



Early national comparison of robotic versus conventional hip replacements for arthritis using National Joint Registry data: target trial emulation study

Hasan R Mohammad,¹ Andrew Judge,^{2,3,4,5} Xavier L Griffin,^{1,6} David W Murray²

¹Bone and Joint Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London E1 2AT, UK

²Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Nuffield Orthopaedic Centre, Oxford, UK

³Musculoskeletal Research Unit, Bristol Medical School, University of Bristol, Southmead Hospital, Westbury-on-Trym, Bristol, UK

⁴National Institute for Health and Care Research Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, Bristol, UK

⁵National Institute for Health and Care Research Applied Research Collaboration South West (NIHR ARC South West) at University of Bristol, Bristol, UK

⁶Barts Health NHS Trust, The Royal London Hospital, London, UK

Correspondence to: H R Mohammad
hasan.mohammad@qmul.ac.uk
(or @HasanRMohammad on X; ORCID 0000-0002-2146-8601)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2026;394:e100709
<http://dx.doi.org/10.1136/bmj-2026-100709>

Accepted: 25 August 2026

ABSTRACT

OBJECTIVES

To evaluate the early comparative effectiveness of robotic versus conventional total hip replacement (THR) in the UK.

DESIGN

Target trial emulation study.

SETTING

Public and private hospitals using data from the National Joint Registry of England, Wales, Northern Ireland, the Isle of Man, Guernsey, and Jersey.

PARTICIPANTS

666 283 THRs performed between 2018 and 2024; 656 080 were conventional THRs and 10 203 were robotic THRs.

MAIN OUTCOME MEASURES

Five year THR implant survival (mean 2.5 year follow-up), all cause revision risk, cause specific revision risk, malpositioning revision risk, intraoperative complications, and patient survival up to five years. Kaplan-Meier survival analysis, Cox proportional hazards regression, and Fine and Gray models were used. Propensity score matching created comparable treatment groups to estimate the average treatment effect in the treated population.

RESULTS

The five year implant survival of the matched conventional and robotic THR groups was 98.8% (95% confidence interval (CI) 98.6% to 99.0%) and 98.8% (CI 98.3% to 99.1%), respectively, with no difference in revision risk between groups (hazard ratio 0.96, 95% CI 0.75 to 1.22; P=0.73). Cause specific revision risk and intraoperative complication

risks did not differ between groups. However, a significantly lower risk of revisions related to implant malposition (dislocation, leg length discrepancy, and malalignment) was observed in the robotic group (hazard ratio 0.53, 95% CI 0.30 to 0.93; P=0.03). The five year patient survival of the matched conventional and robotic THR groups was 96.3% (95% CI 95.8% to 96.8%) and 96.2% (95.2% to 97.0%), respectively, with no difference in all cause mortality risk (hazard ratio 1.12, 95% CI 0.95 to 1.34; P=0.17).

CONCLUSIONS

No significant differences were found in overall THR or patient survival, cause specific revision, or intraoperative complication risks between robotic and conventional groups. Robotic THR was associated with a lower revision risk from causes related to implant malpositioning. However, the possibility of unmeasured and residual confounding cannot be ruled out owing to the observational design. These results highlight the importance of careful evaluation of robotic technology in publicly funded healthcare systems because of the substantially higher capital and procedural costs.

Introduction

The main treatment for end stage hip arthritis is a total hip replacement (THR). More than 100 000 THRs are performed annually in the UK, and this number is projected to increase because of an increasing ageing population and higher prevalence of risk factors such as obesity.^{1,2}

Although long term survival for THRs is good, the revisions after surgery follow a bimodal pattern. Early revisions within the first two years after surgery dominate, with leading causes including instability, malpositioning, periprosthetic fractures, and early loosening.³⁻⁶ The later peak is often after 10 years, mainly because of loosening, wear, and implant failure.^{7,8}

Robotic surgery was introduced to improve acetabular and femoral implant positioning and therefore reduce the early revision risk of THRs. These systems use preoperative and intraoperative imaging to provide guidance to the surgeon in theatre so that the implants are placed more accurately according to the operative plan. More precise implant positioning may also improve longer term revision risks and hip function for patients after surgery.

The uptake of robotic technology has been rapid. The NHS recently published its 10 year plan to use robotic surgery to revolutionise patient care,⁹ aiming for a sevenfold increase in such surgeries by 2035.¹⁰ Robotic

WHAT IS ALREADY KNOWN ON THIS TOPIC

The NHS has recently announced plans for a sevenfold increase in the number of robotic surgeries by 2035

Robotic assisted total hip replacement (THR) improves implant positioning, but it is unclear whether this translates into better outcomes and reduced revision risk

The National Institute for Health and Care Excellence recommended urgent research to study the effect of robotic surgery on revision risk after joint replacement

WHAT THIS STUDY ADDS

In this national study, robotic assisted surgery was not associated with lower overall or cause specific revision risk, or improved patient survival

Robotic surgery was associated with a lower risk of revisions for THR malpositioning

These early findings provide real world evidence during planned rapid NHS adoption of robotic surgery

systems cost around £1.0m (\$1.4m; €1.2m) to deploy initially,^{11 12} followed by an increased cost for each patient of about £1000-2500.^{13 14} If robotic systems were used for all THRs in the NHS, the additional cost would be at least £100m annually.

Currently it is unknown whether robotic surgery for THRs improves clinical outcomes. The National Institute for Health and Care Excellence highlighted an urgent need for research into robotic orthopaedic research focused on failure risk.¹⁵ Existing studies have single centre designs with small patient groups.^{16 17} RACER-Hip, a UK randomised controlled trial comparing conventional and robotic THR surgery, is currently under way to assess patient reported outcome measures.¹²

To make decisions on robotic procurement, we need answers on how the early failure and mortality risks compare between robotic and conventional surgery. The National Joint Registry¹ is a large orthopaedic registry that collects information on all joint replacements performed in the UK outside of Scotland. We used nationwide data to study whether patients having robotic THR surgery have any differences in early revision risk and all cause mortality compared with those having conventional surgery.

Methods

Data sources

The National Joint Registry was established in April 2003 and is a large joint replacement register with more than two million operations recorded.¹ Case ascertainment is 99.6% and linkage to follow-up procedures is 95%.¹⁸ Mortality data were obtained from the Office for National Statistics. Data collection forms are available on the NJR website.¹⁹

Design

We obtained retrospective observational data from the National Joint Registry for all THRs performed in England, Northern Ireland, and the Isle of Man between 1 January 2018 (year in which the first robotic procedure was recorded) and 31 December 2024. We used a target trial emulation design²⁰ to compare the effectiveness of robotic THR surgery with conventional surgery using the endpoints of revision risk, cause specific revision risk, malpositioning revision risk, and all cause mortality risk. This approach uses the design principles of a hypothetical randomised controlled trial and applies them to routine data (table 1).²⁰ Our estimand is the average treatment effect in the treated population. We also assessed the average treatment effect for a different target population (wider population) in another analysis.

Table 1 | Details of target trial emulation design applied to dataset

Design	Target trial protocol	Emulation of target trial using real world data
Eligibility criteria	Inclusion: adults (>18 years) with primary hip replacements. All indications for surgery will be included Exclusion: revision hip replacements, hip resurfacing, hip hemiarthroplasty, metal on metal articulations, patients without consent	Inclusion: adults (>18 years) with primary hip replacements. All indications for surgery will be included Exclusion: revision hip replacements, hip resurfacing, hip hemiarthroplasty, metal on metal articulations, patients who did not give consent to the National Joint Registry to use their data for research
Treatment strategies	Conventional total hip replacement (THR; control) versus robotic THR (intervention). All robotic systems in use in the UK	Same as target trial
Treatment assignment	Patients who meet inclusion criteria are randomised using computer system to each surgical arm at 1:1 ratio. Clinicians are not masked to surgical allocation but patients and analysts are masked In a randomised controlled trial, randomisation ensures that any difference in outcomes is attributable to effect of treatment and so represents the causal effect of treatment in study population (average treatment effect in treated and untreated participants)	Randomisation emulated using propensity score methods (propensity score matching and inverse probability of treatment weighting) to balance groups according to known and measured confounders Propensity score matching to estimate average treatment effect in treated population. Inverse probability of treatment weighting estimates average treatment effect in entire population (not treated participants only)
Follow-up	Start (time zero): time of treatment allocation (randomisation). Baseline is when a patient is randomised before undergoing surgery on day of surgery End: patients will be followed up from baseline until date of death, loss to follow-up (last date of contact with service provider), or end of study period (last date of data collection), whichever comes first	Start (time zero): initiation of treatment (date of surgery) End: same as target trial
Outcomes	All outcomes collected at one year, two years, three years, four years, and five years postoperatively Primary: implant survival, all cause revision surgery risk, implant malpositioning revision risk (malalignment, dislocation, and limb length discrepancy), all cause mortality risk Secondary: cause specific revision risk, intraoperative complications	Same as target trial
Estimands	Effect of receiving robotic versus conventional THR Clinical outcomes: all cause revision risk, implant malpositioning revision risk, cause specific revision risk, and all cause mortality risk compared between groups with hazard ratios	Effect of receiving robotic versus conventional THR and retained in matched analyses Clinical outcomes: all cause revision risk, implant malpositioning revision risk, cause specific revision risk, and all cause mortality risk compared between groups with hazard ratios
Statistical analysis	Multivariable regression models (Cox for time to event, logistic for binary outcome). As patients receive surgery, no requirement to further adjust for factors associated with adherence In a randomised controlled trial, randomisation ensures that any difference in outcomes is attributable to effect of treatment, and so represents causal effect of treatment in study population (average treatment effect in treated and untreated participants)	Same as target trial with adjustment for baseline confounders using propensity score methods. Propensity score matching to estimate average treatment effect in treated population. Inverse probability of treatment weighting estimates average treatment effect in entire population (not treated participants only)

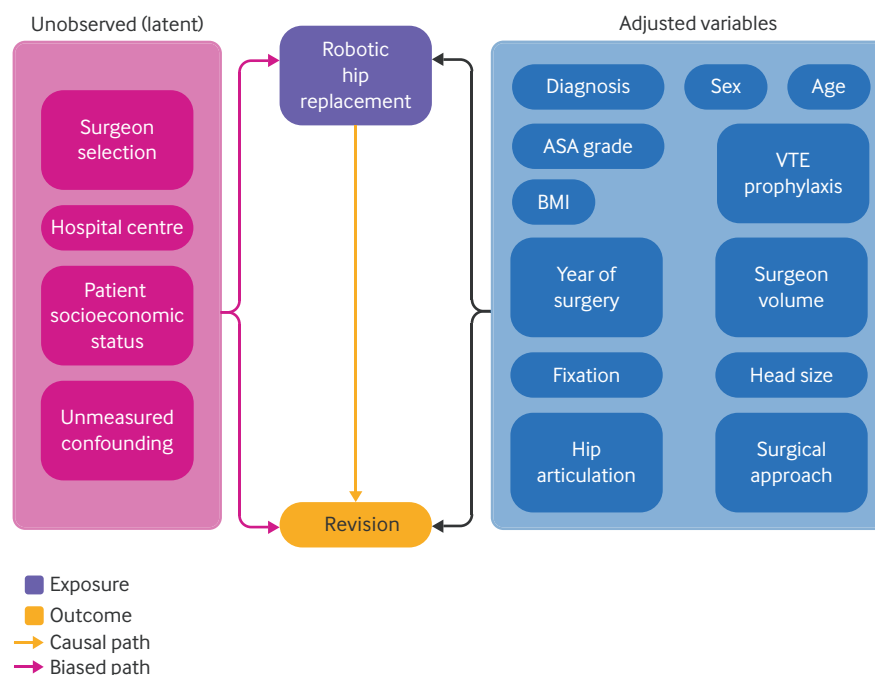


Fig 1 | Directed acyclic graph for study. Direct arrow from exposure to outcome represents total effect mediated through factors including implant positioning, soft tissue balancing, gait, and surgical skill. This study's focus is total effect. To preserve visual clarity and maintain focus on blocking the primary exposure-outcome backdoor pathways, arrows between nodes representing baseline covariates are not shown because they do not have implications for the study's confounding adjustment strategy. ASA=American Society of Anesthesiologists; BMI=body mass index; VTE=venous thromboembolism

A directed acyclic graph was used to visualise measured and unmeasured confounders of surgery type and outcome in our study (fig 1).

Propensity score matching and rationale

Model parameters were examined using routine statistical summaries of central tendency and variance to identify baseline imbalance between conventional and robotic groups. Because of the potential for confounding, we a priori matched conventional and robotic groups for known patient, surgical, and implant confounders using propensity scores. The covariates were matched on age, sex, ASA grade (American Society of Anesthesiologists physical status classification), diagnosis, surgical approach, implant fixation, surgeon caseload (number of THRs performed annually), year of surgery, head size, hip articulation, and venous thromboembolism prophylaxis. The rationale for these covariates is described in the supplementary appendix (table S1). We did not match on body mass index (BMI) in our primary analysis owing to missing data (19%) but include BMI as a confounder in a sensitivity analysis for patients with complete data. We used logistic regression to generate a propensity score representing the probability that a patient received a robotic THR. The algorithm used matched on the logit of the propensity score with a 0.02 standard deviation (SD) calliper width. The matching ratio was 4:1 owing to the excess numbers

of conventional THRs. We chose a maximum of four controls for each patient because using additional controls is known to have limited benefit in improving precision and power.²¹ We used greedy matching without replacement.²² Standardised mean differences were examined before and after matching to assess for any covariate imbalance between conventional and robotic groups, with standardised mean differences $\geq 10\%$ considered to suggest covariate imbalance.²³

Outcomes

Revision surgery (all cause) was defined as any implant component removed, exchanged, or added since the original operation. Cause specific revisions were those involving specific indications. Revisions considered to be a result of implant malpositioning related to dislocation, leg length discrepancy, and malalignment (malpositioning revision risk). Patient survival used an endpoint of all cause mortality. Secondary outcomes were intraoperative complications and complexity of the revision surgery. Two stage revision procedures are those requiring more than one operation to correct the failed THR.²⁴

Statistical analyses

Cumulative survival functions were determined using the Kaplan-Meier method. The endpoint for implant survival was revision surgery (defined as any implant component removed, exchanged, or added since

the original operation). Malpositioning revision risk was defined as revisions from dislocation, leg length discrepancy, and malalignment. The endpoint for patient survival was all cause mortality. Implant and

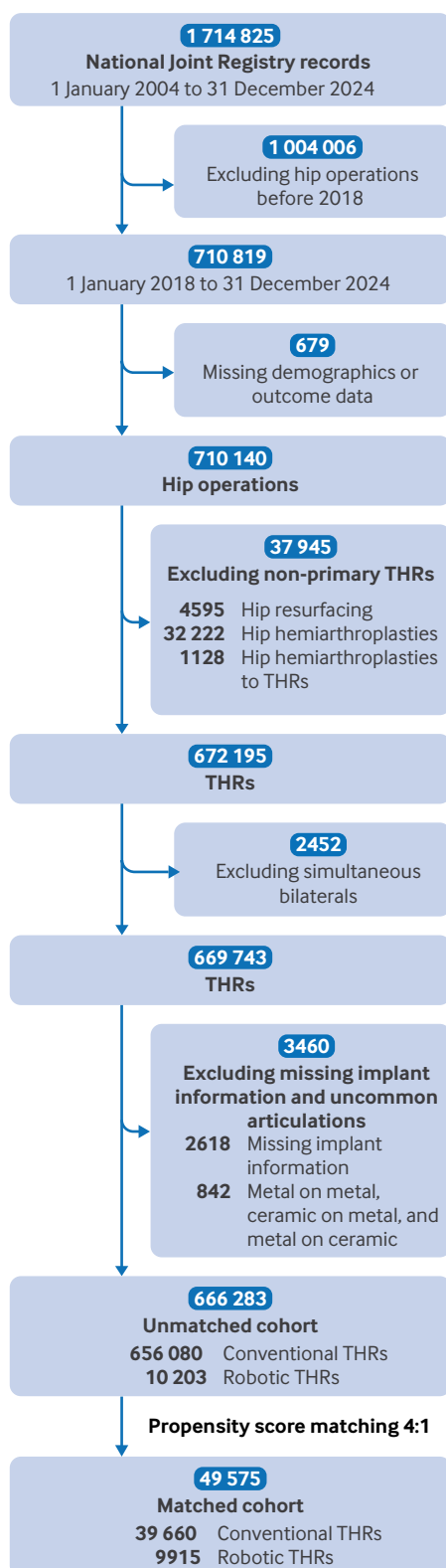


Fig 2 | Data flowchart of dataset cleaning and merging. THR=total hip replacement

patient survival were compared between conventional and robotic surgery groups using Cox regression models to generate hazard ratios, with the proportional hazard's assumptions assessed in all analyses. Hazard ratios >1 indicated higher risk in the robotic group. To account for clustering within the matched cohort, we tested a robust variance estimator in regression models. We performed a multilevel frailty model sensitivity analysis to control for patient clustering within surgeons, but this did not influence the results presented. Adjusted Cox models included covariates (BMI) with residual imbalance after matching (standardised mean difference $\geq 10\%$), which did not alter the results. Fine and Gray regression analysis was also used, which considers the competing mortality risk and creates sub hazard ratios. To compare indications for revision, Cox regression models were used and hazard ratios generated. The risk of intraoperative complications and complexity of revision surgery (risk of two stage revision) were compared using the proportional χ^2 test with Yates' correction unless frequencies were <5 when Fisher's exact test was used.

All statistical analyses were performed using Stata (version 15.1; StataCorp, College Station, TX) except for propensity score matching, which was performed using R (version 3.4.0; R Foundation for Statistical Computing, Vienna, Austria).

Sensitivity analyses

We performed E value analyses²⁵ to quantify and transparently report exactly how strong an unmeasured confounder would need to be to alter our point estimates and their confidence intervals. A surgeon's learning curve for performing robotic THR is described to be 12 cases.^{11 26} We carried out a sensitivity analysis in which we rematched the cohort after excluding the first 12 patients who underwent a robotic procedure for every surgeon (supplementary appendix, tables S2 and S3).

The principal missingness covariate of interest in our directed acyclic graph was BMI (19%). We compared the characteristics of all patients versus those with complete BMI data (supplementary appendix, table S4). This showed no clinically relevant differences in measured characteristics. For the primary analysis, BMI was not included in the propensity score model because of missing data. However, the distribution of BMI between conventional and robotic groups before and after matching only had a small imbalance (standardised mean difference 0.12); we adjusted for this imbalance in the Cox models and the results did not change. We performed a separate sensitivity analysis where we repeated the propensity score matching but included the BMI covariate as a confounder in the model and therefore only patients with complete BMI data were included (see supplementary appendix, tables S5 and S6).

Inverse probability treatment weighting analyses for full cohort

Propensity score matching estimates the average treatment effect in treated patients: the closest

Table 2 | Model parameters before and after matching THRs

Covariate	Unmatched cohort			Matched cohort		
	Conventional THR (n=656 080)	Robotic THR (n=10 203)	SMD	Conventional THR (n=39 660)	Robotic THR (n=9 915)	SMD
Sex:						
Female	395 578 (60.3)	5699 (55.9)	0.10	22 240 (56.1)	5528 (55.8)	0.007
Male	260 502 (39.7)	4504 (44.1)		17 420 (43.9)	4387 (44.2)	
Mean age (SD) at surgery (years)	69.2 (11.1)	66.5 (11.7)	0.24	66.4 (10.9)	66.3 (11.7)	0.01
Mean BMI (SD)	28.8 (5.3) (n=527 510)	27.9 (5.3) (n=8329)	0.17	28.5 (5.1) (n=33 281)	27.9 (5.3) (n=8045)	0.12
Primary diagnosis:						
Primary osteoarthritis	576 383 (87.9)	9500 (93.1)	0.32	36 921 (93.1)	9219 (93.0)	0.007
Trauma	36 037 (5.5)	40 (0.40)		146 (0.4)	40 (0.4)	
Avascular necrosis	19 119 (2.9)	180 (1.8)		722 (1.8)	180 (1.8)	
Other	24 541 (3.7)	483 (4.7)		1871 (4.7)	476 (4.8)	
Surgical approach:						
Posterior	518 204 (79.0)	9707 (95.1)	0.56	37 717 (95.1)	9419 (95.0)	0.01
Anterolateral	119 767 (18.2)	221 (2.2)		894 (2.2)	221 (2.2)	
Anterior	6287 (1.0)	147 (1.4)		577 (1.5)	147 (1.5)	
Other	11 822 (1.8)	128 (1.3)		472 (1.2)	128 (1.30)	
Mean (SD) surgeon caseload/year*	72.0 (49.2)	99.9 (52.1)	0.55	95.9 (60.7)	96.9 (49.8)	0.02
ASA grade:						
1	60 061 (9.2)	1571 (15.4)	0.34	5993 (15.1)	1551 (15.6)	0.03
2	450 803 (68.7)	7531 (73.8)		29 568 (74.6)	7263 (73.3)	
≥3	145 216 (22.1)	1101 (10.8)		4099 (10.3)	1101 (11.1)	
VTE—chemical:						
LMWH (±other)	471 402 (71.9)	8069 (79.1)	0.25	31 309 (79.0)	7781 (78.5)	0.02
Aspirin only	27 962 (4.3)	79 (0.7)		344 (0.8)	79 (0.8)	
Other	154 693 (23.6)	2035 (20.0)		7920 (20.0)	2035 (20.5)	
None	2023 (0.3)	20 (0.2)		87 (0.2)	20 (0.2)	
VTE—mechanical:						
Any	617 431 (94.1)	9872 (96.8)	0.13	38 417 (96.9)	9584 (96.7)	0.01
None	38 649 (5.9)	331 (3.2)		1243 (3.1)	331 (3.3)	
Fixation:						
All cemented	141 493 (21.6)	45 (0.44)	0.77	158 (0.4)	45 (0.5)	0.04
All cementless	243 464 (37.1)	5052 (49.5)		19 473 (49.1)	5049 (50.9)	
Hybrid (cemented stem and cementless cup)	255 412 (38.9)	5104 (50.2)		20 019 (50.5)	4819 (48.6)	
Hybrid (cementless stem and cemented cup)	15 711 (2.4)	2 (0.02)		10 (0.03)	2 (0.02)	
Head size:						
≤28	122 217 (18.6)	874 (8.6)	0.42	3425 (8.6)	831 (8.4)	0.02
32	327 054 (49.9)	4254 (41.7)		16 543 (41.7)	4249 (42.9)	
≥36	206 809 (31.5)	5075 (49.7)		19 692 (49.7)	4835 (48.8)	
Hip articulation:						
Ceramic on poly	311 287 (47.5)	8916 (87.4)	0.95	34 313 (86.5)	8628 (87.0)	0.02
Ceramic on ceramic	29 464 (4.5)	28 (0.3)		105 (0.3)	28 (0.3)	
Metal on poly	315 329 (48.1)	1259 (12.3)		5242 (13.2)	1259 (12.7)	
Year of surgery:						
2018	98 309 (15.0)	10 (0.1)	0.93	56 (0.1)	10 (0.1)	0.02
2019	100 201 (15.3)	84 (0.8)		312 (0.8)	84 (0.9)	
2020	54 909 (8.4)	567 (5.6)		2123 (5.4)	567 (5.7)	
2021	84 134 (12.8)	1628 (16.0)		6335 (16.0)	1565 (15.8)	
2022	98 884 (15.1)	2187 (21.4)		8383 (21.1)	2092 (21.1)	
2023	107 990 (16.5)	2670 (26.2)		10 485 (26.4)	2590 (26.1)	
2024	111 653 (17.0)	3057 (30.0)		11 966 (30.2)	3007 (30.3)	

Data are numbers (%) unless stated otherwise. All variables had complete sets of data except BMI where the number is presented in brackets.

ASA=American Society of Anesthesiologists; BMI=body mass index; LMWH=low molecular weight heparin; SD=standard deviation; SMD=standardised mean difference; THR=total hip replacement; VTE=venous thromboembolism.

*Primary THR surgery practice.

untreated and treated people are matched and the remaining untreated people who were not matched are excluded from the analysis. Inverse probability of treatment weighting estimates the treatment effect in a population whose distribution of risk factors is equal to that found in all study participants (not treated participants only).²⁷ We included the inverse probability of treatment weighting analyses as an additional analysis referring to a different target

population, and the estimand for this analysis was the average treatment effect. A concern of inverse probability of treatment weighting is that people with large weights (extremely high or low probability of treatment) can influence the result by introducing high variance. We addressed this concern by trimming patients at the extreme 1% distribution of the propensity score (supplementary appendix, table S7).

Patient and public involvement

The authors ran public advisory group (PAG) workshops that helped shape this study. The PAG consisted of people from London with lived experience of joint replacement from diverse ethnic backgrounds and of different ages. The PAG were concerned over NHS plans to make robotic surgery widespread without reliable evidence. The group felt questions about robotic joint replacements should focus on key outcomes, including failure risk and cost effectiveness to ensure good value for money. We therefore focused our article on implant survival and revision risk. The PAG favoured using routinely collected data given this is cost effective and the data have already been collected.

Results

Overall, 1 714 825 hip operations were performed in England, Northern Ireland, and the Isle of Man between 1 January 2004 and 31 December 2024; 710 819 procedures were performed between 1 January 2018 and 31 December 2024. After excluding hip resurfacing, hip hemiarthroplasties, and data cleaning, 666 283 THRs were available for analysis. Figure 2 summarises the data cleaning process. Propensity matching resulted in well balanced groups (table 2). The robotic system most used was Mako (Stryker, MI; 96.9%; see table 3).

Unmatched analyses

The unmatched analysis included 656 080 conventional THRs and 10 203 robotic THRs. Baseline characteristics differed substantially between groups (table 2). Robotic procedures were performed by higher volume surgeons, had a healthier population from ASA grades, were younger on average by three years, had >10 times fewer trauma procedures, larger head sizes, and different fixations and hip articulations. Many of these factors are known to influence outcomes.

The mean follow-up was 3.7 years (SD 2.1) in the conventional group and 2.5 years (SD 1.3) in the robotic group. The mean annual surgeon caseload for THR was 72.0 (SD 49.2) in the conventional group and 99.9 (SD 52.1) in the robotic group. In the conventional group, there were 7807 revisions (1.2%) at a mean of 1.2 years (SD 1.6) after surgery. In the robotic group, there were 80 revisions (0.8%) at a mean of 1.0 years (SD 1.1) after surgery.

The five year implant survival of the unmatched conventional and robotic groups was 98.6% (95% CI 98.5% to 98.6%) and 98.8% (98.4% to 99.1%), respectively. The risk of revision was significantly lower in the robotic group (hazard ratio 0.79, 95% CI 0.63 to 0.98; $P=0.03$; fig 3).

Table 3 | List of robotic systems used in the UK for THRs

Robotic system	Number (%) of THRs (n=10 203)
Mako	9889 (96.9)
CORI	166 (1.6)
Other	148 (1.4)

THR=total hip replacement.

The risk of revision from THR malpositioning was significantly lower in the robotic group (hazard ratio 0.40, 95% CI 0.23 to 0.69; $P=0.001$). The absolute risk for conventional and robotic groups was 0.38% (2507 revisions) and 0.13% (13 revisions), respectively. Using this endpoint, the five year survival for unmatched conventional and robotic groups was 99.5% (95% CI 99.5% to 99.6%) and 99.8% (99.6% to 99.9%), respectively.

Table 4 presents cause specific revision risk. The risk of revision for dislocation was significantly lower in the robotic group (0.11% v 0.33%; $P=0.002$). No other differences in cause specific revisions were found between groups (table 4). Intraoperative complications of femoral and acetabular fractures were significantly lower in the robotic group (table 5).

Among the revisions in the conventional group, 7000 (89.7%) were single stage revisions and 807 (10.3%) were two stage revisions. In the robotic group, there were 64 (80.0%) single stage revisions and 16 (20.0%) two stage revisions. The risk of two stage revision did not differ between groups ($P=0.34$).

The five year patient survival of the unmatched conventional and robotic THR groups was 91.7% (95% CI 91.6% to 91.8%) and 96.2% (95.3% to 96.9%), respectively. The risk of all cause mortality was significantly lower in the robotic group (hazard ratio 0.49, 95% CI 0.43 to 0.57; $P<0.001$; fig 4).

Propensity score matched comparison analyses

Overall, 39 660 conventional THRs and 9915 robotic THRs were included in the matched analysis. The mean follow-up in both groups was 2.5 years (SD 1.3). The mean annual surgeon caseload for THRs was 95.9 (SD 60.7) in the conventional group and 96.9 (SD 49.8) in the robotic group. In the conventional group, there were 334 revisions (0.84%) at a mean of 0.8 years (SD 1.0) after surgery. In the robotic group, there were 80 revisions (0.81%) at a mean of 0.9 years (SD 1.1) after surgery.

The five year implant survival of the matched conventional and robotic THR groups was 98.8% (95% CI 98.6% to 99.0%) and 98.8% (98.3% to 99.1%), respectively, with no difference in revision risk between groups (hazard ratio 0.96, 95% CI 0.75 to 1.22; $P=0.73$; fig 5). Fine and Gray models showed a sub hazard ratio of 0.96 (95% CI 0.75 to 1.22; $P=0.73$).

The point estimate E value for the hazard ratio of 0.96 (95% CI 0.75 to 1.22) was 1.25, reflecting a stable, near null finding. The E values for the lower and upper confidence bounds were 2.0 and 1.74, respectively. To shift these boundaries and achieve a statistically significant directional outcome, an unmeasured confounder would require a minimum joint hazard ratio association of 2.0 to shift the lower confidence bound towards a potential benefit, or 1.74 to shift the upper bound towards a potential harm.

The risk of revision related to THR malpositioning was significantly lower in the robotic group (hazard ratio 0.53, 95% CI 0.30 to 0.93; $P=0.03$). The absolute risk for conventional and robotic groups was 0.25%

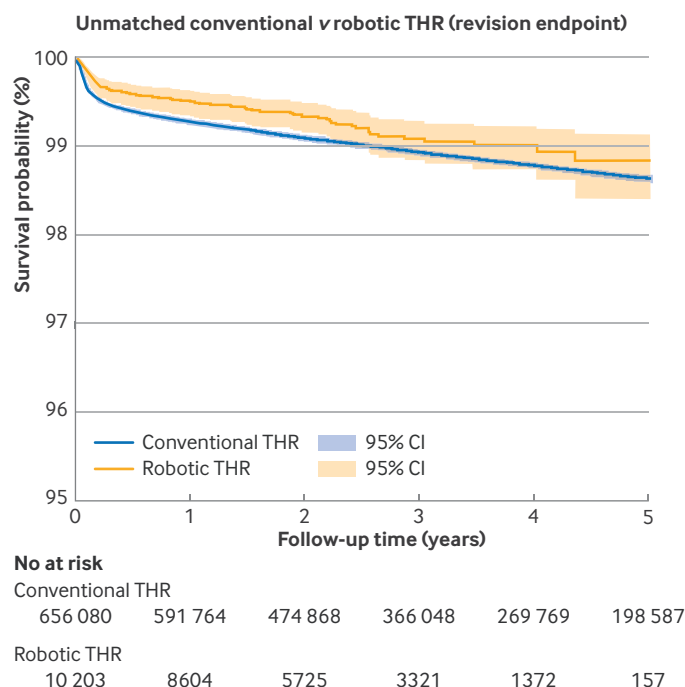


Fig 3 | Kaplan-Meier graph comparing unmatched implant survival of robotic and conventional total hip replacements (THRs). CI=confidence interval

(99 revisions) and 0.13% (13 revisions), respectively. Using this endpoint, the five year survival for matched conventional and robotic groups was 99.6% (95% CI 99.5 to 99.7) and 99.8% (99.6 to 99.9), respectively.

No differences were found in cause specific revision, intraoperative complications, or two stage revision risk between groups (table 4 and table 5). Of the revisions in the conventional group, 298 (89.2%) were single stage revisions and 36 (10.8%) were two stage revisions. In the robotic group there were 64 (80.0%) single stage revisions and 16 (20.0%) two stage revisions. Evidence of a difference in the risk of two stage revisions between groups was weak ($P=0.05$).

The five year patient survival of the matched conventional and robotic THR groups was 96.3% (95% CI 95.8% to 96.8%) and 96.2% (95.2% to 97.0%) with no difference in all cause mortality risk between groups (hazard ratio 1.12, 95% CI 0.95 to 1.34; $P=0.17$; fig 6).

The point estimate E value for the hazard ratio of 1.12 (95% CI 0.95 to 1.34) was 1.49, reflecting a stable, near null finding. The E values for the lower and upper confidence bounds were 1.29 and 2.01, respectively. To shift these boundaries and achieve a statistically significant directional outcome, an unmeasured confounder would require a minimum joint hazard ratio association of 1.29 to shift the lower confidence bound towards a potential benefit, or 2.01 to shift the upper bound towards a potential harm.

The sensitivity analyses performed for surgeons' learning curves and complete BMI data showed similar results to the primary analyses (supplementary appendix, tables S3 and S6).

Inverse probability weighting analyses

Overall, the cohort consisted of 656 080 conventional THRs and 10 203 robotic THRs. The mean follow-up for THRs was 3.7 years (SD 2.1) in the conventional group and 2.5 years (SD 1.3) in the robotic group. Baseline characteristics differed substantially between the groups (table 2). Inverse probability treatment weighting with 1% trimming created well balanced groups (supplementary appendix, table S7). No significant differences were found in revision risk between conventional and robotic THRs with a hazard ratio of 0.95 (95% CI 0.76 to 1.20; $P=0.76$). Fine and Gray models showed a sub hazard ratio of 0.91 (95% CI 0.72 to 1.14; $P=0.41$). No differences in mortality risk were observed, with a hazard ratio of 0.86 (95% CI 0.72 to 1.05; $P=0.15$).

Discussion

Principal findings

This study compared differences in early revision risk between conventional and robotic surgery with more

Table 4 | Comparisons of cause specific revision in THR unmatched and matched cohorts

Cause specific revision indication	Unmatched conventional THR (n=656 080)	Unmatched robotic THR (n=10 203)	Hazard ratio (95% CI), P value	Matched conventional THR (n=39 660)	Matched robotic THR (n=9915)	Hazard ratio (95% CI), P value
Aseptic loosening	892 (0.14)	7 (0.07)	0.71 (0.34 to 1.50), 0.38	31 (0.09)	7 (0.07)	0.91 (0.40 to 2.06), 0.81
Leg length discrepancy	82 (0.01)	2 (0.02)	2.34 (0.57 to 9.53), 0.24	7 (0.02)	2 (0.02)	1.14 (0.24 to 5.52), 0.87
Lysis	135 (0.02)	1 (0.01)	0.77 (0.11 to 5.52), 0.80	6 (0.02)	1 (0.01)	0.67 (0.08 to 5.53), 0.71
Pain	180 (0.03)	0 (0)	NA	7 (0.02)	0 (0)	NA
Infection	2725 (0.42)	41 (0.40)	1.08 (0.79 to 1.47), 0.64	134 (0.34)	41 (0.41)	1.22 (0.86 to 1.74), 0.26
Periprosthetic fracture	1575 (0.24)	20 (0.20)	1.0 (0.64 to 1.55), 0.99	66 (0.17)	20 (0.20)	1.21 (0.73 to 1.99), 0.45
Wear	88 (0.01)	0 (0)	NA	3 (0.01)	0 (0)	NA
Dislocation subluxation revision	2161 (0.33)	11 (0.11)	0.39 (0.21 to 0.70), 0.002	81 (0.20)	11 (0.10)	0.54 (0.30 to 1.02), 0.06
Implant component dissociation	134 (0.02)	0 (0)	NA	6 (0.02)	0 (0)	NA
Malalignment	517 (0.08)	2 (0.02)	0.31 (0.08 to 1.25), 0.10	18 (0.05)	2 (0.02)	0.44 (0.10 to 1.92), 0.28
Implant fracture	129 (0.02)	1 (0.01)	0.72 (0.10 to 5.16), 0.74	5 (0.01)	1 (0.01)	0.80 (0.09 to 6.85), 0.84
Incorrect sizing	43 (0.01)	0 (0)	NA	2 (0.01)	0 (0)	NA
Other	289 (0.04)	5 (0.05)	1.26 (0.52 to 3.04), 0.61	7 (0.02)	5 (0.05)	2.86 (0.91 to 9.02), 0.07

Data are numbers (%). Comparisons between revision indications conducted using Cox regression analyses with hazard ratio presented and P values. CI=confidence interval; NA=not applicable; THR=total hip replacement.

Table 5 | Comparisons of intraoperative complications in THR unmatched and matched cohorts

Intraoperative complications	Unmatched conventional THR (n=656 080)	Unmatched robotic THR (n=10 203)	P value	Matched conventional THR (n=39 660)	Matched robotic THR (n=9915)	P value
None	651 525 (99.3)	10 173 (99.7)	0.78	39 456 (99.5)	9885 (99.7)	0.89
Femoral fracture	2970 (0.45)	21 (0.21)	<0.001	134 (0.34)	21 (0.21)	0.05
Acetabular fracture or penetration	380 (0.06)	0 (0)	0.02	4 (0.01)	0 (0)	1.00
Other	1205 (0.18)	9 (0.09)	0.03	66 (0.17)	9 (0.09)	0.08

Data are numbers (%). Comparisons between complications were performed with χ^2 test with P values presented. When frequencies were <5 the Fisher exact test was used.
THR=total hip replacement.

than 650 000 THRs included in the analyses. The key finding from the matched study was that robotic surgery was associated with a lower revision risk related to implant malpositioning than conventional surgery. However, no evidence was found of any differences in robotic THR overall implant survival or patient survival compared with conventional surgery over a mean of 2.5 years after surgery. No differences were observed in intraoperative complications or individual cause specific revision indications between matched robotic and conventional groups.

Robotic surgery is known to provide better preoperative planning and implant positioning, particularly for the acetabular implant component.²⁸ This study found a significantly lower revision risk for THR malpositioning in the robotic group. From a biological and mechanistic perspective, this is arguably the outcome most plausibly influenced by robotic assistance. However, this did not translate into a difference in overall revision risk.

The unmatched analyses showed that the robotic THR population differed substantially from the

conventional THR population. Robotic procedures were performed by higher volume surgeons, had a healthier population according to ASA grades, were younger on average by three years, had 10 times fewer trauma cases, larger head sizes, and different fixations and hip articulations (table 1). This appears to represent a selected care pathway or system rather than a technical adjunct applied uniformly across routine THR practice. In the independent sector, robotic surgery seems to be more attractive for surgeons and patients, and these populations tend to be fitter and younger. Additionally, anaesthetics for patients with higher ASA grades are less likely in the independent sector owing to the absence of intensive care units.²⁹

In the unmatched analyses, it is unsurprising that implant survival was superior for the robotic group because all these factors influence the revision rate. Trauma cases and smaller head sizes are associated with higher revision risk from dislocation,^{30 31} which was reflected by a significantly lower risk of dislocation in the robotic group than in the conventional group (hazard ratio 0.39, P=0.002). Patient survival was also significantly higher in the robotic group, which is probably because the proportion of patients with ASA grade ≥ 3 in the conventional group was more than double the proportion in the robotic group. Because of the marked differences in apparent benefits of robotic surgery in the matched and unmatched groups, future studies of robotic surgery should include matched groups. Additionally, we found that in the unmatched analyses, all cause mortality was significantly lower (hazard ratio 0.49) in the robotic group. Although, there is no plausible biological mechanism for this lower mortality rate, it shows selection bias in the robotic group towards healthy patients.

The revision risk for dislocation was lower in the robotic group in the unmatched analyses but evidence of an association was weaker in the matched analyses (hazard ratio 0.53, P=0.06); the reasons for this are unclear. This variable was included in the THR malpositioning outcome, which was significantly lower in the robotic group for the unmatched and matched analyses.

More research is needed to study the effect of robotic surgery on complex hips and patients with hip dysplasia, which was beyond the scope of this study. Patients with these conditions may benefit more from robotic surgery.

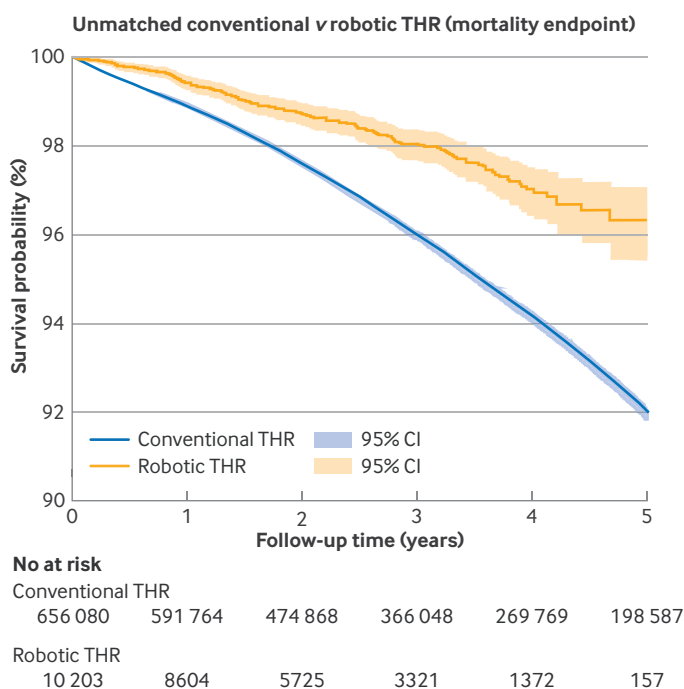


Fig 4 | Kaplan-Meier graph comparing unmatched patient survival of robotic and conventional total hip replacements (THRs). CI=confidence interval

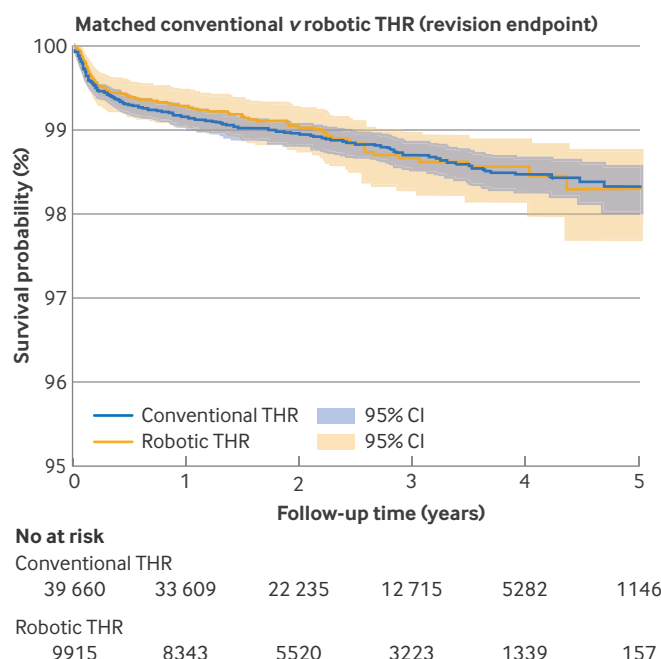


Fig 5 | Kaplan-Meier graph comparing matched implant survival of robotic and conventional total hip replacements (THRs). CI=confidence interval

Strengths and limitations

The main strength of this national study is that it uses a target trial emulation approach to compare the clinical outcomes of robotic and conventional THRs using a dataset with excellent case ascertainment and linkability. Our approach allowed direct comparison

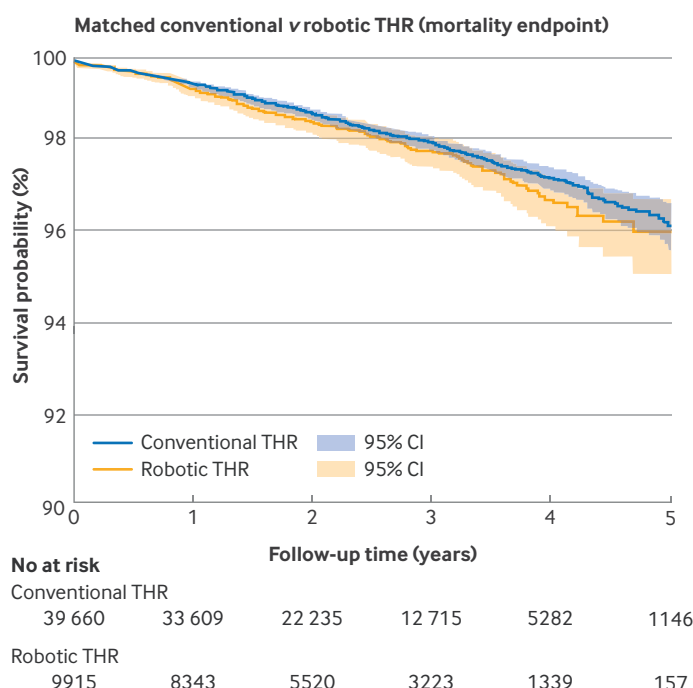


Fig 6 | Kaplan-Meier graph comparing matched patient survival of robotic and conventional total hip replacements (THRs). CI=confidence interval

of similar population cohorts. All matched variables achieved a standardised mean difference $<10\%$, which suggests covariate balance between comparative groups for known and measured confounders. We assessed the average treatment effect in the treated population (propensity matching) and the average treatment effect in the whole population (inverse probability of treatment weighting analyses) estimands using different target populations and found consistent results. Our observational registry study is large enough to assess early implant survival, cause specific revision indications, and all cause mortality—outcomes that are not feasible to assess in a traditional trial. Research from these studies is crucial to help guide commissioners when planning care.

The main study limitation is that our study design is a retrospective observational design. Despite matching, we can only control for known and measured confounders collected in the registry dataset. Potential exists for residual confounding from unmeasured variables, including centre level effects, patient selection, surgeon preference, private versus NHS care pathways, and technology adoption bias. We are also limited by follow-up time and longer term implementation effects may become apparent. Matching can also reduce the generalisability of findings—although we were able to match virtually all the robotic THRs, our findings are relevant to the treated population. Additionally, patient reported outcome measures were not included and so we do not know if pain and function differed between the robotic and conventional surgery groups. This information will be provided in the RACER-hip trial.

Another limitation is that despite our large study size our outcomes were limited by the overall number of events, particularly for analyses of cause specific revision indications. Some BMI data were missing and so we only matched on BMI as a complete case sensitivity analysis. The robotic systems also differed, which adds some clinical heterogeneity. Our E value sensitivity analysis suggests that generally strong unmeasured confounders would be needed to explain a potential benefit or harm of robotic surgery. However, E values also have important limitations.

Comparison with other studies

National studies or emulated trials comparing the effectiveness of robotic THRs with conventional surgery are lacking. Current studies are limited to single centre experiences with a maximum of a few hundred patients and a short duration of follow-up.^{17 32}

The ongoing RACER-hip trial will provide a definitive assessment of the early patient reported outcome measures at one year after surgery. Our observational registry study is complimentary by being large enough to assess early implant survival, cause specific revision indications, intraoperative complications, and mortality—outcomes that are not feasible to assess in a clinical trial design. Research from these studies is crucial to help guide commissioners when planning care.

Robotic surgery is estimated to cost around £1m to deploy initially^{11 12} followed by an increased cost for each THR of about £1000-2500.^{13 14} This cost is substantial and needs to be justified in a healthcare setting. One way this cost could be justified is when robotic surgery improves implant survival and decreases early hip revisions. Revision procedures are expensive: a THR revision on average costs £14 000 for that episode of care, with higher indirect costs.¹³ However, our matched study found no differences in overall revision risk. The results of this study question the 10 year NHS plan for widespread adoption of robotics in the absence of a clear benefit.

Conclusions

This national study found no significant differences in early robotic and conventional THR results when comparing overall implant or patient survival. No differences in cause specific revision or intraoperative complication risks were found between groups. Robotic THR was associated with a lower revision risk from causes related to implant malpositioning. We cannot rule out the possibility of unmeasured and residual confounding. These results highlight the importance of careful evaluation of robotic technology in publicly funded healthcare systems given the substantially higher capital and procedural costs.

The authors thank the patients and staff of all the hospitals who have contributed data to the National Joint Registry (NJR); and the Healthcare Quality Improvement Partnership, the NJR steering committee, and staff at the NJR Centre for facilitating this work. The authors have conformed to the NJR's standard protocol for data access and publication. The views expressed represent those of the authors and do not necessarily reflect those of the NJR steering committee or the Healthcare Quality Improvement Partnership who do not vouch for how the information is presented. HRM and XLG acknowledge the support of the NIHR Barts BRC (NIHR203330). AJ would like to thank the NIHR Bristol BRC and NIHR ARC South West for their ongoing support. DWM would like to thank the NIHR Oxford BRC for their support.

Contributors: All authors contributed to the study conception and design. Data analysis was performed by HRM and AJ. HRM is the manuscript's guarantor. Data interpretation was performed by all authors. All authors read and approved the final manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: No specific funding was received for this study.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: The study was approved by the Queen Mary University of London Research Ethics Committee (reference QME25.1847).

Data sharing: Data will not be publicly available owing to restrictions imposed by the ethics committee to protect participant confidentiality. Access will require a signed data sharing agreement and institutional ethics approval where applicable. Access to the data analysed in this study required permission from the National Joint Registry research subcommittee. Information on research data access requests to the National Joint Registry is available at <https://www.njrcentre.org.uk/research/research-requests/>. No additional data are available.

Transparency: The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: The National Joint Registry and its patient

representative group have been made aware of the results of this study and will support the dissemination of results through their networks. Dissemination to the wider public will be supported by Queen Mary University of London, University of Oxford, and the University of Bristol. The (NIHR) Biomedical Research Centre (BRC) at these institutions will also support dissemination through the press, media, and charities. Surgeons and policy makers will be informed of the results through presentations at national and international scientific meetings.

Provenance and peer review: Not commissioned; externally peer reviewed.

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

- 1 National Joint Registry. *National Joint Registry 22nd Annual Report*. National Joint Registry for England, Wales, Northern Ireland and Isle of Man. 2025
- 2 Kurtz S, Ong K, Lau E, et al. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am* 2007;89:780-5. doi:10.2106/JBJS.F.00222
- 3 Melvin JS, Karthikeyan T, Cope R, et al. Early failures in total hip arthroplasty—a changing paradigm. *J Arthroplasty* 2014;29:1285-8. doi:10.1016/j.arth.2013.12.024
- 4 Ulrich SD, Seyler TM, Bennett D, et al. Total hip arthroplasties: what are the reasons for revision? *Int Orthop* 2008;32:597-604. doi:10.1007/s00264-007-0364-3
- 5 Aqil A, Shah N. Diagnosis of the failed total hip replacement. *J Clin Orthop Trauma* 2020;11:2-8. doi:10.1016/j.jcot.2019.11.003
- 6 Hailer NP, Weiss RJ, Stark A, et al. The risk of revision due to dislocation after total hip arthroplasty depends on surgical approach, femoral head size, sex, and primary diagnosis. An analysis of 78,098 operations in the Swedish Hip Arthroplasty Register. *Acta Orthop* 2012;83:442-8. doi:10.3109/17453674.2012.733919
- 7 Park KT, Lee DH, An JH, et al. Causes of reoperations after primary total hip arthroplasty: a retrospective cohort study over 20 years. *J Arthroplasty* 2025;40:2338-2346.e1. doi:10.1016/j.arth.2025.02.032
- 8 Kelmer G, Stone AH, Turcotte J, et al. Reasons for revision: primary total hip arthroplasty mechanisms of failure. *J Am Acad Orthop Surg* 2021;29:78-87. doi:10.5435/JAAOS-D-19-00860
- 9 Department of Health & Social Care. Fit for the future: 10 Year Health Plan for England. GOV.UK. <https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future/fit-for-the-future-10-year-health-plan-for-england-executive-summary> 2025.
- 10 NHS England. Millions to benefit from NHS robot drive. <https://www.england.nhs.uk/2025/06/millions-to-benefit-from-nhs-robot-drive/#:~:text=Millions%20of%20patients%20will%20benefit,24%2C%20according%20to%20NHS%20projections.> 2025.
- 11 Griffin J, Davis ET, Parsons H, et al. Robotic arthroplasty clinical and cost effectiveness randomised controlled trial (RACER-knee): a study protocol. *BMJ Open* 2023;13:e068255. doi:10.1136/bmjopen-2022-068255
- 12 Griffin J, Davis ET, Parsons H, et al. UK robotic arthroplasty clinical and cost effectiveness randomised controlled trial for hips (RACER-Hip): a study protocol. *BMJ Open* 2023;13:e079328. doi:10.1136/bmjopen-2023-079328
- 13 Wrightington WaLNFT. Self pay tariff 2023/2024. <https://www.wvl.nhs.uk/media/FOI/April%202024/9730%20-%20Self%20Pay%20Tariff%2023.24.pdf>. 2024.
- 14 Parsons H, Metcalfe A, Griffin J, et al. Robotic-arm-assisted versus conventional total knee replacement (RACER-Knee): a pragmatic, multicentre, participant-masked and assessor-masked, superiority, randomised controlled trial. *Lancet* 2026;408:924-34. doi:10.1016/S0140-6736(26)00986-4
- 15 National Institute of Health and Care Excellence. Robot-assisted surgery for orthopaedic procedures: early value assessment. Recommendations for research. <https://www.nice.org.uk/guidance/hte22/chapter/1-Recommendations>. NICE 2025.
- 16 Barakat N, Ramamurti P, Duensing IM, et al. Financial conflicts of interest and industry funding are associated with conclusions favorable to new technologies: a review of published economic analyses in hip and knee arthroplasty. *J Arthroplasty* 2024;39:S299-S305.e9. doi:10.1016/j.arth.2024.02.054
- 17 Mostafa O, Malik M, Qayum K, et al. Robotic-assisted versus conventional total knee arthroplasty: a systematic review and meta-analysis of alignment accuracy and clinical outcomes. *Annals of Medicine & Surgery* 2025;87:867-79. doi:10.1097/MS9.0000000000002919

- 18 National Joint Registry. Data Completeness and Quality. <https://reports.njrcentre.org.uk/Data-Completeness-and-quality#:~:text=The%20linkability%20rate%20for%202024,registry%20since%202003%20is%2094.39%25.> 2026.
- 19 National Joint Registry. NJR data collection forms. <https://www.njrcentre.org.uk/healthcare-providers/njr-data-collection-forms-mdsv8/>. 2026.
- 20 Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA* 2022;328:2446-7. doi:10.1001/jama.2022.21383
- 21 Kirkwood BR, Sterne JA. *Essential medical statistics*: John Wiley & Sons 2010.
- 22 Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. *Biom J* 2009;51:171-84. doi:10.1002/bimj.200810488
- 23 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009;28:3083-107. doi:10.1002/sim.3697
- 24 Lenguerrand E, Whitehouse MR, Kunutsor SK, et al. Mortality and re-revision following single-stage and two-stage revision surgery for the management of infected primary knee arthroplasty in England and Wales: evidence from the National Joint Registry. *Bone Joint Res* 2022;11:690-9. doi:10.1302/2046-3758.1110.BJR-2021-0555.R1
- 25 VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the e-value. *Ann Intern Med* 2017;167:268-74. doi:10.7326/M16-2607
- 26 Kayani B, Konan S, Huq SS, et al. The learning curve of robotic-arm assisted acetabular cup positioning during total hip arthroplasty. *Hip Int* 2021;31:311-9. doi:10.1177/1120700019889334
- 27 Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006;163:262-70. doi:10.1093/aje/kwj047
- 28 Bensa A, Pagliazzi G, Miele A, et al. Robotic-assisted total hip arthroplasty provides greater implant placement accuracy and lower complication rates, but not superior clinical results compared to the conventional manual approach: a systematic review and meta-analysis. *J Arthroplasty* 2025;40:1921-31. doi:10.1016/j.arth.2024.12.014
- 29 Van Dooren BJ, Bos P, Peters RM, et al. Time trends in case-mix and risk of revision following hip and knee arthroplasty in public and private hospitals: a cross-sectional analysis based on 476,312 procedures from the Dutch Arthroplasty Register. *Acta Orthop* 2024;95:307-18. doi:10.2340/17453674.2024.40906
- 30 Kunutsor SK, Barrett MC, Beswick AD, et al. Risk factors for dislocation after primary total hip replacement: a systematic review and meta-analysis of 125 studies involving approximately five million hip replacements. *Lancet Rheumatol* 2019;1:e111-21. doi:10.1016/S2665-9913(19)30045-1
- 31 Wu VJ, Ross BJ, Sanchez FL, et al. Complications following total hip arthroplasty: a nationwide database study comparing elective vs hip fracture cases. *J Arthroplasty* 2020;35:2144-48. doi:10.1016/j.arth.2020.03.006
- 32 Zhang F, Li H, Ba Z, et al. Robotic arm-assisted vs conventional unicompartmental knee arthroplasty: A meta-analysis of the effects on clinical outcomes. *Medicine (Baltimore)* 2019;98:e16968. doi:10.1097/MD.00000000000016968

Supplementary file: Appendix