

Functional Versus Mechanical Alignment in Robotic-Assisted Primary Total Knee Arthroplasty: An Updated Systematic Review and Meta-Analysis of Randomized Controlled Trials

Review began 10/05/2026
Review ended 10/21/2026
Published 10/22/2026

© Copyright 2026
Zin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118404

Nay Aung Zin ¹, Sai Tun ², Min Maung Maung ³, Phay Phone Paing ⁴, Phyo Thant ^{5,6}, Htoo Aung Naing Shwe ⁷, Thi La Si Thu ⁸, Thu Ya Aung ⁹, Brang Jar ⁹, Phyu S. Thwin ¹⁰, Paing Ye Kyaw ¹¹, Than Htay ¹²

1. Orthopaedics and Traumatology, Kulhudhuffushi Regional Hospital, Kulhudhuffushi, MDV 2. Orthopaedics and Traumatology, Royal Devon University Healthcare NHS Foundation Trust, Barnstaple, GBR 3. Orthopaedics and Traumatology, Central Clinic, Malé, MDV 4. Orthopaedics and Traumatology, Raja Isteri Pengiran Anak Saleha Hospital, Bandar Seri Begawan, BRN 5. Orthopaedic and Traumatology, University of Medicine 1, Yangon, Yangon, MMR 6. Orthopaedic Surgery, Kulhudhuffushi Regional Hospital, Kulhudhuffushi, MDV 7. Orthopaedics and Traumatology, K. Himmaphushi Health Centre, Himmaphushi, MDV 8. Orthopaedics and Traumatology, Yangon Orthopaedic Hospital, Yangon, MMR 9. Orthopaedics and Traumatology, University of Medicine 1, Yangon, Yangon, MMR 10. Emergency Department, Kulhudhuffushi Regional Hospital, Kulhudhuffushi, MDV 11. Emergency Department, Nalhivaram Health Centre, Nalhivaram, MDV 12. Orthopaedics and Traumatology, 200-Bedded Teaching Hospital, Magway, Magway, MMR

Corresponding author: Nay Aung Zin, dr.nayaungzin1987@gmail.com

Abstract

Functional alignment (FA) in robotic-assisted total knee arthroplasty (TKA) aims to individualize component positioning and soft-tissue balance. However, whether these technical differences improve patient-reported outcomes compared with mechanical alignment (MA) remains uncertain.

A systematic review and meta-analysis of randomized controlled trials comparing FA and MA in primary robotic-assisted TKA was conducted. MEDLINE, Embase, and CENTRAL were searched, supplemented by trial registries and citation searching. The primary outcome was the Forgotten Joint Score (FJS) at approximately 12 months. Secondary outcomes included other patient-reported measures, soft-tissue release, radiological alignment, and complications. Risk of bias was assessed using RoB 2, and evidence certainty using Grading of Recommendations Assessment, Development and Evaluation (GRADE). Random-effects meta-analysis was performed where appropriate, using restricted maximum likelihood estimation and Hartung-Knapp inference.

Of 406 database records, 11 reports representing seven independent randomized trial families were included. A definitive 12-month FJS meta-analysis was inappropriate because of methodological and reporting heterogeneity. Two randomized comparisons contributed to the early FJS analysis ($\leq$ six months). The pooled mean difference favoured FA by 8.62 points, but the Hartung-Knapp 95% confidence interval was wide (-16.90 to 34.14 ; $I^2=0\%$). A DerSimonian-Laird sensitivity analysis yielded a narrower interval (95% CI 4.56 – 12.68). FA was associated with fewer soft-tissue releases across several included trials. Radiological findings generally favoured preservation of individualized alignment, while patient-reported outcomes showed no consistent clinically important improvement. Serious complications and revisions were uncommon, and certainty of evidence for principal outcomes was very low.

Robotic FA reduces the requirement for soft-tissue release and produces distinct radiological alignment patterns compared with MA. However, current randomized evidence does not establish superior patient-reported outcomes. Early FJS findings remain uncertain because of the limited evidence base and imprecision. Larger randomized trials with standardized protocols and longer follow-up are required.

Categories: Medical Education, Orthopedics, Anatomy

Keywords: functional alignment, mechanical alignment, meta-analysis, randomized controlled trial, robotic-assisted surgery, robotic-assisted total knee arthroplasty, total knee arthroplasty

Introduction And Background

Osteoarthritis is a major and increasing global health burden, with the knee among the most commonly affected joints [1]. For patients with advanced symptomatic knee osteoarthritis that is unresponsive to non-operative treatment, total knee arthroplasty (TKA) is an established surgical treatment, providing substantial pain relief and functional improvement. Long-term implant durability is generally excellent, with contemporary studies reporting high survivorship at extended follow-up [2]. Nevertheless, implant survival does not fully reflect the patient experience. Systematic reviews have shown that a clinically important proportion of patients remain dissatisfied following primary TKA despite technically successful surgery [3,4]. Persistent postoperative pain and dissatisfaction may arise from multiple factors beyond component positioning, including soft-tissue dysfunction, altered pain processing, and extra-articular conditions such as coexisting hip or lumbar spine pathology. Consequently, postoperative patient-reported outcomes should not be attributed exclusively to implant alignment or surgical technique.

Component alignment and soft-tissue balance are important modifiable aspects of TKA. Mechanical alignment (MA) has traditionally aimed to position the femoral and tibial components relative to their mechanical axes and restore a neutral limb axis. Although this strategy has demonstrated reliable long-term

How to cite this article

Zin N, Tun S, Maung M, et al. (October 22, 2026) Functional Versus Mechanical Alignment in Robotic-Assisted Primary Total Knee Arthroplasty: An Updated Systematic Review and Meta-Analysis of Randomized Controlled Trials. Cureus 18(10): e118404. DOI 10.7759/cureus.118404

implant survival, it applies a relatively uniform alignment target to patients with diverse constitutional anatomy. Alternative personalized alignment strategies have therefore been developed to better reproduce individual joint-line orientation and native knee characteristics [5,6].

Functional alignment (FA) is a contemporary individualized alignment strategy that incorporates patient-specific anatomy and intraoperative soft-tissue balance to guide component positioning within predefined alignment boundaries, commonly described as guardrails. Rather than pursuing neutral mechanical alignment in every patient, FA permits controlled adjustments of component positioning to optimize flexion and extension gaps, maintain ligament tension, and reduce unnecessary soft-tissue release [5,6]. These alignment boundaries vary between protocols and should not be assumed to be identical across studies. FA should therefore be distinguished from unrestricted kinematic alignment and other personalized alignment strategies. Conventional MA generally aims to restore neutral mechanical limb alignment, whereas adjusted MA permits defined modifications to conventional targets to improve intraoperative balance.

The development of robotic-assisted TKA has facilitated the implementation and investigation of FA. Robotic systems enable three-dimensional planning, precise bone preparation, incremental component adjustment, and real-time assessment of soft-tissue balance [7]. Randomized evidence comparing robotic-assisted with conventional TKA has demonstrated improved precision of component positioning and fewer alignment outliers, although consistent superiority in short- to medium-term patient-reported outcomes has not been established [7]. When both treatment groups in a randomized controlled trial (RCT) undergo surgery using the same robotic technology, the comparison can more directly evaluate alignment philosophy while reducing variability attributable to surgical execution. This rationale was also specified in the review protocol.

An increasing number of randomized controlled trials have compared FA with MA in robotic-assisted primary TKA, but their findings have not been uniform. Young et al. reported substantially fewer soft-tissue releases with FA, although Forgotten Joint Score (FJS) was not significantly superior at two years [8]. Manara et al. demonstrated differences in balancing and kinematic characteristics between FA and adjusted mechanical-axis alignment [9]. The multicentre RASKAL factorial randomized trial similarly demonstrated substantially fewer soft-tissue releases with FA but no clinically meaningful superiority in major patient-reported outcomes through two years [10]. Bilateral randomized studies by Masilamani et al. [11] and Chompoosang et al. [12] also reported reduced soft-tissue release requirements with FA, with Chompoosang et al. additionally reporting higher FJS at six months [12]. Wang et al. reported favourable early recovery and intraoperative outcomes [13], while Airapetov et al. reported greater preservation of native coronal-plane alignment phenotype with FA [14].

Consequently, randomized evidence appears to demonstrate a relatively consistent technical effect of FA but a less certain patient-reported clinical effect. Reduced soft-tissue release, altered component orientation, or greater preservation of constitutional alignment does not necessarily translate into an improvement that patients perceive as clinically meaningful. Interpretation is further complicated by differences in FA definitions and guardrails, conventional versus adjusted MA comparators, robotic platforms, follow-up periods, and outcome reporting. The available evidence also includes parallel-group, paired bilateral, and factorial randomized designs, requiring design-appropriate statistical methods to avoid unit-of-analysis errors or inappropriate pooling.

Previous evidence syntheses have generally reported favourable technical outcomes with FA, with some suggesting improvements in patient-reported outcomes [15,16]. However, these reviews differ in study eligibility, comparator definitions, follow-up periods, and statistical handling. Gao et al. conducted a meta-analysis of randomized trials [15], whereas Jamaledine et al. specifically evaluated functional versus mechanical alignment in robotic primary TKA [16]. An updated RCT-focused synthesis is therefore warranted to incorporate newer randomized evidence and explicitly account for comparator definition, trial design, follow-up period, and complex randomized structures.

The central unresolved question is whether the reproducible technical effects of FA - particularly reduced soft-tissue release and greater preservation of individualized joint-line and limb alignment - translate into clinically meaningful improvements in patient-reported outcomes. We therefore conducted an updated systematic review and meta-analysis of randomized controlled trials comparing functional with mechanical alignment in robotic-assisted primary TKA. The primary clinical outcome was FJS at approximately 12 months, with earlier and longer-term FJS evaluated separately. Secondary outcomes included other patient-reported measures, soft-tissue release and intraoperative outcomes, radiological alignment, complications, reoperation, and revision.

Review

Method

Study Design and Registration

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement. The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD420261505183). The registration date and any subsequent protocol amendments will be reported according to the official registration record. Multiple publications arising from the same randomized population were linked as a single trial family to prevent double-counting of participants and outcomes.

The review was designed to evaluate randomized evidence comparing FA with MA in robotic-assisted primary TKA. Multiple publications arising from the same randomized population were linked as a single

trial family to prevent double-counting of participants or outcomes.

Eligibility Criteria

Studies were selected according to prespecified population, intervention, comparator, outcome, and study-design criteria.

Adults aged >18 years undergoing elective primary TKA, principally for symptomatic or end-stage knee osteoarthritis, were eligible. Studies of revision TKA, unicompartmental knee arthroplasty, patellofemoral arthroplasty, cadaveric procedures, or simulation studies were excluded.

The intervention of interest was FA performed using a robotic-assisted platform, including MAKO (Stryker Orthopaedics, Mahwah, NJ, USA), ROSA (Zimmer Biomet, Warsaw, IN, USA), CORI (Smith & Nephew, Inc., Memphis, TN, USA), or an equivalent robotic system. The primary comparator was MA performed using the same or a comparable robotic platform. Conventional MA, generally targeting neutral coronal mechanical limb alignment, was considered the principal comparator. Trials employing adjusted MA were eligible but classified separately because their permitted component-positioning and limb-alignment adjustments may differ from conventional MA. In particular, Manara et al. were evaluated as a distinct comparator category. The exact alignment boundaries and balancing rules applied in that trial will be reported from its original surgical protocol rather than assumed to be equivalent to conventional MA.

Randomized controlled trials were eligible, including parallel-group, paired bilateral, and factorial randomized designs. Observational cohorts, retrospective comparative studies, case series, case reports, editorials, reviews, protocol-only publications without outcome data, and non-randomized studies were excluded. Reports evaluating kinematic alignment, restricted kinematic alignment, inverse kinematic alignment, or another personalized strategy were excluded unless the intervention was explicitly defined as FA and fulfilled the prespecified eligibility criteria.

Studies were required to report at least one extractable patient-reported, clinical, intraoperative, radiological, or safety outcome.

Information Sources

The electronic literature search included MEDLINE via PubMed and the Cochrane Central Register of Controlled Trials (CENTRAL), with both databases searched on 13 September 2026. The search log documented 192 records from PubMed and 30 records from CENTRAL. An Embase search via Ovid was also reported in the original manuscript as having been conducted on 14 September 2026, with database coverage from 1974 to 10 September 2026. However, the corresponding search history and retrieval records were not available for independent verification. Therefore, the Embase search was considered unverified pending reconciliation with the original search documentation.

Supplementary searches were conducted in ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP). Reference lists of eligible studies and relevant systematic reviews were examined, and forward citation searching of included trial families was undertaken using Google Scholar. Additional relevant systematic reviews were screened to identify potentially eligible randomized trials.

Scopus and Web of Science Core Collection were prespecified information sources but could not be searched because database access was unavailable. This departure from the planned search strategy was documented as a protocol deviation. Complete reproducible search strategies for verified databases and trial registries are presented in Appendix 1.

Search Strategy

Search strategies combined controlled vocabulary, where available, with free-text terms relating to total knee arthroplasty, functional alignment, and robotic-assisted surgery. Robotic search terms included robot, robotic, MAKO, ROSA, and CORI. No randomized-trial filter was applied because randomized design may not be consistently indexed, particularly for paired bilateral and factorial studies.

The PubMed search strategy was:

```
("Arthroplasty, Replacement, Knee"[Mesh] OR "total knee arthroplasty"[Title/Abstract] OR "total knee replacement"[Title/Abstract] OR TKA[Title/Abstract])
AND ("functional alignment"[Title/Abstract])
AND ("Robotic Surgical Procedures"[Mesh] OR robot*[Title/Abstract] OR MAKO[Title/Abstract] OR ROSA[Title/Abstract] OR CORI[Title/Abstract])
```

Equivalent concepts were adapted for Embase and CENTRAL. The complete electronic database and trial-registry search strategies are presented in Appendix 1.

Study Selection

The available search log documented 192 records from MEDLINE via PubMed and 30 from CENTRAL. The previous manuscript reported 406 records across PubMed, Embase, and CENTRAL, followed by removal of 177 duplicates and screening of 229 unique records. However, the Embase search history and export were unavailable for independent verification. Consequently, the reported total of 406 records and the associated

deduplication and screening counts could not be fully reconciled against the available database records.

The existing screening summary reported 24 full-text reports assessed, of which 13 were excluded and 11 reports representing seven independent randomized trial families were included. These downstream counts remain provisional pending reconciliation with the complete bibliographic search exports and screening records.

Supplementary trial-registry searches were documented separately. No additional independent eligible randomized trial was identified in the recorded supplementary searches.

Data Extraction

Data were extracted using a standardized framework covering study design, participant characteristics, sample sizes, robotic platform, alignment protocols, comparator definitions, follow-up duration, and clinical outcomes. Continuous outcomes were recorded as means, standard deviations, and sample sizes, while dichotomous outcomes were extracted as event counts and denominators. Numerical estimates were checked against original publications and supplementary materials.

For paired bilateral trials, within-patient effect estimates were preferred to preserve statistical dependence. Factorial trials were analyzed using relevant robotic-only comparisons, while adjusted-MA comparators were considered separately. Missing standard deviations were derived from appropriate reported statistics when feasible. Median-to-mean conversions were restricted to sensitivity analyses, with assumptions documented. Inconsistent or insufficiently reported data were excluded from quantitative pooling. Multiple reports from the same trial were cross-checked to avoid duplication.

Outcomes

The FJS is a validated patient-reported outcome measure designed to assess joint awareness during daily activities following arthroplasty [17]. The prespecified primary clinical outcome was the FJS at approximately 12 months [17]. Follow-up periods were considered separately to avoid inappropriate combination of clinically different recovery stages. FJS measured at ≤six months was categorized as an early outcome, while longer-term FJS was considered separately.

Secondary patient-reported and clinical outcomes included the Oxford Knee Score (OKS) [18], Knee injury and Osteoarthritis Outcome Score (KOOS) [19], KOOS-JR [20], Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [21], Knee Society Score (KSS) [22], pain, satisfaction, and range of motion.

Intraoperative outcomes included the requirement for soft-tissue release, type of release, gap balance, operative duration, blood loss, and intraoperative component adjustments where reported.

Radiological outcomes included hip-knee-ankle alignment, joint-line orientation, component positioning, alignment outliers, and preservation of constitutional alignment or Coronal Plane Alignment of the Knee (CPAK) phenotype [23]. The study-selection process is presented according to PRISMA 2020 [24].

Safety outcomes included complications, infection, manipulation under anaesthesia, fracture, reoperation, revision, instability, and loosening.

Risk-of-Bias Assessment

The Cochrane Risk of Bias tool for randomized trials (RoB 2) tool was applied at the level of individual trial results, accounting for outcome type and follow-up period. Seven consensus assessments were available, including one result judged at high risk of bias and six with some concerns. The high-risk judgment applied specifically to the 24-month FJS result reported by Young et al., primarily because of concerns regarding post-randomization exclusions and deviations from the intended intervention.

For the six-month FJS and intraoperative soft-tissue release outcomes reported by Young et al. [8], additional outcome-specific verification was required. In particular, assessment of six-month FJS depended on the completeness of follow-up data and the prespecified analysis population, whereas assessment of soft-tissue release required consideration of outcome definitions, surgeon awareness of allocation, and protocol-defined balancing techniques.

The risk of bias of included randomized controlled trials was assessed using the revised Cochrane RoB 2, considering bias arising from the randomization process, deviations from intended interventions, missing outcome data, measurement of outcomes, and selection of the reported result [25].

The lower frequency of soft-tissue release observed with FA was interpreted as a technical or process outcome, because the surgical strategies differed in their intended balancing mechanisms. This design-related difference was distinguished from bias arising from deviations, outcome measurement, or selective reporting.

Data Synthesis and Statistical Analysis

Meta-analysis was performed only when studies were considered sufficiently comparable with respect to intervention definition, comparator, outcome, follow-up period, and randomized design.

For continuous outcomes measured using the same instrument, treatment effects were expressed as mean

differences (MDs) with 95% confidence intervals (CIs). Standardized mean differences were planned when conceptually similar outcomes were measured using different scales. Dichotomous outcomes were summarized using risk ratios (RRs) with 95% CIs when quantitative pooling was methodologically appropriate.

A random-effects framework was prespecified because clinical and methodological variation was anticipated across robotic platforms, FA boundaries, implants, comparator definitions, and trial populations. Although the protocol specified a random-effects framework, it did not define the between-study variance estimator or confidence-interval method. Restricted maximum likelihood (REML) estimation with Hartung-Knapp inference was therefore selected for the final primary analysis to account for uncertainty in the pooled effect estimate, particularly when only two studies contributed. DerSimonian-Laird analysis using conventional normal-based confidence intervals was performed as a sensitivity analysis.

Statistical heterogeneity was assessed using I^2 and τ^2 . Decisions regarding quantitative pooling were not based on I^2 alone but also considered the magnitude and direction of treatment effects and clinical and methodological differences among studies. A prediction interval was not calculated for the early FJS meta-analysis because only two randomized comparisons were available. With such a limited evidence base, estimation of between-study variance is unstable, making prediction intervals insufficiently reliable for clinical interpretation. The absence of observed statistical heterogeneity was therefore interpreted cautiously.

For the early FJS meta-analysis, two sufficiently compatible randomized comparisons contributed to the core quantitative synthesis. One comparison, reported by Chompoosang et al., involved bilateral randomization of contralateral knees within the same patients. Although the published report provided group-level means, standard deviations, and a treatment-effect confidence interval, the within-patient covariance required to independently verify the paired variance could not be established from the available summary data.

To assess the influence of this uncertainty, an exploratory correlation-based sensitivity analysis was performed. The variance of the paired mean difference was reconstructed using the reported marginal standard deviations and the number of paired observations, assuming within-patient correlations (ρ) of 0, 0.25, 0.50, 0.75, and 0.90. Under each assumption, the pooled effect and its 95% confidence interval were recalculated using REML estimation and unmodified Hartung-Knapp inference. These correlation values represented hypothetical sensitivity scenarios rather than observed or estimated patient-level correlations.

The original extraction-based pooled estimate was retained for comparison, while the correlation-based analyses were used to evaluate the robustness of the direction and precision of the pooled effect. Because the true within-patient covariance was unavailable, the resulting estimates were interpreted cautiously, and no correlation assumption was considered definitively correct.

Special Trial Designs

Paired bilateral randomized trials were not analyzed as though individual knees were statistically independent. Where studies reported paired mean differences and corresponding confidence intervals or other measures of uncertainty, these paired estimates were used directly. When the within-patient information required for a valid paired binary analysis was unavailable, such outcomes were synthesized narratively rather than treating contralateral knees as independent observations.

For factorial randomized trials, the randomized structure was preserved. In the RASKAL trial, the strict primary review comparison required robotic FA versus robotic MA. Overall marginal alignment estimates combining robotic-assisted and computer-assisted surgery were therefore not treated as equivalent to a purely robotic comparison and were retained for appropriately labelled narrative or sensitivity interpretation.

Adjusted-MA comparators were similarly distinguished from conventional MA. The Manara trial was therefore retained as randomized evidence but considered separately where its adjusted MA comparator limited direct comparability with conventional robotic MA.

Missing and Non-standard Data

Where standard deviations were not reported, they could be derived from standard errors, confidence intervals, or other appropriate statistics when mathematically valid.

Outcomes reported as medians and interquartile ranges were retained in their original form for the principal synthesis. Where a conversion to approximate means and standard deviations was undertaken for sensitivity analysis, the conversion method, assumptions, source statistics, and resulting estimates were required to be documented explicitly. Converted observations were excluded from the core analysis unless their use was justified and the sensitivity of the findings to the conversion assumptions could be assessed.

Where outcome definitions, numerical reporting, or directional conventions could not be verified with sufficient confidence, the data were not forced into quantitative synthesis. This applied, for example, to internally inconsistent soft-tissue release data and radiological measurements for which the sign convention could not be established reliably.

Sensitivity and Additional Analyses

Prespecified or methodologically relevant sensitivity analyses considered comparator definition, trial design,

data format, and statistical assumptions. These included conventional MA versus adjusted MA, parallel versus paired bilateral designs, exclusion of converted or imputed data, and exclusion of mixed-technology estimates that did not provide a strict robotic FA-versus-robotic MA comparison.

For the early FJS meta-analysis, the DerSimonian-Laird model was used as a sensitivity analysis to the primary REML/Hartung-Knapp analysis. Conversion of median and interquartile-range data from Airapetov et al. was also examined only as a secondary sensitivity analysis.

Formal assessment of small-study effects using funnel plots or statistical tests was planned only when at least 10 sufficiently comparable studies contributed to an outcome. This threshold was not reached for the principal quantitative outcomes; therefore, formal funnel-plot assessment was not undertaken.

Certainty of Evidence

The certainty of evidence for the principal outcomes was assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework, which considers risk of bias, inconsistency, indirectness, imprecision, and publication bias when determining confidence in effect estimates [26,27].

The certainty of evidence was rated very low across all six principal outcome categories in the current GRADE evidence profile (Table 1). For early FJS, the primary random-effects analysis suggested a numerical advantage with FA (MD +8.62 points; 95% CI -16.90 to +34.14). However, the confidence interval was wide and included no effect. The analysis included only two randomized comparisons, one of which used a paired bilateral design requiring assumptions about within-patient correlation. These limitations contributed to very serious imprecision.

Table with 4 columns: Outcome, Main findings, GRADE domains: RoB / Inconsistency / Indirectness / Imprecision, and Certainty. Rows include FJS at approximately 12 months, Early FJS (≤6 months), Soft-tissue release, Oxford Knee Score, Radiological alignment, and Safety outcomes.

TABLE 1: GRADE Summary of Findings for Functional Versus Mechanical Alignment in Robotic-Assisted Primary Total Knee Arthroplasty

Certainty was assessed using GRADE [26,27], considering risk of bias, inconsistency, indirectness, imprecision, and publication bias. Risk of bias was evaluated using RoB 2 [25]. All six outcome categories were rated very low certainty.

Early Forgotten Joint Score (FJS) [17] data from Young et al. [8] and Chompoosang et al. [12] were pooled using a random-effects model with restricted maximum likelihood estimation and Hartung–Knapp inference (MD +8.62; 95% CI -16.90 to +34.14). Although I²=0%, only two comparisons contributed, so clinically important heterogeneity could not be excluded.

Soft-tissue releases occurred in 20/123 FA knees versus 78/121 MA knees in Young et al. [8], corresponding to approximately 482 fewer releases per 1,000 knees. This is a single-trial estimate, not a pooled effect.

Other evidence came from Manara et al. [9], RASKAL [10], Masilamani et al. [11], Wang et al. [13], and Airapetov et al. [14]. Oxford Knee Score is described by Dawson et al. [18]. Outcomes were not pooled when differences in design, comparators, measurement, or available data prevented appropriate synthesis.

Risk-of-bias downgrading remains provisional where outcome-specific assessments are incomplete; certainty judgments require reconciliation after their completion.

Abbreviations: CI, confidence interval; FA, functional alignment; FJS, Forgotten Joint Score; GRADE, Grading of Recommendations Assessment, Development and Evaluation; MA, mechanical alignment; MD, mean difference; RCT, randomized controlled trial; RR, risk ratio.

For FJS at approximately 12 months, differences in comparator definitions, trial designs, follow-up periods, and reporting formats prevented a defensible pooled estimate. Evidence for soft-tissue release suggested fewer releases with FA in several trials, but differences in operative definitions and analytical structures limited quantitative synthesis. The absolute and relative effects reported from Young et al. represent a single randomized comparison rather than an overall pooled effect.

OKS outcomes did not establish a consistent clinically important advantage for either alignment strategy. Radiological outcomes demonstrated differences in reconstruction characteristics, but heterogeneous measurements prevented estimation of a single pooled effect. Safety outcomes were limited by rare events and short follow-up, resulting in very serious imprecision.

Seven outcome-specific RoB 2 consensus assessments were documented, with one result classified as high

risk of bias and six as having some concerns. Because assessments for several other contributing results remained incomplete, the risk-of-bias components of the GRADE ratings were considered provisional. These limitations should be considered when interpreting the overall certainty of the findings.

Protocol Deviations

The systematic review was conducted according to the registered protocol (PROSPERO registration number CRD420261505183), with the following deviations from the planned methods.

First, Scopus and Web of Science Core Collection were prespecified as additional bibliographic information sources but could not be searched because institutional database access was unavailable despite repeated attempts to obtain access. The completed bibliographic searches therefore comprised MEDLINE via PubMed, Embase via Ovid, and the Cochrane Central Register of Controlled Trials (CENTRAL). Supplementary searches included ClinicalTrials.gov, the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP), backward and forward citation searching, and citation mining of relevant systematic reviews.

Second, manual searching of recent journal issues was not performed as a separate formal search. Reference-list screening and citation searching were undertaken as supplementary identification methods.

Third, although contacting study authors was planned when essential methodological or outcome information was unavailable or incomplete, author contact was not undertaken systematically. Consequently, data that could not be reliably extracted or derived from available reports were not incorporated into the principal quantitative synthesis.

Finally, the registered statistical plan specified a random-effects framework but did not define the between-study variance estimator or confidence-interval method. The final analysis implementation used restricted maximum likelihood estimation with Hartung-Knapp inference, with DerSimonian-Laird analysis using conventional confidence intervals as a sensitivity analysis. This represents a post-protocol specification of the statistical implementation rather than a change in the planned random-effects framework.

These deviations were documented to maintain transparency. The prespecified eligibility criteria were not changed as a consequence of the deviations. The incomplete coverage of planned information sources and limitations in obtaining additional study data were considered when interpreting the completeness and certainty of the evidence.

Results

Study Selection

The searches of MEDLINE via PubMed, Embase, and CENTRAL identified 406 records. After removal of 177 duplicates, 229 unique records underwent title and abstract screening. Of these, 205 records were excluded, leaving 24 reports for full-text assessment. All 24 reports were retrieved and assessed for eligibility.

Thirteen reports were excluded after full-text assessment: seven because of an ineligible study design, four because they were protocol-only reports, and two because they evaluated an ineligible technology. Consequently, 11 reports representing seven independent randomized trial families were included in the systematic review. Reports excluded after full-text assessment and the reasons for exclusion are presented in Appendix 2.

Supplementary searches of ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform identified 16 registry records, corresponding to 15 unique records after removal of one duplicate between registry searches. Twelve were already represented in the bibliographic database search set. The remaining three registry-only records were assessed separately and were ineligible. Backward and forward citation searching and citation mining of relevant systematic reviews identified no additional independent eligible randomized trial.

The complete study-selection process is shown in Figure 1.

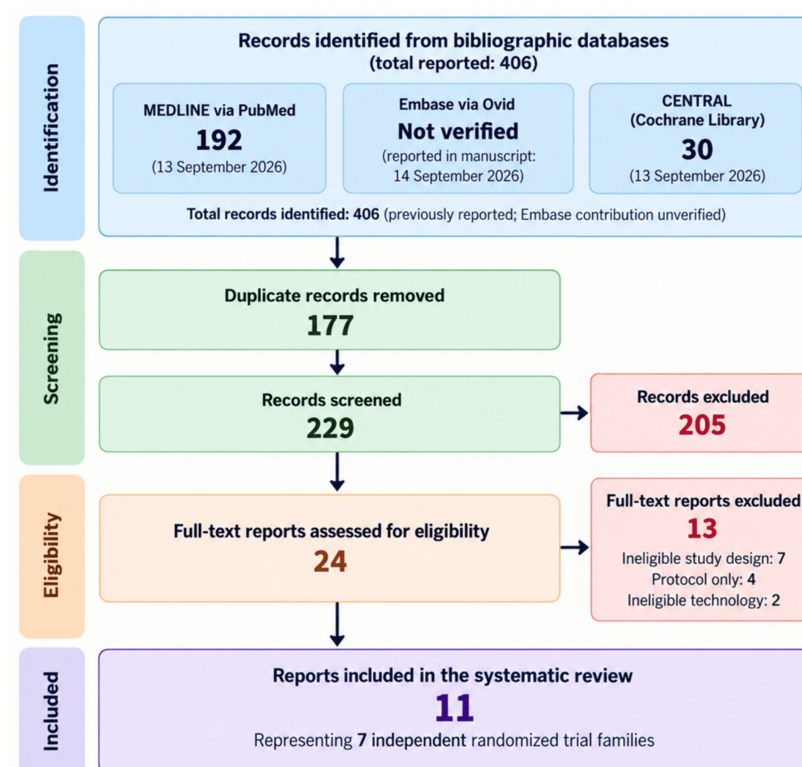


FIGURE 1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 Flow Diagram of Study Identification, Screening, Eligibility Assessment, and Inclusion

The PRISMA 2020 flow diagram summarizes the study-selection counts reported in the review. The manuscript recorded 406 bibliographic records, 177 duplicate records removed, 229 records screened, and 205 records excluded during title and abstract screening. Twenty-four full-text reports were assessed, of which 13 were excluded (seven ineligible study designs, four protocol-only reports, and two ineligible technologies). Eleven reports representing seven independent randomized trial families were included.

Although these counts are arithmetically consistent, the original Embase search export and complete deduplication records remain unavailable for verification. The diagram therefore represents the previously reported screening flow and requires source-level reconciliation before final submission. The study-selection flow diagram is presented according to PRISMA 2020 [24].

Characteristics of Included Studies

The 11 included reports represented seven independent randomized trial families: Young et al. [8], Manara et al. [9], MacDessi et al. (RASKAL) [10], Masilamani et al. [11], Chompoosang et al. [12], Wang et al. [13], and Airapetov et al. [14]. The evidence base comprised parallel-group RCTs, paired bilateral randomized trials, and a registry-nested 2×2 factorial RCT.

Young et al. randomized 278 knees to robotic FA or MA; the per-protocol population comprised 123 FA and 121 MA knees. The trial used the MAKO robotic system (Stryker Orthopaedics, Mahwah, New Jersey, USA) and reported patient-reported outcomes, soft-tissue release, operative outcomes, and safety through two years [8]. Manara et al. included 60 participants (29 FA and 31 adjusted MA) and evaluated clinical, radiological, kinematic, and soft-tissue balancing outcomes at 12 months using MAKO (Stryker Orthopaedics) [9].

The multicentre RASKAL trial used a registry-nested 2×2 factorial design, with 303 participants undergoing surgery. The directly relevant robotic cells comprised 76 FA and 77 MA participants [10]. Wang et al. randomized 99 participants (50 FA and 49 MA) and reported outcomes through six months [13].

Two studies used paired bilateral designs. Masilamani et al. analyzed 65 patients (130 knees), with FA performed in one knee and MA in the contralateral knee [11]. Chompoosang et al. analyzed 30 patients, with 28 paired observations available for FJS at six months [12]. Airapetov et al. randomized 100 participants equally between FA and MA using the CORI robotic platform (Smith & Nephew, Inc.) and reported outcomes through 12 months [14].

The characteristics of the included randomized trial families are summarized in Table 2.

Study	Study design	Sample analyzed	Robotic system	FA comparator	Follow-up	Main outcomes relevant to review
Young et al. [8]	Parallel-group RCT	123 FA; 121 MA (per-protocol)	MAKO	Conventional MA	2 years	FJS, OKS, KOOS, satisfaction, soft-tissue release, operative time, complications
Manara et al. [9]	Randomized substudy	29 FA; 31 adjusted MA	MAKO	Adjusted MA	12 months	FJS, OKS, KOOS-JR, ROM, pain, kinematics, soft-tissue release, radiological alignment
MacDessi et al. (RASKAL) [10]	Multicentre registry-nested 2 × 2 factorial RCT	303 overall; robotic cells: 76 FA, 77 MA	MAKO for robotic arm	MA	2 years	FJS, OKS, KOOS, soft-tissue release, complications
Masilamani et al. [11]	Paired bilateral RCT	65 patients; 130 knees	MAKO	Contralateral MA	Early/intraoperative	Pain, gap balance, soft-tissue release
Chompoosang et al. [12]	Paired bilateral RCT	30 patients; 60 knees; 28 paired at 6 months	MAKO	Contralateral MA	6 months	FJS, KOOS, satisfaction, ROM, soft-tissue release, radiological outcomes
Wang et al. [13]	Parallel-group RCT	50 FA; 49 MA	Not reliably extractable from available report	MA	6 months	Early clinical outcomes, operative time, soft-tissue release, radiological outcomes
Airapetov et al. [14]	Parallel-group RCT	50 FA; 50 MA	CORI	MA	12 months	FJS, KSS, pain, radiological alignment, CPAK phenotype, complications

TABLE 2: Characteristics of the Seven Independent Randomized Trial Families

Outcome instruments and classifications: FJS, Forgotten Joint Score [17]; OKS, Oxford Knee Score [18]; KOOS, Knee injury and Osteoarthritis Outcome Score [19]; KOOS-JR, short-form knee arthroplasty outcomes survey [20]; KSS, Knee Society Score [22]; CPAK, Coronal Plane Alignment of the Knee classification [23].

Robotic system manufacturers: MAKO (Stryker Orthopaedics, Mahwah, NJ, USA); CORI (Smith & Nephew, Inc., Memphis, TN, USA).

FA: functional alignment, MA: mechanical alignment, RCT: randomized controlled trial, ROM: range of motion

Risk of Bias

The final consensus RoB 2 assessment classified one assessed result as High risk of bias and six as Some concerns; no assessed result was judged Low risk overall.

The Young trial's 24-month FJS result was judged at High overall risk of bias, principally because post-randomization exclusions from the per-protocol primary analysis could have affected estimation of the effect of assignment to intervention. Manara, RASKAL, Wang, Masilamani, Chompoosang, and Airapetov were judged as having Some concerns overall. The principal residual concerns included missing outcome data, limited methodological information in some reports, and uncertainty regarding prespecification of the exact reported analyses or outcomes.

The overall pattern of the consensus RoB 2 assessments is presented in Figure 2. Domain-level judgments are summarized in Table 3, and detailed supporting justifications are provided in Appendix 4.



FIGURE 2: Risk-of-Bias Assessment of Randomized Trial Results Using the Cochrane RoB 2 Tool

Traffic-light plot showing consensus risk-of-bias judgments for seven selected randomized trial results. Green (+) indicates low risk of bias, yellow (?) indicates some concerns, and red (!) indicates high risk of bias.

The five domains are D1 (randomization process), D2 (deviations from intended interventions), D3 (missing outcome data), D4 (measurement of the outcome), and D5 (selection of the reported result).

One assessed result, Young et al. at 24 months, was judged to have high overall risk of bias, whereas the remaining six results were judged to have some concerns. Judgments apply only to the specified outcomes and follow-up periods. They should not be generalized to other outcomes reported within the same trials.

Risk-of-bias assessments used the Cochrane RoB 2 tool [25]. Trial sources were Young et al. [8], Manara et al. [9], MacDessi et al. (RASKAL) [10], visual analogue scale (VAS) for pain, as reported by Masilamani et al. [11], Chompoosang et al. [12], Wang et al. [13], and Airapetov et al. [14]. FJS, Forgotten Joint Score [17].

Trial family	Outcome / timepoint assessed	D1: Randomization process	D2: Deviations from intended interventions	D3: Missing outcome data	D4: Measurement of outcome	D5: Selection of reported result	Overall RoB 2 judgment	Main justification
Young et al. [8]	FJS, 24 months	Some concerns	High	Some concerns	Low	Some concerns	High	Randomized design was described, but the primary analysis was per-protocol with post-randomization exclusions. This created concern regarding deviations from the assigned intervention and the effect of assignment. Some uncertainty also remained regarding missing outcome data and selection of the reported analysis.
Manara et al. [9]	FJS, 12 months	Some concerns	Low	Low	Low	Some concerns	Some concerns	Randomized allocation was reported, with no major concern regarding deviations, missing data, or outcome measurement. However, available reporting provided limited information for fully excluding concerns regarding allocation procedures and prespecification of the reported analysis.
MacDessi et al. (RASKAL) [10]	FJS, 24 months	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns	The multicentre factorial randomized design was robust, but interpretation was complicated by the factorial structure and analysis populations. Some uncertainty remained regarding missing outcome data and selection of the reported alignment analysis relevant to this review.
Wang et al. [13]	FJS , 6 months	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Available reporting did not provide sufficient methodological detail to confidently classify all RoB 2 domains as low risk, including allocation procedures, deviations, completeness of outcome data, outcome assessment, and prespecification of analyses.
Masilamani et al. [11]	Early postoperative pain / clinical outcome	Some concerns	Low	Some concerns	Low	Some concerns	Some concerns	The paired bilateral randomized design reduced between-patient confounding, but reporting of randomization procedures and completeness of paired outcome data was limited. Prespecification of the selected analyses could not be fully established.
Chompoosang et al. [12]	FJS, 6 months	Low	Low	Some concerns	Low	Some concerns	Some concerns	Randomization and within-patient bilateral comparison were appropriate, with no major concern regarding deviations or measurement. Some participants were unavailable at six months, and there was residual uncertainty regarding selection of the reported analysis.
Airapetov et al. [14]	FJS, 12 months	Low	Low	Low	Low	Some concerns	Some concerns	Randomization, intervention adherence, completeness of follow-up, and outcome measurement were generally adequately reported. Some concern remained because the available information was insufficient to fully establish prespecification of the reported analyses.

TABLE 3: Cochrane Risk of Bias 2 Assessments for Selected Outcomes of the Included Randomized Trials

Domain-level judgments for specified outcomes and follow-up intervals. D1, randomization process; D2, deviations from intended interventions; D3, missing outcome data; D4, outcome measurement; D5, selection of the reported result. Overall judgments were classified as low risk, some concerns, or high risk. Assessments for additional outcomes contributing to the quantitative synthesis and Grading of Recommendations Assessment, Development and Evaluation (GRADE) evaluation should be reported separately. Assessments used the Cochrane RoB 2 tool [25]. FJS, Forgotten Joint Score [17].

Forgotten Joint Score at Approximately 12 Months

A definitive pooled estimate for the prespecified primary outcome, FJS at approximately 12 months, was not considered methodologically appropriate.

Young et al. reported one-year FJS of 58.4 ± 29.5 in the FA group and 55.4 ± 29.8 in the MA group [8]. Manara et al. reported 12-month FJS of 64.9 ± 32.3 with FA and 68.4 ± 23.5 with adjusted MA; however, the comparator represented adjusted rather than conventional MA [9].

RASKAL reported an overall marginal alignment comparison that incorporated both robotic-assisted and computer-assisted procedures and therefore was not considered equivalent to a strict robotic-only comparison [10]. Airapetov et al. reported 12-month FJS as medians and interquartile ranges: 92 (87.5–98.5) with FA and 91 (86–96) with MA [14].

Because of differences in comparator definition, randomized design, technology, and data format, these results were synthesized narratively rather than combined into a single primary meta-analysis. Overall, the

available randomized evidence did not establish a clinically important difference in FJS at approximately 12 months.

The principal FJS and other patient-reported findings are summarized together with the intraoperative and radiological outcomes in Table 4.

Trial	Clinical outcomes	Soft-tissue release	Other findings
Young et al.	FJS at 6 months: 53.3 ± 29.2 FA vs 47.6 ± 27.9 MA	20/123 FA vs 78/121 MA; RR 0.25 (95% CI 0.17–0.38)	Operative time: 77.4 ± 13.7 vs 78.6 ± 13.1 min
Manara et al.	FJS at 12 months: 64.9 ± 32.3 vs 68.4 ± 23.5	1/29 FA vs 11/31 adjusted MA	HKA: -2.4 ± 2.4° vs -0.7 ± 1.5°
RASKAL	FJS at 1 year: 65.6 ± 29.5 vs 65.0 ± 30.3	8.1% FA vs 44.8% MA*	Marginal alignment comparison
Masilamani et al.	Early postoperative pain assessed	Fewer specific releases with FA	Paired bilateral design
Chompoosang et al.	FJS at 6 months: 67.8 ± 9.5 vs 57.8 ± 10.2	Posteromedial release: 7/30 vs 23/30	KJLO: 3.0 ± 1.8° vs 4.7 ± 1.5°
Wang et al.	Early clinical outcomes favoured FA in some measures	32.0% FA vs 59.2% MA	Operative time: 74.6 ± 11.2 vs 80.1 ± 14.0 min
Airapetov et al.	FJS at 12 months: median 92 (IQR 87.5–98.5) vs 91 (86–96)	Reported release data inconsistent	Native CPAK phenotype preserved in 42/50 FA vs 0/50 MA

TABLE 4: Principal Clinical, Intraoperative, and Radiological Findings of the Included Randomized Trials

Selected clinical, intraoperative, and radiological outcomes from seven randomized trial families comparing functional alignment (FA) with mechanical alignment (MA) in robotic-assisted primary TKA: Young et al. [8], Manara et al. [9], RASKAL [10], Masilamani et al. [11], Chompoosang et al. [12], Wang et al. [13], and Airapetov et al. [14].

RASKAL results are marginal alignment comparisons combining robotic-assisted and computer-assisted procedures, rather than robotic-only estimates. Manara et al. used adjusted MA; Masilamani et al. and Chompoosang et al. used paired bilateral designs. Differences in design, outcome definitions, or reporting precluded pooling.

Abbreviations: CI, confidence interval; CPAK, Coronal Plane Alignment of the Knee classification [23]; FJS, Forgotten Joint Score [17]; HKA, hip–knee–ankle angle; IQR, interquartile range; KJLO, knee joint-line obliquity; MD, mean difference; RR, risk ratio; TKA, total knee arthroplasty.

Early Forgotten Joint Score

Two randomized comparisons, Young et al. and Chompoosang et al., provided sufficiently compatible data for quantitative synthesis of early FJS outcomes at six months [8,12].

At six months, Young et al. reported mean FJS values of 53.3 ± 29.2 with FA and 47.6 ± 27.9 with MA, corresponding to a mean difference (MD) of 5.7 points favouring FA [8]. In the bilateral randomized trial by Chompoosang et al., the corresponding FJS values were 67.8 ± 9.5 and 57.8 ± 10.2, respectively, with a reported MD of 10.0 points (95% CI 4.84–15.16) [12]. Because the latter trial randomized contralateral knees within patients, the precision of its treatment effect depended on within-patient correlation, which could not be independently verified from the available summary data.

The original extraction-based random-effects meta-analysis using REML estimation with Hartung-Knapp inference yielded a pooled MD of 8.62 FJS points favouring FA (95% CI -16.90 to 34.14; I²=0%; τ²=0). Although the point estimate favoured FA, the confidence interval was wide and included the null value. The DerSimonian-Laird sensitivity analysis produced the same pooled MD with a narrower 95% CI (4.56–12.68). These differences in interval width reflected the sensitivity of statistical inference to the selected method when only two comparisons contributed. The original pooled estimate remained conditional on the extracted variance assumption for the bilateral trial.

To evaluate the influence of the unverified paired variance, an exploratory correlation-based sensitivity analysis was conducted using assumed within-patient correlations (ρ) of 0, 0.25, 0.50, 0.75, and 0.90. Across these assumptions, the pooled MD ranged from 8.53 to 9.32 FJS points favouring FA. The corresponding Hartung-Knapp 95% confidence intervals ranged from -17.38 to 34.44 at ρ=0 to -10.60 to 29.24 at ρ=0.90, with all intervals crossing zero. Thus, the direction of the pooled effect was consistent across the examined assumptions, but a statistically significant difference was not established.

Although observed statistical heterogeneity was low, the inclusion of only two randomized comparisons limited the reliability of I² and between-study variance estimates. Furthermore, the clinical importance of the observed mean differences remained uncertain because a definitive minimum clinically important difference applicable to these trial populations was not established.

The original extraction-based meta-analysis is presented in Figure 3, and the correlation-based sensitivity analyses are presented in Appendix 3, including the figure. Overall, the available evidence suggested a possible improvement in early FJS with FA, but the magnitude and precision of this effect remained uncertain.

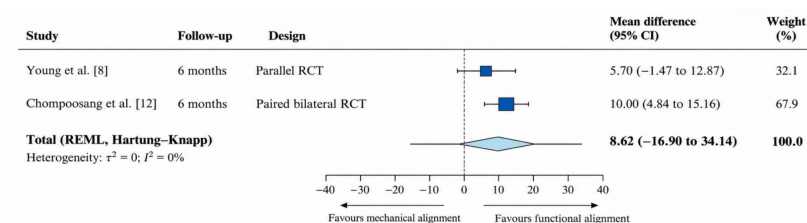


FIGURE 3: Forest Plot of Early Forgotten Joint Score Following Robotic-Assisted Total Knee Arthroplasty: Functional Versus Mechanical Alignment

Forest plot comparing six-month Forgotten Joint Score (FJS) outcomes between functional alignment (FA) and mechanical alignment (MA) in two randomized comparisons. Squares represent study-specific mean differences (MDs), with sizes proportional to inverse-variance weights; horizontal lines indicate 95% confidence intervals (CIs). The diamond represents the pooled effect calculated using restricted maximum likelihood (REML) estimation with Hartung–Knapp inference (MD 8.62; 95% CI -16.90 to 34.14). Positive values favour FA, whereas negative values favour MA. The Forgotten Joint Score is a validated patient-reported outcome measure [17]. Data were obtained from Young et al. [8] and Chompoosang et al. [12].

The analysis is conditional on the extracted variance for the paired bilateral trial by Chompoosang et al., because the within-patient covariance could not be independently verified. Statistical heterogeneity was $I^2=0\%$ and $\tau^2=0$. Given that only two randomized comparisons contributed, heterogeneity estimates were interpreted cautiously, and no prediction interval was calculated. The pooled confidence interval crossed zero, indicating no statistically significant difference in early FJS between alignment strategies. Correlation-based sensitivity analyses are presented separately in the figure in Appendix 3.

Airapetov et al. were excluded from the core analysis because continuous outcomes were reported as medians and interquartile ranges; conversion of these data was considered only in sensitivity analysis [14].

Soft-Tissue Release and Intraoperative Outcomes

A reduction in the requirement for soft-tissue release was one of the most consistent technical findings associated with FA.

In Young et al., soft-tissue release was required in 20 out of 123 FA knees compared with 78 out of 121 MA knees, corresponding to an RR of 0.25 (95% CI 0.17-0.38) [8]. Manara et al. similarly reported release in one out of 29 FA participants compared with 11 out of 31 adjusted-MA participants ($p=0.002$) [9].

RASKAL demonstrated substantially fewer releases with FA. In the factorial alignment comparison, release occurred in 8.1% of FA procedures compared with 44.8% of MA procedures, with a reported OR of 9.2 (95% CI 4.6-18.3) for MA relative to FA [10]. Because this marginal comparison incorporated both robotic-assisted and computer-assisted procedures, it was not pooled with strict robotic-only estimates.

The paired bilateral trials demonstrated the same directional pattern. Masilamani et al. reported more knees achieving balance without additional release with FA and fewer posteromedial, posterior capsular, tibial reduction osteotomy, and superficial medial collateral ligament procedures [11]. Chompoosang et al. reported posteromedial release in seven out of 30 FA knees versus 23 out of 30 MA knees ($p<0.001$) [12].

Wang et al. reported soft-tissue release in 32.0% of FA procedures compared with 59.2% of MA procedures ($p=0.007$) [13]. Operative duration was also shorter with FA (74.6 ± 11.2 versus 80.1 ± 14.0 minutes; $p=0.033$).

An overall pooled estimate for soft-tissue release was not calculated because definitions and thresholds for release differed among trials, Manara used an adjusted-MA comparator, the RASKAL marginal estimate incorporated mixed technology, and valid paired binary synthesis of the bilateral trials required discordant-pair data that were unavailable. Airapetov's release data were excluded from quantitative synthesis because the reported counts and percentages were internally inconsistent [14].

The principal findings across individual randomized trials are summarized in Table 4.

Other Patient-Reported and Clinical Outcomes

Other patient-reported outcomes did not demonstrate a consistent clinically important advantage for either alignment strategy.

In Manara et al., 12-month OKS was 41.9 ± 7.2 with FA and 41.8 ± 5.7 with adjusted MA, while KOOS-JR was 80.7 ± 14.9 and 82.9 ± 13.3 , respectively [9]. Range of motion was also similar between groups.

In RASKAL, the marginal adjusted difference in OKS was 0.5 points (95% CI -1.2 to 2.1) at one year and -0.2 points (95% CI -1.8 to 1.5) at two years [10]. Corresponding KOOS estimates similarly showed no clear clinically important difference.

Young et al. reported no clear overall two-year FJS advantage, although some secondary patient-reported domains and satisfaction measures differed at individual timepoints [8]. Chompoosang et al. reported higher

FJS with FA at three and six months, while differences in other clinical measures were less consistent [12].

Because of differences in outcome instruments, follow-up periods, comparator definitions, and randomized designs, these outcomes were synthesized narratively rather than pooled across clinically heterogeneous comparisons (Table 4).

Radiological Outcomes

Radiological outcomes demonstrated expected differences between the two alignment philosophies. Across the included trials, FA generally permitted greater preservation of constitutional limb alignment, joint-line orientation, or patient-specific component positioning, whereas MA more consistently targeted neutral mechanical alignment [8,9,12,14].

In Chompoosang et al., knee joint-line obliquity was $3.0 \pm 1.8^\circ$ with FA compared with $4.7 \pm 1.5^\circ$ with MA ($p < 0.001$) [12]. Airapetov et al. reported preservation of the native CPAK phenotype in 42 out of 50 (84%) FA participants compared with 0 out of 50 MA participants [14].

A pooled radiological estimate was not calculated because studies differed in angle definitions, sign conventions, reporting formats, follow-up periods, and randomized designs. Wang's HKA measurements were not quantitatively synthesized because the directional convention could not be verified with sufficient confidence [13].

These findings indicate that FA and MA produced different radiological reconstructions; they should not, by themselves, be interpreted as demonstrating clinical superiority of either alignment strategy (Table 4).

Safety Outcomes

Safety events were uncommon across the randomized trials, limiting the precision of comparative estimates.

In Young et al., manipulation under anaesthesia was required in one out of 123 FA knees compared with five out of 121 MA knees. One participant in the FA group developed a deep methicillin-sensitive *Staphylococcus aureus* infection requiring revision [8].

Within the directly relevant robotic cells of RASKAL, manipulation under anaesthesia occurred in one out of 76 FA and one out of 77 MA participants, surgical-site infection in one out of 76 and one out of 77, respectively, and no revisions were reported in either robotic alignment group during the reported follow-up [10]. A discrepancy was identified between the RASKAL table and narrative regarding the alignment group in which one pin-site fracture occurred; therefore, this event was not used to derive a comparative quantitative estimate.

Given the low event counts and limited follow-up, no pooled RR for serious adverse events, infection, reoperation, or revision was considered reliable. The available evidence did not demonstrate a clear comparative safety signal, but clinically important differences in rare adverse outcomes could not be excluded.

Certainty of Evidence

The certainty of evidence was rated very low for all six principal outcome categories (Table 1). For early FJS, serious risk-of-bias concerns and very serious imprecision were identified, particularly because only two randomized comparisons contributed and the Hartung-Knapp confidence interval was wide. At approximately 12 months, methodological differences prevented a defensible pooled FJS estimate. Evidence for soft-tissue release, OKS, and radiological outcomes was limited by varying combinations of risk of bias, inconsistency, indirectness, and imprecision. Safety evidence was particularly limited by the low frequency of adverse events and insufficient precision to exclude clinically important differences. The outcome-specific downgrading judgments are summarized in Table 1.

Discussion

Principal Findings

This systematic review and meta-analysis synthesized 11 reports representing seven independent randomized trial families comparing FA with MA in robotic-assisted primary TKA. The principal findings were that FA consistently produced a technically and radiologically different reconstruction from MA and was associated with a lower requirement for soft-tissue release in several randomized trials. However, these technical differences did not translate into a clearly established patient-reported benefit. The certainty of evidence for the principal outcomes was very low, and evidence concerning complications, revision, and long-term implant durability remained limited.

The most consistent finding concerned soft-tissue balancing. Young et al. reported soft-tissue release in 20 out of 123 FA knees compared with 78 out of 121 MA knees [8], while Manara et al., RASKAL, Chompoosang et al., and Wang et al. demonstrated findings in the same general direction [9,10,12,13]. Masilamani et al. similarly reported fewer specific release procedures with FA in a paired bilateral design [11]. Although methodological differences prevented a valid overall pooled estimate, the consistency of direction across different randomized designs suggests that FA can achieve intraoperative balance with less reliance on soft-tissue release than MA.

In contrast, the evidence for improved patient-reported outcomes was considerably less certain. For early FJS, the pooled point estimate favoured FA by 8.62 points. However, when REML estimation with Hartung-Knapp inference was applied, the 95% CI was wide (-16.90 to 34.14) and crossed the null despite $I^2=0\%$. The corresponding DerSimonian-Laird sensitivity analysis produced a substantially narrower CI of 4.56 to 12.68. This difference is important because only two randomized comparisons contributed to the core analysis. With such a small number of studies, the absence of observed heterogeneity should not be interpreted as evidence that the treatment effect is known precisely. The wider Hartung-Knapp interval more appropriately reflects uncertainty surrounding the pooled estimate.

At approximately 12 months, the available randomized evidence did not support a methodologically defensible single pooled estimate. Young et al. showed only a small numerical difference in FJS, while Manara et al. reported similar scores between FA and adjusted MA [8,9]. RASKAL provided marginal alignment estimates incorporating both robotic and computer-assisted procedures [10], whereas Airapetov et al. reported FJS using medians and interquartile ranges [14]. Pooling these estimates without qualification would have obscured clinically important differences in comparator definition, technology, trial design, and statistical reporting. Therefore, narrative synthesis was considered more appropriate for the prespecified primary outcome.

Interpretation of the Clinical and Technical Findings

The apparent dissociation between improved intraoperative balancing and uncertain patient-reported outcome measure (PROM) benefit is clinically important. FA uses individualized component positioning and bone-resection adjustments within defined boundaries to reproduce aspects of the patient's native alignment and achieve soft-tissue balance. MA, by contrast, generally seeks a neutral mechanical target and may require additional soft-tissue modification when the patient's native anatomy differs substantially from that target.

The reduced requirement for soft-tissue release with FA is biomechanically plausible and reflects fundamental differences between alignment strategies. Robotic technology enables intraoperative assessment of planned bone resections, component positioning, limb alignment, and compartment balance, allowing surgeons to optimize ligament tension through individualized component adjustments rather than relying primarily on soft-tissue releases to achieve predetermined MA targets. However, this difference should be interpreted within the context of the surgical protocols themselves, as FA inherently prioritizes component-position adjustments for balancing. Furthermore, decisions regarding soft-tissue release may be influenced by surgeon judgment and the lack of intraoperative blinding.

Importantly, fewer soft-tissue releases do not necessarily translate into superior clinical outcomes. Although reduced ligamentous intervention may theoretically preserve native soft-tissue integrity and physiological knee kinematics, soft-tissue release remains an intermediate surgical outcome rather than a direct indicator of postoperative pain, function, satisfaction, or implant survivorship. The available randomized evidence does not yet establish whether these technical differences produce clinically meaningful or sustained patient-reported benefits.

The early FJS meta-analysis included only two randomized comparisons and therefore provides limited information regarding the magnitude and consistency of treatment effects. Although the pooled point estimate favoured FA by 8.62 points, the primary Hartung-Knapp 95% confidence interval (-16.90 to 34.14) crossed the null and included both potentially important benefit and harm. The apparent difference between the two trials may reflect differences in patient populations, trial design, outcome variability, or assessment procedures. In particular, the paired bilateral design of Chompoosang et al. allowed patients to compare their operated knees directly, potentially influencing subjective outcome assessment if treatment allocation was known. These findings should therefore be regarded as exploratory rather than definitive evidence of clinical superiority.

Radiological Alignment

Radiological findings consistently demonstrated that FA and MA produced different alignment phenotypes, as would be expected from their underlying philosophies. FA generally permitted greater preservation of constitutional limb alignment, joint-line orientation, and patient-specific component positioning, whereas MA more consistently targeted neutral mechanical alignment [8,9,12,14].

Chompoosang et al., for example, demonstrated differences in knee joint-line obliquity between FA and MA [12], while Airapetov et al. reported preservation of the native CPAK phenotype in 84% of FA participants compared with none in the MA group [14]. These findings confirm that robotic FA is not simply a minor technical modification of MA; it results in a measurably different reconstruction.

However, radiological restoration of constitutional anatomy represents another surrogate outcome. Greater preservation of a patient's pre-arthritis or constitutional alignment cannot itself establish clinical superiority. The relationship between postoperative alignment phenotype, subjective knee function, implant loading, polyethylene wear, instability, and long-term survivorship remains incompletely defined. Consequently, the radiological findings in this review should primarily be interpreted as confirmation that the intended alignment strategies were biomechanically distinct rather than as evidence that one radiographic target is preferable.

Safety and Long-Term Outcomes

The current evidence is particularly limited regarding safety and implant durability. Serious complications, reoperations, and revisions were uncommon across the randomized trials. Young et al. reported one deep

infection requiring revision in the FA group and fewer manipulations under anaesthesia with FA, while the directly relevant robotic cells of RASKAL contained very few adverse events [8,10]. These small numbers do not permit reliable comparative estimates.

Importantly, the absence of an observed difference in rare adverse events should not be interpreted as evidence of equivalent safety. The included trials were primarily designed to assess functional, alignment, or intraoperative outcomes rather than revision risk, and their sample sizes and follow-up durations were inadequate for detecting uncommon complications or differences in implant survival.

This limitation is especially important for individualized alignment strategies. Allowing component and limb alignment to deviate from traditional mechanical targets raises legitimate questions regarding long-term fixation, load distribution, polyethylene wear, instability, and aseptic loosening. Conversely, preservation of constitutional anatomy and reduced soft-tissue disruption could theoretically offer long-term functional advantages. Neither possibility can currently be resolved from the available randomized evidence. Long-term follow-up of existing trials and adequately powered prospective studies are therefore required before conclusions regarding implant durability can be drawn.

Comparison With Previous Evidence

Previous systematic reviews and meta-analyses have examined FA versus MA in TKA, including recent analyses by Gao et al. and Jamaledine et al. [15,16]. The present review differs in several important respects. It focused specifically on randomized evidence relevant to robotic-assisted primary TKA, linked multiple reports from the same randomized population into trial families, and explicitly accounted for paired bilateral and factorial trial designs.

These distinctions materially affected the synthesis. The RASKAL trial, for example, used a 2×2 factorial design in which some published alignment estimates combined robotic-assisted and computer-assisted procedures [10]. Treating those marginal estimates as purely robotic FA-versus-MA evidence would not preserve the specific comparison defined by this review. Similarly, Masilamani et al. and Chompoosang et al. randomized contralateral knees within the same patients [11,12]. Treating these knees as independent observations would ignore within-patient correlation and could produce artificially precise estimates.

The present review also avoided forcing clinically heterogeneous outcomes into a single meta-analysis. Adjusted MA in Manara et al. was distinguished from conventional MA [9], Airapetov's median-based outcomes were not converted for the core analysis [14], and radiological outcomes with uncertain directional conventions were not quantitatively combined. These decisions reduced the amount of data contributing to pooled estimates but preserved the randomized structure and clinical interpretation of the evidence.

The resulting conclusions are consequently more cautious. Although some individual trials and conventional meta-analytic approaches may suggest an early functional advantage with FA, the primary Hartung-Knapp random-effects analysis demonstrates that substantial uncertainty remains regarding the magnitude of this effect.

Strengths

This review was strengthened by its restriction to randomized evidence, identification of independent trial families, and explicit consideration of paired bilateral and factorial randomized designs. Comparator definitions and follow-up intervals were evaluated before pooling, reducing the risk of inappropriate synthesis. The use of sensitivity analyses, outcome-level risk-of-bias assessment, and GRADE further supported a cautious interpretation of the available evidence.

Limitations

This systematic review and meta-analysis has several important limitations. First, although seven independent randomized trial families were included, only two randomized comparisons contributed to the quantitative synthesis of early FJS. The small number of studies limited statistical precision and the reliability of heterogeneity estimation. Although the observed I^2 was 0%, clinically important between-study differences could not be excluded. The wide confidence interval obtained using Hartung-Knapp inference further highlights the uncertainty surrounding the pooled effect estimate.

Second, clinical and methodological heterogeneity limited the feasibility of quantitative synthesis across several outcomes. Differences in FA protocols, robotic platforms, comparator definitions, follow-up durations, and outcome reporting reduced comparability. In particular, trials using adjusted MA could not be considered directly equivalent to those using conventional MA. Furthermore, paired bilateral and factorial randomized designs required specific analytical considerations to avoid inappropriate assumptions of independence or the combination of non-equivalent treatment groups.

Third, incomplete or inconsistent reporting of outcome data restricted quantitative analyses. Some studies reported continuous outcomes as medians and interquartile ranges, while others lacked sufficient information for reliable effect estimation. Variations in soft-tissue release definitions and radiological measurements also prevented meaningful pooling of certain outcomes. Although sensitivity analyses were undertaken where appropriate, these limitations reduced the strength of the available evidence.

Fourth, the risk of bias and certainty of evidence require cautious interpretation. Methodological concerns included post-randomization exclusions, incomplete outcome data, and uncertainty regarding

prespecification of reported analyses. The certainty of evidence was assessed as very low across the principal outcome categories, limiting confidence in the estimated treatment effects.

Fifth, the literature search did not cover all prespecified databases. Scopus and Web of Science Core Collection could not be searched because of access limitations. Although MEDLINE, Embase, CENTRAL, trial registries, and citation searches were used, potentially relevant studies may have been missed. Formal journal handsearching was not undertaken, and study authors were not contacted systematically to clarify missing or incomplete information.

Finally, most included trials reported relatively short-term outcomes, with follow-up extending to approximately two years. Consequently, the available randomized evidence was insufficient to establish whether the technical advantages of FA translate into sustained improvements in patient-reported outcomes, implant survivorship, revision risk, or long-term safety.

These limitations emphasize the need for adequately powered, multicentre randomized trials using standardized alignment definitions, consistent patient-reported outcome measures, appropriate handling of complex trial designs, and longer-term follow-up. Future research should determine whether the observed reductions in soft-tissue release and greater preservation of individualized alignment translate into clinically meaningful and durable benefits for patients.

Implications for Clinical Practice and Future Research

The present evidence suggests that robotic FA can reduce the requirement for soft-tissue release and more closely preserve individualized alignment characteristics. These findings may be relevant when surgeons seek to achieve balanced TKA while limiting soft-tissue disruption. However, the evidence does not currently establish that these technical advantages translate into a clinically important and durable improvement in patient-reported outcomes.

Future trials should use clearly standardized definitions of FA and MA, report the boundaries permitted for component and limb alignment, and distinguish conventional from adjusted MA comparators. Core outcome reporting should include FJS and other validated PROMs at standardized early, one-year, and longer-term timepoints. Reporting of soft-tissue release should also be standardized according to structure, indication, and extent.

For bilateral randomized trials, paired effect estimates and discordant-pair data should be reported to permit statistically appropriate synthesis. Factorial trials should provide estimates for the individual randomized cells when these are clinically relevant. Prospective registration and publication of detailed statistical analysis plans would reduce uncertainty regarding selective reporting.

Most importantly, longer-term follow-up is required. Large multicentre randomized trials with sufficient duration to assess revision, loosening, instability, and implant survival are needed to determine whether the radiological and intraoperative differences observed with FA have meaningful long-term consequences. Future randomized trials should standardize the definitions and permitted boundaries of FA and MA, prospectively specify patient-reported outcomes and clinically important differences, and report outcomes at comparable follow-up intervals. Baseline assessment should also document relevant coexisting hip and lumbar spine disorders, which may influence postoperative pain and patient-reported function independently of knee alignment. Longer follow-up is necessary to determine whether the intraoperative and radiological differences between alignment strategies translate into durable functional improvements, lower revision rates, or improved implant survivorship.

Conclusions

Robotic-assisted FA produces a distinct technical and radiological reconstruction compared with MA in primary TKA and is associated with a reduced requirement for soft-tissue release across several randomized trials. Although the pooled point estimate suggested a possible improvement in early FJS with functional alignment, this effect was highly imprecise when appropriate small-study inference was applied. At approximately 12 months, the available randomized evidence does not establish a clinically important patient-reported advantage of FA over robotic MA.

Evidence regarding complications, revision, and long-term implant durability remains insufficient. Given the very low certainty of evidence for the principal outcomes, the observed technical advantages of FA should not be interpreted as established clinical superiority. Larger, methodologically standardized randomized trials with consistent outcome reporting and longer follow-up are required to determine whether the intraoperative and radiological differences associated with FA translate into meaningful and durable benefits for patients.

Appendices

Appendix 1: Electronic database and trial-registry search strategies

A1. MEDLINE via PubMed

Search date: 13 September 2026

Records retrieved: 192

Restrictions: No search filters, publication-date restrictions, or language restrictions recorded.

Exact search strategy:

("Arthroplasty, Replacement, Knee"[Mesh] OR "total knee arthroplasty"[Title/Abstract] OR "total knee replacement"[Title/Abstract] OR TKA[Title/Abstract]) AND ("functional alignment"[Title/Abstract]) AND ("Robotic Surgical Procedures"[Mesh] OR robot*[Title/Abstract] OR MAKO[Title/Abstract] OR ROSA[Title/Abstract] OR CORI[Title/Abstract])

The search log records successful retrieval of all five prespecified known randomized studies used for search validation.

A2. Embase via Ovid - Search Documentation Status

The manuscript previously reported an Embase search conducted on 14 September 2026, with database coverage extending from 1974 to 10 September 2026. However, the available search-log workbook lists Embase as pending and contains no executed Ovid search history, retrieval count, or exported record set.

The Embase search could therefore not be independently verified. No historical search strategy or record count was reconstructed or inferred. This limitation must be resolved through recovery of the original search history or execution of a newly documented Embase search before the search methods and PRISMA flow diagram can be finalized.

A3. Cochrane Central Register of Controlled Trials (CENTRAL)

Platform: Cochrane Library
Search date: 13 September 2026
Records retrieved: 30
Restrictions: No filters or limits recorded.

Exact recorded search strategy:

#1 ("total knee arthroplasty" OR "total knee replacement" OR TKA) #2 "functional alignment" #3 (robot* OR MAKO OR ROSA OR CORI) #4 #1 AND #2 AND #3

All 30 retrieved trial records were recorded as exported.

A4. ClinicalTrials.gov

Search date: 13 September 2026
Records retrieved: 7
Restrictions: No filters or restrictions recorded.

Recorded search expression:

("total knee arthroplasty" OR TKA) AND "functional alignment" AND (robot OR robotic OR "robot-assisted" OR "robotic-assisted" OR MAKO OR ROSA OR CORI)

All seven retrieved records were documented as exported in RIS format.

A5. World Health Organization International Clinical Trials Registry Platform (WHO ICTRP)

Search date: 13 September 2026
Recruitment status: All
Synonyms: Enabled
Country/date restrictions: None recorded.

Search 1:

Condition: "total knee arthroplasty" AND Intervention: "functional alignment"

Records retrieved: 3. Export format: XML.

Search 2 - supplementary sensitivity search:

Title: "functional alignment" AND Intervention: "robotic"

Records retrieved: 6. Export format: XML.

The two searches were recorded separately; overlap between their results has not been established.

A6. Additional Information Sources

The review search log documents forward citation searches of included randomized trial families using PubMed and Google Scholar. These searches were intended to identify additional eligible trials and companion publications while avoiding double-counting of reports from the same randomized population.

Scopus and Web of Science Core Collection were identified as planned information sources in the

manuscript but were reported as unavailable because of access limitations. No executed search strategies or record counts are documented for these databases.

A7. Search Reproducibility and Reporting

Search strategies are reproduced from the available electronic search log. Recorded database counts refer to records retrieved before deduplication and should not be interpreted as numbers of unique studies.

The exact Embase search strategy, execution history, and retrieved record count remain outstanding. Final bibliographic totals and PRISMA identification counts must be reconciled against the original search exports before submission.

Appendix 2: Reports excluded after full-text assessment

Reports excluded after full-text eligibility assessment and primary reasons for exclusion

No.	Study or trial identifier	Primary reason for exclusion
1	van de Graaf et al.	Ineligible study design
2	Wang Yijun et al.	Ineligible study design
3	Zhang et al.	Ineligible study design
4	Yamagami et al.	Ineligible study design
5	Wierer/Clatworthy et al.	Ineligible study design
6	Manara cohort study	Ineligible study design
7	Lee et al.	Ineligible study design
8	personalKNEE trial	Protocol-only report; no eligible randomized outcome results
9	Kayani et al. / NCT04092153	Protocol-only report; no eligible randomized outcome results
10	NCT04748510	Protocol-only trial record
11	NCT06259032	Protocol-only trial or registry record
12	MARVEL / NCT07551856	Ineligible technology
13	Coenen et al.	Ineligible technology

TABLE 5: Reports excluded after full-text eligibility assessment and primary reasons for exclusion

Thirteen reports were recorded as excluded following full-text eligibility assessment: seven because of ineligible study designs, four because they were protocol-only reports or trial records without eligible outcome data, and two because of ineligible technology.

The exclusions were based on the review's eligibility criteria, which required randomized comparisons relevant to functional versus mechanical alignment in robotic-assisted primary total knee arthroplasty.

The study identifiers and primary exclusion categories were transcribed from the reconciled PRISMA workbook. Full bibliographic citations and detailed eligibility justifications should be cross-checked against the original screening records before final submission.

Abbreviations: NCT, ClinicalTrials.gov trial registration identifier; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Appendix 3: Statistical sensitivity analyses for early Forgotten Joint Score

C1. Purpose

An exploratory sensitivity analysis was conducted to evaluate the influence of the unverified within-patient correlation in the paired bilateral randomized trial by Chompoosang et al. on the pooled six-month Forgotten Joint Score (FJS) estimate.

C2. Data Included

Study	Design	FA, mean ± SD	MA, mean ± SD	Sample size	Mean difference
Young et al.	Parallel-group RCT	53.3 ± 29.2	47.6 ± 27.9	FA 123; MA 121	+5.70
Chompoosang et al.	Paired bilateral RCT	67.8 ± 9.5	57.8 ± 10.2	28 paired patients	+10.00

TABLE 6: Study-level data used for the early FJS meta-analysis

The outcome was the Forgotten Joint Score [17]. Trial data were obtained from Young et al. [8] and Chompoosang et al. [12].

StudyDesignFA, mean ± SDMA, mean ± SDSample sizeMean difference

Young et al. Parallel-group RCT53.3 ± 29.247.6 ± 27.9FA 123; MA 121+5.70

Chompoosang et al.Paired bilateral RCT67.8 ± 9.557.8 ± 10.228 paired patients+10.00

The extracted Chompoosang treatment-effect confidence interval was 4.84-15.16. The within-patient covariance could not be independently established from the available summary data.

C3. Statistical Methods

For Young et al., the variance of the mean difference was calculated as:

$$V_{\text{Young}} = \frac{29.2^2}{123} + \frac{27.9^2}{121}$$

For Chompoosang et al., the variance of the paired mean difference was reconstructed under assumed within-patient correlations:

$$V_{\text{paired}}(\rho) = \frac{9.5^2 + 10.2^2 - 2\rho(9.5)(10.2)}{28}$$

The following correlation assumptions were evaluated: ρ = 0, 0.25, 0.50, 0.75, and 0.90.

For each scenario, a random-effects meta-analysis was recalculated using restricted maximum likelihood (REML) estimation and unmodified Hartung-Knapp inference. The between-study variance (τ²) was re-estimated separately for each correlation assumption. With two randomized comparisons, the Hartung-Knapp confidence interval used one degree of freedom.

The assumed correlations were hypothetical sensitivity parameters and did not represent measured patient-level correlations.

C4. Original Extraction-Based Analysis

The original extraction-based random-effects analysis produced a pooled mean difference of +8.62 FJS points favouring functional alignment (95% CI -16.90 to +34.14) using REML estimation with Hartung-Knapp inference.

The estimated heterogeneity was I² = 0% and τ² = 0.

A DerSimonian-Laird sensitivity analysis produced the same pooled mean difference (+8.62) but a narrower conventional 95% confidence interval (+4.56 to +12.68).

These estimates were conditional on the extracted variance of the paired bilateral comparison.

C5. Correlation-Based Sensitivity Analysis

Table with 5 columns: Assumed p, Paired-trial SE, REML τ², Pooled MD, Hartung–Knapp 95% CI. Rows show data for assumed p values of 0.00, 0.25, 0.50, 0.75, and 0.90.

TABLE 7: Influence of assumed within-patient correlation on the pooled early FJS effect

Positive mean differences favour functional alignment. Correlation assumptions were used to reconstruct the variance of the paired treatment contrast. REML between-study variance estimates were recalculated for each scenario. Confidence intervals were calculated using unmodified Hartung–Knapp inference. The outcome was the Forgotten Joint Score [17]. Trial data were obtained from Young et al. [8] and Chompoosang et al. [12].

C6. Interpretation

Across the five assumed correlations, pooled mean differences ranged from +8.53 to +9.32 FJS points favouring functional alignment. However, every Hartung-Knapp 95% confidence interval included zero.

The direction of the pooled point estimate was consistent, but its precision was insufficient to establish superiority of functional alignment for early FJS.

The analysis was limited by the inclusion of only two randomized comparisons and uncertainty regarding the true paired covariance. Estimates of heterogeneity were consequently unstable, and no prediction interval was calculated.

C7. Conclusion

The correlation-based sensitivity analysis demonstrated that the direction of the early FJS point estimate remained favourable to functional alignment across the examined assumptions. Nevertheless, the wide confidence intervals and uncertainty regarding paired-trial variance precluded a definitive conclusion regarding comparative effectiveness.

These results support cautious interpretation of the early FJS meta-analysis and are consistent with the very low certainty assigned to this outcome in the GRADE assessment.

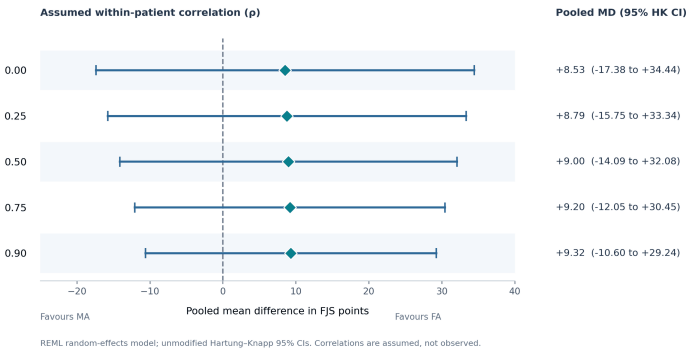


FIGURE 4: Sensitivity of the Pooled Early Forgotten Joint Score Mean Difference to Assumed Within-Patient Correlation in the Paired Bilateral Randomized Trial

Pooled mean differences and 95% Hartung–Knapp confidence intervals for early Forgotten Joint Score (FJS) under assumed within-patient correlations of 0, 0.25, 0.50, 0.75, and 0.90 in the bilateral randomized trial by Chompoosang et al. Random-effects models were fitted using restricted maximum likelihood estimation, with between-study variance recalculated for each scenario. Positive values favour functional alignment. All confidence intervals crossed zero. The correlation values represent exploratory assumptions rather than observed within-patient correlations. The outcome was the Forgotten Joint Score [17]. Data sources were Young et al. [8] and Chompoosang et al. [12].

Appendix 4: Detailed risk-of-bias assessments using the Cochrane RoB 2 tool

D1. Assessment Framework

Risk of bias was assessed using the Cochrane Risk of Bias 2 (RoB 2) tool for randomized trials. Assessments were outcome-specific and considered the following five domains:

D1: Bias arising from the randomization process.

D2: Bias due to deviations from intended interventions.

D3: Bias due to missing outcome data.

D4: Bias in measurement of the outcome.

D5: Bias in selection of the reported result.

Judgments were classified as low risk of bias, some concerns, or high risk of bias. The consensus tracker was used as the source of the final recorded judgments for the seven assessed results.

D2. Consensus Risk-of-Bias Assessments

Trial	Assessed outcome	D1	D2	D3	D4	D5	Overall
Young	FJS, 24 months	SC	High	SC	Low	SC	High
Manara	FJS, 12 months	SC	Low	Low	Low	SC	SC
RASKAL	FJS, 24 months	Low	SC	SC	Low	SC	SC
Wang	FJS-12, 6 months	SC	SC	SC	SC	SC	SC
Masilamani	VAS pain, 72 hours	SC	Low	SC	Low	SC	SC
Chompoosang	FJS, 6 months	Low	Low	SC	Low	SC	SC
Airapetov	FJS, 3 months	Low	Low	Low	Low	SC	SC

TABLE 8: Domain-level RoB 2 judgments for seven randomized trial results

Low indicates low risk of bias; SC indicates some concerns; High indicates high risk of bias. Each assessment applies to the specified outcome and follow-up period rather than to every result reported by the corresponding trial. Assessments used the Cochrane RoB 2 tool [25]. FJS, Forgotten Joint Score [17]. Trial sources: Young et al. [8], Manara et al. [9], MacDessi et al. (RASKAL) [10], VAS, visual analogue scale for pain, as reported by Masilamani et al. [11], Chompoosang et al. [12], Wang et al. [13], and Airapetov et al. [14].

D3. Supporting Justifications

Young et al. - FJS at 24 months

The consensus assessment identified high risk of bias in Domain 2 because the primary per-protocol analysis excluded participants after randomization, potentially affecting estimation of the effect of assignment to intervention. An intention-to-treat sensitivity analysis was reported. Additional concerns were recorded for randomization, missing outcome data, and selection of the reported result. Outcome measurement was judged at low risk. The overall consensus judgment was high risk of bias.

Manara et al. - FJS at 12 months

Some concerns were recorded for the randomization process because allocation concealment and selection into the randomized substudy could not be fully verified. Some concerns were also recorded regarding prespecification of the exact reported analysis. Deviations from intended interventions, missing outcome data, and outcome measurement were judged at low risk. The overall judgment was some concerns.

RASKAL - FJS at 24 months

The randomization process and outcome measurement were judged at low risk. Some concerns were recorded regarding deviations from intended interventions, incomplete 24-month outcome responses, and selection of the reported result. Verification of the exact robotic-only statistical analysis plan remains outstanding. The overall judgment was some concerns.

Wang et al. - FJS-12 at six months

Some concerns were recorded across all five domains because the available source material did not permit complete verification of randomization procedures, intervention deviations, missing data, outcome measurement, or prespecified reporting. The overall judgment was some concerns, based on limited methodological information.

Masilamani et al. - VAS pain at 72 hours

Some concerns were recorded regarding randomization and allocation concealment, missing outcome data, and selective reporting. The consensus record notes that 72 patients were enrolled and 65 were analyzed. Patient blinding was reported, and deviations from intended interventions and outcome measurement were judged at low risk. The overall judgment was some concerns.

Chompoosang et al. - FJS at six months

Randomization, deviations from intended interventions, and outcome measurement were judged at low risk. Some concerns were recorded for missing outcome data because the six-month FJS analysis included 28 observations per treatment strategy, and for selective reporting because the exact prespecification and paired analysis required further verification. The overall consensus judgment was some concerns.

Airapetov et al. - FJS at three months

The randomization process, deviations from intended interventions, missing outcome data, and outcome measurement were judged at low risk. Some concerns were recorded for selection of the reported result because prospective specification of the exact FJS analysis could not be fully verified. The overall judgment was some concerns.

D4. Summary of Consensus Assessments

Among the seven documented outcome-specific assessments, one result was judged at high risk of bias and six were judged to have some concerns. No assessed result received an overall low-risk judgment.

These findings indicate limitations in the methodological certainty of the available randomized evidence. However, the recorded judgments should be interpreted in relation to their specific outcomes and follow-up periods.

D5. Assessments Requiring Further Verification

The existing consensus tracker does not establish completed outcome-specific judgments for every result contributing to the systematic review.

Outstanding assessments include Young et al.'s six-month FJS result, intraoperative soft-tissue release outcomes, Oxford Knee Score outcomes, radiological outcomes, and safety outcomes where applicable.

The six-month FJS result from Young et al. must not automatically inherit the high-risk judgment assigned to its 24-month FJS result.

Additional source verification is also required for the RASKAL robotic-only statistical analysis plan and for methodological details that were unavailable in the Wang and Masilamani assessments.

D6. Implications for Certainty of Evidence

The documented RoB 2 judgments informed the risk-of-bias considerations in the GRADE assessment. Nevertheless, the outcome-level GRADE judgments require reconciliation with all contributing results and their relative importance.

The existing very low certainty ratings were retained provisionally, particularly for early FJS and soft-tissue release, pending completion of the remaining outcome-specific assessments.

D7. Limitations

Although consensus judgments were recorded for seven selected results, complete signaling-question responses and supporting quotations were not available for every domain. Accordingly, this appendix presents the documented consensus assessments and their recorded rationales rather than claiming that every Cochrane RoB 2 signaling question has been independently reverified.

The assessments should be updated if additional trial protocols, statistical analysis plans, or outcome-specific data become available.

Table 9. Outstanding RoB 2 Assessment: Young et al., Six-Month Forgotten Joint Score

Trial: Young et al. (2025)

Outcome: Forgotten Joint Score at six months

Effect of interest: Effect of assignment to functional versus mechanical alignment

Domain	Evidence requiring consideration	Current determination
D1: Randomization	Random sequence generation, allocation concealment, and baseline comparability	Verification required
D2: Deviations from intended interventions	Post-randomization non-receipt of assigned intervention and use of a per-protocol analysis	Important potential bias; verify intention-to-treat analysis
D3: Missing outcome data	Six-month FJS responder counts, exclusions, and reasons for missingness by randomized group	Verification required
D4: Measurement	Patient blinding, FJS administration, and consistency of assessment procedures	Verification required
D5: Selection of reported result	Trial registration, prespecified follow-up time points, and analysis methods	Verification required

TABLE 9: Outstanding RoB 2 Assessment: Young et al., Six-Month Forgotten Joint Score

The published six-month FJS means were 53.3 ± 29.2 for functional alignment and 47.6 ± 27.9 for mechanical alignment. However, the precise analysis population, missing-outcome pattern, and prespecified analysis supporting this time point require further verification. The high-risk consensus judgment for the 24-month FJS result was not automatically transferred to this separate six-month result. The assessment concerns the six-month Forgotten Joint Score [17] reported by Young et al. [8] and uses the Cochrane RoB 2 framework[25].

Assessment status: Pending full source verification and independent reviewer consensus.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Nay Aung Zin, Sai Tun, Than Htay, Min Maung Maung, Phyo Thant, Htoo Aung Naing Shwe, Brang Jar

Acquisition, analysis, or interpretation of data: Nay Aung Zin, Sai Tun, Phay Phone Paing, Thi La Si Thu, Thu Ya Aung, Phyu S. Thwin, Paing Ye Kyaw

Drafting of the manuscript: Nay Aung Zin, Min Maung Maung, Phay Phone Paing, Htoo Aung Naing Shwe, Brang Jar, Phyu S. Thwin, Paing Ye Kyaw

Critical review of the manuscript for important intellectual content: Nay Aung Zin, Sai Tun, Than Htay, Phyo Thant, Thi La Si Thu, Thu Ya Aung

Supervision: Nay Aung Zin, Than Htay

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: Payment/services info: All authors have declared that no financial support was received from any organization for the submitted work. Financial relationships: All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. Other relationships: All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors would like to express their sincere gratitude to Professor Than Htay, Emeritus Professor of Orthopaedics and Trauma, University of Medicine, Magway, Myanmar, for his valuable supervision, guidance, and academic support throughout this systematic review and meta-analysis. We also acknowledge the contributions of all members of the review team who assisted with literature screening, data extraction, risk-of-bias assessment, and methodological review. Finally, we thank Kulhudhuffushi Regional Hospital, Maldives, for supporting the academic and research activities associated with this work.

References

1. Global, regional, and national burden of osteoarthritis, 1990-2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol. 2023, 5:e508-22. 10.1016/S2665-9913(23)00163-7

2. Patiño Contreras JL, Sebastián Pérez V, Antonio CG, Jorge MS, Javier MM: Third-generation posterior-stabilised cemented system for total knee arthroplasty: survivorship and outcomes after 20-year follow-up. J Exp Orthop. 2024, 11:e12063. 10.1002/jeo2.12063

3. Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJ, Robertson BW: Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. J Arthroplasty. 2017, 32:3854-60. 10.1016/j.arth.2017.07.021

4. DeFrance MJ, Scuderi GR: Are 20% of patients actually dissatisfied following total knee arthroplasty? A systematic review of the literature. *J Arthroplasty*. 2023, 38:594-9. [10.1016/j.arth.2022.10.011](#)
5. Hiranaka T: Current concept: personalized alignment total knee arthroplasty as a contrast to classical mechanical alignment total knee arthroplasty. *Arthroplasty*. 2024, 6:23. [10.1186/s42836-024-00246-2](#)
6. Karasavvidis T, Pagan Moldenhauer CA, Lustig S, Vigdorich JM, Hirschmann MT: Definitions and consequences of current alignment techniques and phenotypes in total knee arthroplasty (TKA) - there is no winner yet. *J Exp Orthop*. 2023, 10:120. [10.1186/s40634-023-00697-7](#)
7. Alrajeb R, Zarti M, Shuia Z, Alzobi O, Ahmed G, Elmhiregh A: Robotic-assisted versus conventional total knee arthroplasty: a systematic review and meta-analysis of randomized controlled trials. *Eur J Orthop Surg Traumatol*. 2024, 34:1333-43. [10.1007/s00590-023-03798-2](#)
8. Young SW, Tay ML, Kawaguchi K, van Rooyen R, Walker ML, Farrington WJ, Bayan A: The John N. Insall Award: functional versus mechanical alignment in total knee arthroplasty: a randomized controlled trial. *J Arthroplasty*. 2025, 40:S20-S30.e2. [10.1016/j.arth.2025.02.065](#)
9. Manara JR, Steer R, Whitehouse SL, Collopy D, Clark GW: Functional alignment restores native kinematics more consistently than mechanical axis alignment in total knee arthroplasty. *Bone Joint J*. 2025, 107-B:423-31. [10.1302/0301-620X.107B4.BJJ-2024-0956.R1](#)
10. MacDessi SJ, Wernecke GC, Ghadirinejad K, et al.: Robotic-assisted surgery and functional alignment in total knee arthroplasty: the RASKAL registry-nested 2 × 2 factorial randomized trial. *Bone Joint J*. 2026, 108-B:622-33. [10.1302/0301-620X.108B.BJJ-2025-1472.R2](#)
11. Masilamani AB, Jayakumar T, Mulpur P, Gandhi V, Kikkuri RR, Reddy AV: Functional alignment is associated with increased incidence of pre-balance, reduced soft-tissue release, and post-operative pain compared to mechanical alignment in patients undergoing simultaneous bilateral robotic-assisted TKA. *J Robot Surg*. 2023, 17:2919-27. [10.1007/s11701-023-01732-6](#)
12. Chompoosang T, Ketkaewsuwan U, Ploynumpon P: Comparative effects of mechanical and functional alignment in bilateral robotic total knee arthroplasty: a randomized controlled trial. *Arthroplasty*. 2025, 7:25. [10.1186/s42836-025-00310-5](#)
13. Wang Q, Yan Z, Wang Q, Chen Y, Zhang X: The early outcome of robotic-assisted functional versus mechanical alignment in total knee arthroplasty: a prospective randomized control trial. *J Arthroplasty*. 2025, [10.1016/j.arth.2025.08.030](#)
14. Airapetov G, Prizov A, Daniliyants A, Aliev R, Abdulaye A, Iatsukova V: Functional alignment preserves the native Coronal Plane Alignment of the Knee phenotype while improving early clinical outcomes after robotic-assisted total knee arthroplasty. *Orthop Rev (Pavia)*. 2026, 18:166120. [10.52965/001c.166120](#)
15. Gao XP, Sun XD, Wang QS, et al.: Functional alignment versus mechanical alignment in total knee arthroplasty: a meta-analysis of randomized controlled trials. *BMC Musculoskelet Disord*. 2026, [10.1186/s12891-026-10353-7](#)
16. Jamaledine Y, Maroun R, Assi C, Khoury A, Kouyoumdjian P: Functional versus mechanical alignment in robotic primary total knee arthroplasty: a meta-analysis. *SICOT J*. 2026, 12:50. [10.1051/sicotj/2026061](#)
17. Behrend H, Giesinger K, Giesinger JM, Kuster MS: The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. *J Arthroplasty*. 2012, 27:430-436.e1. [10.1016/j.arth.2011.06.035](#)
18. Dawson J, Fitzpatrick R, Murray D, Carr A: Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br*. 1998, 80:63-9. [10.1302/0301-620X.80b1.7859](#)
19. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD: Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. *J Orthop Sports Phys Ther*. 1998, 28:88-96. [10.2519/jospt.1998.28.2.88](#)
20. Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE: Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res*. 2016, 474:1461-71. [10.1007/s11999-016-4719-1](#)
21. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW: Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988, 15:1833-40.
22. Insall JN, Dorr LD, Scott RD, Scott WN: Rationale of the Knee Society clinical rating system. *Clin Orthop Relat Res*. 1989, 248:13-4. [10.1097/00003086-198911000-00004](#)
23. MacDessi SJ, Griffiths-Jones W, Harris IA, Bellemans J, Chen DB: Coronal Plane Alignment of the Knee (CPAK) classification. *Bone Joint J*. 2021, 103-B:329-37. [10.1302/0301-620X.103B2.BJJ-2020-1050.R1](#)
24. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021, 372:n71. [10.1136/bmj.n71](#)
25. Sterne JA, Savović J, Page MJ, et al.: RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019, 366:l4898. [10.1136/bmj.l4898](#)
26. Hultcrantz M, Rind D, Akl EA, et al.: The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017, 87:4-13. [10.1016/j.jclinepi.2017.05.006](#)
27. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ: GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008, 336:924-6. [10.1136/bmj.39489.470347.AD](#)