference between treatment groups (95% CI) 　○ 0.1 (NR), favours rTHA 　○ *P* < 0.001 　○ Described by the authors as "not… clinically meaningful" (p. 1492)
Kirchner 2021 Retrospective matched cohort study	**Days in hospital** • rTHA (n= 758 pts), mean (SD): 2.69 (1.25) • cTHA (n= 758 pts), mean (SD): 2.82 (1.18) • Statistical difference between treatment groups, favours rTHA: *P* < 0.001
Remily 2021 Retrospective matched cohort study	**Days in hospital** • rTHA (n= 4,630 pts), mean (SD): 3.4 (1.7) • cTHA (n= 4,630 pts), mean (SD): 3.7 (1.4) • Statistical difference between treatment groups, favours rTHA: *P* < 0.001

Study Citation and Design	Outcomes
Discharge Disposition	
Emara 2021 Retrospective matched cohort study	**Routine discharge to home or home health setting** • rTHA (n= 32,092 procedures), n discharges (%): 26,506 (82.6) • mTHA (n= 128,356 procedures), n discharges (%): 104,577 (81.5) • Statistical difference between treatment groups, NS: *P* = 0.205 **Discharge to non-home setting** • rTHA (n= 32,092 procedures), n discharges (%): 5,585 (17.4) • mTHA (n= 128,356 procedures), n discharges (%): 23,727 (18.5) • Statistical difference between treatment groups, NS: *P* = 0.205
Hospital Readmission Rate	
Remily 2021 Retrospective matched cohort study	**Readmissions at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 133 (2.9) • cTHA (n= 4,630 pts), n pts (%): 129 (2.8) • Statistical difference between treatment groups, NS: *P* = 0.806 **Readmissions at 1yr f/u** • rTHA (n= 4,630 pts), n pts (%): 360 (7.8) • cTHA (n= 4,630 pts), n pts (%): 308 (6.6) • Statistical difference between treatment groups, favours cTHA: *P* = 0.037

CI = confidence interval; cTHA = conventional total hip arthroplasty; d = day(s); f/u = follow up; MA = meta-analysis; mo = month(s); mTHA = manual total hip arthroplasty; NR = not reported; NRS = non-randomized study; NS = not significant; P = probability value; pt/pts = patient(s); rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; SR = systematic review; wk = week(s); w/ = with; yr = year(s)

Table 12: Summary of Findings for Clinical Effectiveness — Mortality

Study Citation and Design	Outcomes
Mortality	
Emara 2021 Retrospective matched cohort study	**Patients who died, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 85 (0.1) • Statistical difference between treatment groups, NS: *P* = 0.301
Kirchner 2021 Retrospective matched cohort study	**Patients who died, f/u timing NR** • rTHA (n= 758 pts), n: 0 • cTHA (n= 758 pts), n: 0 • Statistical difference between treatment groups: NA

f/u = follow up; mTHA = manual total hip arthroplasty; NA = not applicable; NR = not reported; NRS = non-randomized study; NS = not significant; P = probability value; pt/pts = patient(s); rTHA = robotic-assisted total hip arthroplasty

Table 13: Summary of Findings for Clinical Effectiveness — Safety

Study Citation and Design	Outcomes
Complications i.e., any/all/combined	
Emara 2021 Retrospective matched cohort study	**Non-mechanical complications, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 11,012 (34.3) • mTHA (n= 128,356 procedures), n (%): 42,090 (32.8) • Statistical difference between treatment groups, NS: *P* = 0.296
Kirchner 2021 Retrospective matched cohort study	**Major perioperative complications** • rTHA (n= 758 pts), n (%): 17 (2.24) • cTHA (n= 758 pts), n (%): 25 (3.30) • Statistical difference between treatment groups, NS: *P* = 0.729 • Adjusted odds ratio (95% CI), NS: 0.732 (0.320 to 1.672) **Minor perioperative complications** • rTHA (n= 758 pts), n (%): 164 (21.64) • cTHA (n= 758 pts), n (%): 97 (12.8) • Statistical difference between treatment groups, favours cTHA: *P* = 0.004 • Adjusted odds ratio (95% CI), NS: 1.294 (0.938 to 1.785)

Ng 2021
SR (MA of 8 NRS)
	MA using random effects (N=3,208 pts), f/u either NR or ranging from 90d to min. 5yr
	• rTHA (n=1,385 pts), n total events: 10
	• mTHA (n=1,823 pts), n total events: 25
	• Pooled odds ratio between treatment groups (95% CI)
	    ◦ 0.61 (0.30 to 1.24), NS
	    ◦ *P* = 0.81

Han 2019 SR (6 studies w/MA)	**Kamara 2017 (N=296 pts), f/u timing NR** • rTHA (n=98 pts), n events: 2 • mTHA (n=198 pts), n events: 7 • Odds ratio (95% CI): 0.57 (0.12 to 2.79), NS
(continued)	**Lim 2015, ROBODOC system (N=49 pts), f/u timing NR** • rTHA (n=24 pts), mean (SD): 0 • mTHA (n=25 pts), mean (SD): 2 • Odds ratio (95% CI): 0.19 (0.01 to 4.21), NS
(continued)	**Nakamura 2010 (N=146 pts), f/u timing NR** • rTHA (n=75 pts), n events: 7 • mTHA (n=71 pts), n events: 10 • Odds ratio (95% CI): 0.63 (0.23 to 1.75), NS
(continued)	**Nakamura 2009 (N=118 pts), f/u timing NR** • rTHA (n=40 pts), n events: 3 • mTHA (n=78 pts), n events: 5 • Odds ratio (95% CI): 1.18 (0.27 to 5.23), NS

Study Citation and Design	Outcomes
(continued)	**Siebel 2005 (N=71 pts), f/u timing NR** • rTHA (n=36 pts), n events: 4 • mTHA (n=35 pts), n events: 3 • Odds ratio (95% CI): 1.33 (0.28 to 6.44), NS
(continued)	**Bargar 1998 (N=127 pts), f/u timing NR** • rTHA (n=65 pts), n events: 9 • mTHA (n=62 pts), n events: 9 • Odds ratio (95% CI): 0.95 (0.35 to 2.57), NS

(continued)	**Quantitative synthesis of 6 studies, MA using fixed effects (N=907 pts)**
	• rTHA (n=388 pts), n events: 26
	• mTHA (n=519 pts), n events: 36
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.83 (0.49 to 1.40), NS
	  ○ *P* = 0.48

Cardiovascular/Blood complications
Emara 2021 Retrospective matched cohort study	**Cardiac complications, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 3,385 (10.5) • mTHA (n= 128,356 procedures), n (%): 12,989 (10.1) • Statistical difference between treatment groups, NS: *P* = 0.423 **Peripheral vascular, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 0 (0) • Statistical difference between treatment groups: NA **Hematoma/seroma, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 55 (0.2) • mTHA (n= 128,356 procedures), n (%): 333 (0.3) • Statistical difference between treatment groups, NS: *P* = 0.185 **Acute post hemorrhagic anemia, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 8,077 (25.2) • mTHA (n= 128,356 procedures), n (%): 30,774 (24.0) • Statistical difference between treatment groups, NS: *P* = 0.426 **Pulmonary embolism and infarction, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 30 (0.1) • mTHA (n= 128,356 procedures), n (%): 30 (0) • Statistical difference between treatment groups, favours mTHA: *P* = 0.015

Study Citation and Design	Outcomes
(continued)	**Lower limb DVT, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 30 (0.1) • mTHA (n= 128,356 procedures), n (%): 110 (0) • Statistical difference between treatment groups, NS: P = 0.849 **Transfusion of blood and products, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 15 (0) • mTHA (n= 128,356 procedures), n (%): 540 (0.4) • Statistical difference between treatment groups, favours rTHA: P < 0.001
Remily 2021 Retrospective matched cohort study	**Patients who experienced blood transfusion at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 204 (4.4) • cTHA (n= 4,630 pts), n pts (%): 147 (3.2) • Statistical difference between treatment groups, favours cTHA: P = 0.001 **Patients who experienced cerebrovascular accident at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 32 (0.7) • cTHA (n= 4,630 pts), n pts (%): 26 (0.6) • Statistical difference between treatment groups, NS: P = 0.510 **Patients who experienced deep vein thrombosis at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 74 (1.6) • cTHA (n= 4,630 pts), n pts (%): 67 (1.5) • Statistical difference between treatment groups, NS: P = 0.611 **Patients who experienced pulmonary embolus at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 50 (1.1) • cTHA (n= 4,630 pts), n pts (%): 36 (0.8) • Statistical difference between treatment groups, NS: P = 0.159
Zhou 2021 Retrospective matched cohort study	**Blood loss in mL (f/u timing NR)** • rTHA (n= 59 pts), mean (SD): 1,050 (399.6) • cTHA (n= 59 pts), mean (SD): 935.9 (344.4) • Statistical difference between treatment groups, NS: P = 0.098
Dislocations
Emara 2021 Retrospective matched cohort study	**Dislocation of prosthetic joint (f/u timing NR)** • rTHA (n= 32,092 procedures), n (%): 45 (0.1) • mTHA (n= 128,356 procedures), n (%): 1,050 (0.8) • Statistical difference between treatment groups, favours rTHA: P < 0.001
Ng 2021 SR (7 NRS with MA)	**Domb 2020 (N=132 pts), min. 5yr f/u** • rTHA (n=66 pts), n events: 1 • mTHA (n=66 pts), n events: 0 • Odds ratio (95% CI): 3.05 (0.12 to 76.13), NS

Study Citation and Design	Outcomes
(continued)	**Kong 2020 (N=230 pts), 3mo f/u** • rTHA (n=106 pts), n events: 0 • mTHA (n=124 pts), n events: 2 • Odds ratio (95% CI): 0.23 (0.01 to 4.85), NS
(continued)	**Kayani 2019 (N=100 pts), 90 days f/u** • rTHA (n=50 pts), n events: 0 • mTHA (n=50 pts), n events: 0 • Odds ratio (95% CI): NE
(continued)	**Lawson 2019 (N=99 pts), 24mo f/u** • rTHA (n=49 pts), n events: 0 • mTHA (n=50 pts), n events: 0 • Odds ratio (95% CI) : NE
(continued)	**Heng 2018 (N=90 pts), f/u timing NR** • rTHA (n=45 pts), n events: 0 • mTHA (n=45 pts), n events: 0 • Odds ratio (95% CI) : NE
(continued)	**Illgen 2017 (N=200 pts), 24mo f/u** • rTHA (n=100 pts), n events: 0 • mTHA (n=100 pts), n events: 3 • Odds ratio (95% CI): 0.14 (0.01 to 2.72), NS
(continued)	**Kamara 2016 (N=296 pts), f/u timing NR** • rTHA (n=98 pts), n events: 1 • mTHA (n=198 pts), n events: 1 • Odds ratio (95% CI): 2.03 (0.13 to 32.82), NS

(continued)	**Quantitative synthesis of 7 studies, MA using random effects (N=1,147 pts)**
	• rTHA (n=514 pts), n events: 2
	• mTHA (n=633 pts), n events: 6
	• Pooled odds ratio between treatment groups (95% CI)
	   ○ 0.66 (0.15 to 3.02), NS
	   ○ $P = 0.60$

Study Citation and Design	Outcomes
Remily 2021 Retrospective matched cohort study	**Patients who experienced dislocation at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 66 (1.4) • cTHA (n= 4,630 pts), n pts (%): 69 (1.5) • Statistical difference between treatment groups, NS: P = 0.862 **Patients who experienced dislocation at 1yr f/u** • rTHA (n= 4,630 pts), n pts (%): 83 (1.8) • cTHA (n= 4,630 pts), n pts (%): 90 (1.9) • Statistical difference between treatment groups, NS: P = 0.645 **Patients who experienced dislocation at 5yr f/u** • rTHA (n= 4,630 pts), n pts (%): 105 (2.3) • cTHA (n= 4,630 pts), n pts (%): 121 (2.6) • Statistical difference between treatment groups, NS: P = 0.312
Samuel 2021 SR (8 studies w/MA)	**Domb 2020, MAKO system (N=132 pts) at 5yr f/u** • rTHA (n=66 pts), n events: 1 • mTHA (n=66 pts), n events: 0 • Odds ratio (95% CI) 　○ 3.05 (0.12 to 76.13), NS 　○ P = NR
(continued)	**Kong 2020b, MAKO system (N=115 pts) at 3mo f/u** • rTHA (n=53 pts), n events: 0 • mTHA (n=62 pts), n events: 2 • Odds ratio (95% CI) 　○ 0.23 (0.01 to 4.82), NS 　○ P = NR
(continued)	**Bargar 2018, ROBODOC system (N=67 pts) at 14yr f/u** • rTHA (n=22 pts), n events: 1 • mTHA (n=45 pts), n events: 0 • Odds ratio (95% CI) 　○ 6.35 (0.25 to 162.34), NS 　○ P = NR
(continued)	**Kamara 2017, MAKO system (N=296 pts) at 1yr f/u** • rTHA (n=98 pts), n events: 1 • mTHA (n=198 pts), n events: 1 • Odds ratio (95% CI) 　○ 2.03 (0.13 to 32.82), NS 　○ P > 0.05

Study Citation and Design	Outcomes

(continued)	**Bukowski 2016, MAKO system (N=200 pts) at 1yr f/u**
	• rTHA (n=100 pts), n events: 0
	• mTHA (n=100 pts), n events: 3
	• Odds ratio (95% CI)
	  ○ 0.14 (0.01 to 2.72), NS
	  ○ $P = 0.101$

(continued)	**Nakamura 2010, ROBODOC system (N=146 pts) at 52mo f/u**
	• rTHA (n=75 pts), n events: 4
	• mTHA (n=71 pts), n events: 1
	• Odds ratio (95% CI)
	  ○ 3.94 (0.43 to 36.17), NS
	  ○ $P > 0.05$

(continued)	**Honl 2003, ROBODOC system (N=141 pts) at 2yr f/u**
	• rTHA (n=61 pts), n events: 11
	• mTHA (n=80 pts), n events: 3
	• Odds ratio (95% CI)
	  ○ 5.65 (1.50 to 21.25), favours mTHA
	  ○ $P < 0.001$

(continued)	**Bargar 1998, ROBODOC system (N=93 pts) at 2yr f/u**
	• rTHA (n=48 pts), n events: 4
	• mTHA (n=45 pts), n events: 4
	• Odds ratio (95% CI)
	  ○ 0.93 (0.22 to 3.97), NS
	  ○ $P > 0.05$

(continued)	**Quantitative synthesis of 8 studies, MA using random effects (N=1,190 pts)**
	• rTHA (n=523 pts), n events: 22
	• mTHA (n=667 pts), n events: 14
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 1.81 (0.71 to 4.58), NS
	  ○ $P = 0.11$
	
	Quantitative synthesis of 4 studies investigating the MAKO system, MA using random effects (N=743 pts)
	• rTHA (n=317 pts), n events: 2
	• mTHA (n=426 pts), n events: 6
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.66 (0.15 to 3.02), NS
	  ○ $P = 0.60$

Study Citation and Design	Outcomes

(continued)	**Quantitative synthesis of 4 studies investigating ROBODOC, MA using random effects (N=447 pts)**
	• rTHA (n=206 pts), n events: 20
	• mTHA (n=241 pts), n events: 8
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 2.87 (1.07 to 7.70), favour mTHA
	  ○ *P* = 0.04

Zhou 2021 Retrospective matched cohort study	**Patients experiencing dislocation (N=158), f/u timing NR** • rTHA (n= 59 pts), n pts: 0 • cTHA (n= 59 pts), n pts: 1 • Statistical difference between treatment groups: NR
Han 2019 SR (6 studies w/MA)	**Kamara 2017 (N=296 pts), f/u timing NR** • rTHA (n=98 pts), n events: 1 • mTHA (n=198 pts), n events: 1 • Odds ratio (95% CI): 2.03 (0.13 to 32.82), NS
(continued)	**Nakamura 2010 (N=146 pts), f/u timing NR** • rTHA (n=75 pts), n events: 4 • mTHA (n=71 pts), n events: 1 • Odds ratio (95% CI): 3.94 (0.43 to 36.17), NS
(continued)	**Nakamura 2009 (N=118 pts), f/u timing NR** • rTHA (n=40 pts), n events: 0 • mTHA (n=78 pts), n events: 2 • Odds ratio (95% CI): 0.38 (0.02 to 8.06), NS
(continued)	**Siebel 2005 (N=71 pts), f/u timing NR** • rTHA (n=36 pts), n events: 2 • mTHA (n=35 pts), n events: 1 • Odds ratio (95% CI): 2.00 (0.17 to 23.11), NS
(continued)	**Honl 2003 (N=141 pts), f/u timing NR** • rTHA (n=61 pts), n events: 11 • mTHA (n=80 pts), n events: 3 • Odds ratio (95% CI): 5.65 (1.50 to 21.25), favours mTHA
(continued)	**Bargar 1998 (N=127 pts), f/u timing NR** • rTHA (n=65 pts), n events: 4 • mTHA (n=62 pts), n events: 4 • Odds ratio (95% CI): 0.95 (0.23 to 3.98), NS

Study Citation and Design	Outcomes

(continued)	**Quantitative synthesis of 6 studies, MA using fixed effects (N=899 pts)**
	• rTHA (n=375 pts), n events: 22
	• mTHA (n=524 pts), n events: 12
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 2.28 (1.12 to 4.67), favours mTHA
	  ○ P = 0.02

Fractures	
Emara 2021 Retrospective matched cohort study	**Periprosthetic fracture, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 70 (0.2) • mTHA (n= 128,356 procedures), n (%): 675 (0.5) • Statistical difference between treatment groups, favours rTHA: P < 0.001
Remily 2021 Retrospective matched cohort study	**Periprosthetic fractures at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 80 (1.7) • cTHA (n= 4,630 pts), n pts (%): 70 (1.5) • Statistical difference between treatment groups, NS: P = 0.459 **Periprosthetic fractures at 1yr f/u** • rTHA (n= 4,630 pts), n pts (%): 98 (2.1) • cTHA (n= 4,630 pts), n pts (%): 90 (1.9) • Statistical difference between treatment groups, NS: P = 0.606 **Periprosthetic fractures at 5yr f/u** • rTHA (n= 4,630 pts), n pts (%): 127 (2.7) • cTHA (n= 4,630 pts), n pts (%): 122 (2.6) • Statistical difference between treatment groups, NS: P = 0.797

Samuel 2021
SR (6 studies w/o MA; all findings described narratively only)
	Kamara 2017, MAKO system (N=NR pts), 1yr f/u
	• Intraoperative fracture
	  ○ rTHA (n=NR pts), n events: 0
	  ○ mTHA (n=NR pts), n events: 3
	  ○ Statistical difference between treatment groups: P > 0.05

(continued)	**Lim 2015, ROBODOC system (N=NR pts), 2yr f/u**
	• Intraoperative fracture
	  ○ rTHA (n=NR pts), %*: 0
	  ○ mTHA (n=NR pts), %*: 8
	  ○ Statistical difference between treatment groups, favours rTHA: P < 0.05**
	The sources of analysis (e.g., number of patients, number of events) was not specified.
	**This finding is presented as reported in the study paper; however, this finding was narratively described in the review paper as being not statistically significant.

Study Citation and Design	Outcomes

(continued)	**Nakamura 2010, ROBODOC system (N=NR pts), 1-5yr f/u**
	• Periprosthetic fracture
	  ○ rTHA (n=NR pts), n events: 0
	  ○ mTHA (n=NR pts), n events: 1
	  ○ Statistical difference between treatment groups: NR

(continued)	**Hananouchi 2007, MAKO system (N=NR pts), 2yr f/u**
	• Intraoperative fracture
	  ○ rTHA (n=NR pts), n events: 0
	  ○ mTHA (n=NR pts), n events: 2
	  ○ Statistical difference between treatment groups, NS: *P* = 0.21

(continued)	**Nishihara 2006, ROBODOC system (N=NR pts), 2yr f/u**
	• Intraoperative fracture
	  ○ rTHA (n=NR pts), n events: 0
	  ○ mTHA (n=NR pts), n events: 5
	  ○ Statistical difference between treatment groups, favours rTHA: *P* < 0.05

(continued)	**Bargar 1998, ROBODOC system (N=NR pts) at 14yr f/u**
	• Intraoperative fracture
	  ○ rTHA (n=NR pts), n events: 0
	  ○ mTHA (n=NR pts), n events: 3
	  ○ Statistical difference between treatment groups: NR
	• Periprosthetic fracture
	  ○ rTHA (n=NR pts), n events: 1
	  ○ mTHA (n=NR pts), n events: 1
	  ○ Statistical difference between treatment groups: NR

Zhou 2021 Retrospective matched cohort study	**Patients experiencing femoral side fracture intraoperatively (N=158), f/u timing NR** • rTHA (n= 59 pts), n pts: 0 • cTHA (n= 59 pts), n pts: 1 • Statistical difference between treatment groups: NR

Study Citation and Design	Outcomes
	Heterotopic Ossification (HO)

Study Citation and Design	Outcomes
Samuel 2021 SR (2 studies w/o MA; all findings described narratively only)	**Nakamura 2010, ROBODOC system (N=NR pts) 67mo f/u** • Grade 1 HO ◦ rTHA (n=NR pts), incidence rate %: 11 ◦ mTHA (n=NR pts), incidence rate %: 11 ◦ Statistical difference between treatment groups: NS • Grade 2 HO ◦ rTHA (n=NR pts), incidence rate %: 7 ◦ mTHA (n=NR pts), incidence rate %: 1 ◦ Statistical difference between treatment groups: NS • Grade 3 HO ◦ rTHA (n=NR pts), incidence rate %: 0 ◦ mTHA (n=NR pts), incidence rate %: 3 ◦ Statistical difference between treatment groups: NS **Nakamura 2010, ROBODOC system (N=NR pts) longer-term f/u (not otherwise described)** • Any Grade HO ◦ rTHA (n=NR pts), incidence rate %: NR ◦ mTHA (n=NR pts), incidence rate %: NR ◦ Statistical difference between treatment groups, NS: 0.20
(continued)	**Honl 2003, ROBODOC system (N=NR pts) 24mo f/u** • Grade 2 or 3 HO ◦ rTHA (n=NR pts), incidence rate %: 10 ◦ mTHA (n=NR pts), incidence rate %: 10 ◦ Statistical difference between treatment groups, NS: 0.31
Han 2019 SR (4 studies w/MA)	**Nakamura 2010 (N=146 pts), f/u timing NR** • rTHA (n=75 pts), n events: 5 • mTHA (n=71 pts), n events: 3 • Odds ratio (95% CI): 1.62 (0.37 to 7.04), NS
(continued)	**Hananouchi 2007 (N=58 pts), f/u timing NR** • rTHA (n=31 pts), n events: 1 • mTHA (n=27 pts), n events: 1 • Odds ratio (95% CI): 0.87 (0.05 to 14.56), NS
(continued)	**Siebel 2005 (N=71 pts), f/u timing NR** • rTHA (n=36 pts), n events: 11 • mTHA (n=35 pts), n events: 6 • Odds ratio (95% CI): 2.13 (0.69 to 6.58), NS
(continued)	**Honl 2003 (N=141 pts), f/u timing NR** • rTHA (n=61 pts), n events: 6 • mTHA (n=80 pts), n events: 8 • Odds ratio (95% CI): 0.98 (0.32 to 2.99), favours mTHA

Study Citation and Design	Outcomes

(continued)	**Quantitative synthesis of 4 studies, MA using fixed effects (N=416 pts)**
	• rTHA (n=203 pts), n events: 23
	• mTHA (n=213 pts), n events: 18
	• Pooled odds ratio between treatment groups (95% CI)
	   ◦ 1.44 (0.74 to 2.80), NS
	   ◦ P = 0.29

Infection	
Emara 2021 Retrospective matched cohort study	**Post-operative infection, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 50 (0) • Statistical difference between treatment groups: NR **Disruption of wound infection, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 15 (0) • Statistical difference between treatment groups: P = 0.803 **Infection/inflammatory reaction, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 0 (0) • Statistical difference between treatment groups: NA **Urinary tract infection, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 351 (1.1) • mTHA (n= 128,356 procedures), n (%): 1,646 (1.3) • Statistical difference between treatment groups, NS: P = 0.223 **Cellulitis/abscess, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 30 (0) • Statistical difference between treatment groups, NS: P = 0.706
Ng 2021 SR (5 NRS with MA)	**Domb 2020 (N=132 pts), min. 5yr f/u** • rTHA (n=66 pts), n events: 2 • mTHA (n=66 pts), n events: 0 • Odds ratio (95% CI): 5.16 (0.24 to 109.46), NS
(continued)	**Kong 2020b (N=230 pts), 3mo f/u** • rTHA (n=106 pts), n events: 0 • mTHA (n=124 pts), n events: 0 • Odds ratio (95% CI): NE
(continued)	**Heng 2018 (N=90 pts), f/u timing NR** • rTHA (n=45 pts), n events: 0 • mTHA (n=45 pts), n events: 0 • Odds ratio (95% CI): 0.33 (0.01 to 8.22)

Study Citation and Design	Outcomes
(continued)	**Illgen 2017 (N=200 pts), 24mo f/u** • rTHA (n=100 pts), n events: 0 • mTHA (n=100 pts), n events: 0 • Odds ratio (95% CI): NE
(continued)	**Kamara 2016 (N=296 pts), f/u timing NR** • rTHA (n=98 pts), n events: 0 • mTHA (n=198 pts), n events: 1 • Odds ratio (95% CI): 0.67 (0.03 to 16.56), NS

(continued)	**Quantitative synthesis of 4 studies, MA using random effects (N=948 pts)**
	• rTHA (n=415 pts), n events: 2
	• mTHA (n=533 pts), n events: 2
	• Pooled odds ratio between treatment groups (95% CI)
	   ◦ 1.10 (0.18 to 6.83), NS
	   ◦ $P = 0.92$

Remily 2021 Retrospective matched cohort study	**Prosthetic joint infections at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 60 (1.3) • cTHA (n= 4,630 pts), n pts (%): 43 (0.9) • Statistical difference between treatment groups, NS: $P = 0.113$ **Prosthetic joint infections at 1yr f/u** • rTHA (n= 4,630 pts), n pts (%): 67 (1.5) • cTHA (n= 4,630 pts), n pts (%): 54 (1.2) • Statistical difference between treatment groups, NS: $P = 0.272$ **Prosthetic joint infections at 5yr f/u** • rTHA (n= 4,630 pts), n pts (%): 82 (1.8) • cTHA (n= 4,630 pts), n pts (%): 86 (1.9) • Statistical difference between treatment groups, NS: $P = 0.815$
Samuel 2021 SR (5 studies w/o MA; all findings described narratively only)	**Domb 2020, MAKO system (N=NR pts) min. 5yr f/u** • rTHA (n=NR pts), n events: 2 • mTHA (n=NR pts), n events: 0 • Statistical difference between treatment groups: NR
(continued)	**Kong 2020b, MAKO system (N=NR pts), 3mo f/u** • rTHA (n=NR pts), n events: 0 • mTHA (n=NR pts), n events: 1 • Statistical difference between treatment groups: NR
(continued)	**Perets 2020, MAKO system (N=NR pts), 2yr f/u** • rTHA (n=NR pts), n events: 6 • mTHA (n=NR pts), n events: 2 • Statistical difference between treatment groups, NS: $P = 0.15$

Study Citation and Design	Outcomes
(continued)	**Kamara 2017, MAKO system (N=NR pts), 1yr f/u** • rTHA (n=NR pts), n events: 0 • mTHA (n=NR pts), n events: 2 • Statistical difference between treatment groups: NR
(continued)	**Honl 2003, ROBODOC system (N=NR pts), 2yr f/u** • rTHA (n=NR pts), n events: 0 • mTHA (n=NR pts), n events: 2 • Statistical difference between treatment groups: NR
	Intraoperative complications
Ng 2021 SR (2 NRS with MA)	**Kamara 2016 (N=296 pts)** • rTHA (n=98 pts), n events: 0 • mTHA (n=198 pts), n events: 3 • Odds ratio (95% CI): 0.28 (0.01 to 5.54), NS
(continued)	**Heng 2018 2018 (N=90 pts)** • rTHA (n=45 pts), n events: 1 • mTHA (n=45 pts), n events: 3 • Odds ratio (95% CI): 0.32 (0.03 to 3.18), NS

(continued)	**Quantitative MA using random effects (N=386 pts)**
	• rTHA (n=143 pts), n total events: 1
	• mTHA (n=243 pts), n total events: 6
	• Pooled odds ratio between treatment groups (95% CI)
	  ◦ 0.30 (0.05 to 1.88), NS
	  ◦ $P = 0.20$

Han 2019 SR (9 studies w/MA)	**Kamara 2017 (N=296 pts)** • rTHA (n=98 pts), n events: 0 • mTHA (n=198 pts), n events: 3 • Odds ratio (95% CI): 0.28 (0.01 to 5.54), NS
(continued)	**Lim 2015 (N=49 pts)** • rTHA (n=24 pts), n events: 0 • mTHA (n=25 pts), n events: 1 • Odds ratio (95% CI): 0.19 (0.01 to 4.21), NS
(continued)	**Domb 2014 (N=100 pts)** • rTHA (n=50 pts), n events: 1 • mTHA (n=50 pts), n events: 0 • Odds ratio (95% CI): 3.06 (0.12 to 76.95), NS
(continued)	**Nakamura 2010 (N=146 pts)** • rTHA (n=75 pts), n events: 0 • mTHA (n=71 pts), n events: 5 • Odds ratio (95% CI): 0.08 (0.00 to 1.48), NS

Study Citation and Design	Outcomes
(continued)	**Nakamura 2009 (N=118 pts)** • rTHA (n=40 pts), n events: 2 • mTHA (n=78 pts), n events: 0 • Odds ratio (95% CI): 10.19 (0.48 to 217.59), NS
(continued)	**Hananouchi 2007 (N=58 pts)** • rTHA (n=31 pts), n events: 0 • mTHA (n=27 pts), n events: 2 • Odds ratio (95% CI): 0.16 (0.01 to 3.53), NS
(continued)	**Nishihara 2006 (N=156 pts)** • rTHA (n=78 pts), n events: 0 • mTHA (n=78 pts), n events: 5 • Odds ratio (95% CI): 0.09 (0.00 to 1.57), NS
(continued)	**Siebel 2005 (N=71 pts)** • rTHA (n=36 pts), n events: 0 • mTHA (n=35 pts), n events: 2 • Odds ratio (95% CI): 0.18 (0.01 to 3.96), NS
(continued)	**Bargar 1998 (N=127 pts)** • rTHA (n=65 pts), n events: 0 • mTHA (n=62 pts), n events: 3 • Odds ratio (95% CI): 0.13 (0.01 to 2.56), NS

(continued)	**Quantitative synthesis of 9 studies, MA using fixed effects (N=1,121 pts)**
	• rTHA (n=497 pts), n events: 3
	• mTHA (n=624 pts), n events: 22
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.32 (0.14 to 0.72), favours rTHA
	  ○ P = 0.006

Study Citation and Design	Outcomes
Mechanical Complications	
Emara 2021 Retrospective matched cohort study	**Any, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 170 (0.5) • mTHA (n= 128,356 procedures), n (%): 4,035 (3.1) • Statistical difference between treatment groups, favours rTHA: *P* < 0.001 **Mechanical loosening, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 15 (0) • mTHA (n= 128,356 procedures), n (%): 1,075 (0.8) • Statistical difference between treatment groups, favours rTHA: *P* < 0.001 **Prosthetic joint implant failure, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 120 (0.1) • Statistical difference between treatment groups, favours rTHA: *P* < 0.001 **Periprosthetic osteolysis, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 295 (0.2) • Statistical difference between treatment groups: NR **Unspecified/other mechanical, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 40 (0.1) • mTHA (n= 128,356 procedures), n (%): 1,009 (0.8) • Statistical difference between treatment groups, favours rTHA: *P* < 0.001
Nerve Injury	
Samuel 2021 SR (3 studies w/o MA; findings reported narratively only)	**Domb 2020, MAKO system (N=NR pts) min. 5yr f/u** • rTHA (n=NR pts), n events: 0 • mTHA (n=NR pts), n events: 4 • Statistical difference between treatment groups: NR
(continued)	**Perets 2020, MAKO system (N=NR pts) 2 yr f/u** • rTHA (n=NR pts), n events: 0 • mTHA (n=NR pts), n events: 3 • Statistical difference between treatment groups: NR
(continued)	**Honl 2003, ROBODOC system (N=NR pts), 2yr f/u** • rTHA (n=NR pts), n events: 4 • mTHA (n=NR pts), n events: 1 • Statistical difference between treatment groups: *P* = 0.31
Han 2019 SR (3 studies w/MA)	**Nakamura 2009 (N=118 pts), f/u timing NR** • rTHA (n=40 pts), n events: 1 • mTHA (n=78 pts), n events: 0 • Odds ratio (95% CI): 5.96 (0.24 to 149.72), NS

Study Citation and Design	Outcomes
(continued)	**Siebel 2005 (N=71 pts), f/u timing NR** • rTHA (n=36 pts), n events: 1 • mTHA (n=35 pts), n events: 0 • Odds ratio (95% CI): 3.00 (0.12 to 76.16), NS
(continued)	**Bargar 1998 (N=127 pts), f/u timing NR** • rTHA (n=65 pts), n events: 2 • mTHA (n=62 pts), n events: 0 • Odds ratio (95% CI): 4.92 (0.23 to 104.58), NS

(continued)	**Quantitative synthesis of 3 studies, MA using fixed effects (N=316 pts)**
	• rTHA (n=141 pts), n events: 4
	• mTHA (n=175 pts), n events: 0
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 4.47 (0.72 to 27.90), NS
	  ○ P = 0.11

Other/unspecified systemic and/or organ complications	
Emara 2021 Retrospective matched cohort study	**Central nervous system, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 120 (0.4) • mTHA (n= 128,356 procedures), n (%): 509 (0.4) • Statistical difference between treatment groups, NS: P = 0.796 **Respiratory, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 140 (0.4) • mTHA (n= 128,356 procedures), n (%): 445 (0.3) • Statistical difference between treatment groups, NS: P = 0.297 **Gastrointestinal, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 69 (0.2) • mTHA (n= 128,356 procedures), n (%): 240 (0.2) • Statistical difference between treatment groups, NS: P = 0.644 **Genitourinary, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 100 (0.3) • mTHA (n= 128,356 procedures), n (%): 365 (0.3) • Statistical difference between treatment groups, NS: P = 0.737 **Other organ-specific complications, f/u timing NR** • rTHA (n= 32,092 procedures), n (%): 0 (0) • mTHA (n= 128,356 procedures), n (%): 15 (0) • Statistical difference between treatment groups, NS: P = 0.261

Study Citation and Design	Outcomes
Remily 2021 Retrospective matched cohort study	**Patients who experienced pneumonia at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 50 (1.1) • cTHA (n= 4,630 pts), n pts (%): 36 (0.8) • Statistical difference between treatment groups, NS: P = 0.159 **Patients who experienced respiratory failure at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 29 (0.6) • cTHA (n= 4,630 pts), n pts (%): 36 (0.8) • Statistical difference between treatment groups, NS: P = 0.455 **Patients who experienced aseptic loosening at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 12 (0.3) • cTHA (n= 4,630 pts), n pts (%): 13 (0.3) • Statistical difference between treatment groups, NS: P = 0.999 **Patients who experienced aseptic loosening at 1yr f/u** • rTHA (n= 4,630 pts), n pts (%): 34 (0.7) • cTHA (n= 4,630 pts), n pts (%): 39 (0.8) • Statistical difference between treatment groups, NS: P = 0.638 **Patients who experienced aseptic loosening at 5yr f/u** • rTHA (n= 4,630 pts), n pts (%): 77 (1.7) • cTHA (n= 4,630 pts), n pts (%): 81 (1.8) • Statistical difference between treatment groups, NS: P = 0.810
	Revisions
Ng 2021 SR (4 NRS with MA)	**Domb 2020 (N=132 pts), min. 5yr f/u** • rTHA (n=66 pts), n events: 3 • mTHA (n=66 pts), n events: 6 • Odds ratio (95% CI): 0.48 (0.11 to 1.99), NS
(continued)	**Bukowski 2019 (N=200 pts), 374 days [IQR 359 to 500] for rTHA pts; 367 days [IQR 353 to 387] for mTHA pts f/u** • rTHA (n=100pts), n events: 0 • mTHA (n=100 pts), n events: 1 • Odds ratio (95% CI): 0.33 (0.01 to 8.20)
(continued)	**Lawson (N=99 pts), 24mo f/u** • rTHA (n=49 pts), n events: 0 • mTHA (n=50 pts), n events: 0 • Odds ratio (95% CI): NE
(continued)	**Kamara 2016 (N=296 pts), f/u timing NR** • rTHA (n=98 pts), n events: 2 • mTHA (n=198 pts), n events: 4 • Odds ratio (95% CI): 1.01 (0.18 to 5.61), NS

Study Citation and Design	Outcomes

(continued)	**MA using random effects (N=727 pts)**
	• rTHA (n=313 pts), n events: 5
	• mTHA (n=414 pts), n events: 11
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.60 (0.21 to 1.71), NS
	  ○ P = 0.34

Remily 2021 Retrospective matched cohort study	**Patients who required revision surgery at 90d f/u** • rTHA (n= 4,630 pts), n pts (%): 60 (1.3) • cTHA (n= 4,630 pts), n pts (%): 57 (1.2) • Statistical difference between treatment groups, NS: P = 0.852 **Patients who required revision surgery at 1yr f/u** • rTHA (n= 4,630 pts), n pts (%): 91 (2.0) • cTHA (n= 4,630 pts), n pts (%): 93 (2.0) • Statistical difference between treatment groups, NS: P = 0.941 **Patients who required revision surgery at 5yr f/u** • rTHA (n= 4,630 pts), n pts (%): 130 (2.8) • cTHA (n= 4,630 pts), n pts (%): 142 (3.1) • Statistical difference between treatment groups, NS: P = 0.498

Samuel 2021
SR (7 studies w/MA)
	Singh 2021, MAKO system (N=556 pts) at 90d f/u
	• rTHA (n=132 pts), n events: 1
	• mTHA (n=424 pts), n events: 12
	• Odds ratio (95% CI)
	  ○ 0.26 (0.03 to 2.03), NS
	  ○ P = 0.839

(continued)	**Domb 2020, MAKO system (N=132 pts) at 5yr f/u**
	• rTHA (n=66 pts), n events: 3
	• mTHA (n=66 pts), n events: 5
	• Odds ratio (95% CI)
	  ○ 0.58 (0.13 to 2.54), NS
	  ○ P = 0.479

(continued)	**Perets 2020, MAKO system (N=158 pts) at 2yr f/u**
	• rTHA (n=80 pts), n events: 1
	• mTHA (n=78 pts), n events: 3
	• Odds ratio (95% CI):
	  ○ 0.32 (0.03 to 3.11), NS
	  ○ P = 0.621

Study Citation and Design	Outcomes

(continued)	**Bargar 2018, ROBODOC system (N=67 pts) at 14yr f/u**
	• rTHA (n=22 pts), n events: 4
	• mTHA (n=45 pts), n events: 6
	• Odds ratio (95% CI)
	  ○ 0.26 (0.06 to 1.05), NS
	  ○ $P > 0.05$

(continued)	**Kamara 2017, MAKO system (N=296 pts) at 1yr f/u**
	• rTHA (n=98 pts), n events: 2
	• mTHA (n=198 pts), n events: 2
	• Odds ratio (95% CI):
	  ○ 2.04 (0.28 to 14.72), NS
	  ○ $P > 0.05$

(continued)	**Bukowski 2016, MAKO system (N=200 pts) at 1yr f/u**
	• rTHA (n=100 pts), n events: 0
	• mTHA (n=100 pts), n events: 1
	• Odds ratio (95% CI)
	  ○ 0.33 (0.01 to 8.20), NS
	  ○ $P = 0.101$

(continued)	**Honl 2003, ROBODOC system (N=141 pts) at 2yr f/u**
	• rTHA (n=61 pts), n events: 9
	• mTHA (n=80 pts), n events: 2
	• Odds ratio (95% CI)
	  ○ 6.75 (1.40 to 32.50), favours mTHA
	  ○ $P = 0.007$

(continued)	**Quantitative synthesis of 7 studies, MA using random effects (N=1,550 pts)**
	• rTHA (n=582 pts), n events: 20
	• mTHA (n=968 pts), n events: 31
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.73 (0.26 to 2.03), NS
	  ○ $P = 0.62$
	Quantitative synthesis of 5 studies investigating the MAKO system, MA using random effects (N=1,342 pts)
	• rTHA (n=476 pts), n events: 7
	• mTHA (n=866 pts), n events: 23
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 0.56 (0.23 to 1.38), NS
	  ○ $P = 0.21$

Study Citation and Design	Outcomes

(continued)	**Quantitative synthesis of 2 studies investigating ROBODOC, MA using random effects (N=208 pts)**
	• rTHA (n=106 pts), n events: 13
	• mTHA (n=102 pts), n events: 8
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 1.30 (0.05 to 32.42), NS
	  ○ P = 0.87

Han 2019 SR (3 studies w/MA)	**Kamara 2017 (N=296 pts), f/u timing NR** • rTHA (n=98 pts), n events: 2 • mTHA (n=198 pts), n events: 4 • Odds ratio (95% CI): 1.01 (0.18 to 5.61), NS
(continued)	**Siebel 2005 (N=71 pts), f/u timing NR** • rTHA (n=36 pts), n events: 2 • mTHA (n=35 pts), n events: 1 • Odds ratio (95% CI): 2.00 (0.17 to 23.11), NS
(continued)	**Honl 2003 (N=141 pts), f/u timing NR** • rTHA (n=61 pts), n events: 9 • mTHA (n=80 pts), n events: 2 • Odds ratio (95% CI): 6.75 (1.40 to 32.50), favours mTHA

(continued)	**Quantitative synthesis of 3 studies, MA using fixed effects (N=508 pts)**
	• rTHA (n=195 pts), n events: 13
	• mTHA (n=313 pts), n events: 7
	• Pooled odds ratio between treatment groups (95% CI)
	  ○ 2.88 (1.11 to 7.50), favours mTHA
	  ○ P = 0.03

CI = confidence interval; cTHA = conventional total hip arthroplasty; d = day(s); DVT = deep-vein thrombosis; f/u = follow up; HO = heterotopic ossification; IQR = interquartile range; m = metre(s); MA = meta-analysis; mL = millilitre; mo = month(s); mTHA = manual total hip arthroplasty; NA = not applicable; NE = not estimable; NR = not reported; NRS = non-randomized study; NS = not significant; P = probability value; pt/pts = patient(s); rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; SR = systematic review; wk = week(s); w/ = with; w/o = without; yr = year(s)

Table 14: Summary of Findings for Cost Effectiveness

Study Citation and Design	Outcomes
	Costs
Maldonado 2021 CEA	**Reference case, cumulative Medicare costs in USD at 5yr f/u** • rTHA, mean (SD): 14,410 (40) • mTHA, mean (SD): 15,355 (115) • Average differential cost saving between groups, favours rTHA: 945 **Reference case, cumulative private payer costs at 5yr f/u** • rTHA, mean (SD): 15,212 (82) • mTHA, mean (SD): 17,022 (204) • Average differential cost saving between groups, favours rTHA: 1,810 **Probabilistic sensitivity analyses, cumulative Medicare costs in USD at 5yr f/u** • rTHA, range: 14,274 to 14,456 • mTHA, range: 14,646 to 15,084 **Probabilistic sensitivity analyses, cumulative private payer costs in USD at 5yr f/u** • rTHA, range: 15,012 to 15.411 • mTHA, range: 16,526 to 17,516
	ICER
Maldonado 2021 CEA	**Reference case, calculated ICER for Medicare costs in USD (considering WTP threshold of $50K)** • -23,625 (favours rTHA)
	QALY
Maldonado 2021 CEA	Reference case, cumulative utilities (QALY) • rTHA, mean (SD): 2.96 (0.58) • mTHA: 2.92 (0.57)

CEA = cost-effectiveness analysis; f/u = follow up; ICER = incremental cost-effectiveness ratio; K = thousand; mTHA = manual total hip arthroplasty; QALY = quality-adjusted life year(s); rTHA = robotic-assisted total hip arthroplasty; SD = standard deviation; USD = US dollars; WTP = willingness-to-pay; yr = year(s)

Appendix 5: Overlap Between Included Systematic Reviews

Note that is appendix has not been copy-edited.

Table 15: Overlap in Relevant Primary Studies Between Included Systematic Reviews

Primary study citation	Ng 2021	Samuel 2021	Han 2019
Clement 2021	No	Yes	No
Singh 2021	No	Yes	No
Domb 2020	Yes	Yes	No
Hadley 2020	No	Yes	No
Kong 2020a	Yes	Yes	No
Kong 2020b	Yes	Yes	No
Perets 2021*	No	Yes	No
Kayani 2019a	Yes	No	No
Kayani 2019b	Yes	No	No
Lawson 2019	Yes	No	No
Peng 2019	No	Yes	No
Banchetti 2018	Yes	No	No
Bargar 2018**	No	Yes	Yes
Heng 2018	Yes	No	No
Nakamura 2018	No	Yes	No
Illgen 2017	Yes	No	No
Kamara 2017***	Yes	Yes	Yes
Bukowski 2016	Yes	Yes	No
Lim 2015	No	Yes	Yes
Domb 2014	No	No	Yes
Nakamura 2010	No	Yes	Yes
Nakamura 2009	No	No	Yes
Hananouchi 2007	No	Yes	Yes
Nishihara 2006	No	Yes	Yes
Siebel 2005	No	No	Yes
Honl 2003	No	Yes	Yes
Bargar 1998	No	Yes	Yes

*In Samuel 2021, Perets 2021 is referred to throughout the paper as 'Perets 2020'.

**In Han 2019, Bargar 2018 is referred to throughout the paper as 'Bargar 2017'.

***In Ng 2021, Kamara 2017 is referred to throughout the paper as 'Kamara 2016'.