Clinical outcomes associated with robotic and computer-navigated total knee arthroplasty: a machine learning-augmented systematic review

Quinlan D. Buchlak
Joe Clair
Nazanin Esmaili
Arshad Barmare
Siva Chandrasekaran

Follow this and additional works at: https://researchonline.nd.edu.au/med_article

This article was originally published as:

Original article available here:
10.1007/s00590-021-03059-0

This article is posted on ResearchOnline@ND at. For more information, please contact researchonline@nd.edu.au.
Clinical outcomes associated with robotic and computer-navigated total knee arthroplasty: A machine learning augmented systematic review
Abstract

Background: Robotic (RTKA) and computer-navigated total knee arthroplasty (CNTKA) are increasingly replacing manual techniques in orthopaedic surgery. This systematic review compared clinical outcomes associated with RTKA and CNTKA and investigated the utility of natural language processing (NLP) for literature synthesis.

Methods: A comprehensive search strategy was implemented. Results of included studies were combined and analysed. A transfer learning approach was applied to train deep NLP classifiers (BERT, RoBERTa and XLNet), with cross-validation, to partially automate the systematic review process.

Results: 52 studies were included, comprising 5,067 RTKA and 2,108 CNTKA. Complication rates were 0-22% and 0-16% and surgical time was 70-116 and 77-102 minutes for RTKA and CNTKA, respectively. Technical failures were more commonly associated with RTKA (8%) than CNTKA (2-4%). Patient satisfaction was equivalent (94%). RTKA was associated with a higher likelihood of achieving target alignment, less femoral notching, shorter operative time and shorter length of stay. NLP models demonstrated moderate performance (AUC=0.65-0.68).

Conclusions: RTKA and CNTKA appear to be associated with similarly positive clinical outcomes. Further work is required to determine whether the two techniques differ substantially with regard to specific outcome measures. NLP shows promise for facilitating the systematic review process.
Introduction

Hundreds of thousands of total knee arthroplasties (TKA) are carried out globally each year[1] to alleviate pain and disability, restoring function, wellbeing and quality of life for patients[2,3]. TKA offers patients relief from unstable, painful and deformed knees and demand for this type of surgery is rising[4]. The success of a TKA is based on restoration of angles, soft tissue balancing, accurate bony resection, appropriate fixation and correct limb alignment[5].

Various high technologies are seeing increased development, implementation and adoption in surgery to improve clinical outcomes including artificial intelligence[6–12], computer-guided navigation systems and robotic assist devices[7,8,10,13]. There are three main types of total knee arthroplasty: conventional or manual (mTKA), computer navigated (CNTKA) and robotic (RTKA)[5]. The main advantage of robotic and computer navigation systems is real time intra-operative assessment of alignment. In mTKA, there is the potential to introduce small degrees of error at each step of the procedure and these minor errors may accumulate over multiple steps, leading to suboptimal results[14]. CNTKA and RTKA have been developed to improve clinical outcomes.

CNTKA involves the use of arrays attached to the bone, an infrared tracking device and a display system, which facilitate the achievement of accurate alignment[5]. CNTKA was designed to reduce complications and may be image-free or image-based. Image-based systems use preoperative radiology, while image-free systems involve collecting and registering information intra-operatively[5]. CNTKA has been associated with lower blood loss[15–18], lower revision rates[19], and less systemic emboli[20–22] than mTKA. It is associated with a higher likelihood of satisfactory alignment in the postoperative period and at five[5,23,24] and eight[25] years follow-up. CNTKA enables surgeons to conduct surgery that is more accurate and precise than the manual technique.
Robotic TKA systems appear to deliver improved alignment compared to other techniques, further decreasing variability and increasing precision[26]. Various RTKA systems have been developed[26,27]. They help to protect soft tissues and have been associated with more accurate bone resection than conventional instrumentation[28]. RTKA enables consistent and accurate implant positioning[29] and has been associated with less blood loss than manual procedures[30,31]. Research has suggested that RTKAs require more time than manual procedures. However, recent evidence suggests that, once surgeons operate through a learning curve of 11-43 cases[32], RTKA is not associated with increased operative times[33]. CNTKA involves the use of systems that provide live on-screen information on knee anatomy and kinematics intraoperatively[34]. A robotic system may provide this information, but it is also capable of guiding and constraining the surgeon’s procedural behaviours. Robotic systems help to accurately execute a patient-specific operative plan[35,36]. Robotic systems may be fully-active or semi-active, depending on the degree of control provided to the operating surgeon[34]. Robotic systems are costly. However, economic analyses suggest that RTKA is associated with lower 30, 60 and 90-day episode of care costs than mTKA, which is driven by a lower risk of readmission and less costly discharge destinations[37,38].

CNTKA appears to be more prevalent than RTKA, comprising 4.5% and 0.4%, respectively, of 6,060,901 TKA procedures carried out in the US between 2005 and 2014[13]. Much research has compared mTKA with RTKA or CNTKA. However, it is unclear how RTKA and CNTKA compare with regard to clinical outcomes.

Machine learning is a branch of artificial intelligence that enables effective prediction and classification[8,9]. Natural language processing (NLP) is based on machine learning and linguistics. It involves converting text information into a quantitative form that can be readily interpreted by computers, enabling a range of useful functions including document
classification, data extraction, topic modelling and translation. NLP technology has reached a level of maturity that enables it to be used to facilitate the systematic review process[8,10].

This systematic review was conducted to synthesize and compare clinical outcomes associated with RTKA and CNTKA. It was guided by the following two research questions:

(1) How do RTKA and CNTKA compare with regard to clinical outcomes? (2) Can NLP be applied to facilitate the systematic review process?
Method

Search strategy

Our method was informed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[39]. A comprehensive and systematic search strategy was developed [40] to identify studies that evaluated outcomes associated with CNTKA and RTKA. The search was conducted using the following databases, in accordance with recommended practice[41]: PubMed, ScienceDirect, Embase, Ovid, Ebsco, Google Scholar and Scopus. The search was guided by the PICO framework (see Supplementary Materials) and included the following terms: “(Knee AND (Arthroplasty OR Replacement)) AND (Robot* OR Computer OR Navigat*)”. Reference lists of included articles were searched. The search was conducted in November 2020. Risk of bias for RCTs was assessed using the Cochrane risk of bias tool (RoB 2)[42]. Risk of bias for non-randomised studies was assessed using the Risk Of Bias In Non-randomised Studies - of Interventions (ROBINS-I) tool[43]. These tools include structured lists of signalling questions to facilitate risk of bias analysis.

Inclusion and exclusion criteria

Studies selected for inclusion involved adult patients who underwent CNTKA or RTKA for any indication (e.g., osteoarthritis, trauma, rheumatoid arthritis, etc.) and were published in English. RCTs and cohort studies that reported clinical outcomes were included. Because technology changes quickly, only studies published within the last 10 years were included. Studies were excluded if they investigated unicompartmental knee arthroplasty or involved other techniques in addition to TKA. Abstracts, reviews and correspondence articles were excluded. Studies were assessed for inclusion by one researcher (QDB) and verified by another (JC). Disagreements were resolved by discussion.
Data collection and extraction

Data extracted from each selected study included: abstract, reference, authors, country, primary institution, study design, number of patients, blinding, randomization, mean follow-up time, type of RTKA or CNTKA system and clinical outcomes.

Patient reported outcome measures

Many patient reported outcome measures were used to assess clinical outcomes including the Hospital for Special Surgery (HSS) score, the Knee Society Score (KSS), the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) score, the Knee Injury and Osteoarthritis Outcome Score (KOOS), the Short Form Health Survey (SF-36 and SF-12), the EuroQol score (EQ-5D), the Oxford Knee Score (OKS), the Forgotten Joint Score (FJS), the University of Los Angeles (UCLA) activity score and the Veterans RAND 12-item (VR-12) health survey. The HSS has a total of 100 points and includes seven categories: pain, function, strength, range of motion (ROM), flexion deformity, instability and subtraction[44]. The KSS[45] is divided into a knee score and a function score. For each of these two scores, 100 points indicates full function. The knee score is based on stability, alignment, pain and range of motion, while the function score is based on activities of daily living. The KSS has a total of 200 points. The WOMAC score is divided into three domains: pain, stiffness and function and has a total score of 96 points[46]. The WOMAC comprises a subset of the questions on the KOOS, which has a total of 100 points and is divided into five domains: pain, symptoms, function in daily living, function in sport and recreation and quality of life[47]. The SF-36 has eight subdomains: function, pain, role limitations due to physical problems, role limitations due to emotional problems, emotional well-being, social function, fatigue and perceptions of general health[48]. The EQ-5D questionnaire[49,50] measures mobility, self-care, daily activities, pain or discomfort, anxiety and depression and contains a visual analogue scale
(VAS). It provides an indication of a patient’s self-perceived health state. The OKS[51] measures pain and function after TKR. The FJS measures a patient’s awareness of a joint such that less awareness suggests successful treatment[52]. The UCLA activity score measures activity and function on a 10-point scale[53]. The VR-12 is a short measurement tool that assesses a patient’s health perceptions, physical function, role limitations, pain, energy level, social function and mental health[54]. These measures are based on clinical history and examination information and for most, the higher the score, the better the outcome[44]. Higher WOMAC and OKS scores indicate poorer clinical outcomes[55]. The minimum clinically important difference for the total WOMAC and OKS scores is 10[56] and 5[57] respectively.

Analysis

Quantitative and qualitative analyses were conducted. Data extracted from each article was used to calculate descriptive statistics and compare surgical techniques. Abstracts were used for an NLP analysis that consisted of two phases: (1) keyword identification, and (2) training document classifiers to automate article screening. Text was tokenized and converted to lower case, numeric characters and English stop words were removed and remaining tokens were lemmatized[8,10]. Multiple document classifiers were trained. The classification outcome was article inclusion in this systematic review after coding. A transfer learning approach for text classification was applied using the BERT[58], RoBERTa[59] and XLNet[60] pre-trained deep language models. Model classification performance was assessed using 3-fold cross-validation, with accuracy, area under the receiver operating characteristic curve (AUC), precision, recall, F1 and Matthew’s correlation coefficient (MCC) metrics. The MCC ranges from negative-one to one and has been recommended as the preferred metric for binary classifiers[61]. All other metrics range between zero and one. Higher scores indicate higher performance. Analyses were
conducted using custom Python scripts and the SciPy[62], Scikit-learn[63], nltk[64], gensim[65,66], Keras[67], transformers[68] and simpletransformers[69] packages.
Results

The search resulted in the retrieval of 628 relevant records. We assessed 88 full-text articles and 52 were included for analysis (Figure 1). Study characteristics are displayed in Table 1. The studies included encompassed 5,067 RTKA and 2,108 CNTKA procedures. The most prevalent primary institutions (affiliated with the first author) in this sample of studies were Singapore General Hospital (n=5), Ortho Rhode Island (n=4) and the University of Louisville (n=3). The number of studies investigating outcomes associated with RTKA has increased markedly over the past five years, in stark contrast to the declining number of studies investigating outcomes associated with CNTKA (Figure 2).

Most studies did not directly compare RTKA and CNTKA but compared one of these techniques with mTKA. The results of these studies are synthesized in Table 2. This synthesis of the literature facilitates a comparison of the two techniques and clarifies research gaps. Surgical time ranged between 70-116 minutes for RTKA and 77-102 minutes for CNTKA. Absolute length of stay (LOS) data was lacking for CNTKA. LOS for RTKA ranged between 1.2-5.2 days. Overall complication rates for RTKA and CNTKA ranged between 0-22% and 0-16%, respectively. Anterior femoral notching appeared to be more commonly associated with CNTKA (5%) than RTKA (0%). Technical failures appeared to be more commonly associated with RTKA (8%) than CNTKA (2-4%). Aborted procedures reached 10% in some RTKA studies. Patient satisfaction associated with each technique appeared to be equivalent (94%). Each technique appeared to be associated with satisfactory PRO results extending out to 13 years. Implant survival was 98.8% at 10 years for RTKA and 95.7-100% at 8-12 years for CNTKA. The most commonly applied PRO measure was the KSS.
Clark and Schmidt (2013)[74] directly compared outcomes associated with RTKA with CNTKA. Their retrospective study involved 52 RTKA and 29 CNTKA procedures. They found that RTKA procedures were associated with a shorter mean navigation time (9 minutes), less intraoperative malalignment, and a shorter mean length of stay (0.6 days less) than CNTKA procedures.

**Natural language processing analysis**

Article abstracts were used to identify keywords (Table 3) across the corpus of included articles and train deep learning models to classify included and excluded articles. The full dataset consisted of 456 article abstracts; the 52 abstracts from included articles and 404 abstracts from articles that were screened out. Document classifiers demonstrated moderate performance: XLNet AUC=0.65, BERT AUC=0.68, RoBERTa AUC=0.67 (Figure 3).
Discussion

This study was designed to synthesize the literature investigating clinical outcomes associated with RTKA and CNTKA. Overall, it appeared that each technique was associated with similarly satisfactory outcomes, although RTKA may be associated with shorter operative time, a higher likelihood of achieving target alignment and a shorter LOS than CNTKA. It appears that mean operating times are decreasing as experience with using each of these technologies grows. Clinical outcome measures demonstrated strong results at many postoperative time points for both RTKA and CNTKA. It appears that most PRO measures improve over the first 2-3 years after surgery and then plateau.

The two system types investigated appear to have their own advantages and disadvantages. Advantages of RTKA systems include accurate placement of hardware, increased alignment accuracy, improved procedural reproducibility and consistency, a lower likelihood of femoral notching, and less blood loss[116–119]. Some RTKA systems incorporate haptic feedback to prevent deviation from the surgical plan (e.g., excessive bone resection)[26]. There is some evidence to suggest that robotic systems reduce the learning curve for some orthopaedic procedures and that the cases that comprise the learning curve represent no increased risk to the patient[120]. CNTKA and RTKA systems can be of benefit to orthopaedic surgeons, particularly those who are exposed to low case volumes, early in their careers, or learning the procedure. Disadvantages of robotic systems include their cost, the difficulty of modifying the surgical plan intraoperatively, the potential for failure and the additional time required for planning, registration and milling[26]. Both system types may encounter technical problems and fail, necessitating conversion to mTKA, although this is rare.

Robotic systems are costly and cost effectiveness is a salient issue for practitioners and health systems. Research suggests that robotic knee surgery is more cost effective than manual surgery if the number of annual cases exceeds 94[121,122]. RTKA appears to be associated
with lower 90-day episode of care costs than mTKA[107]. Some have suggested that the capital
cost of a robotic system may be regained within two years[123]. More cost effectiveness
studies, in addition to those focusing on clinical outcomes, would be useful to guide investment
decision making within health systems.

NLP demonstrated utility for facilitating literature synthesis. Trained document classifiers
were capable of differentiating between included and excluded articles with a moderate level
of performance. These models have the potential to reduce the time required for article
screening for future reviews on this topic. Deep language models and transfer learning may be
used to partially automate the systematic review process, potentially facilitating clinical
improvement and decision making.

Limitations and future research

The synthesis presented offers a succinct overview of the literature and highlights gaps. The
most notable gap is the lack of clinical studies designed to directly compare RTKA with
CNTKA. Only one study of this nature met inclusion criteria. To facilitate a more robust
comparison between CNTKA and RTKA outcomes, additional high quality RCTs are required.
Other identified gaps included: the types of outcome measures applied, follow-up time points,
comparing analgesic use and absolute LOS data for CNTKA. Some studies reported
differences between treatment groups, but not absolute means.

Robotic systems may vary substantially in their design. They include hand-held devices,
cutting guides, miniature bone-mounted systems, and larger robotic arms[26,27,124]. This
variability may limit the validity of data aggregation and comparison between studies. The
need for clearer definitions and differentiation between the terms “robot”, “robotic tools” and
“smart instruments” has been suggested by Jacofsky et al.[26]. These categorizations would
likely facilitate more accurate comparison.
Future research directly comparing clinical outcomes associated with RTKA and CNTKA over time is required. Researchers may consider the possibility of amalgamating augmented reality and robotic technologies to further improve the quality and safety of the TKA[125]. Given the potential for robotic systems to offer greater benefits for less experienced surgeons, future studies may consider controlling for surgeon experience. Future research may explore techniques to improve NLP document classification performance using small datasets and further leverage the capabilities of NLP technologies for literature synthesis.

Conclusion

RTKA and CNTKA appear to be associated with robust clinical outcomes and distinct strengths. There is substantial potential for future researchers to further evaluate outcomes associated with these two techniques. Applied NLP may develop into a powerful literature synthesis toolkit.

Declarations

Funding: This research project received no funding.

Conflicts of interest/Competing interests: The authors have no relevant competing interests to disclose.

Availability of data and material: This research was based on publicly available published literature.

Code availability: The analysis was based on custom code, which may be made available upon request.

Authors’ contributions: All authors were involved in the conception of the work. QB managed the project. QB, JC and NE collected data. QB conducted data analysis with verification by
NE. All authors were involved in analysis interpretation. QB, JC and NE drafted the manuscript. All authors were involved in the manuscript revision and editing process. All authors approved this work for submission.
References


Orthop Relat Res 2011;469:3436–42.


Jones E, Oliphant T, Peterson P. {SciPy}: Open source scientific tools for {Python} 2014.


Řehůřek R. Scalability of Semantic Analysis in Natural Language Processing 2011.


Cotter EJ, Wang J, Illgen RL. Comparative cost analysis of robotic-assisted and jig-


[Gharaibeh MA, Solayar GN, Harris IA, Chen DB, MacDessi SJ. Accelerometer-based, portable navigation (KneeAlign) vs conventional instrumentation for total knee arthroplasty: a prospective randomized comparative trial. J Arthroplasty 2017;32:777–82.]


[Jeon S-W, Kim K-I, Song SJ. Robot-assisted total knee arthroplasty does not improve...


Kinney MC, Cidambi KR, Severs DL, Gonzales FB. Comparison of the iAssist...


Zhu M, Chen JY, Chong HC, Pang HN, Tay DKJ, Chia S-L, et al. No difference in functional outcomes after total knee arthroplasty with or without pinless navigation. J


Coon TM. Integrating robotic technology into the operating room. Am J Orthop (Belle Mead NJ) 2009;38:7–9.


Figure legends

Figure 1: Study screening and selection flow diagram, presenting the number of studies considered, excluded and included in this study.

Figure 2: Number of studies per year investigating clinical outcomes associated with RTKA and CNTKA. One study in 2013 investigated both RTKA and CNTKA. The number of studies investigating clinical outcomes associated with RTKA appears to be increasing over time.

Figure 3: Performance of document classifiers trained using 3-fold cross-validation. XLNet, BERT and RoBERTa are deep natural language processing models that were trained to classify documents for inclusion in this systematic review after the completion of the manual labelling process. Multiple metrics were calculated. Trained models demonstrated moderate classification performance.