Amsterdam Movement Sciences

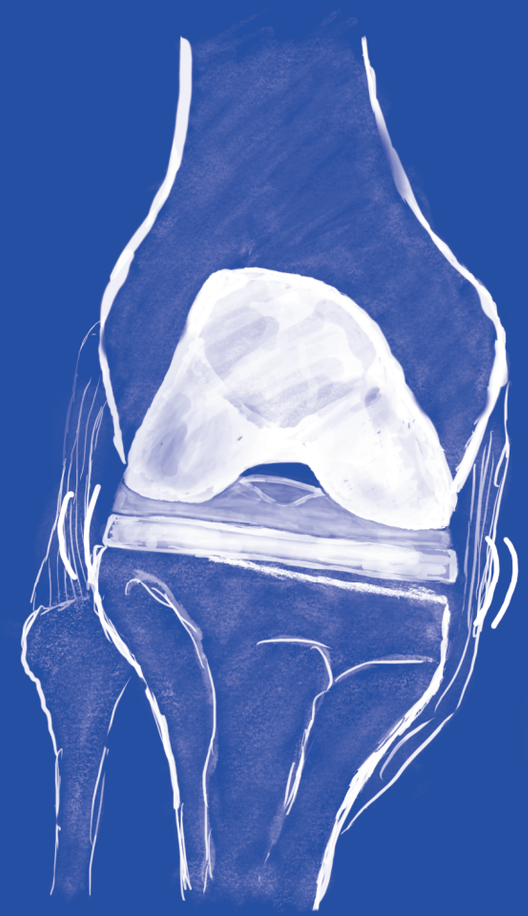
Amsterdam Movement Sciences conducts scientific research to optimize physical performance in health and disease based on a fundameral understanding of human movement in order to contribute to the fulfillment of a meaningful life.

Knee Arthroplasty Loosening

George Samuel Buijs

Knee Arthroplasty Loosening

Movement matters



George Samuel Buijs

Knee Arthroplasty Loosening

Movement matters

George Samuel Buijs

Cover and illustrations:	Laura Buijs, George Buijs
Layout:	George Buijs
Printed by:	Ridderprint, www.ridderprint.nl
ISBN:	978-94-6506-424-6

The research in this thesis was embedded in Amsterdam Movement Sciences Research Institute, at the department of Orthopedic Surgery and Sports Medicine, Amsterdam UMC, Location University of Amsterdam, the Netherlands (use as appropriate).

The research in this thesis was financially supported by an internal TKI-PPP grant from the Amsterdam University Medical Center. Allowance made available by Health-Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

The printing of this thesis was kindly supported by: Department of Orthopedic Surgery and Sports Medicine of the Amsterdam UMC, Amsterdam UMC Research B.V., Amsterdam Movement Sciences, Nederlandse Orthopaedische Vereniging, AtMoves B.V., Vakgroep Orthopedische Chirurgie NWZ, ChipSoft, Huisartsenpraktijk M.Buijs-Wilts, ABN AMRO, Bauerfeind, ImplantCast, Interactive Studios, LeukOrthopedie & Probrace, Mathys, Oudshoorn B.V, Nea International B.V. and Stichting Natuurherstel Peel & Maas.

Knee Arthroplasty Loosening
Movement matters

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 27 mei 2025, te 16.00 uur

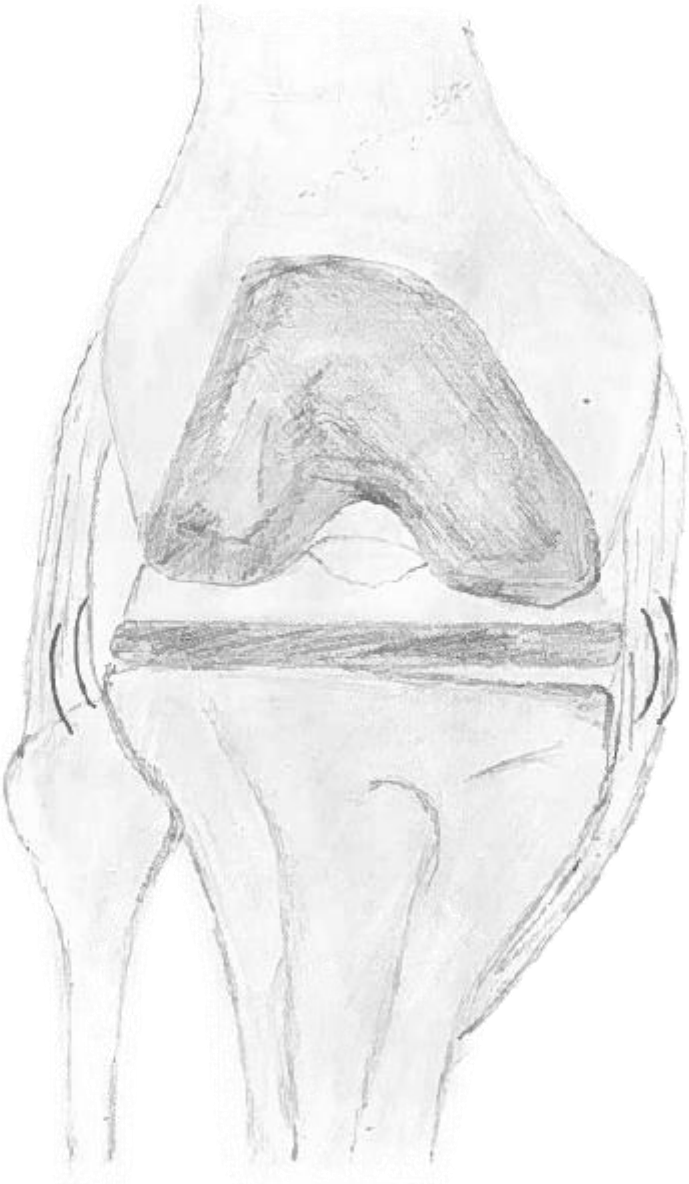
door George Samuel Buijs
geboren te Memphis

Promotiecommissie

Promotor:	dr. ir. L. Blankevoort	AMC-UvA
Copromotor:	dr. M.U. Schafroth	AMC-UvA
Overige leden:	prof. dr. S. Repping	AMC-UvA
	prof. dr. ir. N.J.J. Verdonschot	Universiteit Twente
	dr. B.C.H. van der Wal	Universiteit Utrecht
	prof. dr. J.C. Goslings	AMC-UvA
	prof. dr. D. Eygendaal	AMC-UvA
	prof. dr. M.G.W. Dijkgraaf	AMC-UvA

Faculteit der Geneeskunde

Table of contents



Chapter I

General introduction

Introduction

Osteoarthritis is the most common reason for knee pain and reduced mobility. In 2020, it affected up to 368 million people worldwide [54]. The prevalence of knee osteoarthritis is known to be rising [54]. This rise is primarily attributed to factors such as the aging population and rising rates of obesity [32]. This increasing prevalence of osteoarthritis has led and will lead to an increasing demand for effective treatment options [32]. Knee arthroplasty procedures have become vital and widely used interventions for relieving pain and restoring mobility in patients with end-stage osteoarthritis.

Knee arthroplasty entails the substitution of damaged knee cartilage with prosthetic components, typically including tibial, femoral, and, when indicated, patellar components, to alleviate pain and restore joint function. Knee arthroplasty components are usually made of a titanium- or cobalt chromium-based alloy, with a highly cross-linked polyethylene insert to reduce friction and wear. Knee arthroplasty can be performed using two primary fixation methods: cemented or uncemented. In cemented arthroplasties, the components are secured to the bone using bone cement, also known as polymethyl methacrylate (PMMA). PMMA acts as a filler between the bone and implant, ensuring immediate stability and effective load distribution. Conversely, in uncemented arthroplasties, the components are designed with a porous surface with a hydroxyapatite coating to facilitate osteointegration, where the bone naturally integrates with the prosthesis over time.

After knee arthroplasty, which is generally successful, the likelihood of revision within 10 years after primary implantation is 13% [19]. Revision of a knee arthroplasty is a surgical procedure to replace a failed or problematic knee implant from a primary knee arthroplasty surgery. With increasing primary knee arthroplasty numbers, the number of consequent revisions is also expected to rise. Revision surgery is more complex, costly and carries a higher risk of re-revisions than primary surgery and should therefore be performed only when justified and unavoidable [28, 45].

Primary causes for revision of knee arthroplasties are aseptic loosening, periprosthetic infection and mechanical instability [19]. Aseptic loosening is the debonding of the implant at the implant-bone, implant-cement and/or cement-bone interface, and can occur early, within the first two years due to factors like poor surgical technique or failed osteointegration, or late, after several years, due to long-term mechanical wear or osteolysis from particulate debris. Periprosthetic infection is the results of an inflammatory response to bacterial invasion of the implant site [5]. In some cases, the knee arthroplasty may suffer a persistent but less aggressive low-grade infection. Due to the similarity in the subtle and chronic nature of symptoms for aseptic loosening and low-grade infection, low grade infection can sometimes be misdiagnosed as aseptic loosening [52]. Low-grade infection can also result in implant loosening, as the chronic inflammation and subtle bacterial activity can lead to bone resorption and weakening [52]. Mechanical instability stems from insufficient ligamentous support or suboptimal alignment of knee arthroplasty components, resulting in abnormal knee kinematics and consequent discomfort or pain.

With 20-30%, aseptic loosening is the most frequent cause for revision knee arthroplasty surgery [19]. Notably, tibial component loosening accounts for the majority of indications for revision surgery, representing 20.4%, compared to 8.4% for femoral component loosening [19].

Patients suspected of aseptic loosening typically present with knee pain. After taking the patient's symptoms and history, a physical examination is conducted, followed by anteroposterior and lateral radiographs. In cases of loosening, radiographs may reveal radiolucent lines at the implant interfaces. Radiolucent lines are areas around arthroplasty components that appear darker (less radiopaque) due to lower bone density (osteolysis). Although lower bone density around the implant might suggest loosening, radiolucency is known to be nonspecific and may fail to indicate early loosening [45, 59, 60].

Efforts to enhance the diagnosis of aseptic loosening have spurred extensive research into various expensive and often invasive diagnostic modalities [4, 5, 21, 48, 59]. These include positron emission computed tomography (SPECT) combined with computed tomography (CT), which detects metabolic activity indicative of bone remodeling or inflammation; SPECT/CT which highlights areas of increased bone turnover; and bone scintigraphy, which uses radioactive tracers to detect areas of increased bone activity, helping to identify loosening and infection. However, these modalities often show only indirect signs of loosening rather than direct evidence. Diagnostic test accuracy studies evaluating these modalities yield a wide range of sensitivity and specificity results, which are frequently conflicting [4, 6, 59]. This variability, coupled with the complex presentation of a painful knee post-arthroplasty and the uncertain outcomes of radiographs, contributes to inconsistencies in how the results of different diagnostic tools are evaluated and interpreted. These inconsistencies and the lack of sensitivity and specificity to correctly diagnose aseptic loosening often results in misdiagnosed patients [6, 59]. Accurate diagnosis is essential to avoid unnecessary revision surgeries in patients incorrectly diagnosed with knee arthroplasty loosening and to provide timely treatment for those with undetected loosening.

Given that these diagnostic modalities primarily detect indirect signs, directly measuring implant micro movement could provide a more accurate assessment. Measurement of micro motion is employed when evaluating new implant designs, using model- or marker-based radio stereometric analysis (RSA). Marker-based RSA is a precise technique involving the surgical insertion of tantalum markers into the bone and prosthetic components, allowing for accurate measurement of micro motion using dual angle radiographs. Model-based RSA, which does not require marker implantation, uses preoperative CT or magnetic resonance imaging to create 3-dimensional models of implants and bones which are then superimposed onto radiographs images to track implant migration. While both methods provide very precise (0.1 mm) and insightful information, their use in clinical practice is limited due to the need for specialized equipment, the invasive nature of marker-based RSA and the complex and time-consuming analysis required for model-based RSA [25, 57].

Given these limitations, this thesis proposes a new diagnostic modality that evaluated load-induced implant movement to provide a more practical and direct method for evaluating knee arthroplasty loosening. This new method quantifies and visualizes induced implant movement. A patented loading device applies a consecutive valgus and varus moment over the knee joint during a CT scan, creating a bending moment over the knee joint with the knee in 20 degrees of flexion. This induces compressive forces in the knee compartments, causing the tibial component to displace relative to the bone. Custom semi-automated image analysis software then quantifies and visualizes this displacement by comparing 3-dimensional images of the valgus loading condition to the varus loading condition. The software calculates clinically relevant parameters such as rotation about the screw-axis, the average point displacement of all points in the implant (mean target registration error (mTRE)), and the maximum point displacement observed across the implant (maximum total point motion (MTPM)). Additionally, MTPM is visually represented by a heat map, showing color variations indicating the magnitude of displacement.

Aims and outlines of this thesis

Current state of diagnostics

The diagnostic accuracy of available and scientifically evaluated diagnostic modalities used to aid the diagnosis of aseptic knee arthroplasty loosening is evaluated in **Chapter II**. This chapter describes a comparative diagnostic test review and meta-analysis. Since bone scintigraphy was the most used diagnostic modality for aseptic loosening in our clinic, we evaluated its diagnostic performance in current practice. This evaluation is detailed in **Chapter III**. Both the results of Chapters II and III, along with expert opinion, suggest high variability and a lack of consensus regarding clinical and radiological criteria, as well as the use of intra-operative findings (a loose or fixed prosthesis during revision surgery) as a reference test. Therefore, a Delphi consensus study was conducted to assess both the clinical and radiological criteria associated with the diagnosis of aseptic loosening and the variability and consensus for when the different components of a knee arthroplasty are considered loose. The results of the clinical and radiological criteria associated with aseptic loosening are outlined in **Chapter IV**, while **Chapter V** describes the results of the Delphi study aimed at evaluating and improving intra-operative findings when used as a reference test.

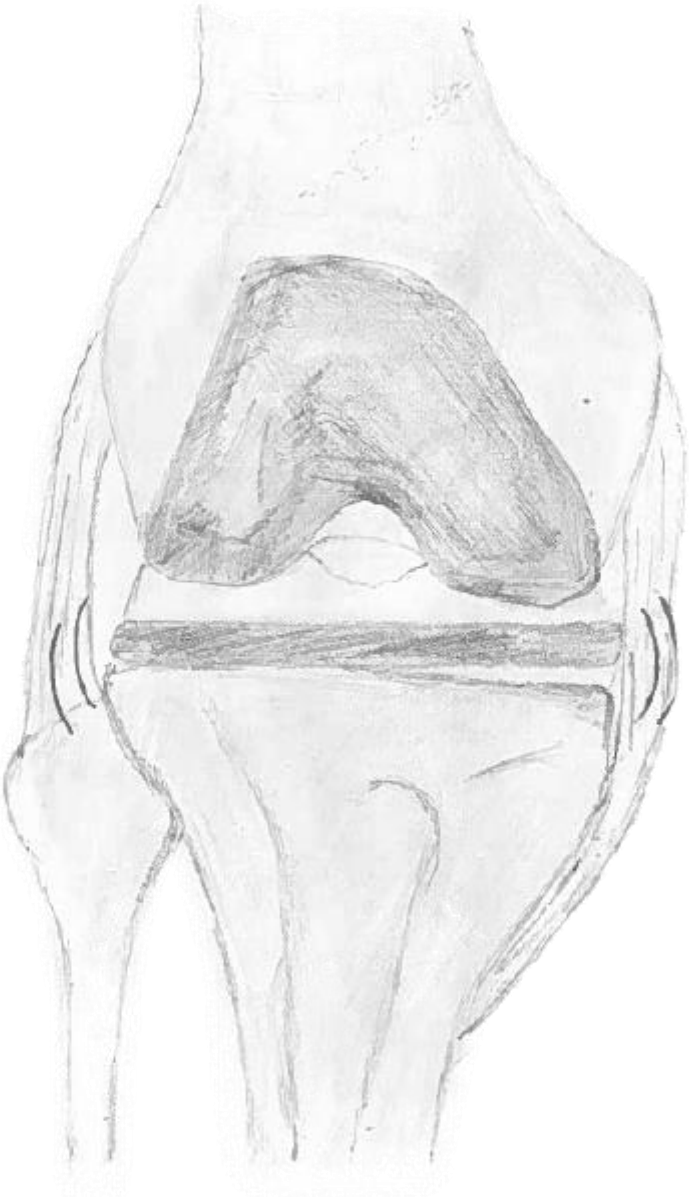
Evaluating induced implant movement

In this thesis, as described previously, a new modality to assist in the diagnosis of aseptic loosening of knee arthroplasties is proposed and evaluated. The outcomes of this method may be subject to noise introduced by the computed tomography scanner or by the semi-automated segmentation and registration. Therefore, the reproducibility, also known as precision, and diagnostic reliability of this method are evaluated in **Chapter VI**. A test-retest study was conducted in patients to assess the extent of imprecision in the use of the loading device. This study is described in **Chapter VII**. The safety, feasibility, and reliability of the semi-automated approach used, along with its diagnostic accuracy, were evaluated in a clinical patient study and are described in **Chapter VIII**.

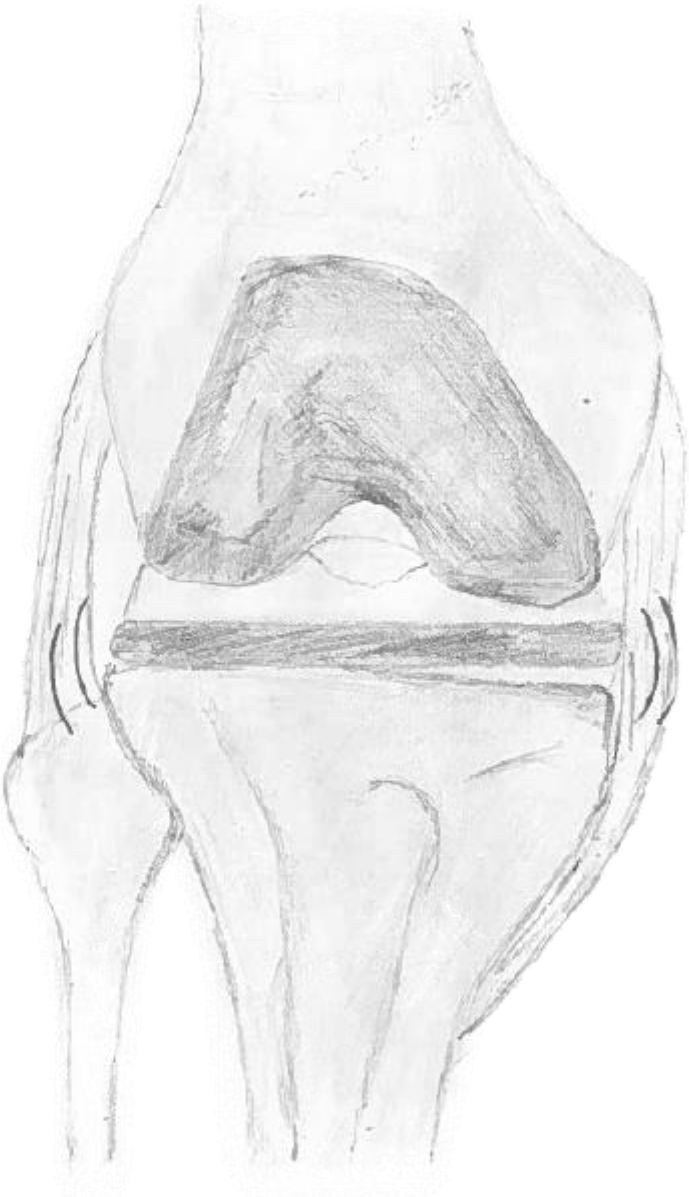
The most important findings of this thesis and future prospects are discussed in **Chapter IX.**

References

1. *Anzola LK, Hernandez N, Rodriguez LF, Sanguino G, Martinez E, Lopez R, et al. (2023) The role of SPECT/CT in painful, noninfected knees after knee arthroplasty: a systematic review and meta-analysis—a diagnostic test accuracy review. J Orthop Surg Res. 18:223*
2. *Awengen R, Rasch H, Amsler F, Hirschmann MT (2016) Symptomatic versus asymptomatic knees after bilateral total knee arthroplasty: what is the difference in SPECT/CT? Eur J Nucl Med Mol Imaging 43:762-772*
3. *Barnsley L, Barnsley L (2019) Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiol 48:1565-1572*
4. *Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284*
5. *Higuera C, Parvizi J. 18 Causes and Diagnosis of Aseptic Loosening After Total Knee Replacement. In: Hirschmann MT, Becker R, eds. The Unhappy Total Knee Replacement: A Comprehensive Review and Management Guide;10.1007/978-3-319-08099-4_23. Cham: Springer International Publishing; 2015:225-237.*
6. *Hochman MG, Melenevsky YV, Metter DF, Roberts CC, Bencardino JT, Cassidy RC, et al. (2017) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty. Journal of the American College of Radiology 14:S421-S448*
7. *Kärrholm J (1989) Roentgen stereophotogrammetry. Review of orthopedic applications. Acta Orthop Scand 60:491-503*
8. *Klug A, Gramlich Y, Rudert M, Drees P, Hoffmann R, Weißenberger M, et al. (2021) The projected volume of primary and revision total knee arthroplasty will place an immense burden on future health care systems over the next 30 years. Knee Surg. Sports Traumatol. 29:3287-3298*
9. *Long H, Liu Q, Yin H, Wang K, Diao N, Zhang Y, et al. (2022) Prevalence Trends of Site-Specific Osteoarthritis From 1990 to 2019: Findings From the Global Burden of Disease Study 2019. Arthritis Rheumatol 74:1172-1183*
10. *O'Donovan P, McAleese T, Harty J (2023) Does lucency equate to revision? A five-year retrospective review of Attune and Triathlon total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 31:4773-4781*
11. *Röhner E, Heinecke M, Matziolis G (2021) Diagnostic algorithm in aseptic TKA failure - What is evidence-based? J Orthop 24:248-253*
12. *Schiffner E, Latz D, Karbowski A, Grassmann JP, Thelen S, Windolf J, et al. (2020) Loosening of total knee arthroplasty - always aseptic? J Clin Orthop Trauma 11:S234-s238*
13. *Steinmetz JD, Culbreth GT, Haile LM, Rafferty Q, Lo J, Fukutaki KG, et al. (2023) Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet Rheumatology 5:e508-e522*
14. *Valstar ER, Vrooman HA, Toksvig-Larsen S, Ryd L, Nelissen RG (2000) Digital automated RSA compared to manually operated RSA. J Biomech 33:1593-1599*
15. *Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454*
16. *Wautier D, Thienpont E (2024) Appearance and evolution of radiolucent lines below the tibial implant in primary total knee arthroplasty. Archives of Orthopaedic and Trauma Surgery 144:1333-1344*



Part I: Current state of diagnostics.



Chapter II

MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening: A Systematic Comparative Diagnostic Test Review and Meta-analysis.

George S. Buijs, Anne C. Kooijenga, Quinten G.H. Rikken, Matthias U. Schafroth, Arthur J. Kievit, Leendert Blankevoort

The content of this chapter is a revised version of the following article:
Knee Surg Sports Traumatol Arthrosc. 2024 Aug;32(8):2061-2074. doi: 10.1002/ksa.12206

Abstract

Purpose: The purpose of this study was to evaluate and compare the diagnostic accuracy of modalities used to aid the diagnosis of aseptic knee arthroplasty loosening.

Methods: A comparative diagnostic test accuracy systematic review and meta-analysis was conducted following the Cochrane and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. PubMed, EMBASE, and Cochrane databases were searched for original articles evaluating diagnostic modalities up to March 2024. Included studies compared the modality (index test) to the intraoperative finding as reference test. The QUADAS-C tool was used to assess the quality of the included studies. The GRADE approach was used to evaluate the certainty of evidence. Level of evidence was evaluated using the Oxford Levels of Evidence tool. The primary outcome is the summary of diagnostic accuracy metrics for each modality as demonstrated by a summary receiver-operating characteristic (SROC) curve.

Results: The search yielded 467 articles. Of these, 14 articles were included. These 14 articles evaluated a total of 5 different diagnostic modalities: bone scintigraphy (3 studies, 146 cases), 18-FDG-PET-CT (2 studies, 50 cases), SPECT/CT (7 studies, 371 cases), radionuclide arthrogram (3 studies, 196 cases), and MRI (1 study, 116 cases). Nine studies exhibited a high risk of bias in patient selection and all studies showed a high risk of bias related to the reference test. The majority of the included studies were classified as Level III evidence, leading to an overall low level of certainty in the evidence. The most accurate tests, as demonstrated by the SROC analysis, were MRI and SPECT/CT, with sensitivities ranging from 0.31-0.81 and 0.43-1.00 and specificities between 0.98-1.00 and 0.63-1.00, respectively.

Conclusions: This review and meta-analysis evaluated available diagnostic modalities to aid the diagnosis of knee arthroplasty loosening and based on a low certainty of evidence suggests that MRI and SPECT/CT are currently the most accurate modalities available to aid the diagnosis aseptic loosening of knee arthroplasty components.

Introduction

About 13% of primary Total Knee Arthroplasties (TKAs) need revision within a decade, often due to aseptic loosening [8, 12]. As TKA rates increase, so do revision TKA (rTKA) surgeries, which are more complex, costly, and carry a higher risk of re-revisions than initial surgeries [19, 37]. This trend stresses the importance of accurately diagnosing aseptic loosening to manage healthcare resources and patient care effectively.

Standard diagnostic pathway involves patient history, physical exams, x-rays, but x-rays are unspecific and may miss early loosening [27, 38]. As a result, various techniques have been suggested to improve the diagnosis of loosening.

Proposed modalities include Computed Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound, and nuclear imaging modalities such as nuclide arthrogram, White Blood Cell (WBC) scanning, bone scintigraphy, and Fluorodeoxyglucose Positron Emission Tomography (18F-FDG-PET) [38].

A systematic review by Barnsley et al. evaluated nuclear imaging modalities and found SPECT/CT most accurate for aseptic TKA diagnosis [4]. A systematic review by Anzola et al. recently evaluated the role of SPECT/CT in painful non-infected knees. Despite variations across studies, their review concluded that SPECT/CT exhibits high sensitivity and specificity in accurately diagnosing various knee conditions, including loosening [3].

No recent high-quality systematic review examines the accuracy of all diagnostic modalities for diagnosing aseptic loosening of knee arthroplasties, both nuclear and non-nuclear. Hence, this review aims to summarize and compare the literature on diagnostic accuracy of all modalities for assessing aseptic knee arthroplasty loosening. Drawing from previous reviews, it is hypothesized that SPECT/CT yields the highest diagnostic accuracy.

Methods

This systematic review and meta-analysis of comparative diagnostic tests was conducted according to the recommendations of the Cochrane Diagnostic Test Accuracy Handbook. [9] Results were reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [24]. The protocol for this systematic review was registered on PROSPERO (registration number: CRD42013005382).

Search strategy

The following PICO algorithm was adopted: "In patients suspected of aseptic knee arthroplasty loosening, which diagnostic modality, compared to intraoperative findings, yields the highest diagnostic accuracy" [20]. A literature search was conducted in PubMed, Embase, and the Cochrane Library. Keyword search terms, MeSH terms, and Boolean operators were tailored for each database (see Appendix 1). The reference lists of the articles retrieved were screened for additional relevant articles. A search for grey literature was not conducted. To ensure the inclusion of all available literature, no start date was set. The last search was completed on March 18, 2024.

Article selection

Articles evaluating modalities used for diagnosing knee arthroplasty component loosening in living patients were included if they met the specified inclusion criteria. These inclusion criteria were: (1) articles must report the number of positive and negative outcomes based on diagnostic evaluations and intraoperative findings. Randomized controlled trials, retrospective and prospective cohort studies were included. As the review team comprised Dutch and English-speaking natives, articles in both Dutch and English were included. Animal studies, ex vivo experiments, case reports, case series, and conference papers were excluded.

All potential articles were gathered, and duplicates were then eliminated. Two reviewers (G.B. and A.K.) independently screened the articles for eligibility based on their titles and abstracts. Following this, the same reviewers performed a full-text screening according to the predefined inclusion and exclusion criteria. Any discrepancies were resolved through discussion until a consensus was achieved. Rayyan was utilized for the collection, removal of duplicates, and screening of articles [28].

Data collection

The following explanatory variables were gathered using a data extraction table including: first author, year of publication, country of publication, study design and method, total number of patients and/or knees, sample size, participant demographics (age and sex), the index test used (type of diagnostic modality) with associated thresholds, reference standard(s) with associated thresholds, the time interval between the index and reference tests.

The following diagnostic performance characteristics were extracted: the number of positive outcomes (True Positives (TP), False Positives (FP)) and negative outcomes (True Negatives (TN) and False Negatives (FN)). If possible, results for tibial and femoral components were extracted separately. Diagnostic performance characteristics were only extracted for subgroups suspected of aseptic knee arthroplasty loosening evaluated with intraoperative findings as reference test.

Methodological quality

To assess the risk of bias and applicability, both the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool and QUADAS- Comparative (QUADAS-C) tool were utilized [39, 40]. Two reviewers (G.B. and A.K.) independently evaluated potential bias using signaling questions across four domains, categorizing risk as high, low, or unclear. A domain was deemed low risk if all signaling questions were affirmed; any negative response indicated potential bias. Applicability concerns regarding included studies were assessed based on similarity to our review's question in patient population, index test, and reference standard. Differences were resolved through discussion to reach consensus. Evidence level was determined using the Oxford Level of Evidence tool [15]. Certainty of evidence was evaluated using the GRADE approach [5]. GRADEproGDT (The Cochrane Collaboration, 2020, London, United Kingdom) was used to generate GRADE certainty of evidence tables.

Statistical analysis

The primary outcomes evaluated were sensitivity and specificity. These metrics, along with their 95% confidence intervals, were presented in forest plots and in a constructed Receiver Operating Characteristic (ROC) curve to assess and compare the diagnostic accuracy reported in the included studies. Each symbol represents the sensitivity and specificity of a single study. Symbols size represents sample size [9].

A meta-analysis was conducted using bivariate logit normal models with random effects to enable visualization of pooled estimates of sensitivity and specificity for all modalities [32]. A Summery ROC (SROC) plot was constructed, featuring summary values (curves and points) for sensitivity and specificity for each modality, along with 95% prediction regions [9, 32]. When two or more curves from different modalities were close to each other, summary points were visualized to further emphasize the proximity to the left upper corner.

When feasible, heterogeneity between the studies was evaluated using visual observation of the 95% prediction region (represents the region within which one has 95% confidence that the true sensitivity and specificity of any future study should lie) [9, 13]. Efforts to avoid publication bias included conducting a comprehensive search to identify all relevant studies. No statistical tests to quantify publication bias were conducted, as such tests are deemed inappropriate for comparative diagnostic test accuracy reviews [9].

Forest plots and SROC curves were constructed using Review Manager (RevMan) version 5.4 (The Cochrane Collaboration, 2020, London, United Kingdom). To establishing 95% prediction regions and summary points, calculations for the bivariate logit normal model were conducted using MetaBayesDTA v1.5.1. [29].

Results

Article screening and inclusion

We identified 467 articles, including 14 in our analysis (Figure 1) [1, 2, 6, 7, 10, 14, 18, 21-23, 25, 30, 34, 36]. None of the included articles reported accuracy metrics for CT with or without contrast. An article by Foti et al., evaluating aseptic loosening with single and dual energy CT, was excluded.

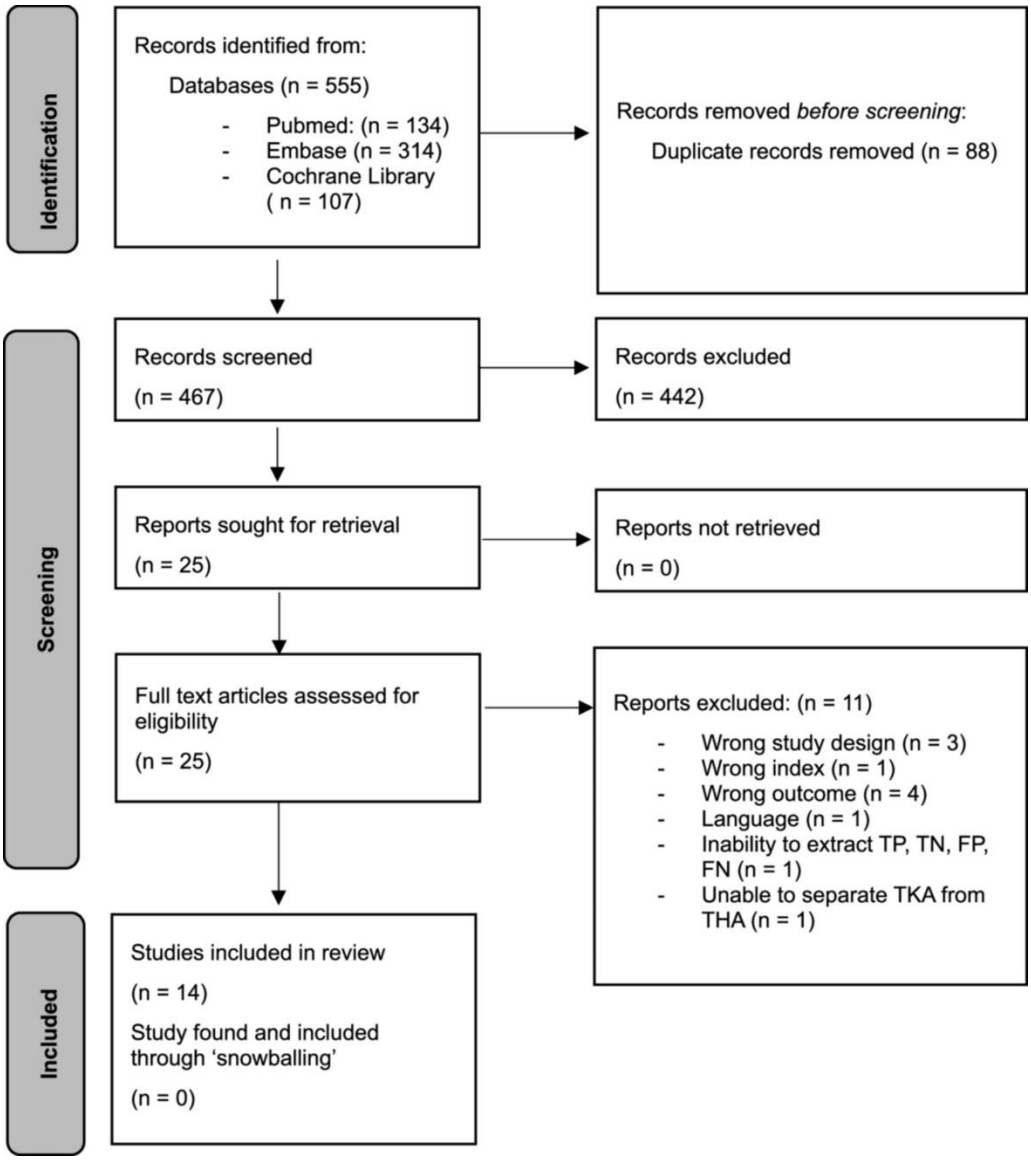


Figure 1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of study selection. TP – True Positive, TN – True Negative, FP – False positive, FN – False Negative, THA – Total Hip Arthroplasty, TKA – Total Knee Arthroplasty.

Characteristics of included studies

Fourteen articles published between 2000 and 2024 assessed the diagnostic accuracy of five modalities (Table 1). SPECT/CT was the most studied modality, with seven articles [1, 2, 6, 14, 21, 25, 30]. Abele et al. retrospectively evaluated SPECT/CT in 17 knees, eight of which used intraoperative findings as a reference standard [1]. Al-Nabhani et al. included 69 patients in their retrospective study, with 24 cases evaluated for suspected loosening using SPECT/CT [2]. Hirschmann et al. prospectively included 33 patients, evaluating both SPECT/CT and intraoperative findings, with the presence of implant toggling as the threshold for intraoperative loosening [14]. Chew et al. studied 44 patients with SPECT/CT, dichotomizing results for tibial and femoral components [6]. Murer et al. conducted a retrospective cohort study with 83 patients, specifying results per component [25]. The thresholds employed for the index and reference tests were identical to the threshold used by Hirschmann et al. [14, 25]. Puijk et al retrospectively evaluated aseptic loosening in 19 patients [30]. Mandegaran et al. assessed SPECT/CT in 33 cases and bone scintigraphy in 29 cases [21].

Claassen et al. and Smith et al. retrospectively identified 46 and 71 cases of aseptic TKA loosening, respectively, using bone scintigraphy [7, 34]. Kitchener et al. evaluated radionuclide arthrograms of 66 cases with suspected aseptic loosening, while Marx et al. retrospectively assessed 23 cases with radionuclide arthrogram [18, 22].

Mayer-Wagner et al. and Sterner et al. studied 44 and 6 cases, respectively, of suspected loosening using 18F-FDG-PET [23, 36]. Endo et al. was the only study reporting MRI. Endo et al. reported on MRI, evaluating 116 cases retrospectively using intraoperative findings as a reference test [10].

A summary of study characteristics is presented in Table 1. Thresholds for the determination of loosening for both the index and reference test are specified in Appendix II.

Methodological quality and level of evidence

Nine studies showed a high risk of bias in patient selection due to including patients where aseptic loosening was one of several potential diagnoses, but another diagnosis was considered more likely. Two studies were identified as having a high risk of bias for the index test, as the threshold for component loosening was not well-defined, and/or raters were not blinded to the reference test results. All included studies were deemed to have a high risk of bias regarding the reference test, primarily due to the absence of blinding for the index test and/or inadequate specification of the reference test threshold. As only subgroups evaluating the index test against intraoperative findings in patients with a differential diagnosis of aseptic loosening were included for meta-analysis, any bias concerning flow and time mainly resulted from the unspecified interval between the index and reference tests and no concerns regarding the applicability to the research questions were registered (Figure 2). According to the Oxford Level of Evidence tool, 11 out of 14 studies were classified as Level III (Table 1).

Study	Country	Population	Data acquisition	Number of cases eligible	Mean age (years)	Percentage of female sex	Time interval between index and reference test	Modality (index test)	Reference standards employed	Level of Evidence
Mayer-Wagner et al., 2009	Germany	Symptomatic knee and hip arthroplasty suspected for aseptic loosening	Prospective cohort study	44	70	65%	NR	18F-FDG-PET	Intraoperative findings	II
Sterner et al., 2006	Germany	Symptomatic knee arthroplasty with aseptic loosening as differential diagnosis	Prospective cohort study	6	70	43%	NR	18F-FDG-PET	Intraoperative and clinical findings	II
Claassen et al., 2014	Germany	Symptomatic knee arthroplasty with uncertainty after clinical and radiological examination, infection excluded	Retrospective cohort study	46	69	78%	3.2 ± 2.2 months	Bone scintigraphy	Intraoperative findings	III
Smith et al., 2000	United Kingdom	Symptomatic knee arthroplasty with aseptic loosening as differential diagnosis	Retrospective cohort study	71	66	59%	NR	Bone scintigraphy	Intraoperative and clinical findings	III
Mandegaran et al., 2018	United Kingdom	Symptomatic knee arthroplasty with aseptic loosening as differential diagnosis	Retrospective cohort study	62	66	63%	NR	Bone scintigraphy and SPECT/CT	intraoperative and clinical findings	III
Endo et al., 2022	United States of America	Symptomatic knee arthroplasty suspected of loosening	Retrospective cohort study	116	63	52%	34 ± 26 days	MRI	Intraoperative findings	III
Kitchener et al., 2005	Australia	Symptomatic knee arthroplasty suspected of loosening	Retrospective cohort study	66	62	52%	5 (1-24) months	Radionuclide arthrogram	Intraoperative findings	III
Marx et al., 2005	Germany	Symptomatic knee arthroplasty with uncertainty after clinical and radiological examination, infection excluded	Retrospective cohort study	23	67	61%	NR	Radionuclide arthrogram	Intraoperative findings	III
Abele et al., 2015	Canada	Symptomatic knee arthroplasty suspected for aseptic loosening	Retrospective cohort study	8	NR	NR	1.1 ± 1.0 years	SPECT/CT	Intraoperative, clinical or radiological findings	III
Al-Nabhani et al., 2013	United Kingdom	Symptomatic knee arthroplasty with aseptic loosening as differential diagnosis	Retrospective cohort study	24	71	72%	NR	SPECT/CT	Intraoperative, clinical or radiological findings	III
Hirschmann et al., 2015	Switzerland	Symptomatic knee arthroplasty suspected of loosening, infection excluded	Prospective cohort study	33	70	66%	0.8 ± 1.8 years	SPECT/CT	Intraoperative findings	II
Murer et al., 2019	Switzerland	Symptomatic knee arthroplasty suspected of loosening, infection excluded	Retrospective cohort study	83	69	61%	NR	SPECT/CT	Intraoperative findings	III
Chew et al., 2010	Australia	Symptomatic hip and knee arthroplasty suspected of loosening	Retrospective cohort study	44	NR	NR	NR	SPECT/CT	Intraoperative findings	III
Puijk et al. 2024	The Netherlands	Symptomatic knee arthroplasty with uncertainty after clinical and radiological examination, infection excluded	Retrospective cohort study	19	66	59%	NR	SPECT/CT	Intraoperative and clinical findings	III

Abbreviations: 18F-FDG-PET, 18-fluorodeoxyglucose positron emission tomography; MRI, magnetic resonance imaging; NR, not reported; SPECT/CT, single-photon emission computed tomography combined with computed tomography

Study	Test	Risk of bias (QUADAS-2)				Applicability concerns (QUADAS-2)			Risk of bias (QUADAS-C)			
		P	I	R	FT	P	I	R	P	I	R	FT
Claassen et al. 2014	Bone scintigraphy	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓
Mandegaran et al. 2018	Bone scintigraphy	✗	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗
	SPECT/CT	✗	✓	✗	✗	✓	✓	✓				
Smith et al. 2000	Bone Scintigraphy	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓
Mayer-Wagner et al. 2009	18F-FDG-PET	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓
Sterner et al. 2006	18F-FDG-PET	✓	✗	✗	✗	✓	✓	✓	✓	✗	✗	✗
Abele et al. 2015	SPECT/CT	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗
Al-Nabhani et al. 2013	SPECT/CT	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓
Chew et al. 2010	SPECT/CT	✗	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗
	Radionuclide Arthrogram	✗	✓	✗	✗	✓	✓	✓				
Hirschmann et al. 2015	SPECT/CT	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓
Murer et al. 2019	SPECT/CT	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	✓
Puijk et al. 2024	SPECT/CT	✓	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗
Kitschener et al. 2005	Radionuclide Arthrogram	✗	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗
Marx et al. 2005	Radionuclide Arthrogram	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓
Endo et al. 2022	MRI	✓	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗

Figure 2a: Quality assessment of included studies using the QUADAS-C tool. P = patient selection; I = index test; R = reference standard; FT = flow and timing. ✓ indicates low risk; ✗ indicates high risk; ? indicates unclear risk. MRI - Magnetic Resonance Imaging, 18F-FDG-PET - Fluorodeoxyglucose Positron Emission Tomography, SPECT/CT - Single-Photon Emission Computed Tomography combined with Computed Tomography.

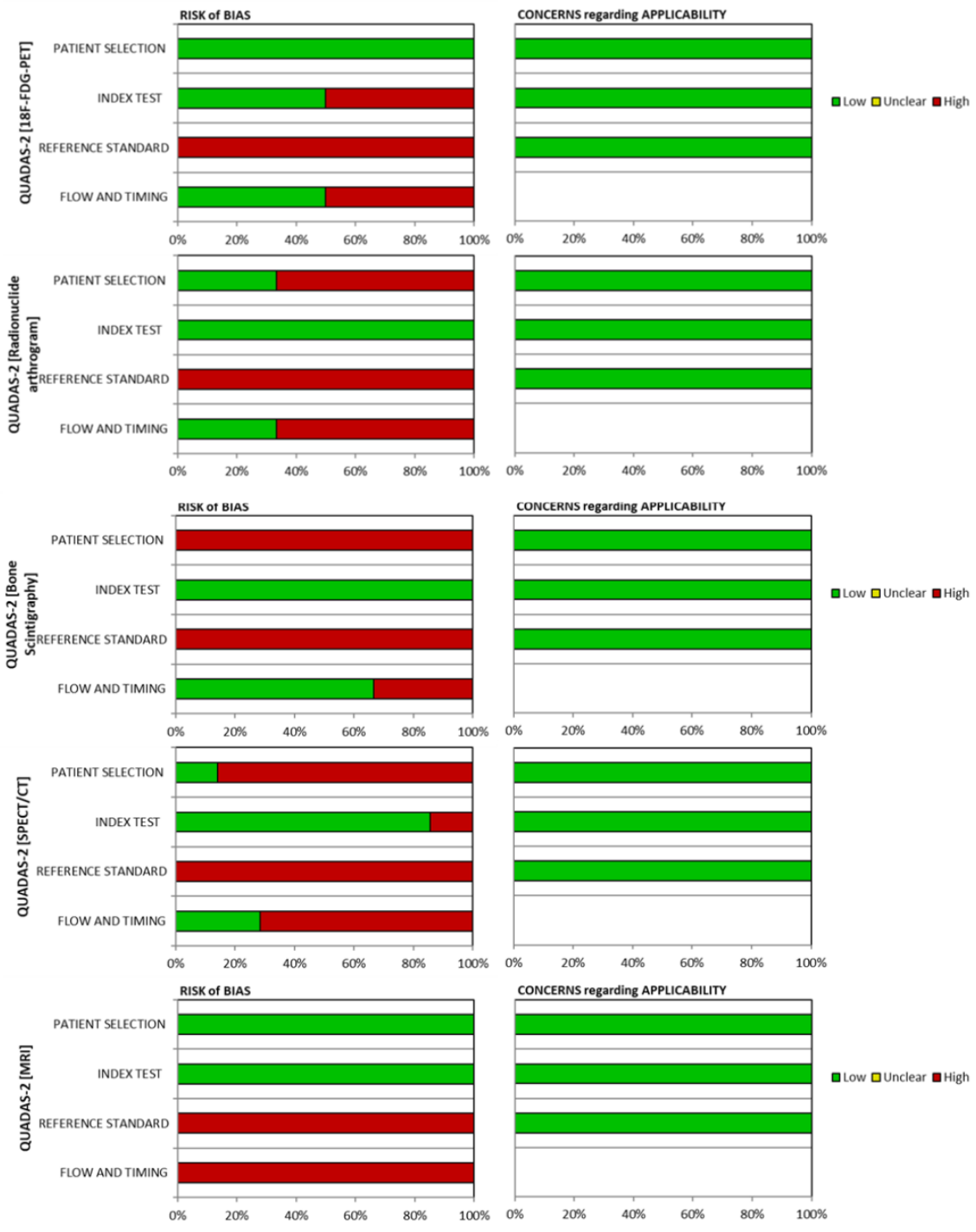


Figure 2b: Summary of risk of bias and concerns about applicability using the Quality Assessment of Diagnostic Accuracy Studies-Comparative (QUADAS-C) tool. Assessment by reviewers of each domain for every study, displayed as percentages indicate the count of studies rated as high, unclear, or low in terms of risk of bias or applicability. MRI - Magnetic Resonance Imaging, 18F-FDG-PET - Fluorodeoxyglucose Positron Emission Tomography, SPECT/CT - Single-Photon Emission Computed Tomography combined with Computed Tomography.

As assessed by visual examination of 95% prediction regions, high heterogeneity is observed between studies evaluating SPECT/CT and slightly less for radionuclide arthrogram and bone scintigraphy. Heterogeneity could not be tested for MRI and 18F-FDG-PET due to the insufficient of number of studies and sample sizes.

Evaluation of certainty of evidence using the GRADE approach resulted in a low certainty of evidence for all included diagnostic modalities. Certainty of evidence was mainly reduced due to high risk of bias across different domains of the QUADAS-C tool and the serious imprecision. The serious imprecision was a result of the reported wide confidence intervals for sensitivity and specificity (Appendix III).

Comparative diagnostic accuracy meta-analysis

Reported confidence intervals around the point estimates of sensitivity and specificity were wide (Figure 3 and 4). The SROC curves and summary points indicate that the combined findings both for SPECT/CT (7 studies) and MRI (1 study) exhibit proximity to the upper left corner, denoting their stature as the most accurate diagnostic modalities included in this meta-analysis (Figure 3).

Reported sensitivities and specificities for SPECT/CT ranged from 0.43 – 1.00 and 0.63 – 1.00, respectively. Reported sensitivity and specificity for MRI were 0.81 and 0.98 (tibia component) and 0.31 and 1.00 (femoral component), respectively. Bone scintigraphy sensitivities and specificities ranged from 0.76 – 0.88 and 0.30 – 0.83, respectively. 18F-FDG-PET sensitivity ranged 0.64 – 1.00 and specificity ranged 0.00 – 0.77. Radionuclide arthrogram curves showed to lowest proximity to the upper left corner, with sensitivities ranging 0.08 – 0.82 and specificities ranging 0.70 – 0.97 (Figure 4).

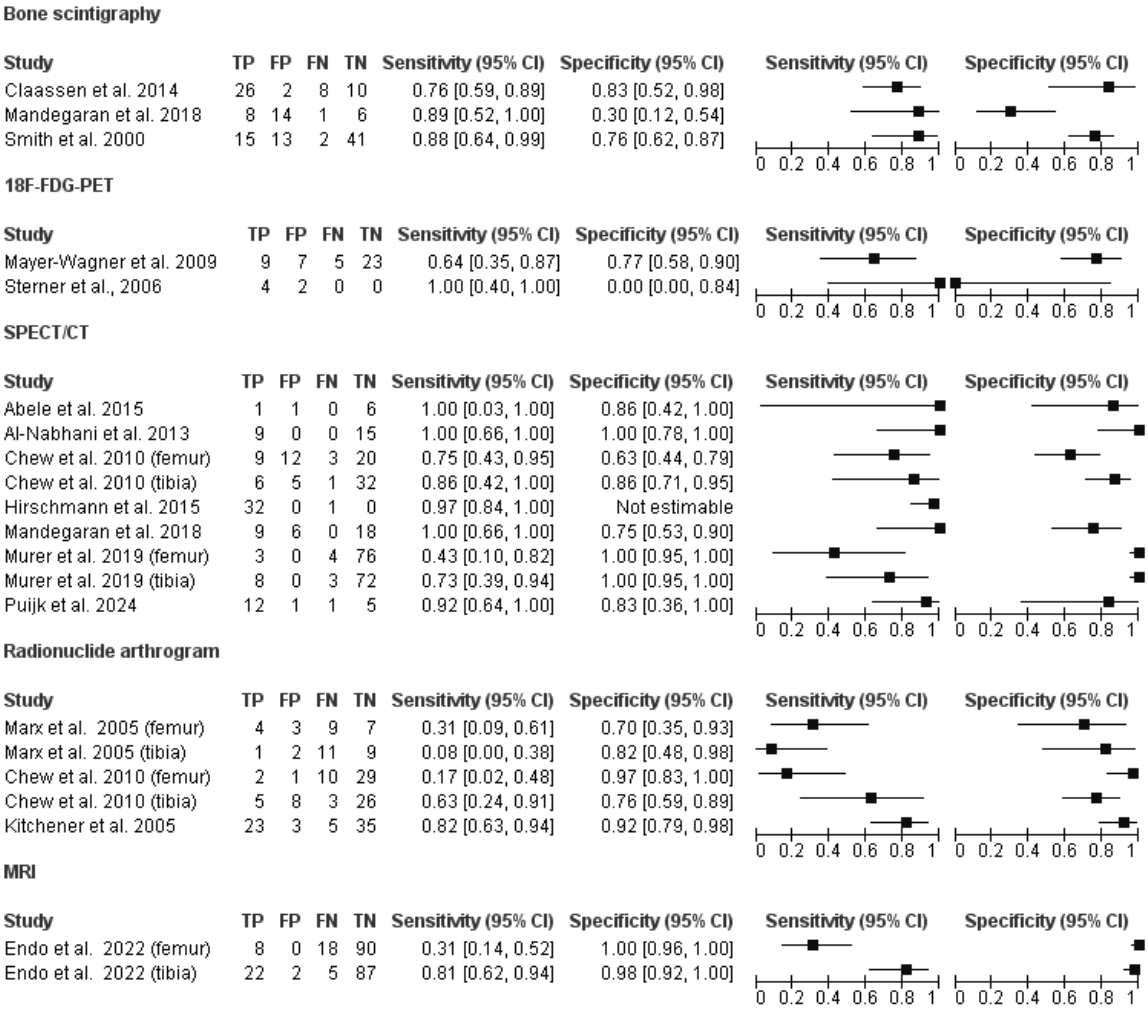


Figure 3: Forest plot displaying provided counts of True Positives (TP), False positives (FP), False Negatives (FN) and True Negatives (TN) and consequent Sensitivity and Specificity with 95% Confidence Intervals (CI) for each study, stratified per modality. MRI - Magnetic Resonance Imaging, 18F-FDG-PET - Fluorodeoxyglucose Positron Emission Tomography, SPECT/CT - Single-Photon Emission Computed Tomography combined with Computed Tomography.

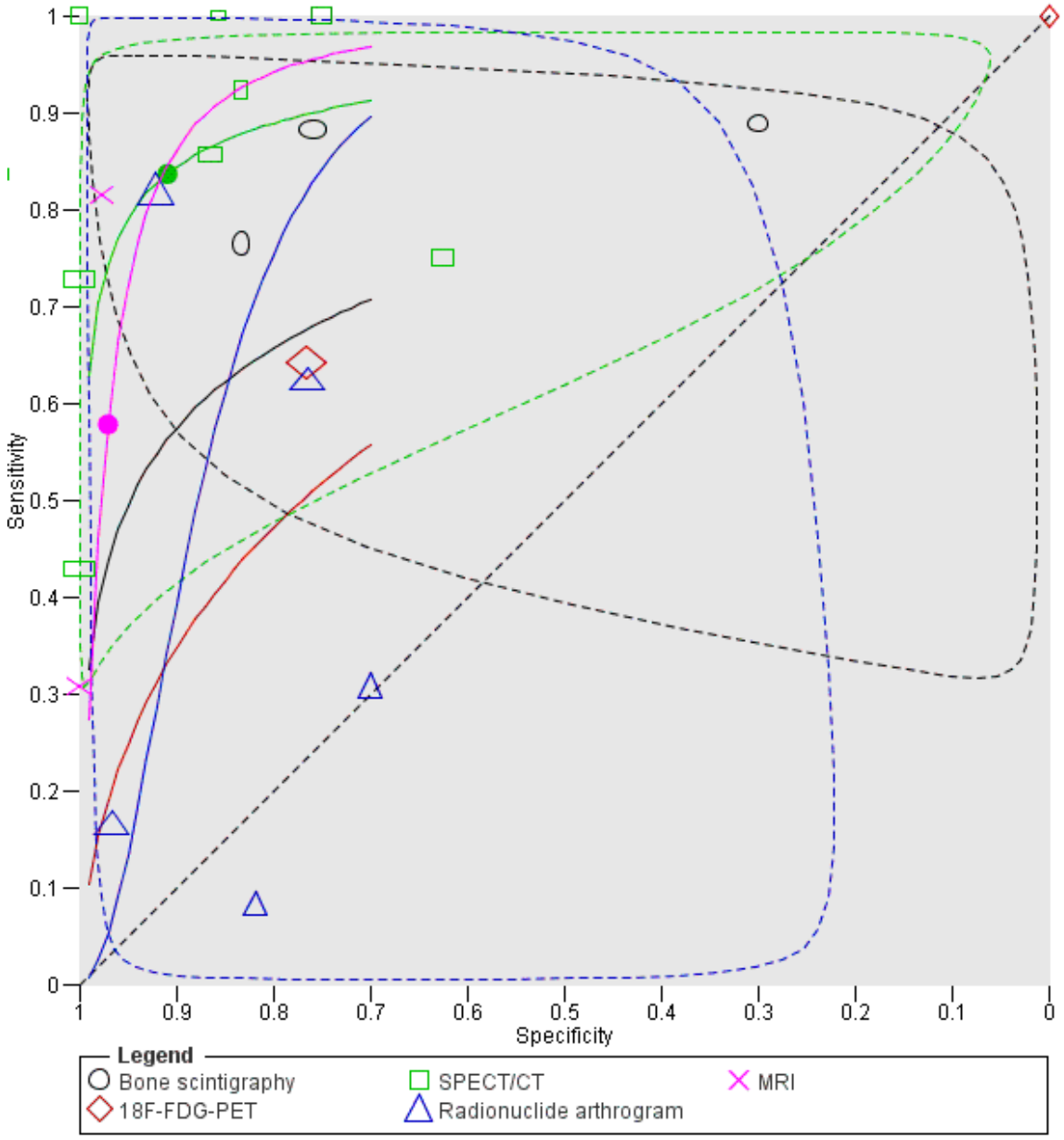


Figure 4: Summary Receiver-Operating Characteristic (SROC) for each modality. Each symbol represents a single study, with the symbol shape corresponding to the type of diagnostic test used (see legend). The position of each symbol reflects the sensitivity and specificity of the modality evaluated in that study—the closer the proximity to the top-left corner a symbol is, the better the modalities diagnostic performance in that study. Sample size per study is represented by symbol size. The curved lines represent the trade-off between sensitivity and specificity for each modality. The summary point (fully colored round circles) marks the average sensitivity and specificity for both MRI and SPECT/CT modality, providing a single estimate of the modality's accuracy. The curve and/or summary point closest to the upper left corner of the graph, indicating the largest area under the curve, represents the modality with the highest accuracy. The dashed line indicates the 95% prediction region (the area where there's a 95% probability that the outcomes of a new study will fall). MRI - Magnetic Resonance Imaging, 18F-FDG-PET - Fluorodeoxyglucose Positron Emission Tomography, SPECT/CT - Single-Photon Emission Computed Tomography combined with Computed Tomography.

Discussion

This review indicates MRI and SPECT/CT as the most accurate modalities for diagnosing knee arthroplasty loosening However, it is crucial to interpret these results cautiously. This caution is mainly due to the significant heterogeneity across most included modalities, the broad confidence intervals observed in many studies, and the high risk of bias leading to a generally low certainty of evidence according to the GRADE approach.

These are consistent with those found in a prior diagnostic test comparison conducted by Barnsley et al., which exclusively investigated nuclear imaging techniques for assessing aseptic loosening in knee arthroplasty components [4]. Their research highlighted that SPECT/CT stands out as the most accurate diagnostic tool. Additionally, a more recent analysis by Anzola et al. focused on the effectiveness of SPECT/CT in identifying pain sources in patients with non-infected knees post-arthroplasty, revealing an overall sensitivity of 0.86 (95% CI: 0.79-0.93) and specificity of 0.90 (95% CI: 0.79-0.96) [3]. Our review builds upon these prior studies by offering an updated, comprehensive evaluation of both nuclear and non-nuclear imaging modalities, specifically concerning the diagnostic accuracy for aseptic loosening in knee arthroplasty.

The American College of Radiology's Expert Panel on Musculoskeletal Imaging stated that CT scans without contrast can assess post-TKA pain effectively. However, our review found no studies specifically assessing CT alone for diagnosing aseptic knee arthroplasty component loosening [38]. Foti et al. conducted a prospective blinded study evaluating single and dual energy CT against intraoperative findings, reporting 88% sensitivity and 91% specificity for the tibial component and 81% sensitivity and 94% specificity for the femoral component. Unfortunately, this study was excluded from our meta-analysis due to unavailability of TP, TN, FP, and FN values [11].

The findings in this review should be interpreted within the context of its limitations. It solely assessed accuracy, overlooking factors such as radiation exposure, availability, and cost-effectiveness of modalities, potentially limiting applicability across various settings. Furthermore, Additionally, it only focused on accuracy for detecting aseptic loosening, omitting differentiation ability for other diagnoses. The meta-analysis's primary limitation lies in the small number of studies, high risk of bias, and consequent low certainty of evidence. This uncertainty is primarily due to bias in the index and reference tests, hindering comparability among studies. Consequently, no single preferred diagnostic modality could be identified, as MRI and SPECT/CT showed similar performance. The wide 95% prediction regions indicate vulnerability to change, especially with potential future, larger and better-conducted studies. Nevertheless, these findings represent the best evidence available. Future studies should adopt prospective blinded designs to assess diagnostic accuracy effectively. In addition to the scope of this review, advanced and/or less cumbersome and/or expensive diagnostic modalities,

including advanced machine learning models, load-bearing image acquisition, and techniques to quantify and visualize implant displacement, are being explored [16, 17, 33]. Moreover, initiatives to mitigate the risk of loosening in total knee arthroplasty are crucial. Recent studies suggest that thicker inserts (>13mm) may elevate early failure risk, while the adoption of cementless deep dish rotating platform knees and unrestricted kinematic alignment may lower loosening incidence [26, 31, 35].

Considering all findings and limitations of this review and meta-analysis, clinicians should consider using MRI and/or SPECT-CT to assist in the diagnosis of aseptic loosening. However, they should refrain from solely relying on the outcomes of these modalities, as their preference is grounded on a low certainty of evidence. It is imperative for clinicians to also contemplate alternative diagnoses and evaluate the impact (e.g., radiation dosage) on patients.

Conclusions

This review and meta-analysis assessed available diagnostic techniques to aid the diagnosis of knee arthroplasty loosening and based on a low certainty of evidence suggests that MRI and SPECT/CT are currently the most accurate modalities used to aid the diagnosis aseptic loosening of knee arthroplasty components. However, it primarily emphasizes the scarcity of high-quality evidence for modalities assessing aseptic loosening. Thus, clinicians should not solely rely on the results of a single modality for their diagnosis.

References

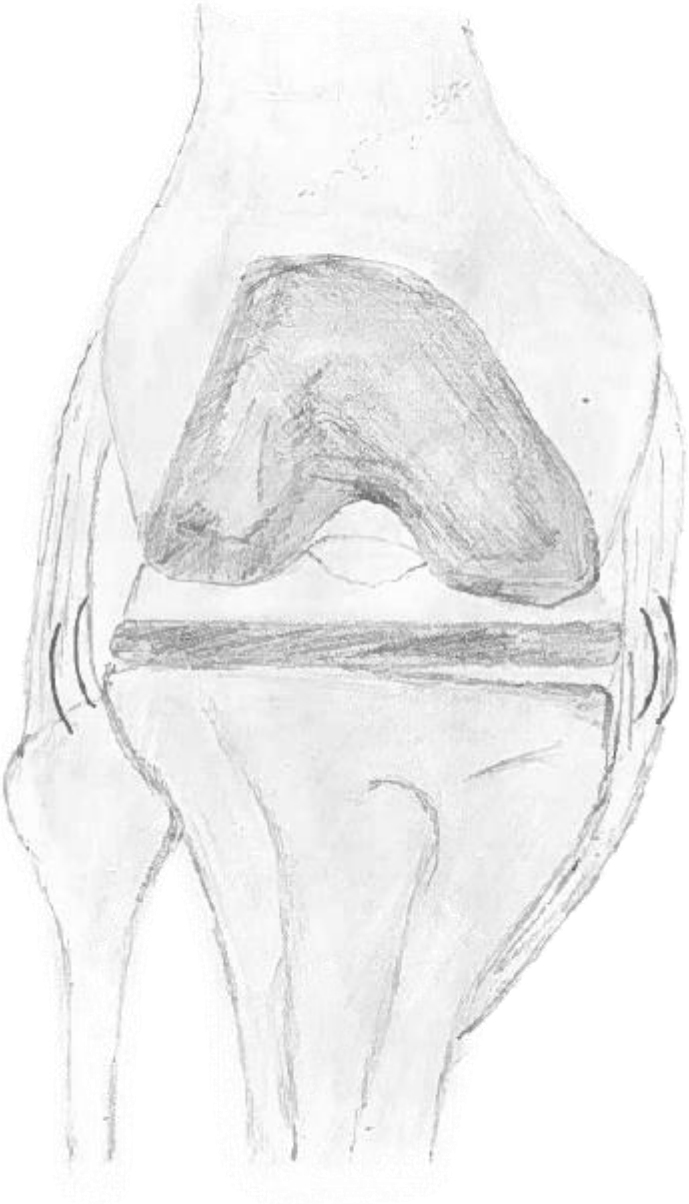
1. Abele JT, Swami VG, Russell G, Masson EC, Flemming JP (2015) The Accuracy of Single Photon Emission Computed Tomography/Computed Tomography Arthrography in Evaluating Aseptic Loosening of Hip and Knee Prostheses. J Arthroplasty 30:1647-1651
2. Al-Nabhani K, Michopoulou S, Allie R, Alkalbani J, Saad Z, Sajjan R, et al. (2014) Painful knee prosthesis: can we help with bone SPECT/CT? Nucl Med Commun 35:182-188
3. Anzola LK, Hernandez N, Rodriguez LF, Sanguino G, Martinez E, Lopez R, et al. (2023) The role of SPECT/CT in painful, noninfected knees after knee arthroplasty: a systematic review and meta-analysis—a diagnostic test accuracy review. Journal of Orthopaedic Surgery and Research 18:223
4. Barnsley L, Barnsley L (2019) Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiol 48:1565-1572
5. Brozek JL, Canelo-Aybar C, Akl EA, Bowen JM, Bucher J, Chiu WA, et al. (2021) GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence-An overview in the context of health decision-making. J Clin Epidemiol 129:138-150
6. Chew CG, Lewis P, Middleton F, van den Wijngaard R, Deshaies A (2010) Radionuclide arthrogram with SPECT/CT for the evaluation of mechanical loosening of hip and knee prostheses. Ann Nucl Med 24:735-743
7. Claassen L, Ettinger M, Plaass C, Daniilidis K, Calliess T, Ezechieli M (2014) Diagnostic value of bone scintigraphy for aseptic loosening after total knee arthroplasty. Technol Health Care 22:767-773
8. Dalury DF, Pomeroy DL, Gorab RS, Adams MJ (2013) Why are total knee arthroplasties being revised? J Arthroplasty 28:120-121
9. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0. . Cochrane; 2023.
10. Endo Y, Burge AJ, Koff MF, Lin B, Westrich GH, Boettner F, et al. (2022) Diagnostic Performance of MRI for Component Loosening in Total Knee Arthroplasty Compared with Radiography. *Radiology 304:128-136*
11. *Foti G, Longo C, D'Onofrio M, Natali S, Piovan G, Oliboni E, et al. (2023) Dual-Energy CT for Detecting Painful Knee Prosthesis Loosening. Radiology 306:e211818*
12. *Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284*
13. *Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 8:239-251*
14. *Hirschmann MT, Amsler F, Rasch H (2015) Clinical value of SPECT/CT in the painful total knee arthroplasty (TKA): a prospective study in a consecutive series of 100 TKA. Eur J Nucl Med Mol Imaging 42:1869-1882*
15. *Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. (2011) The 2011 Oxford CEBM evidence levels of evidence (introductory document). Oxford Center for Evidence Based Medicine*
16. *Kievit AJ, Buijs GS, Dobbe JGG, Ter Wee A, Kerkhoffs G, Streekstra GJ, et al. (2023) Promising results of an non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon) 104:105930*
17. *Kim M-S, Cho R-K, Yang S-C, Hur J-H, In Y. Machine Learning for Detecting Total Knee Arthroplasty Implant Loosening on Plain Radiographs. Bioengineering. 2023;10(6).*
18. *Kitchener MI, Coats E, Keene G, Paterson R (2006) Assessment of radionuclide arthrography in the evaluation of loosening of knee prostheses. Knee 13:220-225*
19. *Klug A, Gramlich Y, Rudert M, Drees P, Hoffmann R, Weißenberger M, et al. (2021) The projected volume of primary and revision total knee arthroplasty will place an immense burden on future health care systems over the next 30 years. Knee Surg Sports Traumatol Arthrosc 29:3287-3298*
20. *Luijendijk HJ (2021) How to create PICO questions about diagnostic tests. BMJ Evid Based Med 26:155-157*
21. *Mandegaran R, Agrawal K, Vijayanathan S, Gnanasegaran G (2018) The value of 99mTc-MDP bone SPECT/CT in evaluation of patients with painful knee prosthesis. Nucl Med Commun 39:397-404*
22. *Marx A, Saxler G, Landgraeber S, Löer F, Holland-Letz T, von Knoch M (2005) Comparison of subtraction arthrography, radionuclide arthrography and conventional plain radiography to assess loosening of total knee arthroplasty. Biomed Tech (Berl) 50:143-147*
23. *Mayer-Wagner S, Mayer W, Maegerlein S, Linke R, Jansson V, Müller PE (2010) Use of 18F-FDG-PET in the diagnosis of endoprosthetic loosening of knee and hip implants. Arch. Orthop. Trauma Surg. 130:1231-1238*
24. *Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339:b2535-b2535*
25. *Murer AM, Hirschmann MT, Amsler F, Rasch H, Huegli RW (2020) Bone SPECT/CT has excellent sensitivity and specificity for diagnosis of loosening and patellofemoral problems after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 28:1029-1035*
26. *Niesen AE, Tirumalai PA, Howell SM, Hull ML (2024) Risk of tibial baseplate loosening is low in patients following unrestricted kinematic alignment total knee arthroplasty using a cruciate-retaining medial conforming insert: A study using radiostereometric analysis at 2 years. Knee Surg Sports Traumatol Arthrosc 32:693-703*
27. *O'Donovan P, McAleese T, Harty J (2023) Does lucency equate to revision? A five-year retrospective review of Attune and Triathlon total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 31:4773-4781*
28. *Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. Systematic Reviews 5:210*
29. *Patel A, Cooper N, Freeman S, Sutton A (2021) Graphical enhancements to summary receiver operating characteristic plots to facilitate the analysis and reporting of meta-analysis of diagnostic test accuracy data. Res Synth Methods 12:34-44*
30. *Puijk R, Lubbe P, Sierevelt IN, Kristians AE, Boer J, Nolte PA (2024) The Diagnostic Characteristic and Reproducibility of Bone Scintigraphy Single-Photon Emission Computed Tomography/Computed Tomography*

for Diagnosing Aseptic Loosening of Uncemented Total Knee Arthroplasty. The Journal of Arthroplasty;https://doi.org/10.1016/j.arth.2024.01.005

31. *Rajamäki A, Niemeläinen M, Junnila M, Lehtovirta L, Karsikas M, Ponkilainen V, et al. (2023) Thicker polyethylene inserts (≥ 13 mm) increase the risk for early failure after primary cruciate-retaining total knee arthroplasty (TKA): a single-centre study of 7643 TKAs. Knee Surg Sports Traumatol Arthrosc 31:1018-1025*
32. *Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology 58:982-990*
33. *Schröder FF, Post CE, van Raak SM, Simonis FFJ, Wagenaar FBM, Huis In't Veld R, et al. (2020) The diagnostic potential of low-field MRI in problematic total knee arthroplasties - a feasibility study. J Exp Orthop 7:59*
34. *Smith SL, Wastie ML, Forster I (2001) Radionuclide Bone Scintigraphy in the Detection of Significant Complications after Total Knee Joint Replacement. Clinical Radiology 56:221-224*
35. *Stadler C, Hofstätter M, Luger M, Stöbich M, Ruhs B, Gotterbarm T, et al. (2023) No component loosening of a cementless deep dish rotating platform knee at a 5-year follow-up. Knee Surg Sports Traumatol Arthrosc 31:969-978*
36. *Sterner T, Pink R, Freudenberg L, Jentzen T, Quitmann H, Bockisch A, et al. (2007) The role of [18F]fluoride positron emission tomography in the early detection of aseptic loosening of total knee arthroplasty. Int. Surg. J. 5:99-104*
37. *Sundfeldt M, Carlsson LV, Johansson CB, Thomsen P, Gretzer C (2006) Aseptic loosening, not only a question of wear: a review of different theories. Acta Orthop 77:177-197*
38. *Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454*
39. *Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann. Intern. Med. 155:529-536*
40. *Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, et al. (2021) QUADAS-C: A Tool for Assessing Risk of Bias in Comparative Diagnostic Accuracy Studies. Ann. Intern. Med. 174:1592-1599*

Supplementary files

Please see: https://esskajournals.onlinelibrary.wiley.com/doi/10.1002/ksa.12206



Chapter III

An evaluation of the diagnostic performance of the triphasic bone scintigraphy in patients suspected of aseptic total knee arthroplasty loosening.

George S. Buijs, Arthur J. Kievit, Matthias U. Schafroth, Leendert Blankevoort

J. Orthop. 2024, 11 Augustus. doi: 10.1016/j.jor.2024.08.007

Abstract

Background: Total Knee Arthroplasty (TKA) is a widely conducted and successful orthopedic procedure. However, aseptic loosening, a common cause of TKA failure, necessitates revision surgery. Diagnostic accuracy of triphasic bone scintigraphy, a common imaging modality for aseptic loosening detection, remains controversial. This study investigated the diagnostic accuracy of bone scintigraphy when separately evaluated by a nuclear physicist and an orthopedic surgeon, and the interrater reliability between the two.

Methods: Patients undergoing knee revision surgery due to suspected aseptic loosening at three medical centers from 2006 to 2023 were included. Relevant demographic, clinical, and procedural data were extracted from the records. The bone scintigraphy results as noted by the nuclear physicist and orthopedic surgeon were used as index test and intraoperative findings of loosening were used as reference tests. Accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were calculated, and kappa's agreement was assessed.

Results: Out of 611 revision TKAs, 59 cases were analyzed. The nuclear physicist's evaluation of bone scintigraphy had a sensitivity of 73%, specificity of 0%, positive predictive value of 93%, negative predictive value of 0%, and diagnostic accuracy of 69%. The orthopedic surgeon's evaluation showed higher sensitivity, specificity, positive and negative predictive values, and an accuracy of 84%. Agreement levels were moderate (kappa = 0.46) between the nuclear physicist's and orthopedic surgeons' evaluation.

Interpretation: The diagnostic accuracy of bone scintigraphy for aseptic loosening is 84% when evaluated by the orthopedic surgeon compared to 69% for the nuclear physicist's evaluation. Kappa's agreement between the two was moderate.

Introduction

Total Knee Arthroplasty (TKA) is one of the most frequent and successful orthopedic procedures currently being conducted [14]. Several published studies and registries report and project a rising incidence of knee arthroplasty [15, 30, 43]. Although successful in most patients, persistent pain after arthroplasty is a common complication that affects up to 27% of TKA patients, with reported revision rates up to 13% within 10 years [31, 46, 49, 62].

Aseptic loosening is reported as one of the most frequent modes of TKA failure, with reported incidences of 20.4% due to loosening of the tibial component and 8.8% due to loosening of the femur component [19]. Aseptic implant loosening often requires major revision surgery and places a significant burden on patients and healthcare systems worldwide [24, 26]. Therefore, accurate and timely detection is crucial in guiding management decisions in patients with persistent post-arthroplasty pain.

A triphasic bone scintigraphy is a commonly used nuclear imaging modality for aiding the diagnosis of aseptic loosening of knee implants [6]. A bone scintigraphy is evaluated for radioisotope uptake around the prosthetic component indicating focally increased bone activity in this area. Increased uptake might indicate component loosening but also infection or healing of micro fractures [21]. Bone scintigraphy is widely available and relatively inexpensive, making it an attractive modality for orthopedic surgeons to aid the diagnosis of aseptic loosening. However, with conflicting results reported in literature, its diagnostic accuracy as an imaging modality on its own is still a matter of debate [22].

Available literature usually reports solely the judgment of component loosening by the nuclear physicist in a blinded research setting. Yet in daily practice, the decision to proceed to revision surgery is made by the orthopedic surgeon who is often evaluating the judgment of the nuclear physicist, images, and levels of tracer activity themselves. Therefore, the judgment of implant loosening of the orthopedic surgeon should be considered as well when reporting on the diagnostic accuracy of the bone scintigraphy in clinical practice. Moreover, the agreement between the nuclear physicist and the orthopedic surgeon should be reported as this may be of clinically relevant as diagnostic accuracy results may be similar while agreement on individual cases can be poor.

Therefore, this study aimed to answer two research questions: (1) what is the diagnostic accuracy of the bone scintigraphy in a daily practice when evaluated separately by the nuclear physicist and the orthopedic surgeon? and, (2) what is the interrater variability between the nuclear physicist and the orthopedic surgeon?

The hypothesis is that diagnostic accuracy of the bone scintigraphy is sufficient to aid the diagnosis of aseptic loosening as its diagnostic accuracy is expected to be > 0.5. Based on

experience of the co-authors, the interrater reliability between nuclear physicist and orthopedic surgeon is hypothesized to be fair to moderate.

Methods

Ethical statement

The institutional review board of the Amsterdam University Medical Centers (AUMC) approved this retrospective study (W22_231) and waived the need for informed consent due to the retrospective nature of this study. This study was performed in accordance with the recommendations for Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals and the European General Data Protection Regulation (GDPR) [1, 22].

Patient screening and inclusion

The data from all patients receiving knee revision surgery for aseptic loosening at the Amsterdam University Medical Center, Academisch Medisch Centrum and the Vrije Universiteit medical center between 2006 and 2023 were extracted from a national registry for orthopedic interventions (Nederlandse Registratie voor Orthopedische Interventies [LROI]) and screened for eligibility [33]. Patients were eligible and therefore included within this study if the bone scintigraphy was performed at least twelve months after previous knee arthroplasty surgery, to reduce the chance of false positives due to positive uptake after surgery [18]. Also, bone scintigraphy needed to be performed at least twelve months prior to revision surgery. Patients with a positive synovial fluid culture after preoperative arthrocentesis were excluded because these patients were considered septic.

If a single patient had multiple bone scintigraphy’s and revision surgeries within the inclusion period, the subsequent scintigraphy and revision surgery were grouped as one case and the two were included as a separate case in the study.

Data collection

All records of included patients were screened and demographic data (gender, year of birth), latest preoperative weight and height, smoking status, and date of latest joint specific arthroplasty surgery were extracted.

For the index test, (three phase bone scintigraphy) dosage and type of used radiopharmaceutical, date of execution of bone scintigraphy and the assessment of loosening of all components (separate and overall) by both the nuclear physicist and (if available) by the orthopedic surgeon were extracted. If applicable, the use of any other nuclear modalities (e.g. (Single Photon Emission Computed Tomography) scan with a CT (Computed Tomography) scan

[SPECT/CT]) were registered. In relation to the reference test, date of revision surgery, intraoperative assessment of loosening of all components (separate and overall), and the results of perioperative cultures were extracted.

Index test

In this study, the results of a three-phase bone scintigraphy (and if applicable, combined with SPECT/CT) as reported in a radiology report (RR) by the nuclear physicist and the preoperative orthopedic surgeon's prediction (SP) were used as index test.

RR was defined as positive for prosthesis component loosening if reported so for any component by the nuclear physicist. The diagnosis was based on the comprehensive evaluation of all images by the academic nuclear physicist, observing increased activity around the prothesis in both the early and later phases. The bone scintigraphy's were evaluated and reported by nuclear physicists who were not blinded to previous imaging, clinical history, pattern of complaints and physical examination.

SP was defined as positive for prosthesis component loosening if reported so for any component based in the notes of the orthopedic surgeon following the bone scintigraphy. Although notes were screened for SP based on solely the evaluation of the bone scintigraphy, orthopedic surgeons were not blinded for RR nor for previous imaging, clinical history, pattern of complaints and physical examination. SP was registered as missing if no separate evaluation of the bone scintigraphy was registered in the notes by the orthopedic surgeon.

The bone scintigraphy scans were evaluated by one of the academic nuclear physicists (RR), and orthopedic staff members (SP).

Imaging protocol

Three Phase Bone Scintigraphy

A dual-headed gamma camera (Symbia Intevo, Siemens Healthcare GmbH, Munich, Germany), equipped with a low energy high-resolution (LEHR) or low energy all-purpose (LEAP) collimator, was used. Patients were intravenously injected with 99mTc-oxidronate ([99Tc-m] HDP), with dosages dependent on body weight, ranging from 60 MBq to 900 MBq. Approximately 2 minutes after administration of the [99Tc-m] HDP), the first-phase dynamic and second-phase static images were taken. After approximately three hours, the third-phase static scan was made. Between the dynamic phase and second-phase static scan, patients were encouraged to walk, drink at least a liter (33.8 oz) of water and use the toilet. All three phases were performed in a supine, feet-first orientation. In each phase, standard anterior, posterior, and lateral images

were taken using a 15% window at a 140 keV photopeak (Tc99m-NMG) with an image matrix size of 128x128 pixels for the dynamic phase and 256x256 pixels for the static phases.

SPECT/CT

If deemed necessary by the nuclear physicist, SPECT/CT images were acquired using a hybrid SPECT/CT system (Siemens Symbia, Munich, Germany) equipped with a LEHR collimator. The CT parameters used were 30 mA, 130 keV, 512 × 512 matrix size, and 1 mm slice thickness. SPECT/CT was performed with a matrix size of 128 × 128, 1.0 zoom, 40 seconds per frame, and 120 frames at 3° intervals. Image reconstruction was performed using vendor-recommended iterative reconstruction algorithms with attenuation correction applied. No post-reconstruction filter was used.

Reference test

Intraoperative findings (IF) of loosening by the orthopedic surgeon performing the revision TKA (rTKA) was used as the only reference test. A component was considered loose when it was reported as such by the orthopedic surgeon in the surgical report. If there was no statement of evidence of loosening of any or both components within the surgical report, then the concerning components were scored as not loose. If any unclear statement was made regarding loosening the applicable component was scored as not loose. Orthopedic surgeons were not blinded for the RR and/or SP.

Statistical analysis

Post-hoc sample size evaluations were performed for both with and without the use of SPECT/CT using ClinCalc Post-hoc Power Calculator (Alpha: 0.05), resulted in an estimated post-hoc power of 100% and 6.5%, respectively [2].

Frequencies of true positives (TP), false positives (FP), true negatives (TN) and true positive (TP) findings by both the nuclear physicist and orthopedic surgeon were determined. Accuracy, sensitivity, specificity, positive predictive value (PPV) negative predictive value (NPV) were calculated.

Inter-observer agreement was calculated using Kappa's between the verdict of component loosening in the RR and SP. Due to insufficient post hoc power, no further logistic regression modeling was employed to assess the potential impact of SPECT/CT on the accuracy of the RR or SP.

Parametric data were presented as mean with standard deviation (SD) and non-parametric as median with interquartile range (IQR). Categorical data were presented as frequencies with

proportions. A p-value < 0.05 was considered statistically significant. Data were analyzed with R for Windows, version 4.2.3, using the and “irr”, “dplyr”, “glm” packages [3].

Results

Patient screening and inclusion

In the three hospitals, 611 rTKAs were performed, of which 175 were for suspected aseptic loosening. Among these, in 79 knee cases, an adjacent bone scintigraphy was deemed necessary by the treating orthopedic surgeon and reported in the patient records. After assessment of eligibility against inclusion and exclusion criteria, 59 knee cases were included for analysis (Figure 1).

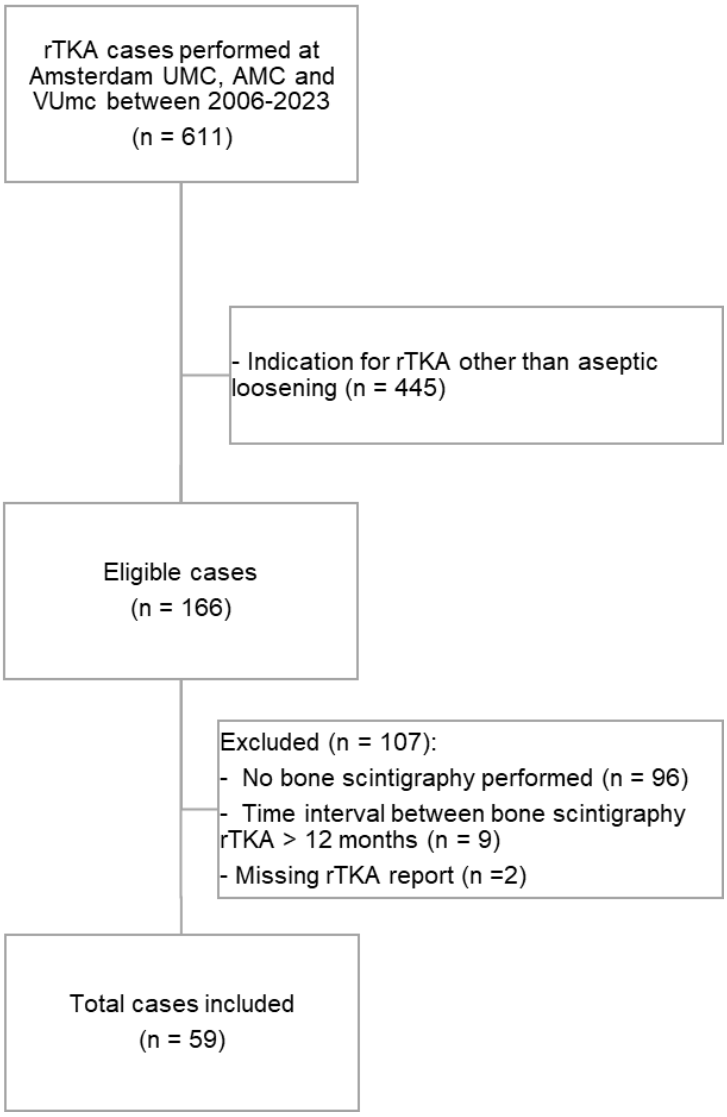


Figure 1: Flowchart of eligible and included cases. Amsterdam University Medical Center (Amsterdam UMC), Academisch Medisch Centrum (AMC), Vrije Universiteit Medisch Centrum (Vumc), revision total knee arthroplasty (rTKA).

Characteristics

The mean age at the latest knee arthroplasty surgery prior to the revision surgery was 59.9 years (SD 9.1). The mean age at the time of the bone scan and revision TKA were 65.5 years (SD 9.2) and 66.0 years (SD 9.1), respectively. For all included cases, there was no pre-operative suspicion of infection. In 9 cases, a two-stage revision TKA was performed due to intra-operatively suspected infection, yet in all cases, intra-operatively taken cultures were proven negative. The median dosage of [99Tc-m] HDP was 524 MBq (IQR: 503 – 573). In 18 cases (30%), an additional SPECT/CT was performed (Table 1).

Patient characteristics	**Overall (n=59), n (%)**
Age at latest knee arthroplasty surgery, mean (SD)	59.9 (9)
Age at revision surgery after bone scintigraphy, mean (SD)	66 (9)
Age at bone scintigraphy, mean (SD)	65.5 (9)
Time interval between bone scintigraphy and revision surgery in months, median (IQR)	4.0 (2-4)
Gender	
Female	30 (51)
Male	29 (49)
Smoking status	
Yes	11 (13)
No	48 (87)
Body Mass Index, median (IQR)	28.0 (27-34)
Laterality	
Right	27 (46)
Left	32 (54)

Table 1: Demographics and baseline characteristics. SD, Standard Deviation. IQR: Interquartile range.

Accuracy

Radiology report (RR)

Analyzing 59 cases of suspected TKA loosening, comparing RR and IF yielded: 41 TP, 0 TN, 3 FP, and 15 FN. These findings indicate a sensitivity of 73% (95% CI: 59.70%- 84.17%), specificity of 0% (95% CI:0.00%-70.76%), PPV of 93% (95% CI: 92.10% - 94.12%), NPV of 0%, and diagnostic accuracy of 69% (95% CI:56.13% - 80.81%) (Table 2a).

		Intra-operative findings (IF)	
		Fixed	**Loose**
Radiology Report (RR)	**Fixed**	0	15
	Loose	3	41

Table 2a: 2x2 contingency table comparing the radiology report (RR) with intra-operative findings (IF).

Surgeons' prediction (SP)

SP was not registered in 13 cases, so these were therefore excluded from the accuracy analysis. Consequently, evaluating 46 cases of suspected TKA loosening, comparing SP and IF revealed: 38 TP, 1 TN, 2 FP, and 5 FN. This yielded sensitivity of 88% (95% CI: 75% - 96%), specificity of 33% (95% CI: 1% - 91%), PPV of 95% (95% CI: 90% - 98%), NPV of 17% (95% CI: 3% - 55%), and accuracy of 85% (95% CI: 71% - 94%) (Table 2b).

		Intra-operative findings (IF)	
		Fixed	**Loose**
Prediction by Orthopedic Surgeon (SP)	**Fixed**	1	5
	Loose	2	38

Table 2b: 2x2 contingency table comparing the prediction of loosening by the orthopedic surgeon (SP) to the intra-operative findings (IF).

Agreement

Overall agreement

The determination of agreement levels for component loosening between RR and SP resulted in a kappa statistic of 0.46 (p = 0.0006; n = 46), denoting a moderate level of agreement. Conversely, the kappa value between RR and IF was -0.09 (p = 0.3). An assessment of agreement between SP and IF yielded a kappa value of 0.15 (p = 0.28).

Per component

Among the cases examined, the RR suspected no loosening in 15 cases (25%), while the IF reported no loose components in only 2 cases (3%). Specifically, the tibial component was predicted as loose in 27 cases (46%), with the IF reporting loosening in 37 cases (63%). On the other hand, the femur component was predicted as loose in only 1 case (2%) but was reported as loose in 4 cases (7%) by the IF. Both the femur and tibia components were predicted as loose in 16 cases (27%), and the IF reported their loosening in 16 cases (27%) (Table 3). The analysis of agreement levels resulted in a kappa statistic between the RR and IF for the loose tibia component of 0.04 (p = 0.72). In contrast, the kappa statistic for the loose femur component was 0.49 (p = 0.0002).

Component diagnosed as loose	Radiologist Report (RR) (n = 59)	Intra-operative Findings (IF) (n = 59)
None	15 (25.4%)	2 (3.4%)
Tibia	27 (45.8%)	37 (62.7%)
Femur	1 (1.7%)	4 (6.8%)
Tibia/femur	16 (27.1%)	16 (27.1%)

Table 3: frequencies and percentages of components reported as loose by the RR, radiologist report and SP, surgeon prediction.

Discussion

This pragmatic study evaluated the diagnostic accuracy of bone scintigraphy when used in clinical practice. It reveals that bone scintigraphy has a diagnostic accuracy of 84% in detecting aseptic loosening when evaluated by an orthopedic surgeon, compared to 69% when evaluated by a nuclear physicist. There is moderate agreement (Kappa = 0.46; p = 0.0006) between the orthopedic surgeons and nuclear physicists' preoperative predictions of TKA component looseness. These findings support our hypothesis that the diagnostic accuracy of bone scintigraphy is > 50%, thus adding value to the diagnostic pathway of knee implant loosening. The results also underscore the hypothesized fair to moderate agreement between the preoperative assessments of the orthopedic surgeon and the nuclear physicist.

These results suggest poorer specificity when the nuclear physicist interpreted the bone scan, compared to the orthopedic surgeon. This implies that bone scintigraphy, as interpreted by the nuclear physicist, is more likely incorrectly labels some patients without aseptic loosening as loose (false positives). However, such false positives were less frequent with the surgeon's interpretation. Our assumption is that the orthopedic surgeons slightly superior accuracy arises from their ability to access a more extensive range of clinical details. This assumption is supported by the reported moderate level of agreement between the two.

Limitations of our study include the lack of blinding for the surgeons regarding the results of the bone scintigraphy as noted by the nuclear physicist, which may have introduced bias in their intraoperative assessments of implant loosening. It is also likely that the orthopedic surgeons, as the primary responsible care providers, were more aware and therefore more inclined to consider the results of clinical and other radiological tests in their diagnostic decision-making. Additionally, potential variability in the surgeons' experience and technique might have influenced the intraoperative findings of what is determined as loose or fixed, potentially affecting the accuracy and agreement results [11]. Furthermore, the impact of additional use of SPECT/CT could not be evaluated in the current study due to insufficient power.

The findings of this study align with the results of previous systematic reviews. Barnsley et al. conducted a comprehensive review of nuclear imaging modalities for evaluating aseptic TKA loosening, reporting sensitivity and specificity rates for bone scintigraphy between 0.76 and 1.00 and 0.33 and 1.00, respectively [6]. While our study did not confirm these findings, Barnsley et al. found that SPECT/CT provided improved accuracy over bone scintigraphy. This conclusion is further supported by Anzola et al.'s recent systematic review and meta-analysis, which assessed SPECT/CT's effectiveness in diagnosing knee conditions, including loosening, in non-infected knees. They reported a pooled sensitivity of 0.86 and specificity of 0.90 across eight studies for SPECT/CT [4]. A more recent diagnostic test accuracy review and meta-analysis by Buijs et al. expands on these findings by incorporating both nuclear and non-nuclear imaging

modalities, highlighting that MRI and SPECT/CT are currently the most accurate diagnostic tools for assessing aseptic loosening in knee arthroplasty, despite the overall low certainty of evidence due to high risk of bias and heterogeneity among studies [13].

While the outcomes derived from both bone scintigraphy and SPECT/CT appear commendable, less cumbersome and cost-effective alternatives, such as quantified and visualized assessment of induced implant motion, have been proposed as potential aids in the diagnosis of implant loosening [13, 23-27].

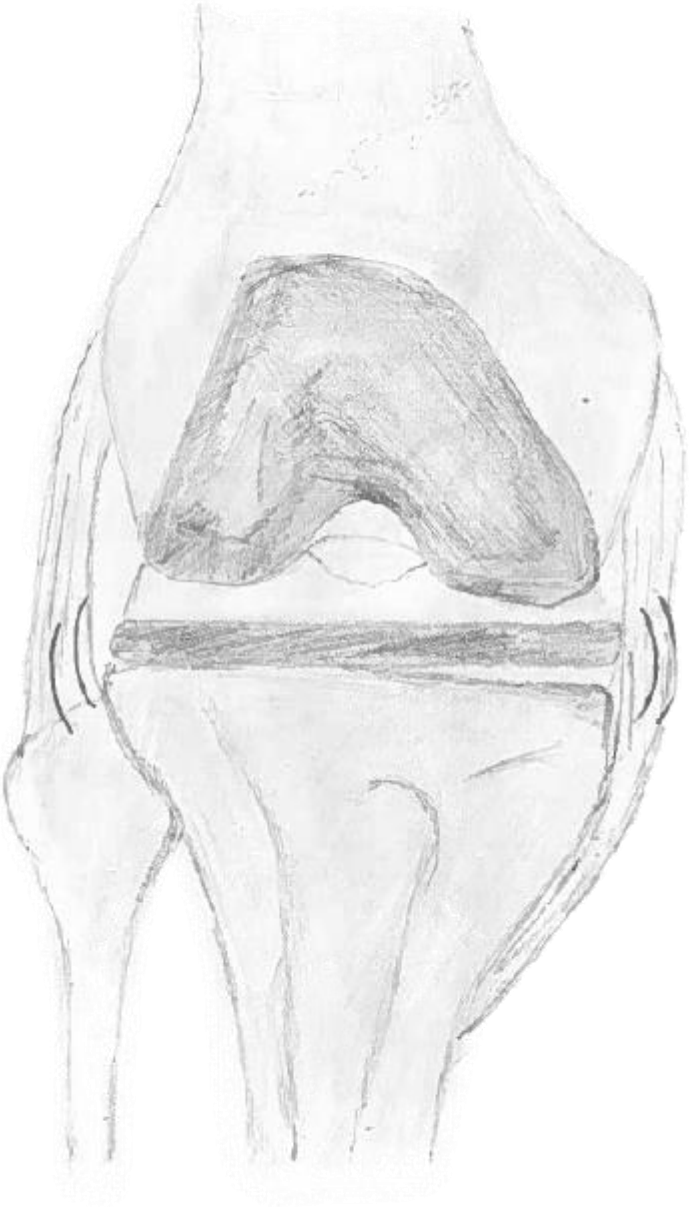
Overall, the results of this pragmatic study provide insights into the utility of bone scintigraphy in diagnosing aseptic loosening post-TKA when evaluated in a clinical routine, but also highlight the importance of research and development of additional diagnostic methods.

Conclusions

The primary results from our study show that bone scintigraphy, whether evaluated by a nuclear physicist or an orthopedic surgeon, demonstrates adequate sensitivity but has limited specificity and negative predictive value. This indicates that bone scintigraphy may be useful for ruling in aseptic loosening but is unreliable for ruling out the condition.

References

1. *The general data protection regulation. Council of the European Union; 2023. https://www.consilium.europa.eu/en/policies/data-protection/data-protection-regulation/. Accessed 01-08-2023.*
2. *Post-hoc Power Calculator. Evaluate statistical power of an existing study. Available at: https://clincalc.com/stats/Power.aspx. Accessed 01-03-2023, 2023.*
3. *R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.*
4. *Anzola L.K., Hernandez N., Rodriguez L.F., et al. The role of SPECT/CT in painful, noninfected knees after knee arthroplasty: a systematic review and meta-analysis—a diagnostic test accuracy review. J Orthop Surg Res. 2023/03/21 2023;18(1):223. doi:10.1186/s13018-023-03687-8.*
5. *Barnsley L., Barnsley L. Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiol. Oct 2019;48(10):1565–1572. doi:10.1007/s00256-019- 03215-y.*
6. *Buijs GS, Kievit AJ, Walinga AB, Schafroth MU, Hirschmann M, Blankevoort L. Visible fluid motion on manipulation as the new threshold for intraoperatively-determined knee arthroplasty component loosening – a Delphi study.;10.1002/ksa.12357. Knee Surg Sports Traumatol Arthrosc 2024.*
7. *Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis.10.1002/ksa.12206. Knee Surg Sports Traumatol Arthrosc 2024.*
8. *Carr A.J., Robertsson O., Graves S., et al. Knee replacement. Lancet. Apr 7 2012;379(9823):1331– 1340. doi:10.1016/S0140-6736(11)60752-6.*
9. *Culliford D., Maskell J., Judge A., Cooper C., Prieto-Alhambra D., Arden N.K. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. Osteoarthritis Cartilage. 2015/04/01/2015;23(4):594-600. doi:10.1016/j.joca.2014.12.022.*
10. *Duff G.P., Lachiewicz P.F., Kelley S.S. Aspiration of the knee joint before revision arthroplasty. Clin Orthop Relat Res. Oct 1996;331:132–139. doi:10.1097/00003086-199610000-00018.*
11. *Gademan M.G.J., Van Steenbergen L.N., Cannegieter S.C., Nelissen R., Marang-Van De Mheen P.J. Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop. Jun 2021;92(3):280-284. doi:10.1080/17453674.2021.1876998.*
12. *Hochman M.G., Melenevsky Y.V., Metter D.F., et al. ACR appropriateness Criteria® imaging after total knee arthroplasty. J Am Coll Radiol. 2017/11/01/2017;14(11, Supplement):S421-S448. doi:10.1016/j.jacr.2017.08.036.*
13. *ICMJE. Recommendations for the Conduct, reporting, editing, and publication of scholarly Work in medical Journals ICMJE. https://www.icmje.org/icmje-recommendations.pdf, 2023.*
14. *Kallala R.F., Ibrahim M.S., Sarmah S., Haddad F.S., Vanhegan I.S. Financial analysis of revision knee surgery based on NHS tariffs and hospital costs Does it pay to provide a revision service? Article. Bone Joint Lett J. 2015;97-B(2):197-201. doi:10.1302/0301-620X.97B2.33707.*
15. *Kassam A.M., Dieppe P., Toms A.D. An analysis of time and money spent on investigating painful total knee replacements. Article. BJMP. 2012;5(3):a526.*
16. *Kurtz S., Ong K., Lau E., Mowat F., Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. JBJS. 2007;89(4).*
17. *Labek G., Thaler M., Janda W., Agreiter M., Stöckl B. Revision rates after total joint replacement: cumulative results from worldwide joint register datasets. Article. J Bone Jt Surg. 2011;93 B(3):293– 297. doi:10.1302/0301-620X.93B3.25467.*
18. *Annual report 2023. Landelijk Registratie Orthopedische Interventies (LROI); 2023.*
19. *Nguyen L.-C.L., Lehil M.S., Bozic K.J. Trends in total knee arthroplasty implant utilization. J Arthroplasty. 2015/05/01/2015;30(5):739-742. doi:10.1016/j.arth.2014.12.009.*
20. *Piscitelli P., Iolascon G., Innocenti M., et al. Painful prosthesis: approaching the patient with persistent pain following total hip and knee arthroplasty. Clin Cases Miner Bone Metab. May 2013;10(2):97–110.*
21. *Sadoghi P., Liebensteiner M., Agreiter M., Leithner A., Böhler N., Labek G. Revision surgery after total joint arthroplasty: a complication-based analysis using worldwide arthroplasty registers. J Arthroplasty. 2013/09/01/2013;28(8):1329-1332. doi:10.1016/j.arth.2013.01.012.*
22. *Wylde V., Hewlett S., Learmonth I.D., Dieppe P. Persistent pain after joint replacement: prevalence, sensory qualities, and postoperative determinants. Pain. 2011;152(3).*
23. *Kievit AJ, Buijs GS, Dobbe JGG, et al. Promising results of a non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon). Apr 2023;104:105930. doi:10.1016/j.clinbiomech.2023.105930*
24. *Buijs GS, Kievit AJ, Schafroth MU, Hirschmann MT, Blankevoort L. Weight-bearing pain and implant migration, progressive radiolucencies, radiolucency more than 2 mm, and subsidence on radiographs and CT are generally accepted criteria for knee arthroplasty loosening - An international Delphi consensus study. Knee Surg Sports Traumatol Arthrosc 2024. DOI: 10.1002/ksa.12419*
25. *Magg C, ter Wee M, Buijs G, et al. Towards automation in non-invasive measurement of knee implant displacement. vol 12927. SPIE Medical Imaging. SPIE; 2024.*
26. *ter Wee MA, Dobbe JGG, Buijs GS, et al. Load-induced deformation of the tibia and its effect on implant loosening detection. Scientific Reports. 2023/12/08 2023;13(1):21769. doi:10.1038/s41598-023-49177-z*
27. *Buijs GS, Kievit AJ, Ter Wee MA, Magg C, Dobbe JGG, Streekstra GJ, Schafroth MU, Blankevoort L. Non-invasive quantitative assessment of induced component displacement can safely and accurately diagnose tibial component loosening in patients: A prospective diagnostic study. Knee Surg Sports Traumatol Arthrosc. 2024 May 31. doi: 10.1002/ksa.12299.*



Chapter IV

Weight-bearing pain, and implant migration, progressive radiolucency, radiolucency more than 2 mm, and subsidence on radiographs and CT are generally accepted criteria for knee arthroplasty loosening - An international Delphi consensus study.

George S. Buijs, Arthur J. Kievit, Matthias U. Schafroth, Michael T. Hirschmann, Leendert Blankevoort

Knee Surg Sports Traumatol Arthrosc. 2024 Aug 10. doi: 10.1002/ksa.12419.

Abstract

Purpose: Establishing the diagnosis of loosening in total- or unicondylar knee arthroplasty remains a challenge with different clinical and radiological signs evaluated in study designs with high risk of bias, where few or incomplete criteria are formulated for establishing the diagnosis of implant loosening. This study aimed at evaluating the variability between different clinical and radiological criteria and establish a consensus regarding clinical and radiological criteria for the diagnosis knee arthroplasty loosening.

Methods: Highly specialized knee surgeons focusing on revision arthroplasty were invited to take part in an international panel for a Delphi consensus study. In a first round, the participants were asked to state their most important clinical and radiological criteria for implant loosening. In a second round, the panel's agreement with the collected criteria was rated on a 5-point Likert scale (1-5). High variability was defined by receiving at least one score each indicating complete disagreement and complete agreement. Consensus was established when over 70% of participants rated a criterion as "fully agree" (5) or "mostly agree" (4).

Results: High variability was observed in 56% of clinical criteria and 38% of radiological criteria. A consensus was reached on one clinical (weight-bearing pain (82%)) and four radiological criteria, i.e. implant migration, progressive radiolucencies, subsidence and radiolucencies > 2mm on X-ray or CT (84–100%).

Conclusion: Among specialized knee revision surgeons, there is high variability in clinical and radiological criteria that are seen as important contributing factors to diagnosis of knee implant loosening. A consensus was reached on weight-bearing pain as clinical criterion and on implant migration, progressive radiolucencies, subsidence, and radiolucencies of more than 2mm on X-ray or CT as radiological criteria. The variability rates observed, along with the criteria that reached consensus, offer important insights for the standardization of diagnostic protocols.

Introduction

After Total Knee Arthroplasty (TKA) and Unicondylar Knee Arthroplasty (UKA) the chance for revision TKA (rTKA) within 10 years are approximately 13% and 12%, respectively [33]. The absolute number of rTKA will increase with the upward trend of numbers of primary TKA and UKA, despite the improvements of TKA and UKA designs and surgical procedures [28, 31, 33, 37].

As 20-30% of the revisions are performed due to aseptic loosening, it stands as a significant cause of TKA-failure [16, 47]. Aseptic loosening often requires complex revision surgery, heavily impacting patients, and healthcare systems worldwide [24, 26]. A correct diagnosis is essential, as unnecessary revision surgery should be avoided for patients incorrectly diagnosed with TKA loosening. Alternatively, undetected TKA loosening due to an inadequate diagnosis may lead to patients being erroneously denied revision surgery.

In the quest to enhance diagnostic accuracy for aseptic loosening, numerous studies have investigated various diagnostic criteria and modalities [4-6, 42, 48]. However, these studies have frequently yielded conflicting outcomes, contributing to a significant degree of heterogeneity of diagnostic criteria in meta-analyses [6, 13, 48]. This inconsistency is partly due to the diverse diagnostic techniques employed to establish the diagnosis of aseptic loosening. The diagnosis is based on clinical assessment, clinical laboratory tests and various imaging modalities. Different criteria are used to arrive at the diagnosis [4-6, 13, 42, 48].

Variations in patient demographics, in the duration of follow-up, and in definitions of which components can be seen as loose intraoperatively, which then is used as reference standard have compounded the complexity of drawing conclusive inferences [6]. As a result, the orthopedic community still faces challenges in establishing a standardized diagnostic protocol for aseptic loosening, which is essential for guiding treatment decisions and improving patient outcomes [59].

Amidst all these uncertainties and challenges, no study has yet explored the extent of variability and consensus of the clinical and radiological criteria for the diagnosis of aseptic loosening in knee arthroplasty. In absence of any consensus based on literature data, a Delphi study might help to arrive at a consensus for future use in clinical studies and in daily practice. It was hypothesized that high variability would be observed for several clinical and radiological criteria. No hypothesis was formulated for criteria expected to reach consensus. The aim of this study is to answer the following research questions by using the Delphi Consensus method:

(1) What is the variability in the perceived trustworthiness of preoperative clinical and radiological criteria used to diagnose aseptic knee arthroplasty loosening?

(2) Which preoperative clinical and radiological criteria are considered trustworthy by specialized knee revision experts to establish a consensus for diagnosing aseptic knee arthroplasty loosening?

Materials and methods

For the current Delphi consensus study the methodological criteria and reporting guidelines as recommended by Diamond et al. and Jünger et al. were employed [17, 23]. After the international consensus panel had been assembled, the Delphi study unfolded over two rounds. The first round focused on identifying clinical and radiological criteria that, according to knee revision specialists, are associated with aseptic loosening in knee arthroplasty. The second round involved rating the significance of these identified criteria. If a consensus was not reached within the second round, a third round was available to re-rate the importance of criteria after having seen a summary of the results from the second round. This study only considered loosening of the tibial and femoral components.

The lead author (GSB) acted as the coordinator, responsible for crafting the questionnaires based on participant responses and managing all communications. To avoid moderator bias by influencing the study's outcome, the lead author did not participate in the study as a panel member. Co-authors MUS, AJK and MTH did participate as panel members, but did not participate in the processing and analysis of the responses.

Assembling the international consensus panel

Orthopedic surgeons specialized in knee arthroplasty revision were identified in three different ways; (1) through screening of author lists of high-quality articles dealing with the diagnosis of aseptic loosening in knee arthroplasty, (2) through screening of programs of major orthopedic congresses to identify keynote speakers presenting on knee arthroplasty and/or loosening related subjects, and (3) through the clinical network of the co-authors. These three different approaches were adopted to ensure a diverse, international, representative group of panel participants, and potentially reduce selection bias.

To date, there are no definitive guidelines or recommendations regarding the ideal sample size for Delphi studies. There is no clear definition of what constitutes a too small or too large group. While some researchers suggest that a group of 10–15 experts may suffice for homogeneous groups, a larger sample is often recommended when a larger variety is expected [58]. For this study, authors required an *a priori* minimum of 25 participants. In anticipation of non-responders, a total of 69 potential expert panel participants were identified and invited. All experts who consented to a complete participation were sent the Delphi consensus questionnaires.

Baseline

At first, all participants were asked to state their years of experience as knee revision surgeon and the average number of knee revision surgeries performed per year. Furthermore, sex and country of current practice were registered. Participants were asked if they wanted to be acknowledged for full participation. An acknowledgment as a group author was only granted if a participant successfully completed all consensus rounds.

First round

To gain a better understanding of the different opinions between the participants, the following two questions were asked:

(1) *To what extent do you agree that there is no clear uniform specification in available literature of preoperative signs associated with knee arthroplasty component loosening?*
(2) *To what extent do you agree that there is a need for a uniform specification of which preoperative signs are indicative for knee arthroplasty component loosening?*

Both questions were Single Select Multiple Choice Questions using a 5-point Likert scale (1 being *I fully agree*, 2: *I mostly agree*, 3 *neutral*, 4; *I do not agree completely*, 5; *I do not agree at all*).

The first round identified criteria associated with implant loosening of both the tibial and femoral component of a TKA, UKA or rTKA, separately. Therefore, at first, two yes or no questions were posed:

(1) *In your opinion are there any preoperative clinical signs that are associated with loosening of the **tibial** component (partial/primary and/or secondary)?*
(2) *In your opinion are there any preoperative clinical signs that are associated with loosening of the **femoral** component (partial/primary and/or secondary)?*

If either of the questions were answered with *'yes'*, participants were asked to specify these elements for UKA, TKA and rTKA separately in two separate open questions. One question regarding the loosening of the tibial component and the second question with regards to loosening of the femoral component.

Second round and third round

In preparation for the second round, all statements were collected, and duplicates were removed. All statements were incorporated in the second-round survey *as "'statement' is indicative for aseptic loosening of all components of either partial, primary or revision knee arthroplasties"*. As no distinctive different elements were registered in the first round, all statements were presented as applicable to UKA, TKA and/or rTKA and for both the tibia and

femur component. Participants were asked to rate their agreement with the statements according to a 5-points Likert scale.

Results of the second round were summarized in frequency tables and presented to all participants. High variability regarding a particular statement was defined as the statement having received at least 1 score of complete disagreement and at least 1 score of complete agreement, both must apply. Consensus was defined as > 70% of participants rating the statement either as “I fully agree” or “I mostly agree”. No third round took place because consensus was already achieved in the second round.

Data collection and analysis

Baseline demographic data were presented as averages and standard deviations or medians and interquartile ranges according to their distribution. These included years of experience, surgical volume, and country of residence. The categorical variables ‘consensus’ were expressed as absolute numbers and percentages. If a participant failed to complete the first round, he or she was excluded from further participation and statistical analysis. The Chi-squared test was used to evaluate potential differences between Dutch responders and non-Dutch responders for those statements that received consensus. A p-value <0.05 was considered statistically significant. Data were analyzed using Excel (Microsoft Excel 2018; Microsoft Corp., Redmond, WA, USA) and Python (Python 3.8, Python Software Foundation, Wilmington, DE, USA). Questionnaires were distributed using Castor EDC (Amsterdam, The Netherlands), an online platform for questionnaire dissemination and data collection.

Results

Panel characteristics

A total of 69 eligible expert panel participants were contacted and 38 (55.1%) agreed to participate and were therefore included in this Delphi consensus study. All but two (n=36; 94.7%) who agreed to participate completed the first round. As four participants failed to complete the second round, both rounds were completed by 33 (84.2%) participants (Figure 1).

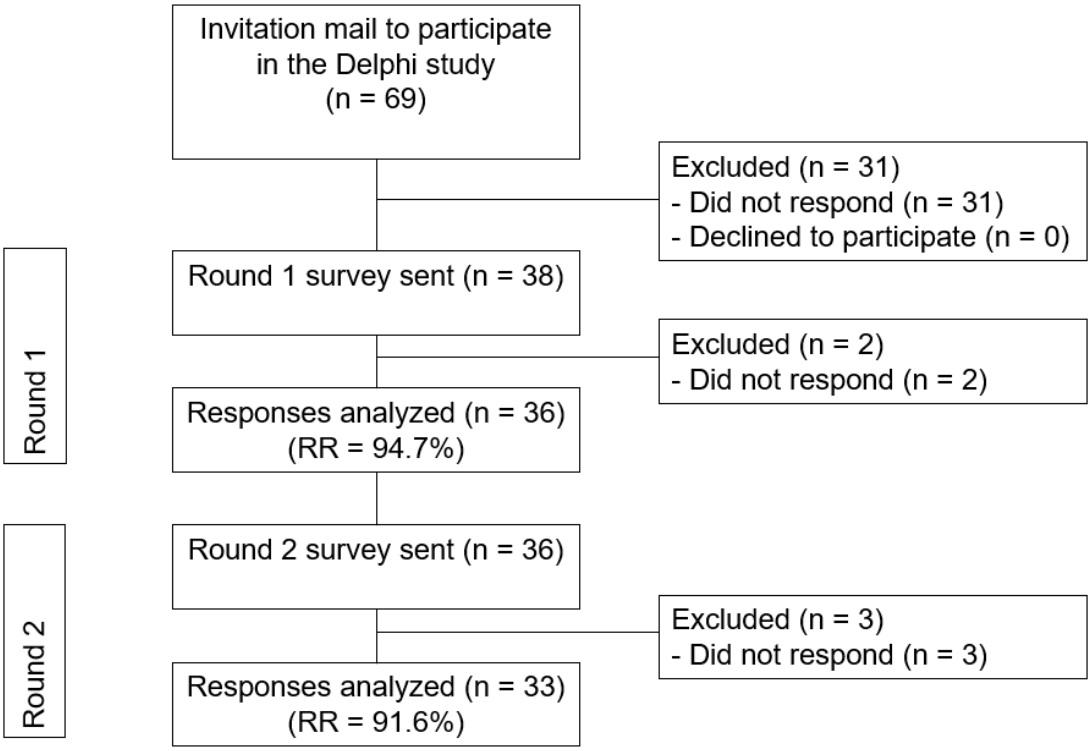


Figure 1: Flowchart of inclusion of panel members and response rates (RR) per round.

Dutch orthopedic surgeons were the predominant group, making up 24.2% of participants, the majority of whom were male (93.9%) and performed over 20 knee arthroplasty revision surgeries annually. Most participants (81.8%) boasted over 10 years of experience with knee arthroplasty revisions (Table 1).

Characteristics	n = 33	Percentages (%)
Sex		
Male	31	93.9
Female	2	6.1
Country / region of practice		
The Netherlands	8	24.2
United States of America	4	12.1
Germany	4	12.1
United Kingdom	4	12.1
Belgium	3	9.1
Switzerland	2	6.1
France	2	6.1
India	2	6.1
Indonesia	1	3.0
Peru	1	3.0
Türkiye	1	3.0
Spain	1	3.0
Experience as knee arthroplasty revision surgeon (years)		
0 - 5 years	2	6.1
5 - 10 years	4	12.1
10 - 15 years	11	33.3
20 - 25 years	9	27.3
> 25 years	7	21.2
Average number of knee arthroplasty revision surgeries per year		
0 - 10	1	3.0
10 - 20	2	6.1
20 - 30	8	24.2
30 - 40	4	12.1
40 - 50	6	18.2
> 50	12	36.4

Table 1: Baseline characteristics of participants completing all rounds of this Delphi consensus study (n = 33).

First round

Between the 36 responders in the first round, 26 (75.0%) agreed fully or mostly with the statement that there is no clear uniform specification in available literature of preoperative signs indicative for knee arthroplasty component loosening. The majority (n=34; 94.4%) either fully agreed or mostly agreed to the statement that there is a need for a clear uniform specification of preoperative signs indicative for knee arthroplasty component loosening.

All responders confirmed the existence of preoperative clinical signs that are associated with loosening of either the tibial or femoral component of a UKA, primary TKA or rTKA. All responders generated the total of 28 statements regarding the UKA tibial component, 27 statements with regards to the tibial component of a primary TKA and 28 statements with regards to the tibial component of a rTKA. Regarding the femoral component, 29 statements concerning UKA, primary TKA and rTKA were proposed. No technically relevant differences were found between statements regarding either UKA, primary TKA or rTKA. Discrimination between these knee arthroplasty types was not further considered.

All statements were then summarized and dichotomized between clinical signs and radiological signs. This resulted in 18 clinical statements (Table 2a). Summarizing the mentioned radiological signs resulted in 16 statements (Table 2b).

Second round

After being presented the summary of round one, the expert panel participants reached consensus on one clinical and four radiological signs (Table 2a/b).

As clinical sign, “Weight-bearing pain” was either fully of mostly agreed upon by 81.8% of the panel experts to be associated with loosening of either the tibial or femoral component in UKA, primary TKA or rTKA. Seven participants (21.2%) fully agreed. High variability was observed in 10 out of 18 (55.6%) statements (Table 2a). “Weight-bearing pain” did not meet the criterion for high variability.

Four radiological signs that surpassed the preset threshold for consensus were: (1) obvious implant migration on X-ray or CT with 100% agreement (23 (69.7%) fully agreed), (2) progressive radiolucency on AP and lateral X-ray or CT with 84.8% of agreement (14 (42%) fully agreed), (3) subsidence on X-ray or CT with 93.9% of agreement (15 (45.5%) fully agreed) and (4) radiolucency on AP and lateral X-ray or CT > 2mm around the implant or bone-cement interface with 90.9% of agreement (13 (39.4%) fully agreed). High variability was observed in 6 of 16 (37.5%) statements (Table 2b). None of the statements with more than 70% agreement met the criteria for high variability.

Statements	**Percentages of agreement (Fully and mostly agreed)**
Weight-bearing pain	81.8*
Change in limb alignment	63.6 ₸
Pain during axial loading, which decreases after a few steps and then increases again	60.6 ₸
Startup pain	57.6 ₸
Femoral shaft pain (for femur components of revision arthroplasties)	51.5 ₸
Tibial shaft pain (for tibia components of revision arthroplasties)	48.5 ₸
Pain on varus-valgus stress	45.5
Transient swelling around the knee after ambulation	36.4 ₸
Tenderness on palpation of the distal thigh (for femoral components) and at proximal tibia (for tibial components)	33.3
Persistent swelling around the knee	30.3 ₸
Antalgic gait	30.3 ₸
(Progressive) unloaded pain	24.2 ₸
Ligament pseudo laxity	21.2
Pain on palpation of the joint line	21.2
Instability of the knee	12.1
Instability of the knee at flexion or extension	9.1
Knee stiffness	6.1
Peri-patellar swelling	3.0

Table 2a: Level of agreement with statements concerning clinical signs. *indicating that preset threshol*d for consensus was met. ₸ indicating that the preset threshold for high variability were met.*

Statements	Percentages of agreement (Fully and mostly agreed)
Obvious implant migration on X-ray or CT	100.0*
Progressive radiolucencies on AP and lateral X-ray or CT	84.8*
Subsidence on X-ray or CT	93.9*
Radiolucency on AP and lateral X-ray or CT > 2mm around implant or bone-cement interface	90.9*
Continuous fluid film around the implant on MRI	51.5
Activity around the implant on PET-CT > 2 years after implantation	36.4 ₸
Cyst formation on X-ray or CT	48.5 ₸
Cortical thickening at the tip of the stem (for revision arthroplasties)	48.5 ₸
Radiolucencies on AP and lateral X-ray or CT < 2mm around implant or bone-cement interface	27.3
Deformity on X-ray or CT	33.3 ₸
Activity around the implant on SPECT-CT > 2 years after implantation	51.5 ₸
Activity around the implant on bone scintigraphy > 2 years after implantation	36.4 ₸
Increased bone density on X-ray or CT	21.2
Activity around the implant on bone scintigraphy < 2 years after implantation	6.1
Activity around the implant on SPECT-CT < 2 years after implantation	6.1
Activity around the implant on PET-CT < 2 years after implantation	0.0

*Table 2b: Level of agreement with statements concerning radiological signs. *indicating that preset threshold for consensus was met. ₸ indicating that the preset threshold for high variability were met.*

Sub analysis based on country of practice

Experts from 13 countries participated, with 24.2% from The Netherlands and the remainder from various other countries. No significant differences were between Dutch and other participants for the criteria that met the consensus threshold (Table 3).

Statements	Number of participants that mostly and fully agreed (The Netherlands) (n=8)	Number of participants that mostly and fully agreed (Other countries) (n=25)	p-Value
Clinical			
Weight-bearing pain	7	20	n.s.
Radiological			
Obvious implant migration on X-ray or CT	8	25	n.s.
Progressive radiolucencies on AP and lateral X-ray or CT	5	23	n.s.
Subsidence on X-ray or CT	8	23	n.s.
Radiolucencies on AP and lateral X-ray or CT > 2mm around implant or bone-cement interface	6	24	n.s.

Table 3: Fully- and-mostly-agree scores for statements receiving consensus dichotomized into groups based on country of practice (The Netherlands vs. Other Countries).

Discussion

The most important finding of the present study was that consensus was achieved on several key preoperative clinical and radiological criteria for diagnosing aseptic loosening in knee arthroplasty. Notably, weight-bearing pain was identified as the clinical sign meeting the preset consensus threshold. Regarding radiological signs, consensus was reached on four specific criteria evident implant migration, progressive radiolucency, subsidence, and radiolucency exceeding 2 mm around the implant or bone-cement interface. Nevertheless, the study revealed considerable variance in expert views, with high variability observed in 55.6% of the clinical statements and 37.5% of the radiological statements. These results emphasize the diverse importance attributed to different clinical and radiological findings, yet they also underscore a general agreement among international experts on certain crucial criteria. Both the observed high variability rates and the criteria that met the consensus threshold should be considered when establishing diagnostic protocols.

Patients with aseptic loosening initially present with post-TKA knee pain [36, 38]. Usually, the standard approach encompasses taking of patient history (presence of weight-bearing pain, minimal pain at full range of motion), physical examination and conventional knee radiography. Nevertheless, the sensitivity and specificity of plain radiography often prove insufficient, particularly in instances of early and subtle yet clinically significant loosening [21, 35, 59].

Whilst these observations are corroborated by the experts in this Delphi consensus, what this study primarily reveals is the lack of consensus and high variability among experts regarding the importance they place on the outcomes of advanced diagnostic modalities such as hybrid Single-Photon Emission Computerized Tomography and CT (SPECT/CT), Magnetic Resonance Imaging and bone scintigraphy. This is somewhat in contrast to results of published articles indicating high sensitivity and specificity for SPECT/CT as a diagnostic tool for knee arthroplasty loosening in specialized centers [4-6]. The high variability and lack of consensus regarding these advanced modalities in the Delphi consensus study underscores the need for further prospective studies evaluating advanced diagnostic modalities with a greater number of patients, as limitations reported for these modalities leave room for diverse opinions and preferences.

This study has several limitations, and therefore its findings should be interpreted considering the following remarks. First, there is no consensus on the most optimal design of a Delphi consensus study. More rounds or an open discussion of results providing statements with nuances and explanations could potentially have increased or broadened the consensus. Nevertheless, this study was conducted based on pre-specified design based on generally accepted methodological criteria [17, 23]. Second, despite the first author's efforts to assemble a well-represented international panel, this was not fully achieved European (particularly Dutch) and American experts are overrepresented within the panel, but a sub analysis revealed no

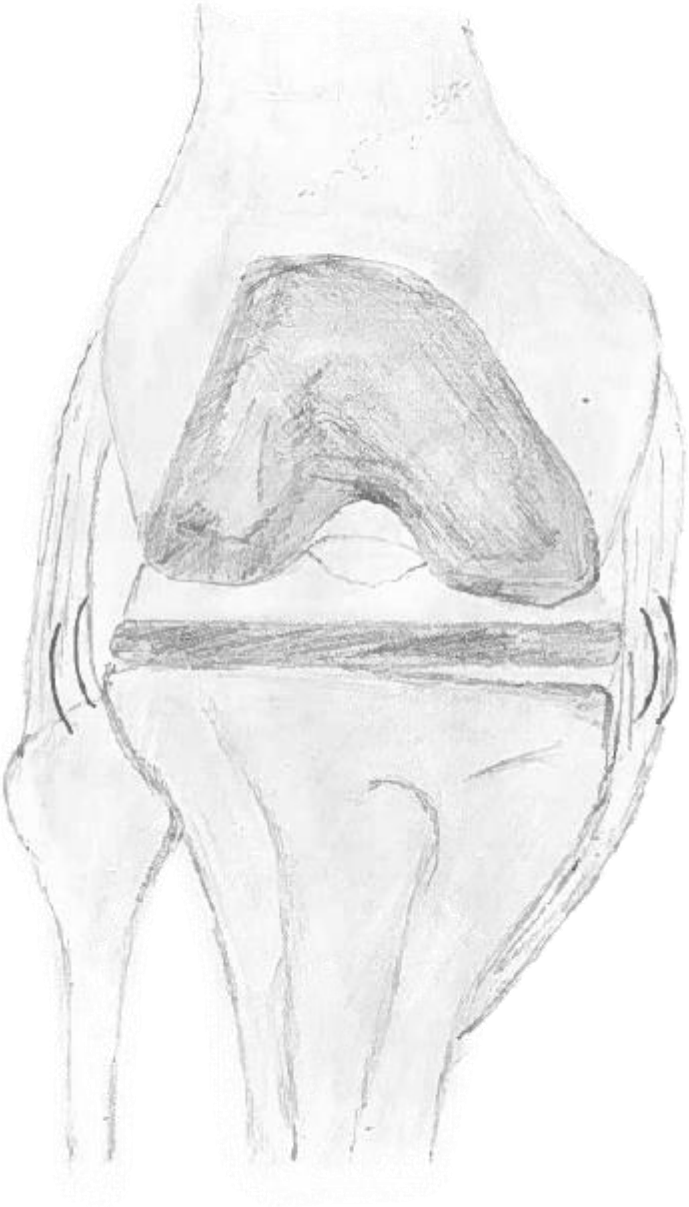
statistically significant differences for statements that received consensus. And lastly, only orthopedic surgeons were involved in this Delphi consensus. Although a definitive diagnosis is made by orthopedic surgeons, diagnosing loosening often requires a multidisciplinary approach. Involvement of radiologists and nuclear physicians may very well have altered the current results.

Conclusion

In conclusion, high variability exists among expert knee revision surgeons regarding the clinical and radiological standards deemed important for identifying loosening of knee arthroplasty components. Consensus was reached on weight-bearing pain as clinical criterion and on implant migration, progressive radiolucency, subsidence, and radiolucency > 2 mm on radiographs or CT as radiological criteria. Both the observed variability rates and the criteria that achieved consensus provide valuable insights to be considered when standardizing diagnostic protocols.

References

1. *Anzola LK, Hernandez N, Rodriguez LF, Sanguino G, Martinez E, Lopez R, et al. (2023) The role of SPECT/CT in painful, noninfected knees after knee arthroplasty: a systematic review and meta-analysis—a diagnostic test accuracy review. J Orthop Surg Res. 18:223. 10.1186/s13018-023-03687-8*
2. *Awengen R, Rasch H, Amsler F, Hirschmann MT (2016) Symptomatic versus asymptomatic knees after bilateral total knee arthroplasty: what is the difference in SPECT/CT? Eur J Nucl Med Mol Imaging 43:762-772*
3. *Barnsley L, Barnsley L (2019) Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiol 48:1565-1572*
4. *Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis. Knee Surg Sports Traumatol. 10.1002/ksa.12206*
5. *Dalury DF, Pomeroy DL, Gorab RS, Adams MJ (2013) Why are total knee arthroplasties being revised? J Arthroplasty 28:120-121*
6. *Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. (2014) Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 67:401-409*
7. *Hochman MG, Melenevsky YV, Metter DF, Roberts CC, Bencardino JT, Cassidy RC, et al. (2017) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty. J Am Coll Radiol 14:S421-S448*
8. *Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG (2017) Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. Palliat Med 31:684-706*
9. *Kallala RF, Ibrahim MS, Sarmah S, Haddad FS, Vanhegan IS (2015) Financial analysis of revision knee surgery based on NHS tariffs and hospital costs Does it pay to provide a revision service? Bone Joint J 97-B:197-201*
10. *Kassam, A., Dieppe, P, Toms, A (2012) An analysis of time and money spent on investigating painful total knee replacements. Orthop. Proc., 94-B(SUPP_XXXIX), pp. 234-234*
11. *Klug A, Gramlich Y, Rudert M, Drees P, Hoffmann R, Weißenberger M, et al. (2021) The projected volume of primary and revision total knee arthroplasty will place an immense burden on future health care systems over the next 30 years. Knee Surg Sports Traumatol. 29:3287-3298*
12. *Labek G, Thaler M, Janda W, Agreiter M, Stöckl B (2011) Revision rates after total joint replacement: Cumulative results from worldwide joint register datasets. J Bone Joint Surg 93 B:293-297*
13. *Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284. 10.1080/17453674.2021.1876998.*
14. *Mandalia V, Eyres K, Schranz P, Toms AD (2008) Evaluation of patients with a painful total knee replacement. J Bone Joint Surg Br 90:265-271*
15. *Mathis DT, Hauser A, Iordache E, Amsler F, Hirschmann MT (2021) Typical Pain Patterns in Unhappy Patients After Total Knee Arthroplasty. J Arthroplasty 36:1947-1957*
16. *Mathis DT, Hirschmann MT (2021) Why do knees after total knee arthroplasty fail in different parts of the world? J Orthop 23:52-59*
17. *Mathis DT, Tschudi S, Amsler F, Hauser A, Rasch H, Hirschmann MT (2022) Correlations of typical pain patterns with SPECT/CT findings in unhappy patients after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 30:3007-3023*
18. *Murer AM, Hirschmann MT, Amsler F, Rasch H, Huegli RW (2020) Bone SPECT/CT has excellent sensitivity and specificity for diagnosis of loosening and patellofemoral problems after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 28:1029-1035*
19. *Robertsson O, Ranstam J, Sundberg M, A WD, Lidgren L (2014) The Swedish Knee Arthroplasty Register: a review. Bone Joint Res 3:217-222*
20. *Röhner E, Heinecke M, Matziolis G (2021) Diagnostic algorithm in aseptic TKA failure - What is evidence-based? J Orthop 24:248-253*
21. *Walinga AB, Janssen SJ, Kievit AJ, de Borgie C, Kerkhoffs G (2024) Consensus on the definition and criteria for failure of surgical treatment in bacterial arthritis of a native joint: An international Delphi study. Knee Surg Sports Traumatol Arthrosc 32:235-242*
22. *Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454*



Chapter V

Visible fluid motion on manipulation as the new threshold for intraoperatively-determined knee arthroplasty component loosening – a Delphi study.

George S. Buijs, Arthur J. Kievit, Alex B. Walinga, Matthias U. Schafroth, Michael T. Hirschmann, Leendert Blankevoort

Knee Surg Sports Traumatol Arthrosc. 2024 Jul 15. doi: 10.1002/ksa.12357

Abstract

Purpose: There is a lack of a clear, uniform definition for intra-operatively assessed component loosening of a knee arthroplasty component, complicating the interpretation and interchangeability of results of diagnostic studies using an intraoperative observation as the reference test. The purpose of this study was to establish a consensus among specialized knee revision surgeons regarding the definition of intraoperatively determined loosening of total or unicondylar knee arthroplasty components.

Methods: Utilizing the Delphi consensus method, an international panel of highly specialized knee revision surgeons was invited to participate in a three-round process. The initiation of the first round involved the exploration of possible criteria for intraoperatively determined loosening with open questions. The second round focused on rating these criteria's importance on a five-point Likert scale. For the third round, criteria with a reached consensus were summarized in consecutive definitions for intraoperatively determined loosening and proposed to the panel. Consensus was established when over 70% of participants agreed with a definition for intraoperatively determined loosening.

Results: The 34 responding panel members described in total 60 different criteria in the first round of which 34 criteria received consensus in the second round. Summarizing these criteria resulted in four different definitions as minimal requirements for intraoperatively determined loosening. Eighty-eight percent of the panel members agreed on defining a component as loose if there is visible fluid motion at the interface observed during specific movements or when gently applying direct force.

Conclusion: This study successfully established a consensus using a Delphi method among knee revision surgeons on the definition of intraoperatively determined component loosening. By agreeing on the visibility of fluid motion as new definition, this study provides a standardized reference for future diagnostic research. This definition will enhance the interpretability and interchangeability of future diagnostic studies evaluating knee arthroplasty component loosening.

Introduction

Following Total Knee Arthroplasty (TKA) and Unicondylar Knee Arthroplasty (UKA), the likelihood for revision TKA (rTKA) within a decade is 13% for TKA and 12% for UKA [8, 16]. Given the rise in numbers of primary TKA and UKA, the incidence of revision surgeries is expected to rise, despite advancements in implant design and surgical techniques [14, 15, 17].

Aseptic loosening accounts for 20-30% of these rTKAs, making it the most prevalent reason for TKA failure [6, 20]. Aseptic TKA or UKA loosening often necessitates complex and demanding revision surgery, presenting substantial challenges to both patients and global healthcare infrastructure [11, 12]. Correct diagnosis is essential to avoid unnecessary revision surgery in patients misdiagnosed with loosening of the TKA, and to ensure that those with undetected loosening receive the required revision interventions. It also eliminates unnecessary delays in providing patients with the proper care based on an accurate diagnosis.

Efforts to refine the diagnostic process for aseptic loosening have led to a multitude of research exploring different diagnostic approaches and modalities [1-3, 13, 21, 23]. These studies often reported conflicting outcomes, contributing to a considerable degree of inconsistency and variability in the importance and value attributed to the results of various available diagnostic modalities [3, 4]. These conflicting results stem largely from a high risk of bias concerning the reference test employed in these studies. In the absence of tools to quantify implant motion, most diagnostic studies employ intraoperative visual assessment of component loosening as the main reference test [1, 3, 5]. A recent systematic review assessed fourteen diagnostic studies evaluating modalities utilized to assist in diagnosing aseptic knee arthroplasty component loosening [5]. Only three of the included studies clearly defined the method by which intraoperative assessment of component loosening was dichotomized as either loose or fixed. Among these, two disparate criteria were identified. One study employed the ability to remove a component with one hand as threshold for component loosening as opposed to the other study were the ability to toggle the implant was deemed sufficient to conclude loosening [5, 9, 18]. The divergence in definitions for loosening across these studies may have contributed to disparate results when applied interchangeably. This supposition is assisted by the high variability in the confidence attributed to various available diagnostic modalities, as reported in a Delphi consensus study resulting from the same Delphi project as the present study [4].

Consequently, the orthopedic community continues to encounter difficulties in formulating a uniform diagnostic standard for aseptic loosening, crucial for improving treatment selection and enhancing outcomes for patients [25].

Amidst these uncertainties, no study has been conducted to evaluate the variation of opinions and to define consensus regarding the definition of an intraoperatively determined loose TKA or

UKA component amongst knee revision specialists. There is need for a clearer clinical guideline regarding aseptic loosening to avoid both overtreatment and under treatment. In absence of any consensus based on literature data, a Delphi study might help to arrive at a consensus on a threshold for component loosening for future use in diagnostic accuracy test studies evaluating diagnostic modalities for aseptic knee arthroplasty loosening. This study aims to answer the following two research questions using a Delphi Consensus method:

(1) Is there consensus amongst highly specialized knee revision surgeons regarding the definition of intraoperatively determined loosening of TKA or UKA components?

(2) What is the extent of variability amongst highly specialized knee revision surgeons for when a TKA or UKA component should be considered loose when intraoperatively determined?

It was hypothesized that consensus could be reached, while high variability would be observed.

Material and Methods

The methods employed in the present Delphi consensus article, are similar to those employed for a recent Delphi consensus article evaluating variability and consensus between diagnostic modalities used to diagnose aseptic knee arthroplasty loosening [4]. For this joint Delphi consensus project, the reporting guidelines and methodological standards as recommended by Jünger et al. and Diamond et al. were adopted and an international consensus panel of highly specialized knee revision surgeons was established [7, 10]. The current Delphi study unfolded over three different rounds. The first round focused on identifying all opinions on criteria for intraoperative loosening according to panel members. The second round involved rating the importance of these identified criteria. In the case uniform consensus has not been achieved by the end of the second round, a last third round was prepared to reassess informed by a summary of the second round's results. This study focused exclusively on the loosening of tibial and femoral components.

The lead author (GSB) with help of a co-author (ABW) took on the role of coordinator, formulating the questionnaires based on feedback from participants and overseeing communications. To eliminate the risk of moderator bias affecting the study's results, GSB and ABW abstained from joining the study as a panelist. Although co-authors AJK, MUS, and MTH were panel members, they were excluded from the response processing and analysis phase to ensure impartiality.

Assembling the international consensus panel

Three different methods were used to identify orthopedic surgeons with expertise in knee arthroplasty revision: (1) by reviewing the authorship of high-quality articles focused on

diagnosing aseptic loosening in knee arthroplasty, (2) by examining the programs of orthopedic conferences to identify keynote speakers discussing or presenting on topics related to knee arthroplasty and loosening, and (3) via the professional networks of the co-authors. This strategy was implemented to guarantee a panel of participants that was varied, international and representative.

Currently, no definitive guidelines or established benchmarks for the optimal number of participants in Delphi studies are available. While some researchers argue that a cohort of 10-15 members might be adequate for more uniform groups, a larger sample is often recommended when a larger variety is expected [22, 24]. For this Delphi consensus study, an a priori minimum of 25 panel members was required. In anticipation of potential non-responders, 69 candidates were identified and invited as potential panel members. Those who agreed to complete participation were sent the Delphi consensus questionnaires.

Baseline

Firstly, participants were requested to disclose their years of experience in knee revision surgery and the annual average number of knee arthroplasty revision surgeries. Additionally, information regarding their gender and the country of their current practice were collected. Participants also had the option to indicate if they wished to be recognized for participation as a panel member. Such acknowledgment as a group author was contingent upon a participant completing all rounds of the consensus process.

First round

To gain a better understanding of the different opinions and standpoints towards the issue of intraoperative assessment of component loosening between the participants, the following two questions were asked:

(1) *To what extent do you agree that there is no clear uniform definition in available literature for what is to be defined as loose and fixed in intraoperative assessment (during revision surgery) of knee arthroplasty component loosening?*

(2) T*o what extent do you agree that there is a need for a uniform definition of what is to be defined as loose and fixed in intraoperative assessment (during revision surgery) of knee arthroplasty component loosening?*

Both questions were Single Select Multiple Choice Questions using a 5-point Likert scale (1 being *I fully agree*, 2: *I mostly agree*, 3 *neutral*, 4; *I do not agree completely*, 5; *I do not agree at all*).

The first round identified criteria associated with intraoperative assessment of loosening of both the tibial and femoral component of a TKA, UKA, or rTKA, separately. Therefore, two questions were posed:

(1) *How would you personally during revision knee arthroplasty define loosening of the **tibia** component? (Please answer separately for partial, primary, and revision surgery)*
(2) *How would you personally during revision knee arthroplasty define loosening of the **femur** component? (Please answer separately for partial, primary, and revision surgery)*

Second round and third round

Ahead of the second round, all statements were collected, and any duplicates were removed. All statements were incorporated in the second-round survey *as:*

*"'statement' should result in the judgement that the **tibial** component of a (either partial, primary or revision) knee arthroplasty is loose".*

and as:

*"'statement' should result in the judgement that the **femoral** component of a (either partial, primary or revision) knee arthroplasty is loose".*

Since no unique differences were noted in the first round, all statements were deemed relevant to UKA, TKA, and/or rTKA in the second round. Panel members were then requested to express their level of agreement with these statements using a 5-point Likert scale.

The outcomes of the second round were compiled into frequency tables and shared with all participants. High variability regarding a particular statement was defined as the statement having received at least 1 score of complete disagreement and at least 1 score of complete agreement, both must apply.

Statements meeting the consensus level (> 70% fully agree or mostly agree) were summarized into different and distinctive definitions for a loose component and presented as consecutive minimal requirements ranked based on the degree of severity of looseness and thus expected mobility of the component. Participants were asked to point out their minimal definition for a component to be considered as loose. It was made clear that by pointing out the minimal definition, all definitions that testify to a larger degree of expected mobility were also interpreted as fitting the definition of a loose component.

Consensus was established when over 70% of participants deemed a proposed definition as fitting the definition of a loose component.

Data collection and analysis

Baseline demographic information was reported using mean and standard deviation (SD) or median and interquartile range (IQR), depending on the distribution of the data. The categorical variable 'consensus' was depicted in terms of absolute numbers and percentages. Participants who did not complete the first round were disqualified from subsequent rounds and their data were not included in the analysis. Data were analyzed using Excel 16.0 (Microsoft Excel 2016; Microsoft Corp., Redmond, WA, USA). Questionnaires were distributed using Castor EDC (2023.4.0.0; Amsterdam, The Netherlands).

Results

Panel characteristics

A total of 69 eligible panel members were contacted by email and 38 (55.1%) agreed to participate in this Delphi consensus study, scheduled from September 2023 to December 2023. All but one (n=37; 97.4%) who agreed to participate completed the first round. Three participants (8.1%) failed to complete the second round. All participants (100%) included in the second round completed the third round. Therefore, all rounds were completed by a total of 34 (89.5%) of the 38 originally included participants (Figure 1).

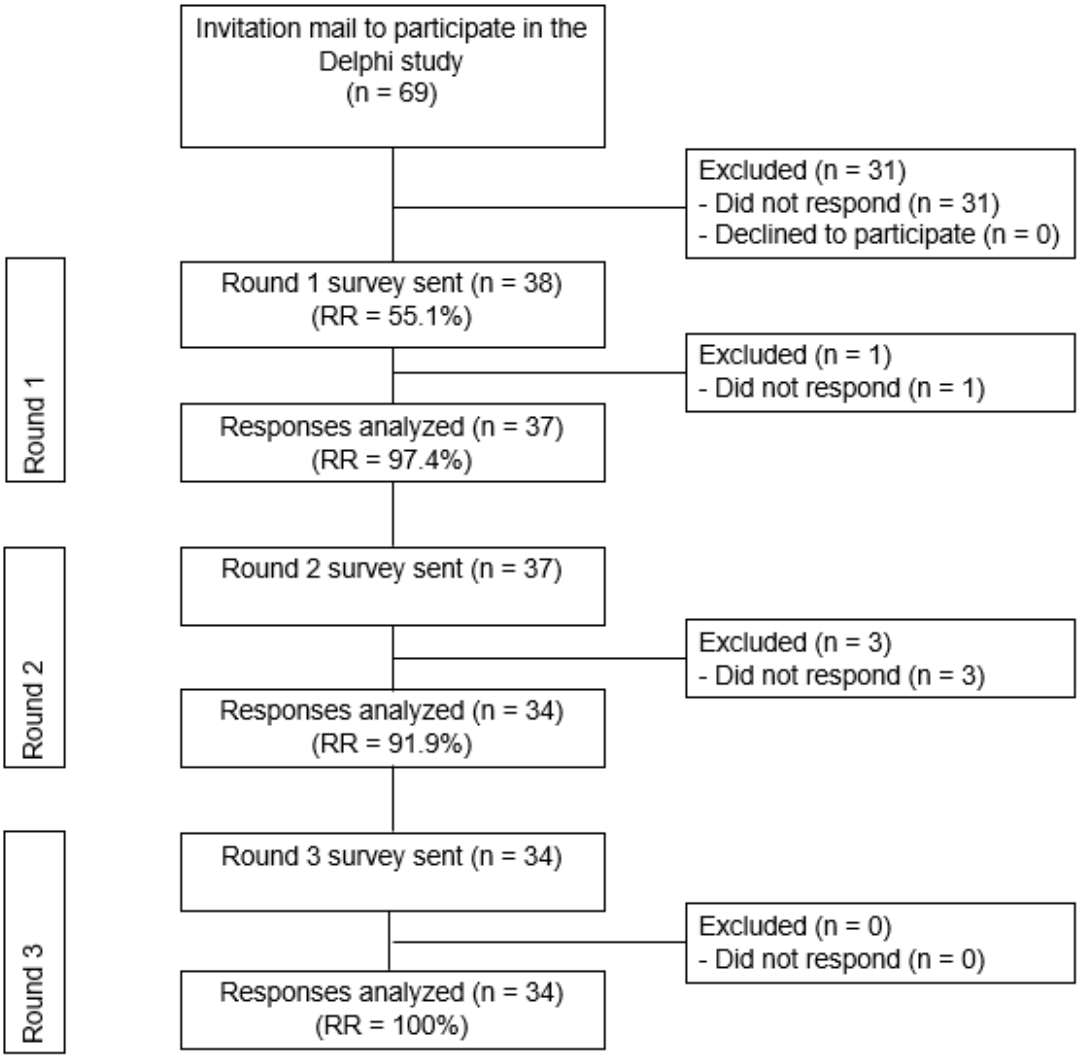


Figure 1: Flowchart of inclusion of panel members and response rates (RR) per round.

The Dutch panel members formed the predominant group, comprising 26.5% (n=9) of all participants (Table 1). Most participants were male (n=32; 94.1%; Table 1). Among the 34 participants, 31 (91.2%) conducted more than 20 knee arthroplasty revision surgeries annually. Most participants (n=28; 82.4%) had more than 10 years of experience as knee arthroplasty revision surgeons (Table 1).

Characteristics	n	Percentages (%)
Gender		
Male	32	94.1
Female	2	5.9
Country / region of practice		
The Netherlands	9	26.5
United States of America	4	11.8
Germany	4	11.8
United Kingdom	4	11.8
Belgium	3	8.8
Switzerland	2	5.8
France	2	5.8
India	2	5.8
Indonesia	1	2.9
Peru	1	2.9
Türkiye	1	2.9
Spain	1	2.9
Experience as knee arthroplasty revision surgeon (years)		
0 - 5 years	2	5.9
5 - 10 years	4	11.8
10 - 15 years	12	35.2
20 - 25 years	9	26.4
> 25 years	7	20.6
Average knee arthroplasty revision surgeries per year		
0 - 10	1	2.9
10 - 20	2	5.9
20 - 30	9	26.5
30 - 40	4	11.8
40 - 50	6	17.7
> 50	12	35.3

Table 1: Baseline characteristics of participants completing all rounds of this Delphi consensus study (n = 34).

First round

Thirty-seven panel members completed the first round. Between these 37 members, 31 (83.8%) agreed fully or mostly with the statement that there is no clear uniform definition of knee arthroplasty component loosening when determined intraoperatively. The majority (n=34; 91.9%) either fully or mostly agreed with the statement that there is a need for a clear uniform definition for intraoperatively determined knee arthroplasty component loosening.

All members in the first round reported a total of 28 distinctive statements regarding the UKA tibial component, 29 statements concerning the TKA tibial component, and 32 statements concerning the rTKA tibial component. Regarding the femoral component, 27 statements concerning UKA, 28 statements concerning primary TKA and 31 statements concerning rTKA were proposed. Distinctions between cemented and uncemented components were made by the participants. Very few technically relevant differences were found between statements regarding either UKA, primary TKA, or rTKA. Therefore, discrimination between different types of knee arthroplasties was not further applied, unless specified otherwise.

Second round

Thirty-four participants completed the second round. In preparation for the second round, all statements were summarized as separate statements for either the tibial or femoral component. This resulted in 31 unique statements regarding the tibial component and 29 unique statements regarding the femoral components, of which “Pedestal formation around the tip of the stem”, and “Perforation of a stem through the cortex”, were solely applicable to suspected loosening of an rTKA component. Ten proposed statements were applicable to cemented components. Six proposed statements were applicable to uncemented components. Forty-four statements were applicable to both cemented and uncemented components. The applicability of a statement to cemented components, uncemented components or both was specified in the summary.

A total of 34 statements received > 70% consensus, of whom 18 applicable to the tibial component and 16 applicable to the femoral component (Table 2a/b).

High variability was observed in 22 (64.7%) statements, of whom 9 applicable to tibial components and 13 applicable to femoral components (Table 2a/b). Between those 22 statements, 8 statements were identical between statements concerning tibial and femoral components.

Between the statements that received consensus, high variability was observed in 3 statements applicable to the tibial component and 2 statements to the femoral component. For the tibial component, these were: “*Both: Macroscopic mobility of the implant with applied force by "punch" of hammer”*, *“Both: Visible fluid motion beneath component when gently lifting”*, and *“Cemented: No cement fixation on component”*. For the femoral components, these were: *“Cemented: No cement fixation on component”*, and *“Cementless: Lack of osteointegration of the component”*.

Statement	Percentages of agreement (Fully and mostly agreed)
Cemented: Complete detachment of bone-cement interface or cement-prosthesis interface, resulting in easy (without force/osteotome) removal of prosthesis with clamp	100*
Uncemented: Complete detachment of prosthesis-bone interface with interposition of fibrous tissue, resulting in easy (without force/osteotome) removal of prosthesis with clamp	100*
Both: lifting of the anterior portion of the prosthesis in deep flexion	97.1*
Both: Macroscopic mobility of the implant during varus or valgus stress	94.1*
Both: Macroscopic mobility of the implant with applied force by finger	94.1*
Both: Macroscopic mobility of the implant during flexion or extension	91.2*
Both: Macroscopic mobility of the implant independent of method force was applied	91.2* ∓
Both: Macroscopic mobility of the implant when pushing implant to bone after single impaction	91.2*
Both: Migrated component	91.2*
Both: Macroscopic mobility of the implant when tested with probe or osteotome	88.2*
Both: Visible fluid motion beneath component when gently lifting	88.2* ∓
Cemented: Visual subsidence of component	88.2*
Both: Ability to tap-off the component with minimal effort	82.4*
Cemented: Visual failure of bone-cement interface	79.4*
Both: Knibbling away anterior bony overgrowth, then stress the component in 10 degrees of flexion with varus and valgus stress and watch for fluid or bubbles going in and out the interface region	73.5*
Both: Macroscopic mobility of the implant with applied force by "punch" of hammer	73.5* ∓
Cemented: No cement fixation on component	70.6* ∓
Uncemented: Lack of osteointegration of the component	70.6*
Both: Release of the component from bone or cement after few strokes with osteotome	64.7
Both: Easy to extract	58.8
Both: Fibrous tissue between implant-bone interface or implant-cement interface	55.9
Uncemented: Clear gap between component and bone	55.9
Both: Easy removal with Chisel	52.9
Both: Clear area of granulation at interface of bone	50.0
Both: presence of a depressed plateau	44.1
Both: After removal; presence of zones of bone resorption around the placement of the component	38.2 ∓
Both: Pedestal formation around the tip of the stem (for revision only)	38.2 ∓
Both: Perforation of a stem through the cortex (for revision only)	38.2 ∓
Both: Mispositioned component	11.8 ∓
Cemented: Visual cement particles	11.8 ∓
Both: Intraosseous antibiotic penetrating directly into the joint space from a tibial injection	2.9

*Table 2a: Level of agreement with statements concerning tibial components. Both: statement applies to both cemented and uncemented components. *indicating that preset threshold for consensus was met. ∓ indicating that the preset threshold for high variability were met.*

Statement	Percentages of agreement (Fully and mostly agreed)
Cemented: Complete detachment of bone-cement interface or cement-prosthesis interface, resulting in easy (without force/osteotome) removal of prosthesis with clamp	100*
Uncemented: Complete detachment of prosthesis-bone interface with interposition of fibrous tissue, resulting in easy (without force/osteotome) removal of prosthesis with clamp	100*
Both: Macroscopic mobility of the implant when tested with probe or osteotome	94.1*
Both: Macroscopic mobility of the implant during flexion or extension	94.1*
Both: Macroscopic mobility of the implant with applied force by finger	94.1*
Both: Macroscopic mobility of the implant independent of method force was applied	91.2*
Both: Macroscopic mobility of the implant during varus or valgus stress	91.2*
Both: lifting of the anterior portion of the prosthesis in deep flexion	91.2*
Both: Migrated component	91.2*
Cemented: Visual subsidence of component	85.3*
Both: Ability to tap-off the component with minimal effort	85.3*
Both: Macroscopic mobility of the implant when pushing implant to bone after single impaction	82.4*
Cemented: Visual failure of bone-cement interface	82.4*
Both: Macroscopic mobility of the implant with applied force by "punch" of hammer	79.4*
Uncemented: Lack of osteointegration of the component	76.5* ₸
Cemented: No cement fixation on component	73.5* ₸
Both: Visible fluid motion beneath component when gently lifting	67.6 ₸
Both: Knibbling away anterior bony overgrowth, then stress the component in 10 degrees of flexion with varus and valgus stress and watch for fluid or bubbles going in and out the interface region	64.7
Both: Release of the component from bone or cement after few strokes with osteotome	61.8
Both: Easy removal with Chisel	61.8 ₸
Both: Easy to extract	58.8 ₸
Both: Clear area of granulation at interface of bone	50.0 ₸
Uncemented: Clear gap between component and bone	50.0 ₸
Both: Perforation of a stem through the cortex (revision only)	47.1 ₸
Both: After removal; presence of zones of bone resorption around the placement of the component	44,1 ₸
Both: Pedestal formation around the tip of the stem (revision only)	41,2 ₸
Both: Mispositioned component	20,6 ₸
Cemented: Visual cement particles	17.6 ₸
Both: Intraosseous antibiotic penetrating directly into the joint space from a tibial injection	11.8 ₸

*Table 2b: Level of agreement with statements concerning femoral components. Both: statement applies to both cemented and uncemented components. *indicating that preset threshold for consensus was met. ₸ indicating that the preset threshold for high variability were met.*

Third round

Thirty-four participants completed the third round. In preparation for the third round, all 34 statements receiving consensus were summarized and duplicates between those applicable to both tibial and femoral components were removed. As very few technically relevant differences were found between statements regarding tibial or femoral components, discrimination between the two different components was not further applied. This resulted in the four consecutive definitions of component loosening that were presented together with the following explanation (Table 3).

"Based on the results of the second round, four statements are formulated, ranked by the degree of severity of the looseness and expected mobility of the component. In your opinion, which of these four statements is the minimum requirement for a component to be considered loose?"

With agreement of 88.2% of the panel members, the preset level of agreement was met for the following definition for intraoperatively determined loosening;

> *"A tibial and/or femoral component is considered loose if there is visible fluid motion at the interface (without macroscopic mobility of the implant) observed during specific movements such as flexion, extension, varus or valgus stress, or when gently applying direct force with a finger or instrument (e.g., probe or osteotome). This applies to both cemented implants (bone-cement or cement-prosthesis interface) and uncemented implants (prosthesis-bone)".*

Results of the third round are shown in Table 3.

#	Definition	Count of times pointed out	Cumulative agreement	Agreement (%)
1	A tibial and/or femoral component is considered loose if there is complete detachment at the interface, which is indicated by the easy removal of the prosthesis without the need for additional force or tools. This applies to both cemented implants (bone-cement or cement-prosthesis interface) and uncemented implants (prosthesis-bone interface).	4	4	11.8
2	A tibial and/or femoral component is considered loose if there is macroscopic mobility of the implant observed during specific movements such as flexion, extension, varus or valgus stress, or when applying direct gentle force with a finger or instrument (e.g. probe or osteotome). This applies to both cemented implants (bone-cement or cement-prosthesis interface) and uncemented implants (prosthesis-bone).	11	15	44.1
3	A tibial and/or femoral component is considered loose if there is visible fluid motion at the interface (without macroscopic mobility of the implant) observed during specific movements such as flexion, extension, varus or valgus stress, or when gently applying direct force with a finger or instrument (e.g. probe or osteotome). This applies to both cemented implants (bone-cement or cement-prosthesis interface) and uncemented implants (prosthesis-bone).	15	30	88.2
4	A tibial and/or femoral component is considered loose if there is macroscopic mobility of the implant after single impaction (e.g. with a hammer). This applies to both cemented implants (bone-cement or cement-prosthesis interface) and uncemented implants (prosthesis-bone).	4	34	100

Table 3: Levels of agreement with proposed definitions for intraoperatively determined loosening, with scores, cumulative scores and percentages (%) of cumulative scores for each proposed definition of intraoperatively determined loosening.

Discussion

The most important finding of the present study was the consensus on the visibility of fluid motion at the interface between the TKA component and/or cement and the bone observed during specific movements, such as flexion, extension, varus, or valgus stress, or when gently applying direct force with a finger or instrument, as the definition for intraoperatively determined knee arthroplasty component loosening. This threshold should be applied when using intraoperatively determined loosening as a reference test in future diagnostic accuracy test studies. This study also revealed considerable variance in expert views, with high variability observed in 64.7% of the statements posed in the second round (Table 2a/b).

These results emphasize the variability in standards applied when testing for TKA or UKA component loosening intraoperatively, yet they also underscore a general agreement among international specialized knee revision surgeons. This resulted in a new definition for intraoperatively determined TKA or UKA component loosening. This new definition will standardize the use of intraoperatively testing of component loosening as reference test in future diagnostic research. Given the crucial role of accurate diagnosis, this new definition will help reduce incomparability between the reported diagnostic accuracy of modalities used to aid the diagnosis of aseptic knee arthroplasty loosening and therefore will help reduce both unnecessary and uncommitted but necessary revision surgeries.

This study is the first to evaluate variability and establish consensus for intraoperatively determined knee arthroplasty component loosening. The need for consensus was identified after conducting a systematic review and meta-analysis. This systematic review evaluated fourteen studies on diagnostic methods for identifying aseptic loosening in knee arthroplasty components. Only three studies provided a clear methodology for distinguishing between loose and fixed components during surgery, revealing inconsistencies with two differing criteria identified. According to Mayer-Wagner et al. a component was considered fixed only if it proved irremovable with one hand during revision surgery [18]. In contrast, studies by Murer et al. and Hirschmann et al. stated that the potential of toggling the implant after standard approach (including synovial debridement and removal of osteophytes) should lead to the determination that the component is loose [9, 19].

This study is subject to several limitations, and as such, the interpretation of its findings should consider the following observations. First, it is important to note the absence of a universally agreed-upon framework for conducting Delphi consensus studies. Incorporating additional rounds or facilitating an open dialogue to discuss results, and providing nuanced statements and explanations, may have enhanced, or expanded the consensus. However, this study proceeded according to a predefined design that aligns with widely accepted methodological standards [7, 10]. Second, despite the lead author's attempts to create a diverse international panel, the

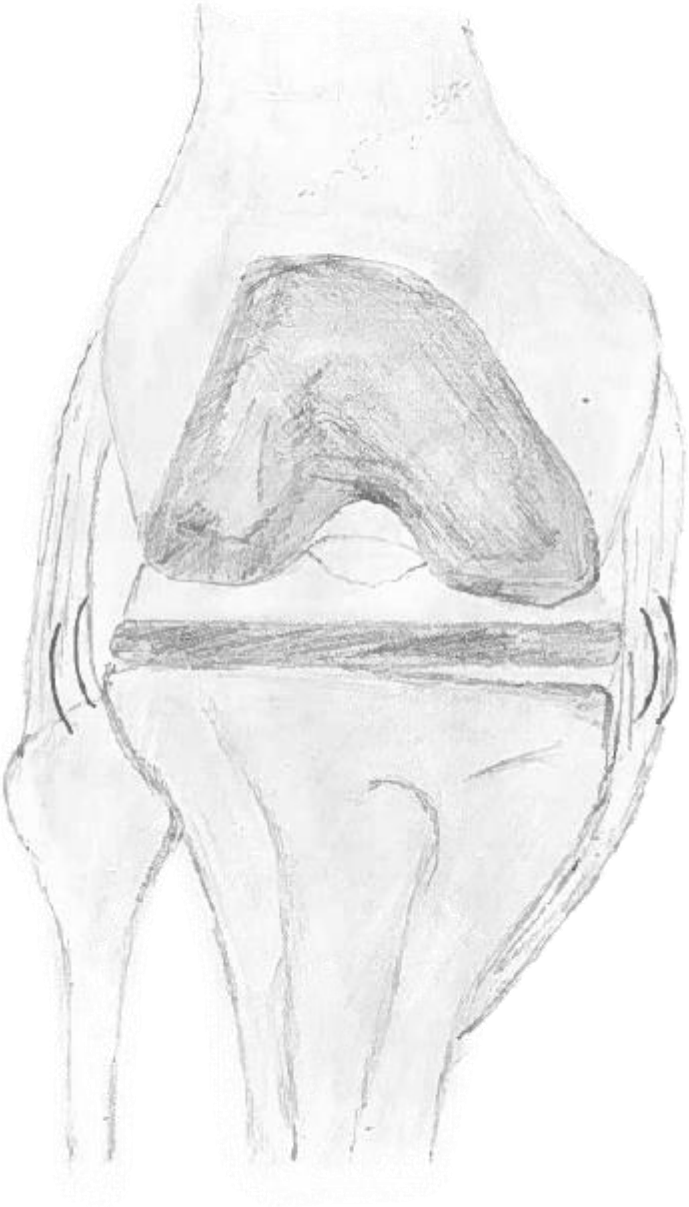
representation was not entirely comprehensive. The panel exhibited an overrepresentation of European (notably Dutch) and American members. This definition should be used solely as threshold for intraoperative observations as a reference test in diagnostic test studies. The clinical decision to revise a component should be based on a comprehensive assessment of the patient's condition, including clinical symptoms, imaging findings, and other diagnostic criteria.

Conclusion

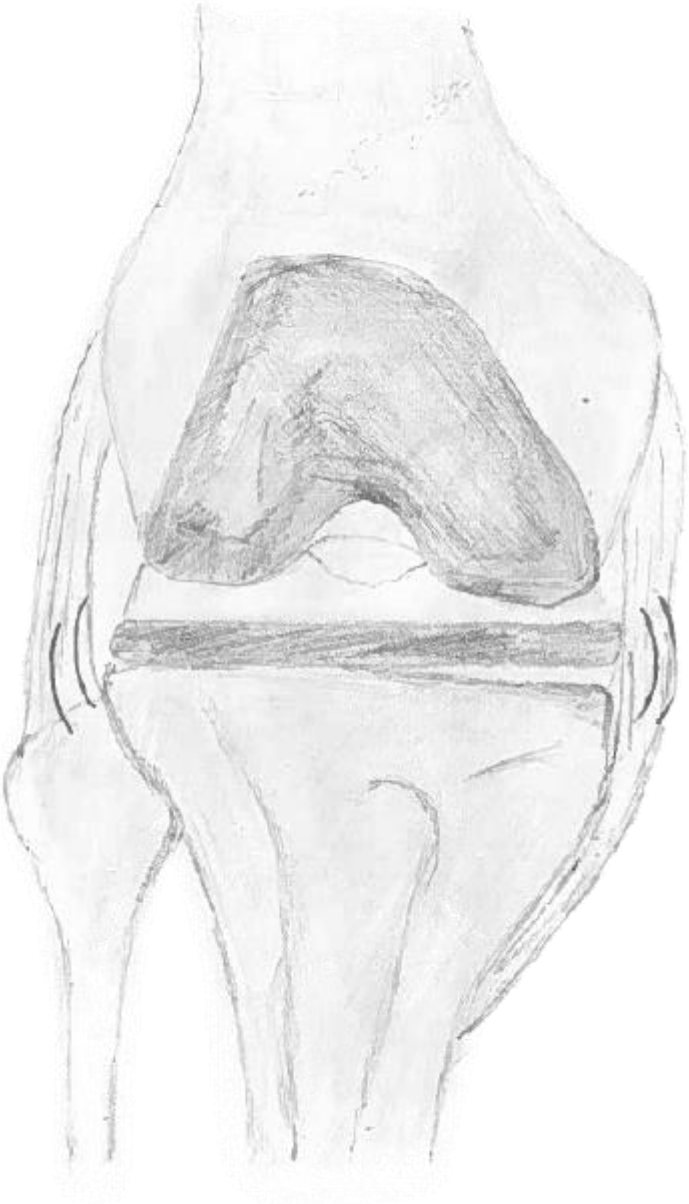
There is high variability in factors contributing to the determination that a TKA or UKA component should be judged as loose, yet using this Delphi method consensus was reached on the visibility of fluid motion at the interface between the TKA component and/or cement and the bone observed during specific movements, such as flexion, extension, varus or valgus stress, or when gently applying direct force with a finger or instrument, as definition for intraoperatively determined knee arthroplasty component loosening. This new definition should be used as a threshold diagnostic test studies were intraoperatively determining of component loosening is employed as reference test.

References

1. *Anzola LK, Hernandez N, Rodriguez LF, Sanguino G, Martinez E, Lopez R, et al. (2023) The role of SPECT/CT in painful, noninfected knees after knee arthroplasty: a systematic review and meta-analysis—a diagnostic test accuracy review. J. Orthop. Surg. Res 18:223*
2. *Awengen R, Rasch H, Amsler F, Hirschmann MT (2016) Symptomatic versus asymptomatic knees after bilateral total knee arthroplasty: what is the difference in SPECT/CT? Eur J Nucl Med Mol Imaging 43:762-772*
3. *Barnsley L, Barnsley L (2019) Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiol 48:1565-1572*
4. *Buijs GS, Kievit AJ, Schafroth MU, Hirschmann MT, Blankevoort L (2024). Weight-bearing pain and implant migration progressive radiolucencies, radiolucencies (> 2 mm) and subsidence on X-ray and CT are generally accepted criteria fitting knee arthroplasty loosening - An international Delphi consensus study. Knee Surg Sports Traumatol Arthrosc.*
5. *Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L (2024). MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis. Knee Surg Sports Traumatol Arthrosc. 10.1002/ksa.12206.*
6. *Dalury DF, Pomeroy DL, Gorab RS, Adams MJ (2013) Why are total knee arthroplasties being revised? J Arthroplasty 28:120-121*
7. *Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. (2014) Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 67:401-409*
8. *Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284*
9. *Hirschmann MT, Amsler F, Rasch H (2015) Clinical value of SPECT/CT in the painful total knee arthroplasty (TKA): a prospective study in a consecutive series of 100 TKA. Eur J Nucl Med Mol Imaging 42:1869-1882*
10. *Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG (2017) Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. Palliat Med 31:684-706*
11. *Kallala RF, Ibrahim MS, Sarmah S, Haddad FS, Vanhegan IS (2015) Financial analysis of revision knee surgery based on NHS tariffs and hospital costs Does it pay to provide a revision service? Bone Joint J 97-B:197-201*
12. *Kassam AM, Dieppe P, Toms AD (2012) An analysis of time and money spent on investigating painful total knee replacements. B J Med Pract 2012;5:a526.*
13. *Kievit AJ, Buijs GS, Dobbe JGG, Ter Wee A, Kerkhoffs G, Streekstra GJ, et al. (2023) Promising results of an non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon) 104:105930*
14. *Klug A, Gramlich Y, Rudert M, Drees P, Hoffmann R, Weißenberger M, et al. (2021) The projected volume of primary and revision total knee arthroplasty will place an immense burden on future health care systems over the next 30 years. Knee Surg. Sports Traumatol. 29:3287-3298*
15. *Labek G, Thaler M, Janda W, Agreiter M, Stöckl B (2011) Revision rates after total joint replacement: Cumulative results from worldwide joint register datasets. J bone Joint Surg. 93 B:293-297*
16. *LROI. Cumulative revision percentage of unicondylar knee arthroplasties by type of revision in the Netherlands in 2007-2022 (n=48,029). 2022; https://www.lroi-report.nl/knee/Survival-UKA/Survival-UKA/. Accessed 13-04-2024, 2024.*
17. *Mathis DT, Hirschmann MT (2021) Why do knees after total knee arthroplasty fail in different parts of the world? J Orthop 23:52-59*
18. *Mayer-Wagner S, Mayer W, Maegerlein S, Linke R, Jansson V, Müller PE (2010) Use of 18F-FDG-PET in the diagnosis of endoprosthetic loosening of knee and hip implants. Arch. Orthop. Trauma Surg. 130:1231-1238*
19. *Murer AM, Hirschmann MT, Amsler F, Rasch H, Huegli RW (2020) Bone SPECT/CT has excellent sensitivity and specificity for diagnosis of loosening and patellofemoral problems after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 28:1029-1035*
20. *Robertsson O, Ranstam J, Sundberg M, A WD, Lidgren L (2014) The Swedish Knee Arthroplasty Register: a review. Bone Joint Res 3:217-222*
21. *Röhner E, Heinecke M, Matziolis G (2021) Diagnostic algorithm in aseptic TKA failure - What is evidence-based? J Orthop 24:248-253*
22. *Hsu C, Standford B (2019). The Delphi Technique: Making Sense of Consensus. Pract. Assess. Res. Eval. 12:10*
23. *Ter Wee MA, Dobbe JGG, Buijs GS, Kievit AJ, Schafroth MU, Maas M, et al. (2023) Load-induced deformation of the tibia and its effect on implant loosening detection. Sci Rep 13:21769*
24. *Walinga AB, Janssen SJ, Kievit AJ, de Borgie C, Kerkhoffs G (2024) Consensus on the definition and criteria for failure of surgical treatment in bacterial arthritis of a native joint: An international Delphi study. Knee Surg Sports Traumatol Arthrosc 32:235-242*
25. *Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454*



Part II: Evaluating induced implant movement.



Chapter VI

Promising results of a non-invasive measurement of knee implant loosening using a loading device, CT-scans, and 3D image analysis.

Arthur J. Kievit*, George S. Buijs*, Johannes G.G. Dobbe, Maaike A. Ter Wee, Gino M.M.J. Kerkhoffs, Geert J. Streekstra, Matthias U. Schafroth, Leendert Blankevoort

** Both authors contributed equally.*

Clin Biomech (Bristol, Avon). 2023 Apr:104:105930. doi:10.1016/j.clinbiomech.2023.105930.

Abstract

Background: After total knee arthroplasty up to 13% requires revision surgery to address loosening. No current diagnostic modalities have a sensitivity or specificity higher than 70-80% to detect loosening, leading to 20-30% of patients undergoing unnecessary, risky, and expensive revision surgery. A reliable imaging modality is required to diagnose loosening. This study presents a new and non-invasive method and evaluates its reproducibility and reliability in a cadaveric study.

Methods: Ten cadaveric specimens were implanted with a loosely fitted tibial components and CT scanned under load towards valgus and varus using a loading device. Advanced three-dimensional imaging software was used to quantify displacement. Subsequently, the implants were fixed to the bone and scanned to determine the differences between the fixed and the loose state. Reproducibility errors were quantified using a frozen specimen in which displacement was absent.

Findings: Reproducibility errors, expressed as mean target registration error, screw-axis rotation and maximum total point motion were 0.073 mm (SD 0.033), 0.129 degrees (SD 0.039) and 0.116 mm (SD 0.031), respectively. In the loose condition, all displacements and rotation changes were larger than the reported reproducibility errors. Comparing the mean target registration error, screw axis rotation and maximum total point motion in the loose condition to the fixed condition resulted in mean differences of 0.463 mm (SD 0.279; $p = 0.001$), 1.769 degrees (SD 0.868; $p < 0.001$) and 1.339 mm (SD 0.712; $p < 0.001$), respectively.

Interpretation: The results of this cadaveric study show that this non-invasive method is reproducible and reliable for detection of displacement differences between fixed and loose tibial components.

Introduction

Total Knee Arthroplasty (TKA) is highly effective in treating pain caused by rheumatoid arthritis or osteoarthritis of the knee [1]. Nevertheless, according to a study which combined the national implant registries of 6 different countries, within 10 years, up to 13% of patients will have undergone revision surgery [2]. In most of the cases the indication is loosening of the tibial component [2]. The main additional tests to aid the diagnosis of aseptic TKA loosening are conventional x-rays, Computed Tomography Scan (CT), white blood cell (WBC) scanning, 3-phase bone scintigraphy and the Positron Emission Tomography combined with CT (PET-CT) [3]. However, according to the American College of Radiology these imaging modalities (with a reported sensitivity and specificity of 70-80%) are insufficiently sensitive and specific and measure secondary and non-specific effects, such as increased bone turnover and osteoclastic activity [4]. These are effects that are seen with loosening but can also be caused by other physiological processes. Therefore, nuclear scanning may put an unsubstantiated burden on patients to merely indicate a suspicion of implant loosening [4].

Yet, if patients present with pain around the knee on ambulation and current imaging modalities raise a suspicion of loosening, the TKA is usually revised. However, 20-30% of the patients who undergo revision surgery for TKA loosening, do not actually need this surgery as the prosthesis appears to be fixed during revision surgery [4]. Furthermore, the same percentage of wrongly conservatively treated patients, would benefit from revision surgery with a correct diagnosis. Detecting actual displacement and rotation of the implant with respect to the bone may be a more reliable and direct approach to detect TKA loosening.

Roentgen stereo photogrammetric analysis (RSA) and model-based RSA are currently the only imaging techniques showing sufficient precision to quantify prosthetic micro-motion. Model-based RSA is a biplane x-ray technique that utilizes known Computer Aided Design (CAD) models of a prosthesis and quantifies its movement with respect to tantalum beads in the bone. These beads need to be implanted, which renders the method invasive. Inducible displacement tests have mostly been performed in small groups of patients with a fixed-knee prosthesis due to the experimental nature, invasiveness, and cost of the RSA technique [5-9].

Therefore, this study presents and evaluates a new non-invasive method to detect induced displacement of the tibial component for potential future clinical use. In this method the displacement of the tibial TKA component is induced by the application of a varus- and valgus load by use of a loading device. With each load a CT-scan is made of the knee. Advanced imaging analysis techniques are applied to process the 3D CT-scan images and to calculate displacement of the tibial component relative to the bone.

The hypothesis is that this non-invasive method can detect implant displacement and rotations with reproducibility and reliability similar to invasive methods. Therefore, the research questions are: 1. What is the reproducibility of the proposed method? 2. Is this method sufficiently reliable to detect displacement differences between fixed and loose TKAs in a laboratory setting using cadaver specimens?

Methods

A two-stage cadaveric study was developed with the aim to evaluate reproducibility and reliability of this new non-invasive method for detection of TKA loosening in a laboratory cadaveric setting. This method consists of a hardware component (loading device) and software component (advanced 3D image analysis of acquired CT-images made under valgus and varus loading, using the loading device.).

Hardware component

For the purpose of this study, a prototype loading device was developed to apply consecutive varus and valgus loading to the knee. This device applies a bending moment up to 20 Nm, similar to loading during walking, in 20 degrees of knee flexion to relieve posterior capsule tension.10 It consists of four contact points, two on the tibia and two on the femur resulting in a four-point bending mechanism. These contact points are connected with a stabilizing frame and equipped with a force application device and force measurement sensor (Figure 1a and 2). The setup ensures that a four-point bending is performed, and no force is applied directly to the prosthesis itself. A contact force between tibia and femur in a compartment, medial or lateral, combined with the tensile forces in the opposite capsule and collateral ligament balances the externally applied bending moment. The application of a bending moment in the frontal plane of the tibia is important for the reproducibility of the induced force and allows for variation of the positioning of the knee along the length of the leg. The prototype is mainly made of aluminum to guarantee a low level of image scattering. The prototype is shown in Figure 1a and 2.

Software component

The non-commercial custom-made 3-D image analysis software specifically developed for this study uses a three-step approach to visualize and quantify prosthesis displacement using CT-images: segmentation, registration and calculation and visualization. C++ programming language (Visual Studio 2013, Microsoft, Redmond, WA). The Qt toolkit [31] was used for GUI programming (Qt 4.8.6, The Qt Company, Espoo, Finland), the Visualization ToolKit [30] was used for 3D visualization (VTK 7.1.0, Kitware Inc., New York, NY), and the Insight ToolKit [26] for level-set segmentation (ITK 4.10.1, Kitware, Inc., Clifton Park, NY). The methods for segmentation and registration were performed in accordance with a protocol described by Dobbe et al [11].

a) b) c)

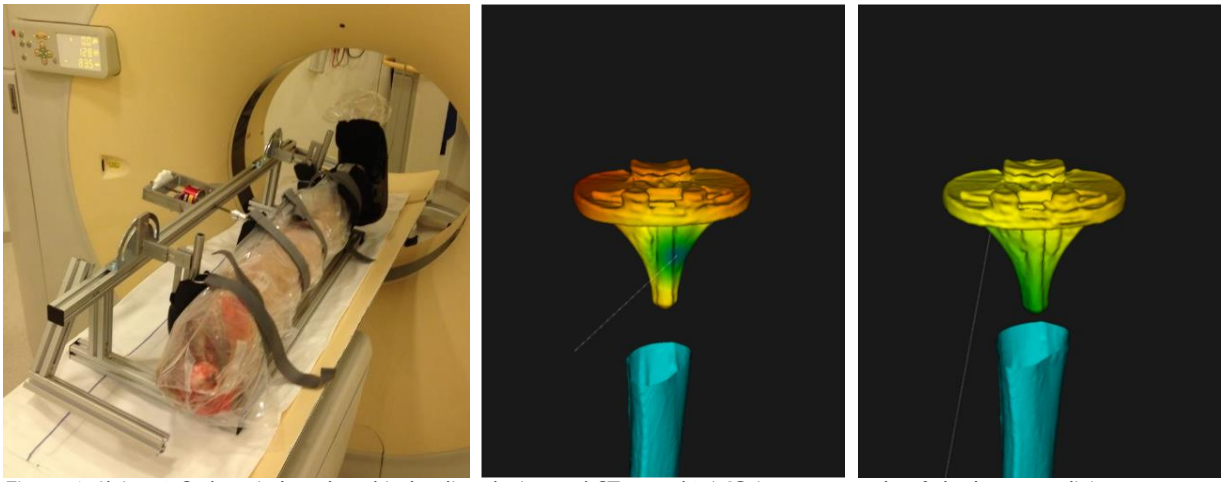


Figure 1a/b/c: a) Cadaveric leg placed in loading device and CT-scan, b) A 3D image example of the loose condition on a scale from 0.0 mm displacement (green) to 0.5 mm displacement (red), c) A 3D image example of the fixed condition with above described color gradations.

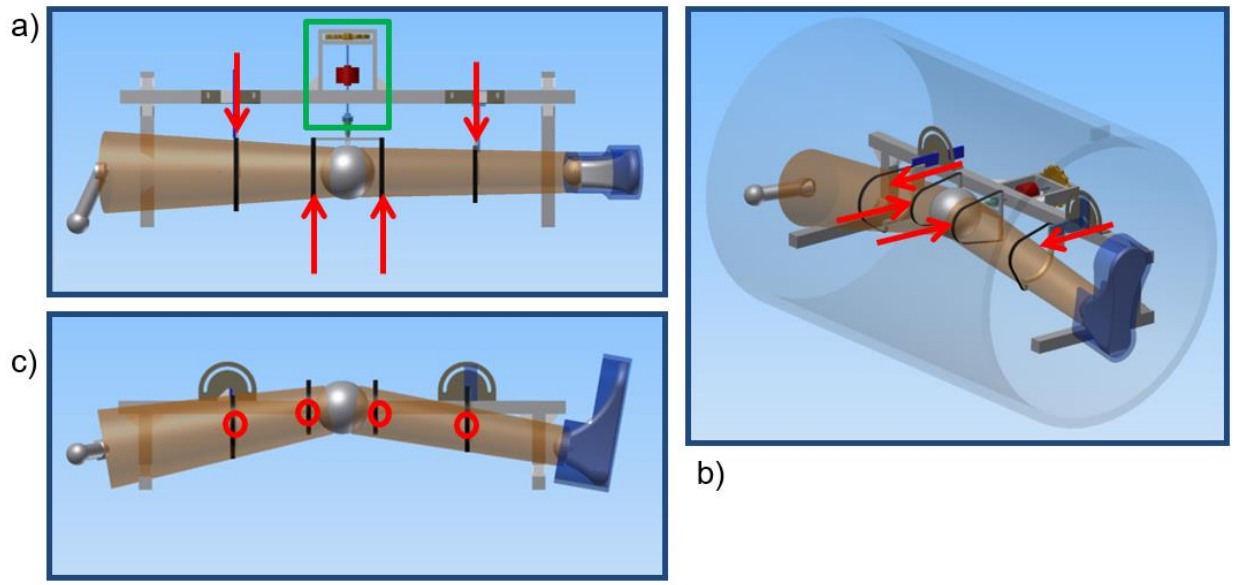


Figure 2a/b/c: schematic drawing of a leg in the loading device, with red arrow pointing out the four points where a bending moment of 20 Nm is applied and a load transducer (green box), measuring the applied moment.

Segmentation

The tibial implant and the tibia were segmented from the valgus CT-scan. Each object was first segmented using threshold-connected region growing. For the implant, a high threshold (2900 HU) was selected to manage metal artifacts as much as possible. For bone segmentation, the chosen threshold was approximately 300 HU. A binary closing algorithm subsequently filled residual holes inside the segmented object and at the surface. This intermediate segmentation result was used to initialize a Laplacian level-set segmentation growth algorithm, which adjusted the edges towards the highest intensity gradient of the implant and bone image.

Finally, a polygon was extracted at the zero-level using the marching cubes algorithm. The tibial implant causes metal artifacts in the reconstruction of the CT image. This hampers segmentation of the proximal segment of the tibia. For this reason, the proximal segment was removed by polygon clipping (Figure 3a, dotted blue tibia segment remains). The resulting polygons were used for 3-D visualization of the implant and the tibia, and for subsequent registration of both virtual objects with the same objects in the varus CT-scan. A visual inspection of the virtual objects was performed to ensure that a complete model of the tibial tray and tibial bone was created.

Registration

Intensity-based point-to-image registration was used for registration of the implant and the tibia to the valgus CT-scan. To this end, points were selected by sampling the gray-level CT image 0.3-mm towards the inside (bright voxels) and outside (dark voxels) of the segmented bone. This resulted in a double-contour polygon, which included the gray-levels at each vertex. Registration resulted in a transformation matrix, MT, describing rotation and translation, which brings the tibia polygon to the varus image (Figure 3a) and a second transformation matrix, MI, which aligns the tibia polygon with the varus image. These matrices were combined to find the loosening matrix, ML= MT-1 MI, which brings the virtual implant from the valgus to the varus position, within the frame of reference of the valgus CT-scan (Figure 3b).

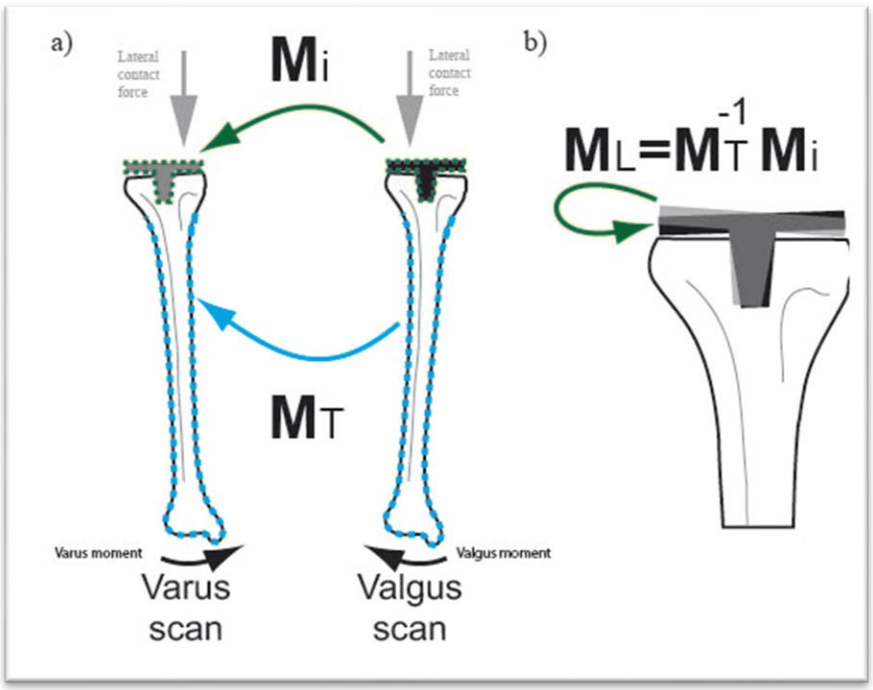


Figure 3a/b: Registration of the tibial bone resulted in a transformation matrix (blue dots; MT) and tibial component (green dots; MI), describing rotation and translation, which brings the tibia polygon to the varus image. These matrices were combined to find the displacement matrix, ML= MT-1 MI, which brings the virtual implant from the valgus to the varus position, within the frame of reference of the valgus scan.

Calculation and visualization

In case of a displaced tibial component, the implant position and orientation with respect to the tibia is different for the valgus and varus images. Implant displacement is quantified using the rotational change along the screw axis in degrees (rotation), the average point displacement of points in the implant mesh between the valgus and varus position (mean target registration error [mTRE]) and the maximum valgus-to-varus displacement of any point across the surface of the implant's polygon-mesh model (maximum total point motion [MTPM]). Calculated displacements are visualized in a heat map, with more reddish colors indicating a large displacement as opposed to more greenish colors indicating a small displacement (0.0 up to 0.5mm).

Experiments

First stage: Reproducibility

To evaluate reproducibility, a whole frozen cadaver leg was used and a TKA was performed on a whole frozen leg in accordance with the standard operative technique. Since the cadaveric leg was scanned in frozen condition, the absence of any motion between implant component and the bone can be assumed. Furthermore, fixation of the implant to the bone was ensured by visual verification after implantation. Therefore, any apparent implant displacement can be attributed to image noise, segmentation and/or registration errors. The leg was CT-scanned in ten slightly different (~ 5 degrees) orientations without application of any load.

Second stage: Reliability

To determine the reliability of the measurements of induced displacement of the tibial component in fixed tibial TKA components as opposed to loose tibial TKA components a second stage experiment was performed. Ten thawed, previously fresh-frozen cadaveric whole leg specimens were used and implanted with a TKA.

First, a loose implant was simulated in all ten legs by inserting an implant and moving it around slightly to simulate an area of bone resorption as found around loose implants. Looseness of the implant was ensured by visual verification. Looseness was defined as confirmed when manual induced movement of the tibial tray was visible after implantation. After which, all legs were scanned twice, first under varus loading and second under valgus loading. The loose condition was assessed first, as it would have been difficult to loosen the cement.

Second, all ten tibial loosely implanted components were fixed to the tibial bone using bone cement. The tibial component was removed, and both the bone and prosthesis were cleaned. Thereafter, bone cement was applied, and the implant was repositioned using pressurizing.

Definitive fixation was defined as the inability to manually induce visual displacement of the tibial tray. All specimens were then again scanned in the same consecution.

Statistics and materials

Results of the loose condition were compared to the fixed condition and statistically tested using a paired sample T-test, as the assumption was made that the data are normally distributed. The results of unloaded first experiment were compared to the results of the fixed condition of the second stage experiment. This comparison was also tested using a paired sample t-test under the same assumption.

Authors' decision to use ten cadaveric legs resulted from convenience sampling. Post-hoc sample size evaluations were performed using ClinCalc Post-hoc Power Calculator (Alpha: 0.05) [12]. The specimens had a median age of 82 years (min – max; 68 – 92; six males and four females). A p-value < 0.05 was considered statistically significant. Statistical analyses were performed using SPSS (IBM Corp. Released 2021. SPSS Statistics for Windows, Version 28, Armonk, NY: IBM Corp).

For all experimental evaluations, the vanguard TKA (Zimmer Biomet, Warsaw, Indiana, United States) was used.13 Palacos bone cement (Heraeus, Hanau, Germany) was used for cementation of the implants.14 All CT scans were made using a Brilliance 64-channel CT scanner (Philips Healthcare, Best, The Netherlands) (isotropic voxel spacing of 0.3 mm).

Ethical statements and source of funding

This study has been conducted following the recommendations of Committee of Ministers as stated in The Recommendation Rec (2006)4 [15,16]. This study was funded by an internal pre-seed grant from the Amsterdam University Medical Centers.

Results

Reproducibility

For reproducibility, the mean error in mTRE was 0.07 mm (SD 0.03 mm). The mean error for rotation was 0.13 degrees (SD 0.04 degrees). The mean error in MTPM was 0,12 mm (SD 0.03 mm) (Table 1, Supplementary Table 1).

Scan #	mTRE (mm)	Rotation (deg)	MTPM (mm)
1	0	0	0
2	0.08	0.09	0.11
3	0.05	0.09	0.10
4	0.05	0.09	0.08
5	0.08	0.20	0.15
6	0.11	0.15	0.13
7	0.04	0.13	0.08
8	0.05	0.15	0.10
9	0.07	0.16	0.13
10	0.14	0.12	0.17
Mean	**0.07**	**0.13**	**0.12**
SD	0.03	0.04	0.03

Table 1: Quantified and calculated values of the reproducibility experiment with all variables; rotation about the screw-axis (rotation), mean Target Registration Error (mTRE) and Maximum Total Point Motion (MTPM).

Reliability

Post-hoc sample size evaluations for mTRE, rotation and MTPM resulted in an estimated post-hoc power of 99.6%, 100% and 96.1% respectively.

In the fixed condition, the mean mTRE was 0.60 mm (SD 0.21 mm) compared to 1.06 mm (SD 0.33 mm) for the loose condition. The mean rotation was 0.67 degrees (SD 0.33 degrees) compared to 2.44 degrees (SD 0.97 degrees) when loose. The mean MTPM when fixed was 0.84 mm (SD 0,31 mm) compared to 2.18 mm (SD 0.86 mm) for the loose condition. (Table 2, Supplementary Table 2a/b). All displacements and rotational changes were larger than the measurement error reported for the reproducibility experiments. An example of the visualizations of the loose and fixed condition are shown in Figure 1b and 1c.

	Loose condition			Fixed condition		
Leg #	**mTRE (mm)**	**Rotation (deg)**	**MTPM (mm)**	**mTRE (mm)**	**Rotation (deg)**	**MTPM (mm)**
1	0.75	1.62	1.28	0.58	0.64	0.74
2	1.21	2.10	2.43	0.60	0.67	0.87
3	0.83	1.27	1.19	0.39	0.23	0.46
4	0.85	2.59	1.96	0.82	0.95	1.14
5	0.83	2.05	1.94	0.38	0.40	0.55
6	1.67	2.08	3.23	1.06	1.22	1.51
7	1.16	3.44	2.63	0.42	0.44	0.67
8	1.52	4.65	3.82	0.58	1.12	0.96
9	0.97	2.55	1.85	0.65	0.67	0.83
10	0.84	2.07	1.48	0.55	0.38	0.69
Mean	1.06	2.44	2.18	0.60	0.67	0.84
SD	0.33	0.97	0.86	0.21	0.33	0.31

Table 2: Calculated values of both the loose and fixed condition with the following variables: mean Target Registration Error (mTRE), rotation about the screw-axis (rotation) and Maximum Total Point Motion (MTPM).

The assumption that the data was normally distributed was considered satisfied as the skew and kurtosis levels were estimated as less than the maximum allowable values for a t-test (i.e. -3.0 > skewness < 3.0 and -10 > kurtosis < 10.0).17 (Table 3a).

	Skewness	**Std. Error**	**Kurtosis**	**Std. Error**
Reproducibility				
mTRE	1.11	0.72	0.69	1.40
Rotation	0.57	0.72	-0.07	1.40
MTPM	0.39	0.72	0.93	1.40
Reliability				
Loose				
mTRE	1.09	0.69	0.02	1.33
Rotation	1.42	0.69	2.31	1.33
MTPM	0.80	0.69	-0.07	1.33
Fixed				
mTRE	1.16	0.69	1.37	1.33
Rotation	0.50	0.69	-0.90	1.33
MTPM	1.14	0.69	1.54	1.33

Table 3a: Distribution of data presented with skewness and kurtosis. mean Target Registration Error (mTRE), rotation about the screw-axis (rotation) and Maximum Total Point Motion (MTPM).

Comparing the mTREs, rotations and MTPMs in the loose condition to the fixed condition resulted in mean differences of 0.46 mm (SD 0.28 mm; p=0.001), 1.77 degrees (SD 0.87 degrees; p<0.001) and 1.34 mm (SD 0.71 mm; p<0.001), respectively. Results of paired sample t-tests are shown in Table 3b and Figure 4.

	Paired Differences				
	Mean	SD	95% CI (Lower)	95% CI (Upper)	*p*-value (2-tailed)
Loose mTRE - Fixed mTRE (mm)	0.46	0.28	-0.26	0.66	0.001
Loose Rotation - Fixed Rotation (deg)	1.77	0.87	1.150	2.39	< 0.001
Loose MTPM - Fixed MTPM (mm)	1.34	0.71	0.83	1.85	< 0.001

Table 3b: Results of paired samples test for means of loose mTRE - Fixed mTRE, Loose rotation - Fixed rotation, Loose MTPM - fixed MTPM, with means, standard deviations (SD), 95% confidence intervals (CI) and p-values. mean Target Registration Error (mTRE), rotation about the screw-axis (rotation) and Maximum Total Point Motion (MTPM).

Comparison of the mTREs, rotations and MTPMs of the results of the non-loaded first stage to the loaded second stage fixed condition resulted in mean differences of 0.54 mm (SD 0.12 mm; p<0.001), 0.56 degrees (SD 0.33 degrees; p<0.001) and 0.74 mm (SD 0.31 mm; p<0.001).

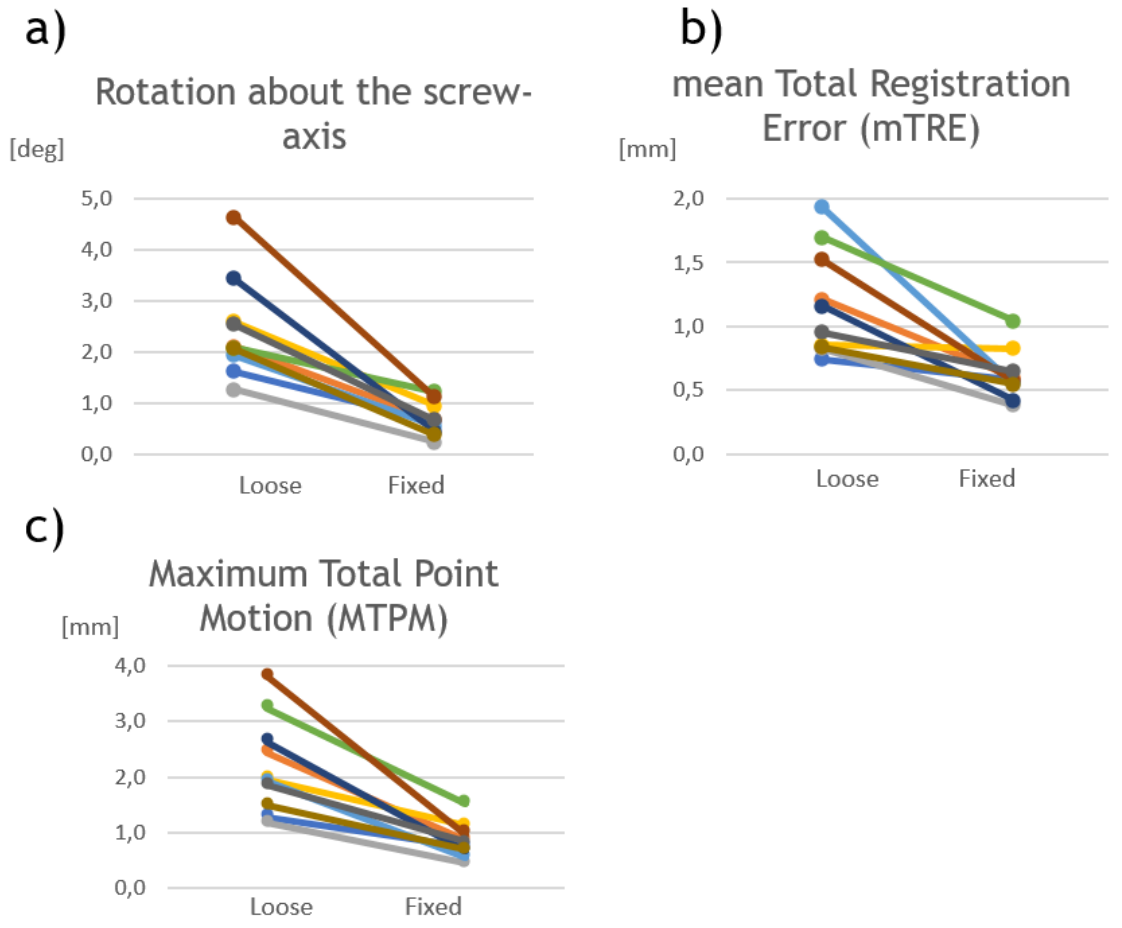


Figure 4a/b/c: Visualization of the individual changes in the rotation about the screw-axis, mTRE and MTPM for each cadaveric leg.

Discussion

The most important finding of this study is that this non-invasive method can significantly detect displacement differences between a loose implant compared to a fixed implant in a reproducible and reliable manner.

Mandalia et al. stated that pain after operation occurs in 1 in 8 patients despite an absence of clinical or radiological abnormalities [18]. Currently, findings from various imaging techniques are used to aid the diagnose of TKA loosening. These are mainly radiolucent lines on X-ray imaging and CT. Because of the low costs and fast processing, radiographs are usually the first diagnostic method, but there are some disadvantages to them. The intra- and inter-observer reliability is low, and the visibility of the radiolucent lines can be poor [19,20]. The reported sensitivity and specificity were 83% and 72% for detecting aseptic loosening of the tibial component [21]. Therefore, Mandalia et al. concluded that the clinical significance of radiolucent lines on x-ray imaging or CT is uncertain [18]. Despite reported low sensitivity and specificity, nuclear scans are one of the first diagnostic tools used by default after X-ray and CT. Nevertheless, nuclear scans measure osteoclastic activity and are therefore only useful after a minimum of one year after the last surgical procedure or else normal post-operative bone remodeling activity can be misinterpreted as signs of loosening [22].

Marker- and model-based RSA are considered the golden standard when it comes to quantifying implant migration, where the gradual migration of a prosthetic component in the bone over time can be measured. In this cadaveric study, the reported measurement error for the proposed method is similar to both RSA methods [23,24]. Theoretically, both RSA methods could be used to assess patients with complaints consistent with the diagnosis implant loosening overtime. However, a patient then will need to undergo surgery to implant the beads. This renders marker-based RSA useless as a non-invasive measurement in patients that were not previously evaluated using RSA. Furthermore, model-based RSA cannot be used for implants for which CAD models are not available or not supplied by the manufacturer. In the here proposed method, none of these disadvantages occur. No additional surgery is required, and the implant type and model need not to be known. This method can be performed in a non-invasive manner on any patient with complaints following TKA surgery, making this method potentially a replacement for current diagnostic methods like bone scanning and PET-CT scanning.

As reported in the results section, the fixed implants still show a displacement and angulation change, albeit being smaller than the loose condition. This may be caused the arbitrary visual confirmation of implant fixation after implantation, although performed similar to the intraoperative assessment of implantation fixation in revision surgery patients. Furthermore, these displacement changes may be caused by the suboptimal cementing technique used in this experimental setting, as the used cementing technique may have resulted in some interposition

of fatty residues between the cement and the tibial tray. With a good cementing technique, however, the differences would only have been greater between the loose and fixed condition. It is, therefore, more likely that these changes are due to elastic deformity of the tibial bone. This is supported by the statistically significant differences for the comparison the results of the non-loaded reproducibility experiment and results of the loaded fixed condition and additionally strengthened by the reported overlap in displacement measures between the loose condition with the fixed condition for different cadaveric legs. Due to these differences, together with the potential effects of the used cementing technique and the cadaveric design of this study, the absolute results of this study cannot be used as a reference for what is to be defined as a loose or a fixed implant. A clinical feasibility study, including both symptomatic- and asymptomatic patients, is needed to evaluate potential clinically significant thresholds.

Conclusions

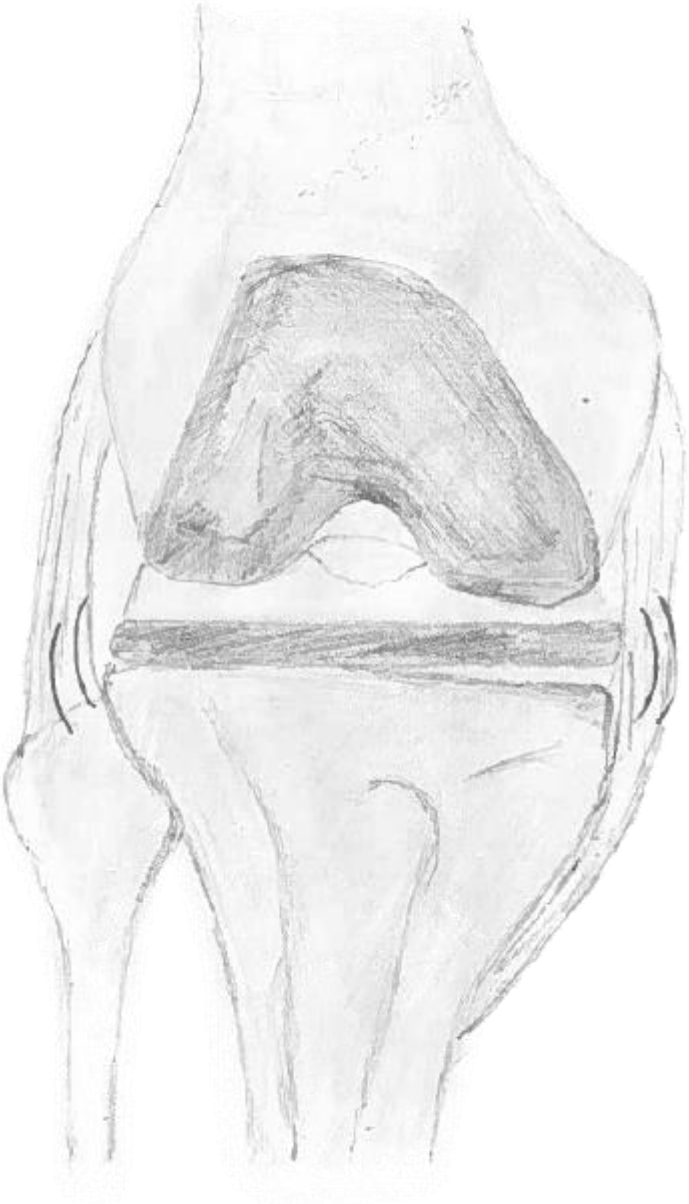
In conclusion, implant displacement can be measured in a reproducible and reliable manner similar as reported for invasive methods using this new method by a combination of induced displacement by a bending moment applied to the knee joint, CT-scans and a software algorithm with segmentation and registration.

References

1. *1.Carr AJ, Robertsson O, Graves S, et al. Knee replacement. Lancet. Apr 7 2012;379(9823):1331-40. doi:10.1016/S0140-6736(11)60752-6*
2. *Labek G, Thaler M, Janda W, Agreiter M, Stockl B. Revision rates after total joint replacement: cumulative results from worldwide joint register datasets. 93-B/3/293 pii ;10.1302/0301-620X.93B3.25467 doi. JBone Joint SurgBr. 2011;93(3):293-297. NOT IN FILE.*
3. *Mandalia V, Eyres K, Schranz P, Toms AD. Evaluation of patients with a painful total knee replacement. 90-B/3/265 pii ;10.1302/0301-620X.90B3.20140 doi. JBone Joint SurgBr. 2008;90(3):265-271. NOT IN FILE.*
4. *Hochman MG, Melenevsky YV, Metter DF, et al. ACR Appropriateness Criteria ® Imaging After Total Knee Arthroplasty. Journal of the American College of Radiology. 2017;14(11):S421-S448. doi:10.1016/j.jacr.2017.08.036*
5. *Lam Tin Cheung K, Lanting BA, McCalden RW, et al. Inducible displacement of cemented tibial components ten years after total knee arthroplasty. Bone Joint J. Feb 2018;100-b(2):170-175. doi:10.1302/0301-620x.100b2.Bjj-2017-0428.R2*
6. *Wojtowicz R, Henricson A, Nilsson KG, Crnalic S. Uncemented monoblock trabecular metal posterior stabilized high-flex total knee arthroplasty: similar pattern of migration to the cruciate-retaining design — a prospective radiostereometric analysis (RSA) and clinical evaluation of 40 patients (49 knees). Acta Orthopaedica. 2019;90(5):460-466. doi:10.1080/17453674.2019.1626097*
7. *Wilson DA, Astephen JL, Hennigar AW, Dunbar MJ. Inducible displacement of a trabecular metal tibial monoblock component. J Arthroplasty. Sep 2010;25(6):893-900. doi:10.1016/j.arth.2009.06.015*
8. *van Hamersveld KT, Marang-van de Mheen PJ, Tsonaka R, Valstar ER, Toksvig-Larsen S. Fixation and clinical outcome of uncemented peri-apatite-coated versus cemented total knee arthroplasty : five-year follow-up of a randomised controlled trial using radiostereometric analysis (RSA). Bone Joint J. Nov 2017;99-b(11):1467-1476. doi:10.1302/0301-620x.99b11.Bjj-2016-1347.R3*
9. *Pijls BG, Valstar ER, Kaptein BL, Fiocco M, Nelissen RG. The beneficial effect of hydroxyapatite lasts. Acta Orthopaedica. 2012;83(2):135-141. doi:10.3109/17453674.2012.665330*
10. *Zhang L, Liu G, Han B, et al. Knee Joint Biomechanics in Physiological Conditions and How Pathologies Can Affect It: A Systematic Review. Applied Bionics and Biomechanics. 2020;2020:1-22. doi:10.1155/2020/7451683*
11. *Dobbe JGG, Kievit AJ, Schafroth MU, Blankevoort L, Streekstra GJ. Evaluation of a CT-based technique to measure the transfer accuracy of a virtually planned osteotomy. Medical Engineering & Physics. 5-13-2014 2014;36(8):1081-1087. doi:http://dx.doi.org/10.1016/j.medengphy.2014.05.012*
12. *Post. KS. ClinCalc. Updated 11-10-2018. Accessed 01-11-2023, 2023. https://clincalc.com/stats/Power.aspx.*
13. *The clinical heritage of AGC® , Maxim® , and Ascent™ Total Knee Systems combined design features have allowed Zimmer Biomet to produce the Vanguard Complete Knee System1-3. Zimmer Biomet. Accessed 17-05-2022, 2022. https://www.zimmerbiomet.com/en/products-and-solutions/specialties/knee/vanguard-knee.html*
14. *PALACOS® bone cements. 17-05-2022, 2022. 2022. https://www.heraeus.com/en/hme/products_solutions_heraeus_medical/arthrosis/palacos_bone_cements.html*
15. *Pierannunzii L. Ethical requirements for musculoskeletal research involving human subjects. Journal of Orthopaedics and Traumatology. 2015;16(4):265-268. doi:10.1007/s10195-015-0371-x*
16. *Ministers Co. Recommendation Rec(2006)4 of the Committee of Ministers to member states on research on bio-logical materials of human origin. 2006.*
17. *Brown TA. Confirmatory Factory Analysis for Applied Research, second edition. . Second ed. The Guilford Press; 2006:462.*
18. *Mandalia V, Eyres K, Schranz P, Toms AD. Evaluation of patients with a painful total knee replacement. 90-B/3/265 pii ;10.1302/0301-620X.90B3.20140 doi. J Bone Joint SurgBr. 2008;90(3):265-271. NOT IN FILE.*
19. *Ecker ML, Lotke PA, Windsor RE, Cella JP. Long-term results after total condylar knee arthroplasty. Significance of radiolucent lines. Clin Orthop Relat Res. Mar 1987;(216):151-8.*
20. *Bach CM, Biedermann R, Goebel G, Mayer E, Rachbauer F. Reproducible assessment of radiolucent lines in total knee arthroplasty. Clin Orthop Relat Res. May 2005;(434):183-8. doi:10.1097/01.blo.0000153077.79573.a4*
21. *Marx A, Saxler G, Landgraeber S, Loer F, Holland-Letz T, von KM. Comparison of subtraction arthrography, radionuclide arthrography and conventional plain radiography to assess loosening of total knee arthroplasty. 10.1515/BMT.2005.021 doi. BiomedTech(Berl). 2005;50(5):143-147. NOT IN FILE.*
22. *Math KR, Zaidi SF, Petchprapa C, Harwin SF. Imaging of Total Knee Arthroplasty. Seminars in Musculoskeletal Radiology. 2006;10(01):047-063. doi:10.1055/s-2006-934216*
23. *Valstar ER, Vrooman HA, Toksvig-Larsen S, Ryd L, Nelissen RGHH. Digital automated RSA compared to manually operated RSA. Journal of Biomechanics. 2000/12/01/ 2000;33(12):1593-1599. doi:https://doi.org/10.1016/S0021-9290(00)00138-X*
24. *Kärrholm J. Roentgen stereophotogrammetry: Review of orthopedic applications. Acta Orthopaedica Scandinavica. 1989;60(4):491-503. doi:10.3109/17453678909149328*

Supplementary files

Please see https://doi.org/10.1016/j.clinbiomech.2023.105930



Chapter VII

Operator variation in applying a loading device for the evaluation of tibial component loosening.

George S. Buijs*, Maaike A. Ter Wee*, Chiel Klein, Lidwine B. Mokkink, Johannes G.G. Dobbe, Mario Maas, Matthias U. Schafroth, Geert J. Streekstra, Leendert Blankevoort, Arthur J. Kievit.

** Both authors contributed equally.*

Submitted for publication.

Abstract

Background: A CT-based method has been developed to aid diagnosis of aseptic loosening by quantifying tibial component displacement through the application of a bending moment to the knee using a loading device. This study evaluates the effect of operator differences in applying the loading device.

Methods: Sixteen subjects underwent repeated CT examinations with valgus and varus loading. Two operators applied the loading device to each patient. With each load, a CT scan was made, and tibial component displacement relative to the tibial bone was quantified as rotation about the screw-axis, Maximum Total Point Motion and mean Target Registration Error. Two protocols were used: (1) analyzing the entire tibia (100%) and (2) the proximal tibia (20%) to mitigate tibia deformation. Reliability and measurement error were assessed using intraclass correlation coefficient ($ICC_{agreement}$), standard error of measurement, standard error of operator and smallest detectable change.

Findings: The 100% tibia protocol showed moderate-to-good $ICC_{agreement}$, i.e. between 0.64-0.84 for the different displacement parameters, with standard error of measurement around 0.15 mm or degree. The 20% tibia protocol showed poor-to-moderate $ICC_{agreement}$, ranging from 0.17-0.31 for the different displacement parameters, with the standard error of measurement around 0.10 mm or degree. This protocol showed smaller measurement errors but poorer $ICC_{agreement}$ due to reduced subject variance explained by smaller apparent implant displacements. Operator related error was statistically and clinically negligible. The smallest detectable change values ranged 0.27-0.44 mm or degree.

Interpretation: The loading device can be operated by different trained operators with negligible inter-operator differences.

Introduction

Aseptic loosening of the tibial component is, after infection, the most common cause of failure in Total Knee Arthroplasty (TKA) [19]. This condition often requires extensive, expensive, and invasive revision TKA. There is absence of consensus in the current literature regarding a standardized diagnostic approach and lack of a definitive diagnostic test [9, 12, 13].

Recently developed CT-based methods that use external loading of the joint to offset relative displacement between the implant and adjacent bone provide a more direct and less cumbersome method, as an alternative to existing imaging modalities for aseptic knee or hip arthroplasty loosening, such as Magnetic Resonance Imaging (MRI), 3-phase bone scintigraphy, Positron Emission Tomography combined with Computed Tomography (PET-CT) and Single-Photon Emission CT (SPECT/CT) [13, 27, 59].

The CT-based solution introduced by Kievit et al. (2023) consists of a hardware and a software component and is currently branded as the 'AtMoves Knee system' (AtMoves BV, Amsterdam). The hardware component is a loading device used to exert a bending moment of up to 20 Nm to the knee in valgus and varus direction, in a four-point bending configuration. The knee is flexed at approximately 20 degrees, to alleviate stress on the posterior capsule and, if applicable, the posterior cruciate ligament. The applied bending moment is internally counterbalanced by forces: (1) a compressive force in either the medial compartment (for a varus moment) or the lateral compartment (for a valgus moment), and (2) forces in a collateral ligament, capsule, and cruciate ligament. The compressive force can cause the tibial component to displace relative to the bone. In each loading condition, a CT-scan is acquired.

The software component is an advanced image analysis software, utilized to quantify the relative displacement between the tibial component and the tibial bone between the varus and valgus loading direction. The software uses a three-step approach: segmentation of the tibial component and the tibial bone, registration, and calculation. Relative displacement between the implant and bone is expressed in three key displacement parameters: (1) rotation about the screw-axis, (2) the maximum point displacement within the implant model, known as Maximum Total Point Motion (MTPM) and (3) the average displacement of all points on the implant surface, referred to as mean Target Registration Error (mTRE) [27, 44]. All displacement parameters are absolute (unsigned) measures.

The findings from their pilot study, conducted in a cadaveric setting, showed reliable results in terms of precision (rotation about the screw-axis: 0.13°, MTPM: 0.12 mm and mTRE: 0.07 mm) and its ability to establish a statistically significant distinction between loose and fixed tibial components [27]. A blinded prospective diagnostic patient study evaluated the image analysis software and showed good to excellent intra- and interrater reliability of the analyses of the CT-

scans [10]. Furthermore, this study reported good diagnostic performance (sensitivity 0.91, specificity 0.70) with optimal thresholds of 0.52˚for rotation about the screw-axis, 0.42 mm for mTRE and 0.70 mm for MTPM to distinguish fixed from loose implants.

Ter Wee et al. (2023) proposed an improvement in the implant analysis protocol, highlighting that using the entire tibia (100%) as a reference object leads to an overestimation of implant displacement measurement caused by tibial bone deformation under applied valgus-varus loading from the loading device [55]. They proposed a solution to compensate for this deformation by only using the proximal tibial part (relative length of 20%) as the reference object for implant displacement.

From these studies, it remained unclear whether variation in placement and operation of the loading device introduces additional uncertainty to the measurement of tibial component displacement.

The current study addresses the following primary research question: What is the inter-operator variation, in terms of reliability and measurement error metrics, of applying the loading device for detecting implant loosening? Secondary, does the choice between 100% tibia versus 20% tibia as a reference object influence the inter-operator variation, and if so, to what extent?

We hypothesize good inter operator reliability, in terms of the practical guideline by Koo, Li [29], because of the loading device's four-point bending configuration that produces a uniform bending moment over the knee-joint and the applied load remains consistent regardless of operator differences.

Methods

Ethical statement

This study was conducted in compliance with the World Medical Association's Code of Ethics (Declaration of Helsinki) for human experimentation. Additionally, it adhered to the guidelines outlined in the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals [22]. This study (2023.1015) received approval from the institutional review board of the Amsterdam University Medical Center (Amsterdam UMC). Informed consent was obtained from all participants prior to their involvement. The prospective trial was registered with the trial registry of the Central Committee for Research with Human Subjects (CCMO; NL85030.018.23).

Subject screening and inclusion

The current study included a subgroup from the PILLAR study, a post-market clinical follow up study. The PILLAR study was conducted to assess the long-term survival of a specific Zimmer Biomet Vanguard femoral component. Subjects were included in the PILLAR study based on the following eligibility criteria: (1) implantation of the Vanguard PS Open Box Porous Femoral component, (2) the procedure took place between 2009 and 2013, (3) the surgery was conducted at Amsterdam UMC, location AMC and (4) the Vanguard PS Open Box Porous Femoral component was not used off-label.

Initially, 50 subjects meeting the in- and exclusion criteria were invited for physical and radiological examinations. From this study, a subset of 16 randomly selected subjects consented to a double CT examination under valgus and varus loading and were therefore included in the current study, with the purpose to evaluate the variation in implant displacement measurement as caused by application of the loading device (Figure 1).

Subject scanning

All 16 included subjects underwent two immediately consecutive evaluations with loading device by two different loading device operators in a single visit. The settings used for the CT scanner (Siemens Somatom Force) were 120 kVp, a tube current of 160 mAs, with isotropic voxel spacing of 0.45 mm.

During the first CT examination, operator 1 (GB) positioned the subjects on the CT scanner’s table and applied the loading device in the valgus direction (Figure 2a) with a target moment of 20 Nm. After the valgus CT scan was acquired, operator 1 reconfigured the loading device to the varus direction (Figure 2b) with the same target moment, and a second CT scan was acquired.

Upon completing these scans, the loading device was removed by operator 1, and the subject was given the opportunity to descend from the CT-scan's table and to move and stretch their legs. Immediately thereafter, operator 1 left the CT room and operator 2 (MW) came into the CT room and repeated the same procedure for the second CT examination. Both operators were blinded for the application of the loading device by the other operator. CT acquisition was performed by an independent technician.

A)

B)

Figure 1A/B: Picture of the knee with the loading device applying a 20 Nm load in the valgus direction (figure 2A) and the varus direction (figure 2B).

Upon completing these scans, the loading device was removed by operator 1, and the subject was given the opportunity to descend from the CT-scan's table and to move and stretch their legs. Immediately thereafter, operator 1 left the CT room and operator 2 (MW) came into the CT room and repeated the same procedure for the second CT examination. Both operators were blinded for the application of the loading device by the other operator. CT acquisition was performed by an independent technician.

Implant displacement measurement

The system's image analysis software was used to quantify and visualize displacement of the tibial component relative to the tibia. As a previous study had already proven excellent inter-rater agreement between different users for the image analysis [10], this step was performed by a single experienced operator (MW).

Image analysis was conducted using two different image-analysis protocols. The first protocol, proposed by Kievit & Buijs et al. (2023), involved segmentation of the complete tibia (100% tibia protocol) [27]. The second protocol, introduced by Ter Wee et al. (2023), employed segmentation of the proximal 20% of the tibia [55]. Segmentation of the tibial cortex was in both protocols performed at the level of the distal edge of the tibia-fibular joint. This dual approach was chosen to determine whether tibial deformation specifically contributes to within-patient variation in implant displacement measurements.

The tibial component and tibial cortex (20% and 100%, separately) were segmented using threshold-connected region growing from the valgus image, creating a polygon mesh model for visualization and registration purposes. Separate image registration of the implant and tibia models to the varus image allowed quantification of displacement relative to the tibia. Key parameters, (rotation about the screw-axis, MTPM and mTRE) were calculated and presented in a diagnostic report alongside visual representation of the tibial component and tibial bone. In this visualization, point displacements across the implant surface were represented in a heat map (Figure 3).

Variance estimates

The variation in implant displacement measurements between two operators applying the loading device and the measurement error were evaluated by repeating the whole system's workflow in all subjects. All components of the system's workflow were kept identical and constant except for the operator handling the loading device. Both operators were trained in the application of the loading device. The first operator was advanced in handling the loading device in a patient setting, the second operator was trained but lacked this experience.

Statistical analysis

The required sample size was calculated using the $ICC_{agreement}$ model with an expected correlation of 0.8 between measurements [39]. No systematic difference between the repeated measurements was anticipated (set to zero), with a target width of 0.4 for the 95% confidence interval of the ICC. To minimize radiation exposure, the number of repeated measurements was limited to two. This resulted in a required sample size of 10-20 subjects. Following this sample size calculation, but limiting the number of subjects exposed to radiation, the required sample size was set to 16 subjects.

The objective of the current study is to determine the reliability and measurement error of implant displacement detection, caused by different operators applying the loading device. The reliability expresses whether or to what extent the instrument can distinguish between situations of interest [40]. In the context of loosening detection, reliability reflects the proportion of total variance in the implant displacement measurement which can be attributed to true inter-patient differences, and not inter-operator differences. Reliability is quantified as the intraclass correlation coefficient (ICC) using a two-way effect model for absolute agreement in scores between operators [53]:

$$ICC_{agreement} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2 + \sigma_{pr,e}^2}, \qquad \text{(Eq. 1)}$$

which describes the ratio of the patient variance (σ_p^2), representing variation between subjects (inter-patient), to the total variance which also includes the inter-operator variance (σ_o^2), representing systematic difference between loading device operators, and the residual error ($\sigma_{pr,e}^2$), accounting for any unknown or random errors [40].

The measurement error represents the systematic and random errors in a score unrelated to changes in the construct being measured [41]. In the context of loosening detection, the measurement error reflects how similar the implant displacement measurements are when the loading device is handled by different operators. This is expressed as the standard error of measurement ($SEM_{agreement}$) in the same units as the displacement parameter:

$$SEM_{agreement} = \sqrt{\sigma_o^2 + \sigma_{pr,e}^2}. \qquad \text{(Eq. 2)}$$

To ease (clinical) interpretation of the extent of the operator variance in relation to the absolute measurement error, the standard error of operator (SEO) was introduced and quantified as,

$$SEO = \sqrt{\sigma_o^2}. \qquad \text{(Eq. 3)}$$

A Bland-Altman plot was employed to visualize the agreement between the two operators over the whole range of implant displacement measurements to highlight any potential proportional bias (larger implant displacement associated with larger inter-operator differences) and to identify potential outliers [8]. A one sample t-test was used to identify potential systematic bias between the operators, with $\alpha < 0.05$ as threshold of statistical significance.

To evaluate to what extent the observed can be trusted, the smallest detectable change is computed as

$$SDC = 1.96 * \sqrt{2} * SEM_{agreement} \ . \qquad \text{(Eq. 4)}$$

This value represents the smallest change in score that can be detected statically with a 95% certainty [56]. While it is similar to the limits of agreement shown in the Bland-Altman plot, it differs by accounting for the systematic change between repeated measurements (between operators), whereas limits of agreement correspond to a two-way mixed effects model for consistency [40].

The reliability and measurement error metrics were quantified for each analysis protocol (100% and 20% tibia segmentation) and for each key parameter, i.e. rotation about the screw-axis, MTPM and mTRE.

Results

Subject examination

A total of 11 females and 5 males were included. Mean age at time of CT scan acquisition was 59. Median follow-up time after primary surgery was 12 years (Table 1). For most of the subjects, the targeted moment of 20 Nm in both the valgus and varus configuration was achieved by both operators (Table 2). The diagnostic reports of one anonymized subject, analyzed with the 20% and 100% tibia analysis protocol, were reported as typical examples (Figure 3).

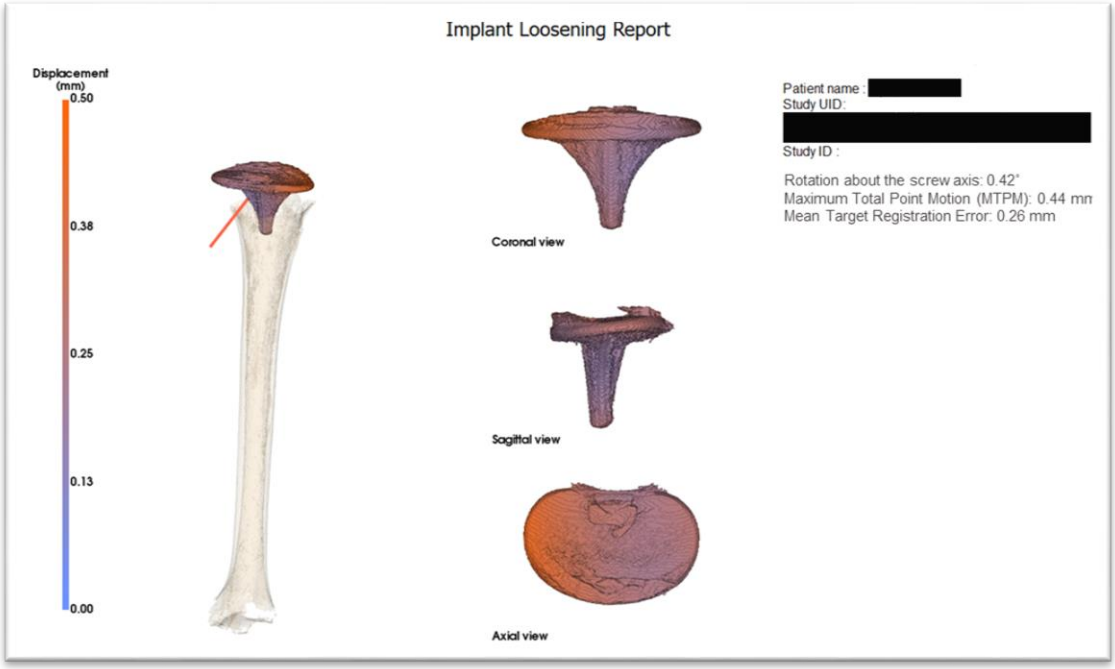
Characteristics	Subjects (n = 16)
Age (mean, SD)	59.1 (9.5)
Sex (n, %)	
Male	5 (31.3)
Female	11 (68.8)
Time interval between arthroplasty and AtMoves Scan (median in years, [IQR])	11.9 (11.4-12.5)
Charnley score (n, %)	
0	8 (50)
1	4 (25)
2	4 (25)
ASA score (n, %)	
1	10 (62.5)
2	6 (37.5)
KOOS sub scores (median, [IQR])	
KOOS-Pain	94.4 (83.3-100)
KOOS-Other Symptoms	85.7 (79.4-93.8)
KOOS-ADL	91.9 (86.4-96.3)
KOOS-Sport/Rec	45.0 (20.0-70.0)
KOOS-QoL	75.0 (60.9-90.6)

Table 1: Baseline characteristics of included subjects. n; Number, ASA; American Society of Anesthesiologists, KOOS; Knee Osteoarthritis Outcome Score, ADL; Activities of Daily Living, QoL; Quality of Life.

Subject	Operator I	Operator II
1	20/15	15*/15
2	20/20	20/20
3	20/20	20/20
4	20/20	20/20
5	20/20	20/20
6	15/20	20*/15*
7	20/20	20/20
8	10/10	10/10
9	20/20	20/20
10	20/20	20/20
11	20/20	20/20
12	20/20	20/20
13	20/20	20/20
14	20/20	20/20
15	20/20	20/20
16	20/20	20/20

Table 2: Applied load (Nm) for each subject for both the valgus and varus position. The asterisks (*) indicate deviating values between operator 1 and 2.

A)



B)

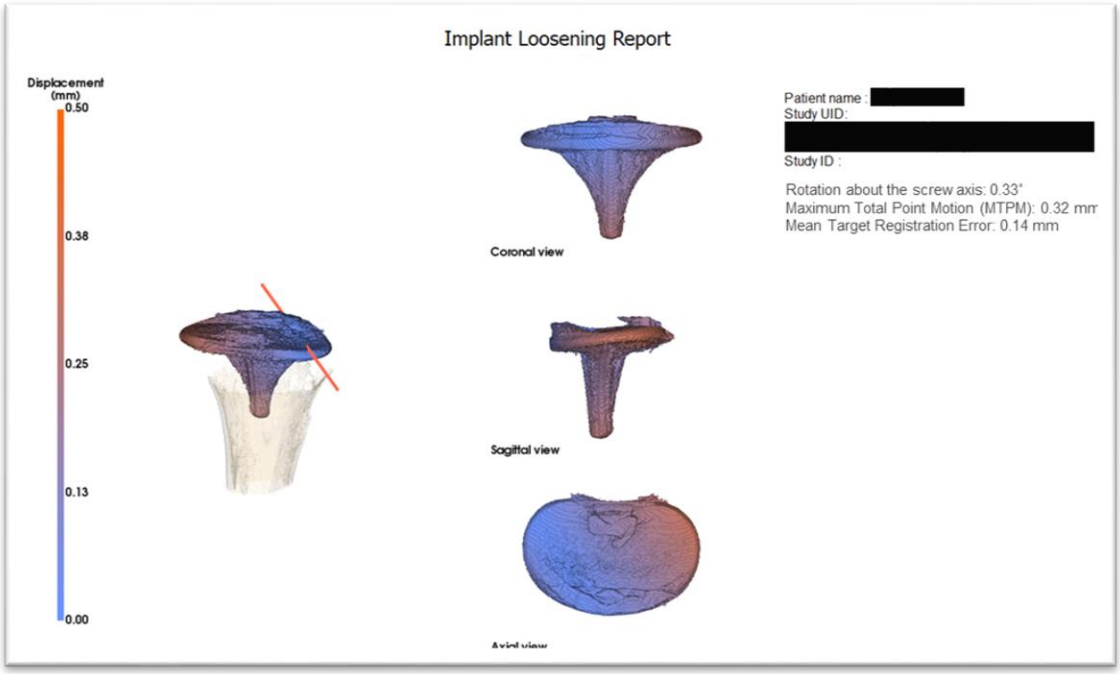


Figure 2A/B: Example of diagnostic report of one patient, analyzed using the 100% tibia (A) and 20% (B) as the reference object to quantify implant displacement. The heatmap visualizes the point displacement across the implant's surface ranging between 0.00 and 0.50 mm. The orange line represents screw-axis, along which rotation is quantified. On the right, the difference implant displacement parameters (rotation about the screw-axis, MTPM and mTRE) are reported.

Variance analysis

100% tibia analysis protocol

The $ICC_{agreement}$ was moderate (0.64 [CI: 0.24-0.86]) for rotation about the screw-axis. Implant displacement quantified as MTPM and mTRE resulted in good $ICC_{agreement}$ with values of 0.81 [95%CI: 0.55-0.93] and 0.84 [95%CI: 0.61 – 0.94], respectively (Table 3). The variance components showed that the operator induced systematic difference is close to zero, with an SEO of 0.06˚for rotation about the screw-axis, 0.05 mm for MTPM and 0.04 mm for mTRE. The SEM was 0.16˚ for rotation about the screw-axis, 0.15 mm for the MTPM and 0.10 mm for the mTRE.

Rotation about the screw-axis

$ICC_{agreement}$	L_ICC	U_ICC	σ_p^2	σ_r^2	$\sigma_{pr,e}^2$	SEM (˚)	SDC (˚)	SEO (˚)
0.6410	0.2425	0.8568	0.04402	0.0032	0.0215	0.1570	0.4352	0.0563

MTPM

$ICC_{agreement}$	L_ICC	U_ICC	σ_p^2	σ_r^2	$\sigma_{pr,e}^2$	SEM (mm)	SDC (mm)	SEO (mm)
0.8093	0.5458	0.9283	0.0934	0.0022	0.0198	0.1484	0.4113	0.0471

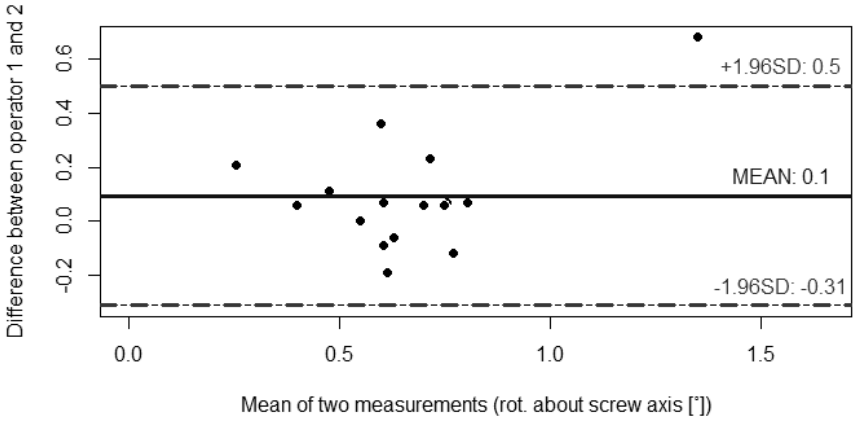
mTRE

$ICC_{agreement}$	L_ICC	U_ICC	σ_p^2	σ_r^2	$\sigma_{pr,e}^2$	SEM (mm)	SDC (mm)	SEO (mm)
0.8398	0.6094	0.9404	0.0512	0.0014	0.0083	0.0988	0.2740	0.0378

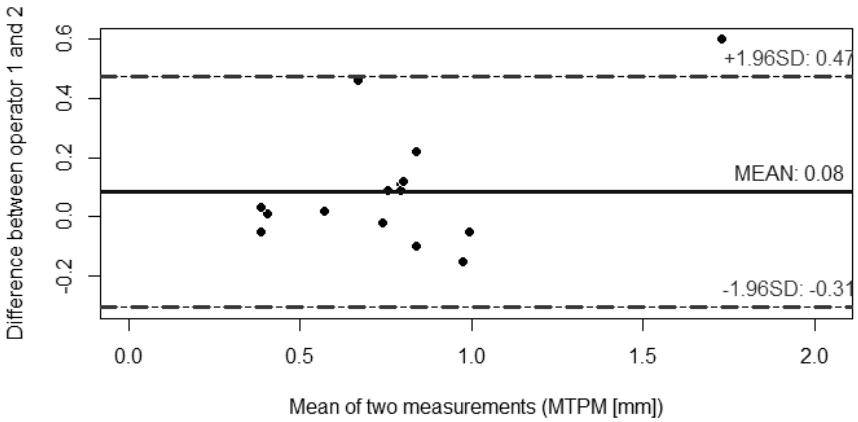
Table 3: Results of ICCagreement, lower and upper bounds of agreement (L_ICC and U_ICC), variance components (σ_p^2 patient variance, σ_o^2 operator of loading device variance and σ_(pr,e)^2 residual error), standard error of measurement (SEM), Smallest detectable change (SDC) and standard error of operator (SEO) of repeated measurements, using the 100% tibia analysis protocol. Values were rounded to two decimal places.

The Bland-Altman plots and SEOs of the rotation about the screw-axis, MTPM and mTRE show that there is no systematic difference between operator 1 and operator 2 (Figure 4). This is supported by the lack of a statistically significant difference between the computed mean and zero. Furthermore, no upwards trend in the plot was identified ruling out the presence of proportional bias. The SDC was 0.44˚for rotation about the screw-axis, 0.41 mm for MTPM and 0.27 mm for mTRE.

A)



B)



C)

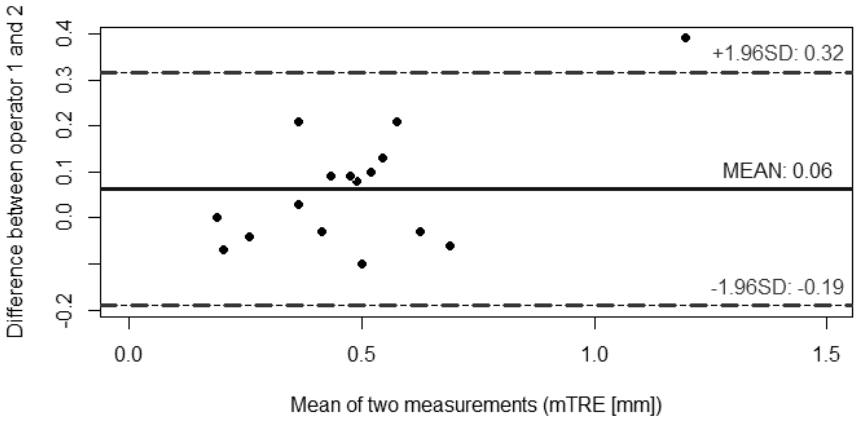


Figure 3 A/B/C: Bland-Altman plots of the mean rotation about the screw-axis (A), MTPM (B) and mTRE (C) of the repeated examinations (horizontal axis) and the difference between them (vertical axis), for the analysis protocol using 100% tibia. The mean difference is shown by the blue line, the red dotted lines represent the limits of agreement (+/-1.96 times standard deviation).

20% tibia analysis protocol

A poor ICCagreement was found when quantifying implant displacement relative to the proximal (20%) tibia, with values of 0.17 [95% CI: -0.33 – 0.60] for rotation about the screw-axis, 0.25 [95%CI: -0.25 – 0.65] for MTPM and 0.31 [95% CI: -0.19 – 0.68] for mTRE (Table 4). The SEO was 0.03˚ for rotation about the screw-axis, 0.01 mm for MTPM and 0.02 mm for mTRE. The SEM was 0.14˚ for rotation about the screw-axis, 0.10 mm for MTPM and 0.06 mm for mTRE. Both the SEO and SEM were smaller than those identified for the 100% tibia analysis protocol.

No systematic differences between operators 1 and 2 were observed, as demonstrated by the SEO, visualized in the Bland-Altman plots, and confirmed by the non-statistically significant mean-difference compared to zero for all implant displacement parameters (Figure 5). No evidence of proportional bias was found. The SDC was 0.37˚for rotation about the screw-axis, 0.29 mm for MTPM and 0.17 mm for mTRE.

Rotation about the screw-axis

$ICC_{agreement}$	L_ICC	U_ICC	σ_p^2	σ_r^2	$\sigma_{pr,e}^2$	SEM (˚)	SDC (˚)	SEO (˚)
0.1683	-0.3298	0.5987	0.0037	0.0009	0.0174	0.1351	0.3746	0.0294

MTPM

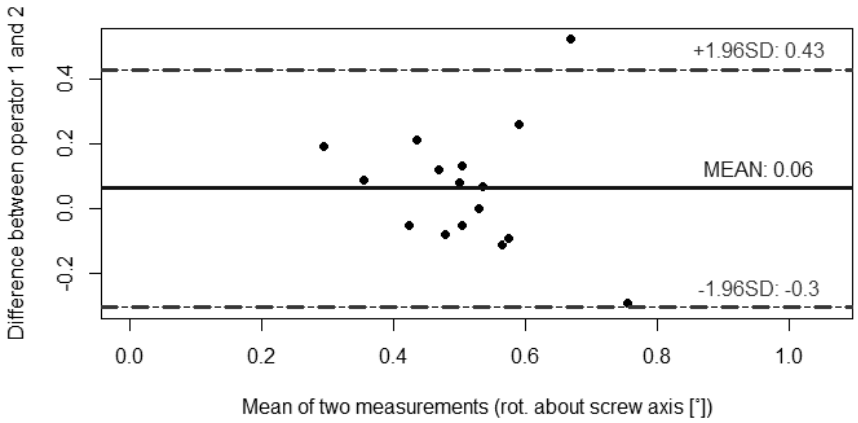
$ICC_{agreement}$	L_ICC	U_ICC	σ_p^2	σ_r^2	$\sigma_{pr,e}^2$	SEM (mm)	SDC (mm)	SEO (mm)
0.2485512	-0.2531	0.6498	0.0036	0.0001	0.0106	0.1036	0.2873	0.0105

mTRE

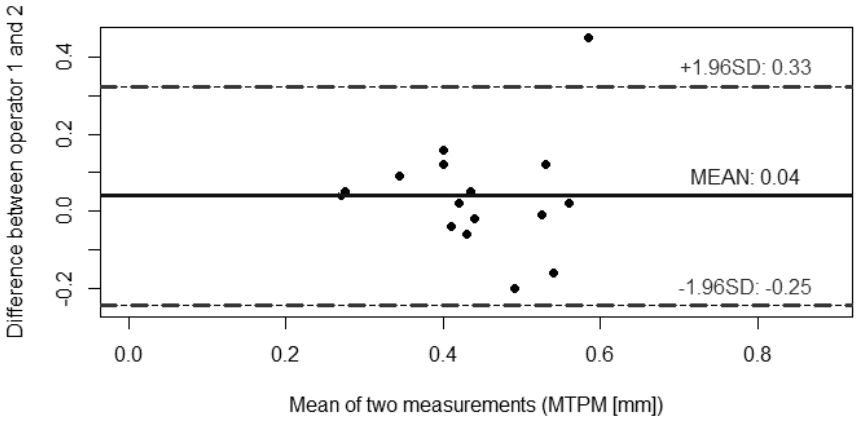
$ICC_{agreement}$	L_ICC	U_ICC	σ_p^2	σ_r^2	$\sigma_{pr,e}^2$	SEM (mm)	SDC (mm)	SEO (mm)
0.3052	-0.1948	0.6839	0.0016	0.0006	0.0030	0.0598	0.1659	0.0237

Table 4: Results of ICCagreement, lower and upper bounds of agreement (L_ICC and U_ICC), variance components (σ_p^2 patient variance, σ_o^2 operator of loading device variance and σ_(pr,e)^2 residual error), standard error of measurement (SEM), smallest detectable change (SDC) and standard error of operator (SEO) of repeated measurements, using the 20% tibia analysis protocol. Values were rounded to two decimal places.

A)



B)



C)

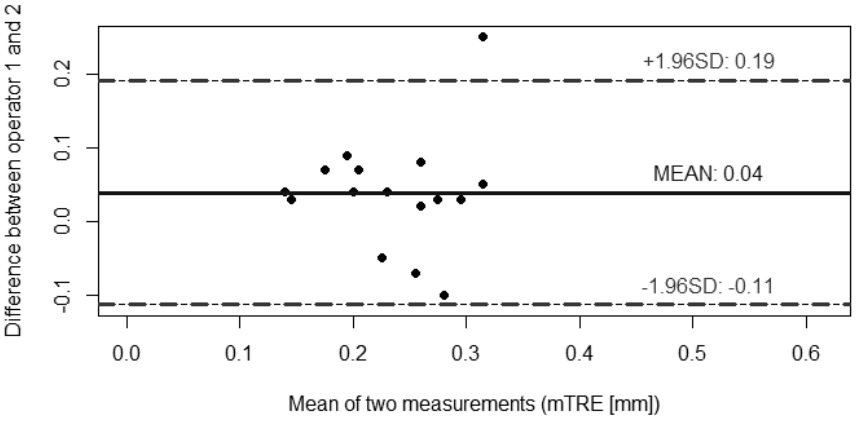


Figure 4 A/B/C: Bland-Altman plots of the mean rotation about the screw-axis(A), MTPM (B) and mTRE (C) of the repeated examinations (horizontal axis) and the difference between them (vertical axis), for the analysis protocol using 20% tibia. The mean difference is shown by the blue line, the red dotted lines represent the limits of agreement (+/-1.96 times standard deviation).

Discussion

This study evaluated the inter operator variation, in terms of reliability and measurement error metrics, of applying a loading device for two different image analysis protocols, and found SEO, SEM and SDC scores of just a few tenths of a mm and degree. Bland-Altman plots visually confirm consistent measurements across operators with no systematic differences or proportional bias. The $ICC_{agreement}$ score was good for the 100% tibia analysis protocol, and poor for the 20% tibia analysis protocol.

The lower $ICC_{agreement}$ scores for the 20% tibia analysis protocol can be explained by a reduction in subject specific variances compared to operator and residual variances (Eq.1.) Focusing on the proximal 20% of the tibia minimizes tibial deformation, leading to smaller apparent implant displacement and reduced differences between subjects compared to the 100% tibia analysis. Although the 20% protocol results in lower $ICC_{agreement}$ values, it also produces smaller SEO, SEM and SDC values, indicating more precise measurements and providing evidence that inter-operator related variance is larger for the 100% tibia analysis protocol. In addressing whether the choice between 100% versus 20% tibia as a reference object influences inter-operator variation, and to what extent, it is important to consider the primary goal of the analysis-loosening detection by differentiating fixed from loose implants. Thus, the relatively low SDC value for the 20% tibia analysis is the most important factor to consider, despite the poor ICC score.

One outlier was identified in each implant displacement parameter for both the 100% and 20% image analysis protocols. A retrospective qualitative comparison of the corresponding CT scans revealed that, during the second examination, the loading device was applied too proximally, positioning the bottom support flange (i.e. the part that connects to the lower leg) such that it covers the joint space. Although the target moment of 20 Nm was achieved in this examination, the forces were not effectively transmitted through the tibial component only, leading to lower induced implant displacement in the second examination compared to the first. To avoid this incorrect placement of the loading device in the future, based on the findings of this study, the loading device's application protocol now includes specific instruction to identify the patella and joint space before applying the loading device.

Recent advancements in implant stability evaluation are shifting from radiostereometric analysis (RSA) to CT-based RSA due to its benefits, such as redundancy of surgical markers and allowing secondary findings on CT scans [51]. Moreover, induced displacement CT enables instantaneous assessment of stability [10, 50, 61]. A recent study highlighted the potential of weight-bearing CT, showing a that the precision-repeated measurements in the seated position- is better than the displacement of the tibial component between loaded and unloaded positions [20]. However, the variability introduced by loading and the reproducibility of load application have

not been evaluated, limiting the ability to directly compare these findings with the present results.

This evaluation of reliability and measurement error reinforces and expands the existing body of evidence for induced tibial displacement assessment. Buijs et al. (2024) demonstrated strong inter-rater reliability among three image analysis operators, with $ICC_{agreement}$ values of 0.98 for rotation about the screw-axis, 0.93 for MTPM, and 0.89 for mTRE [10]. Furthermore, the SDC values determined in this study-0.44˚ for rotation about the screw-axis, 0.41 mm for MTPM, and 0.27 mm for mTRE-fall below the previously established threshold for loosening [10], indicating that the measurement system is sufficiently sensitive to detect changes smaller than these thresholds, even with variability introduced by different loading device operators. Additionally, a recent reproducibility study reported mean errors of 0.13˚for rotation about the screw-axis, 0.12 mm for MTPM, and 0.07 mm for mTRE [27]. These findings, derived from a cadaveric setting, align closely with the $SEM_{agreement}$ values found in the current, in-patient study, further validating the system's precision and consistency.

This overview of previous work highlights the complex interplay of potential sources of error to the system of tibial component displacement assessment. The reported values of intra- and inter-operator variability of the image analysis or reproducibility do not incorporate the error introduced by applying in the loading device. To quantify the true precision of the complete system (the repeatability error), multiple trials on multiple subjects from the same operator should be conducted. However, it can be argued that the intra-rater variability would likely be lower than the inter-rater variability and the current study therefore represents a ‘worst-case-scenario’ [7].

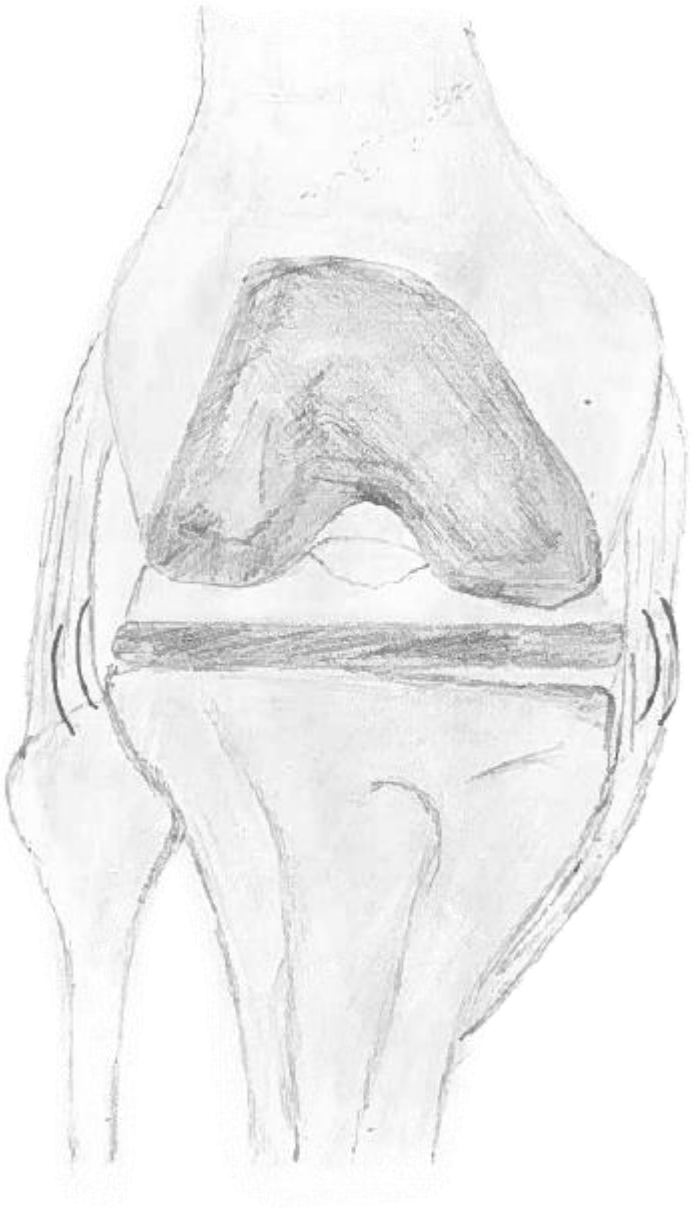
This study has certain limitations, and its findings should be considered in light of the following issues. First, although the study met the preset sample size, the relatively small and uniform population may limit the generalizability of the results to a broader and more diverse patient population as seen in clinical practice. Secondly, the use of only two operators for the ICC analysis limits the generalizability of the reliability findings to a broader population of potential raters. However, this was necessary due to ethical concerns over additional radiation exposure to the subjects. Thirdly, in this study, the target moment of 20 Nm was not achieved in three subjects, leaving the effect of applying a moment lower than 20 Nm on displacement measurements uncertain. To address this knowledge gap, a cadaveric study is warranted, focusing on measuring internal load transfers from the loading device to the implant-bone interface and the effect of different load levels. Additionally, to further reduce the SEM associated with the induced tibial displacement assessment, improvement of the image analysis steps, using deep learning-based segmentation model can and should be considered [34].

Conclusion

The results of this study demonstrate that the loading device for induced tibial component displacement assessment can be operated by both experienced and less experienced trained personnel, with good inter-operator reliability and small measurement errors.

References

1. *Buijs GS, Kievit AJ, Schafroth MU, Hirschmann MT, Blankevoort L. Weight-bearing pain and implant migration, 1. Barnsley L, Barnsley L (2019) Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiology 48:1565-1572*
2. *Bartlett JW, Frost C (2008) Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in Obstetrics & Gynecology 31:466-475*
3. *Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310*
4. *Buijs GS, Kievit AJ, Ter Wee MA, Magg C, Dobbe JGG, Streekstra GJ, et al. (2024) Non-invasive quantitative assessment of induced component displacement can safely and accurately diagnose tibial component loosening in patients: A prospective diagnostic study. Knee Surg Sports Traumatol Arthrosc;10.1002/ksa.12299*
5. *Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis.10.1002/ksa.12206. Knee Surg Sports Traumatol Arthrosc2024.*
6. *Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis.10.1002/ksa.12206. Knee Surg Sports Traumatol Arthrosc 2024.*
7. *Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284*
8. *Hext RA, Kaptein BL, Howard JL, Lanting BA, Teeter MG (2024) Inducible displacement of cementless total knee arthroplasty components with conventional and weight-bearing CT-based radiostereometric analysis. J Orthop Res;10.1002/jor.26017*
9. *ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. https://www.icmje.org/icmje-recommendations.pdf.*
10. *Kievit AJ, Buijs GS, Dobbe JGG, Ter Wee A, Kerkhoffs G, Streekstra GJ, et al. (2023) Promising results of a non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon) 104:105930*
11. *Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine 15:155-163*
12. *Magg C, ter Wee M, Buijs G, Kievit A, Krap D, Dobbe JG, et al. Towards automation in non-invasive measurement of knee implant displacement. Vol 12927: SPIE; 2024.*
13. *Mokkink LB, de Vet H, Diemeer S, Eekhout I (2023) Sample size recommendations for studies on reliability and measurement error: an online application based on simulation studies. Health Services and Outcomes Research Methodology 23:241-265*
14. *Mokkink LB, Eekhout I, Boers M, van der Vleuten CPM, de Vet HCW (2023) Studies on Reliability and Measurement Error of Measurements in Medicine - From Design to Statistics Explained for Medical Researchers. Patient Relat Outcome Meas 14:193-212*
15. *Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 63:737-745*
16. *Niesen AE, Garverick AL, Hull ML (2021) Maximum Total Point Motion of Five Points Versus All Points in Assessing Tibial Baseplate Stability. J Biomech Eng 143:*
17. *Olivecrona H, Olivecrona L, Weidenhielm L, Noz ME, Kardos J, Maguire GQ, Jr., et al. (2008) A new technique for diagnosis of acetabular cup loosening using computed tomography: preliminary experience in 10 patients. Acta Orthop 79:346-353*
18. *Sandberg O, Carlsson S, Harbom E, Cappelen V, Tholén S, Olivecrona H, Wretenberg P (2022) Inducible displacement CT increases the diagnostic accuracy of aseptic loosening in primary total hip arthroplasty. Acta Orthop 93:831-836*
19. *Sandberg OH, Karrholm J, Olivecrona H, Rohrl SM, Skoldenberg OG, Broden C (2023) Computed tomography-based radiostereometric analysis in orthopedic research: practical guidelines. Acta Orthop 94:373-378*
20. *Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420-428*
21. *Ter Wee MA, Dobbe JGG, Buijs GS, Kievit AJ, Schafroth MU, Maas M, et al. (2023) Load-induced deformation of the tibia and its effect on implant loosening detection. Sci Rep 13:21769*
22. *Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, et al. (2021) Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 30:2729-2754*
23. *Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454*
24. *Wretenberg P, Tholen S, Carlsson S, Olivecrona H (2021) Implant Movement Analysis (IMA), A New CT Based Technique for Diagnosis of Aseptic Loosening of Total Knee Arthroplasty. Orthopedic Research Online Journal.*



Chapter VIII

Non-invasive quantitative assessment of induced component displacement can safely and accurately diagnose tibial component loosening in patients: a prospective diagnostic study.

George S. Buijs*, Arthur J. Kievit*, Maaike A. Ter Wee, Caroline Magg, Johannes G.G. Dobbe, Geert J. Streekstra, Matthias U. Schafroth, Leendert Blankevoort

** Both authors contributed equally.*

Clin Biomech (Bristol, Avon). 2023 Apr:104:105930. doi: 10.1016

Abstract

Purpose: Aseptic loosening often requires major, expensive, and invasive revision surgery. Current diagnostic modalities merely show indirect signs of loosening. A recent proof of concept study proposed a non-invasive technique for the quantitative and visual assessment of implant movement as a diagnostic aid for tibial component loosening. The primary research question addressed is whether this novel diagnostic modality can safely and effectively aid the diagnosis of aseptic loosening.

Methods: This clinical study included patients suspected of aseptic total knee arthroplasty (TKA) loosening listed for revision surgery and asymptomatic patients. Safety was evaluated using a numerical rating scale (NRS) for discomfort and by registration of adverse events. Feasibility was assessed by recording the duration and ease of the procedure. Intra- and interrater reliability were evaluated. In symptomatic patients, diagnostic accuracy metrics were evaluated with intra-operative assessment as a reference test.

Results: In total, 34 symptomatic and 38 asymptomatic knees with a TKA were analyzed. The median NRS for discomfort during loading was 6 (interquartile range [IQR]: 3.75-7.00) in symptomatic patients and 2 (IQR: 1.00-3.00) in asymptomatic patients. No adverse events were reported. Most users found the use of the loading device easy. The median time spent in the computed tomography room was 9 min (IQR: 8.00-11.00). Excellent to good intra- and interrater reliabilities were achieved. Diagnostic accuracy analysis resulted in a sensitivity of 0.91 (95% confidence interval [CI]: 0.72-0.97) and a specificity of 0.72 (95% CI: 0.43-0.90).

Conclusions: The proposed diagnostic method is safe, feasible, reliable, and accurate in aiding the diagnosis of aseptic tibial component loosening.

Introduction

Following infection, aseptic loosening is reported to be the second most common cause of Total Knee Arthroplasty (TKA) failure [10, 17]. Aseptic loosening often requires major, expensive, and invasive revision TKA (rTKA) surgery. The diagnosis of aseptic loosening is challenging as there is no consensus in the available literature regarding a standardized diagnostic work-up including a specific diagnostic test [4, 7].

If a patient presents at the outpatient clinic with symptoms indicative of TKA loosening, for instance pain during weight-bearing activities, conventional radiography is typically the first imaging modality used to assess bone resorption and TKA component migration as signs of loosening [3, 6]. The sensitivity and specificity of plain radiography are inadequate, particularly in cases of early and subtle but clinically relevant loosening [12].

Although additional imaging modalities like CT, 3-phase bone scintigraphy, PET-CT, and SPECT/CT are used, the American College of Radiology does not recommend them for evaluating pain post knee arthroplasty once infection is ruled out [23]. Nuclear imaging modalities measure secondary and a-specific effects, such as increased bone turnover and osteoclastic activity while exposing patients to high radiation dosages (4-7mSv) [2, 4, 12, 18]. Accurate diagnosis is crucial to prevent unwarranted revision surgeries in patients misdiagnosed with loosening of the TKA, while also ensuring that patients with unrecognized loosening are not mistakenly excluded from necessary revision procedures. Additionally, it prevents patients from enduring needless delays in receiving appropriate care with a correct diagnosis [20, 23].

As a solution to improve the diagnostic work-up, Kievit et al. proposed and evaluated a less cumbersome alternative [14]. This method employs a loading device to exert up to 20 Nm on the knee in alternating valgus and varus direction. In each direction a CT-scan is made (exposing patients to an estimate of 1.2mSv). Advanced imaging analysis then quantifies and visualizes the tibial component's displacement between the two directions relative to the bone. This pilot study demonstrated reliable and reproducible results in a cadaveric setting [14]. A patient study was deemed necessary to further evaluate the diagnostic potential and clinical usability of this method.

Our primary research question was: What are the diagnostic accuracy metrics of the proposed modality? Our secondary research questions were: (1) Is the method safe for application in a clinical setting? (2) Is the method feasible for clinical use, in terms of duration of evaluations, ease and successfulness of use? (3) What is the intra- and interrater reliability of this method?

For evaluating the diagnostic accuracy metrics, the subsequent questions were posed: (1) What extend of induced tibial component displacement is observed in both symptomatic and asymptomatic TKA patients? (2) What was the most optimal threshold for induced displacement

to classify a tibial component as either fixed or loose, with the intraoperative observation as reference test and what are the consequent diagnostic accuracy metrics? And, additionally, to what extend are Patient Reported Outcome Measures (PROMs) correlated to the induced displacement of the tibial component?

This follow-up study hypothesizes that the proposed method is safe, feasible, reliable, and accurate to aid the diagnosis of loosening of the tibial component.

Methods

Ethical statement

This study was reported in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans, and the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

Patient screening and inclusion

In this prospective diagnostic test accuracy study, conducted by the Amsterdam UMC, seven affiliated hospitals were invited to refer patients scheduled for rTKA for the diagnosis of aseptic TKA loosening for inclusion in the study. The diagnosis of aseptic loosening was established per local protocol, including physical examination and the exclusion of infection by c-reactive protein and if deemed needed joint aspiration [13]. Patients referred solely for the definitive diagnosis of aseptic loosening were included the symptomatic group.

Local registries of TKA patients were screened for eligible patients for the asymptomatic group. Patients who reported complete satisfaction with their existing TKA and had no complaints during their most recent outpatient visit to the Amsterdam UMC were considered asymptomatic. These individuals were subsequently contacted and invited to participate in the study. Those who consented were then included in the asymptomatic group.

Safety

To assess the safety of this method, patients were asked to rate the discomfort they experienced during and after loading of the knee on a Numeric Rating Scale (NRS) of zero to ten, with zero indicating no pain or discomfort and ten indicating extreme pain or discomfort. Additionally, (serious) adverse events were recorded.

Feasibility

To evaluate feasibility, medical staff applying and operating the loading device were asked to rate the ease of positioning of the loading device, attachment of straps and application of moment as either easy, intermediate, or difficult. Additionally, the actual applied moment in both valgus- and varus direction and total time spent by the patient in the CT-room were registered.

Reliability

To evaluate the reliability of the measurements resulting from the 3D analysis software, an intra- and interrater reliability assessment was conducted. Three distinct raters, all trained but two experienced (G.B. and A.W.) and one inexperienced (C.M.) analyzed all the scans separately. One rater (G.B.) analyzed a random sample of ten scans twice, with a 30-day time interval between the analysis. All raters were blinded for the results generated by the other readers, as well as for the outcome of the reference test employed for the diagnostic accuracy analysis.

Diagnostic accuracy

Index test: measurement of induced displacement

Hardware: Loading device

A patented loading device was used to apply a valgus- and varus moment on the knee whilst conducting a CT-scan in both situations (Figure 1) [5, 21]. This device is designed to apply a bending moment on the knee in the frontal plane of the tibia of 20 Nm with the knee in 20 degrees flexion using the principle of four-point bending. The 20-degree flexion is to relax the posterior capsule and cruciate ligaments. The externally applied bending moment is internally balanced by the combination of (1) a compressive force in either the medial compartment (varus moment) or lateral compartment (valgus moment) and (2) the forces in the lateral collateral ligament and capsule or in the medial collateral ligament and capsule, respectively.

The compressive forces on the medial and lateral compartment will induce displacement of the tibial component relative to the bone. The intention was to apply the maximal load of 20 Nm. If the patient indicated that the load was too painful and/or uncomfortable, the load was reduced. The level of the load was registered.

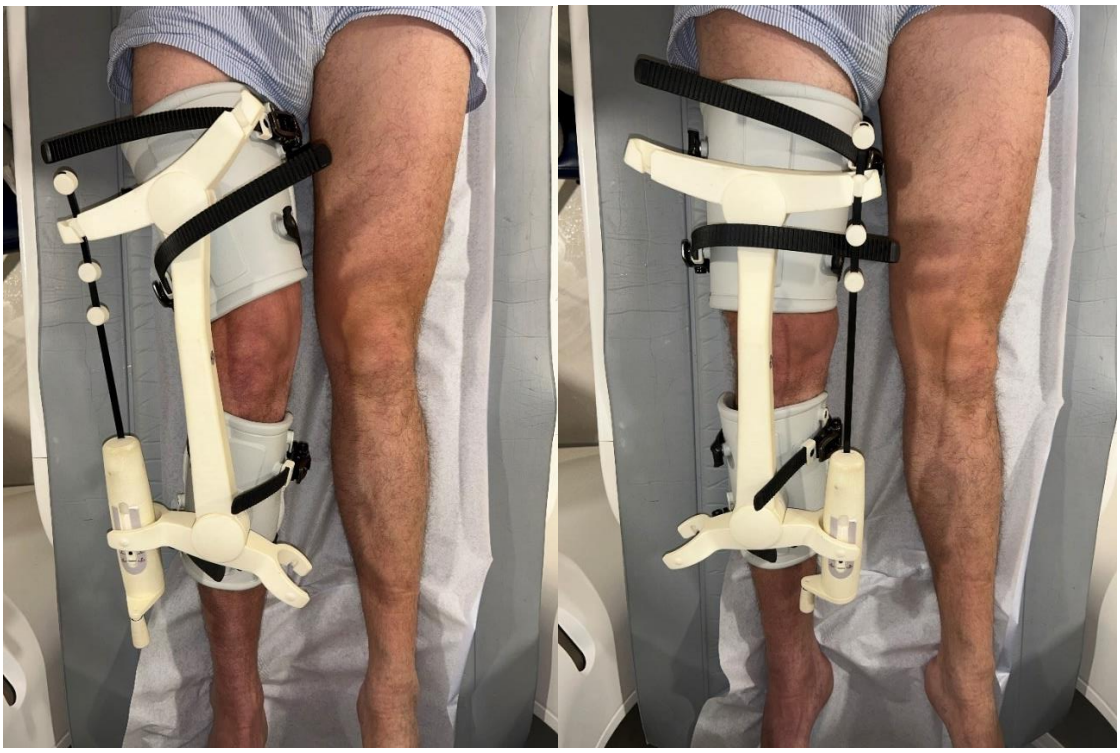
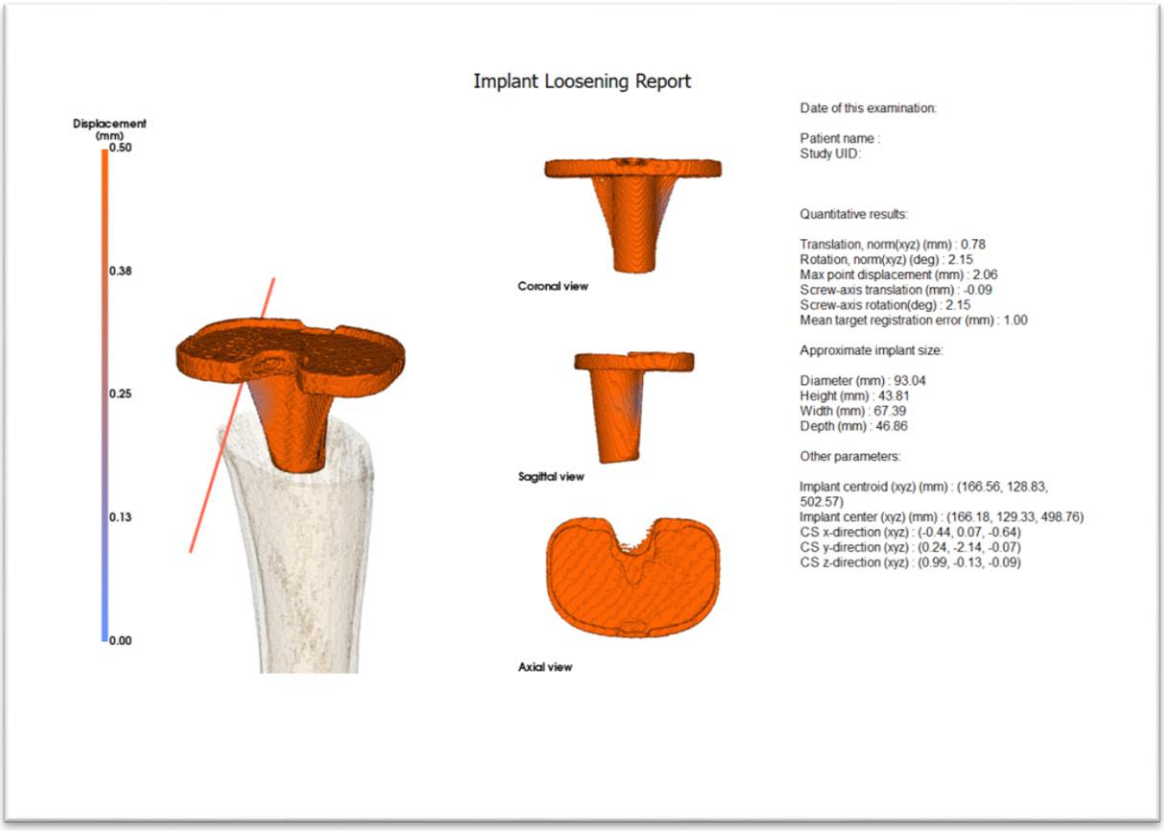


Figure 1 A/B: Picture of knee in CT-scanner with the loading device in valgus and varus.

Software: quantifying and visualizing implant displacement

Custom image analysis software was developed to quantify and visualize implant displacement. The methods incorporated in the software were based on a protocol by Dobbe et al. and were validated in a recently published cadaveric pilot study by Kievit and Buijs et al [8, 14]. In short, the tibial implant and the tibial cortex were segmented by threshold-connected region growing from the valgus image, resulting in a polygon mesh model of the implant and tibia used for visualization purposes. Image registrations from the tibial and implant segmentations to the varus 3D image were subsequently used to find the positions of the implant and tibia in the varus image. This enabled quantification of the relative displacement of the implant with respect to the tibia and expressing this displacement in terms of the following clinically relevant parameters: (1) rotation about the screw-axis [19], (2) the average point displacement of all points in the implant mesh (mean Target Registration Error, mTRE [9]), and (3) the maximum point displacement observed across the implant model (Maximum Total Point Motion, MTPM) [14]. These parameters are shown in a report together with a visual representation of the implant and tibia. In addition, local point displacements of the implant model are visually represented by a heat map, showing color variations indicating the magnitude of the local displacement (Figure 2A/B).



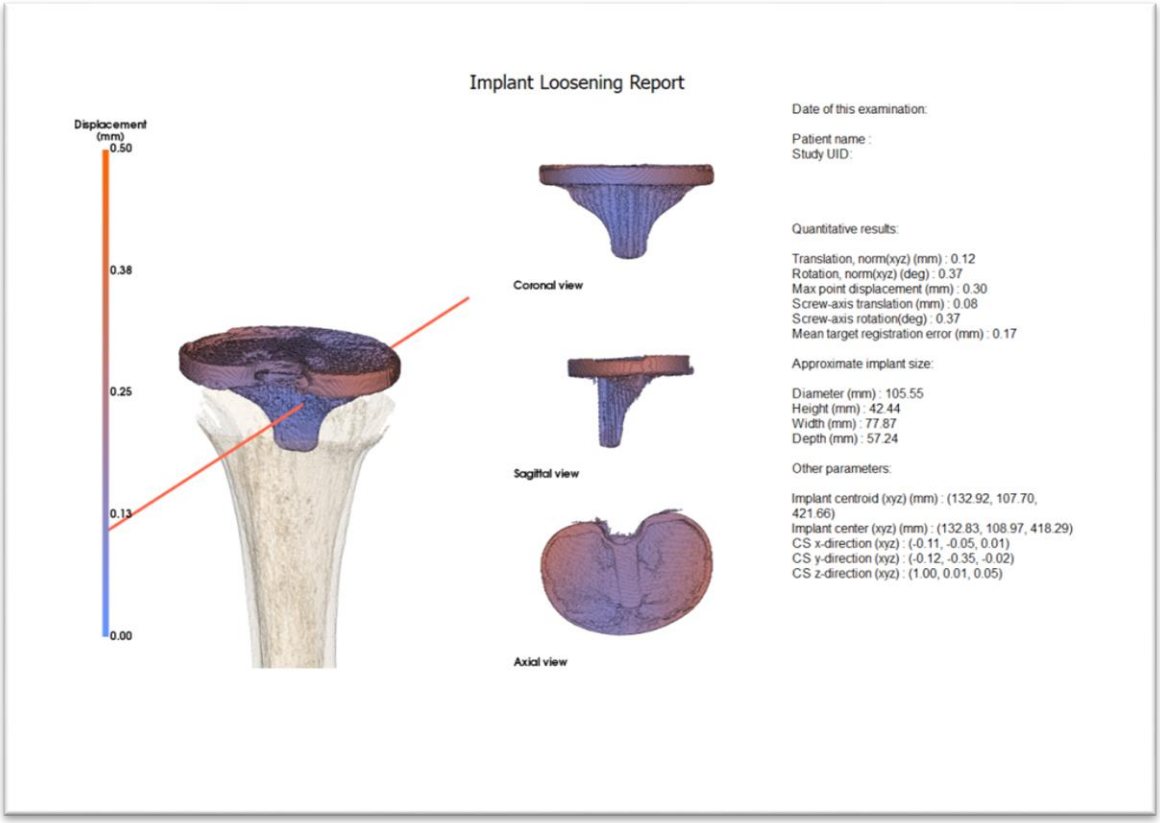


Figure 2A/B: Example of Implant Loosening Report of a fixed tibial (A) and loose (B) component. Mm; millimeter, deg; degrees.

Reference test

Symptomatic patients underwent rTKA at their respective local hospitals. The orthopedic surgeon performing the revision surgery was requested to assess the implant during the procedure. Based on the intraoperative observation, the surgeon classified the component as either 'loose' or 'fixed'. This classification served as the reference for the study. The orthopedic surgeons performing rTKA were blinded for the results of the index test.

Analysis

Diagnostic accuracy was evaluated first for all three selected displacement parameters separately (univariate) and then combined (multivariate). Based on the data from the symptomatic patients, optimal thresholds for individual displacement parameters were determined using univariate Receiver Operating Characteristic (ROC) curves, with results providing sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) for each parameter.

Additionally, a multivariate logistic regression model was developed to determine the optimal threshold when combining the three displacement parameters. A multivariate ROC curve was plotted. Optimal threshold (closest to upper left corner) derived from this multivariate ROC curve was computed and used to categorize predicted outcomes (loose or fixed). Subsequent accuracy metrics were then calculated.

To evaluate robustness, the model was checked for multicollinearity among the three predictors (rotation about the screw axis, mTRE and MTPM) using Variance Inflation Factors (VIFs). A VIF greater than 10 was considered indicative of present multicollinearity that could affect the stability and interpretation of the model coefficients [15].

Statistics

Inter- and Intraclass Correlation Coefficients (ICCs) were calculated for rotation about the screw-axis, mTRE and MTPM to assess interrater reliability, using a two-way mixed-effects (ICC3k) model for assessment of interrater reliability and a two-way random-effects model (ICC2k) for intra-rater reliability. Moderate, good, and excellent reliability were indicated by values of 0.5 to 0.75, 0.75 to 0.90, and >0.90, respectively [16].

Categorical variables were analyzed using frequencies, proportions, and Fisher’s exact test. Parametric data were evaluated using mean, standard deviation (SD), and Student’s T-test. Non-parametric data were assessed with median, Interquartile Range (IQR), and the Mann-Whitney U test.

Correlation between displacement parameters and KOOS (Knee Injury and Osteoarthritis Outcome Score) were analyzed using Spearman correlation coefficient (RS) [1].

Based on pre-liminary results, a logistic regression power analysis was performed (power: 0.85, odds ratio: 4.0, alpha 0.05 and beta 0.15), resulting in a required sample size of 37 symptomatic knees.

Data were analyzed with R for Windows version 4.2.3 (R Foundation for Statistical Computing, Vienna, Austria), using the “car”, “irr”, “psych” and “pROC” packages. A two-sided p-value < 0.05 was considered statistically significant.

Results

Patient inclusion

From the Amsterdam UMC and seven affiliated hospitals, 36 patients including 37 knees suspected of aseptic loosening were referred to the Amsterdam UMC for induced displacement measurements. Of whom, two patients (3 knees) did not receive a revision. Therefore, 34 knees of 34 patients were included for analysis in the symptomatic group (Table 1). In the asymptomatic group, 42 knees of 32 patients were scanned and evaluated. Four knees of two patients were excluded from the analysis because of technical issues (e.g. movement of the leg during scanning) (Figure 3, Table 1).

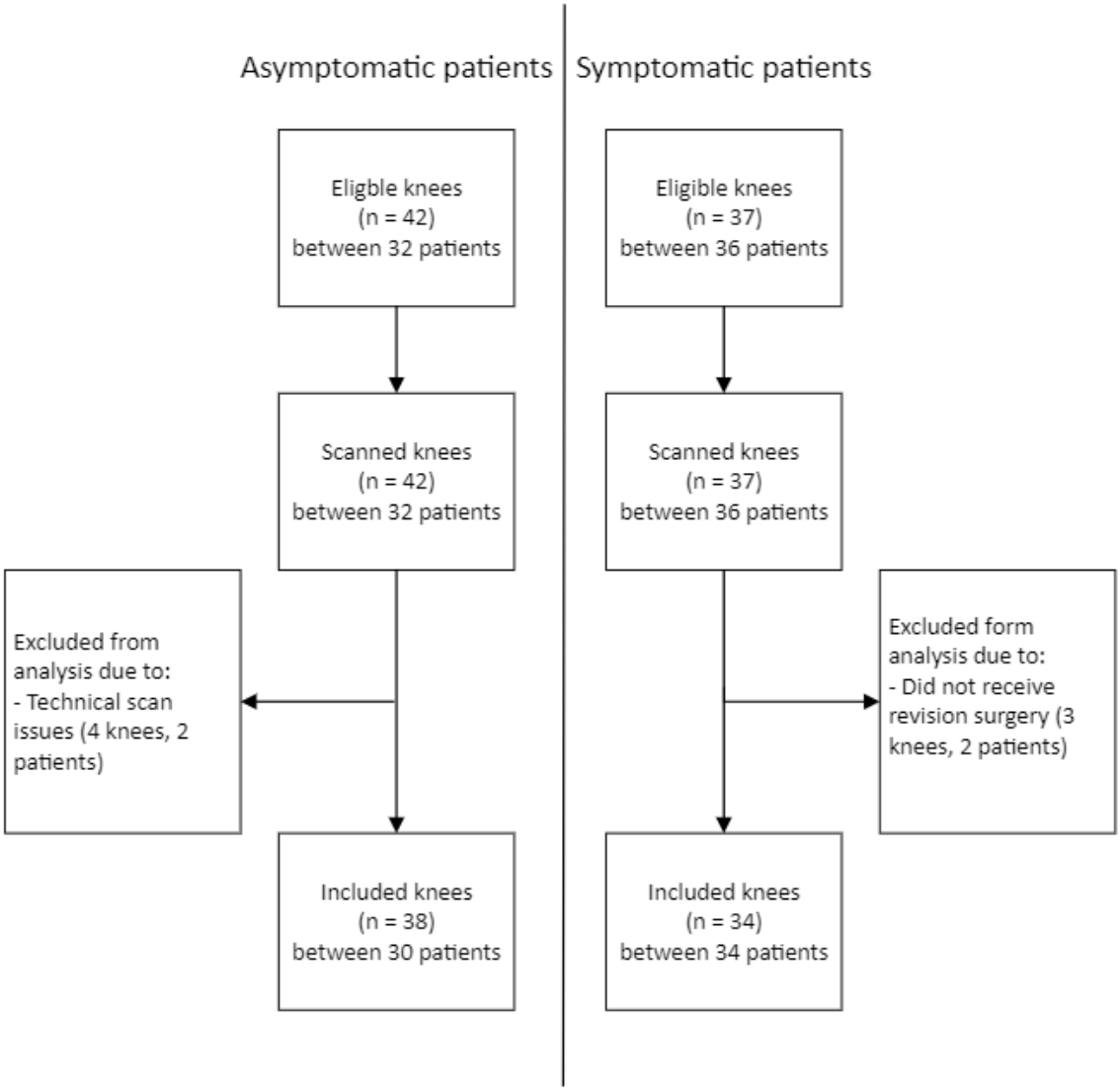


Figure 3: Flowchart of patient selection and inclusion.

	Symptomatic knees (n = 34)	Asymptomatic knees (n = 38)	p-value
Age (mean, SD)	67.9 (8.1)	68.8 (8.7)	0.29
Gender (n, %)			0.85
Male	18 (52.9)	22 (57.9)	
Female	16 (47.1)	16 (42.1)	
Time interval between scan and surgery (months, [IQR])	0.9 (0.2-3.5)		
Type of prosthesis (n, %)			0.25
Regular	26 (76.5)	37 (97.4)	
Revision - short stem	2 (5.8)	0 (0.0)	
Revision - long stem	6 (17.6)	1 (2.6)	
KOOS sub scores (median, [IQR])			
KOOS-Pain	33 (25-55.2)	97 (86.8-99.2)	< 0.001
KOOS-Other Symptoms	50 (36-64)	89 (82-95.2)	< 0.001
KOOS-ADL	41 (31.2-48)	91 (87-99)	< 0.001
KOOS-Sport/Rec	2.5 (0-15)	67 (45-85)	< 0.001
KOOS-QoL	18 (6.5-33)	81 (69-92.5)	< 0.001
Magnitude of Displacement (median, [IQR])			
MTPM (mm)	0.86 (0.69-1.32)	0.64 (0.49-0.82)	0.001
mTRE (mm)	0.53 (0.14-0.82)	0.40 (0.29-0.50)	0.002
Rotation about screw-axis (deg)	0.52 (0.51-1.28)	0.52 (0.40-0.67)	0.02

Table 1: Table of baseline characteristics. Symptomatic knees are the groups "Symptomatic Loose" and Symptomatic Fixed" grouped together. IQR: Interquartile range; KOOS; Knee injury and Osteoarthritis Outcome Score; ADL: Function in daily living; Sport/Rec; Function in Sport/Recreation; QoL; Knee Related Quality of Life; MTPM; Maximum Total Point Motion, mTRE; Mean Target Registration Error, mm; millimeter, deg; degrees.

Safety

Median NRS for pain and/or discomfort during loading was 6 (IQR: 3.75-7.00) and 2 (IQR: 1.00-3.00) and for symptomatic and asymptomatic patients, respectively. The median NRS for pain and discomfort after loading was 0 (IQR: 0.00-1.00) in asymptomatic patients and 3 (IQR: 0.75-5.25) in symptomatic patients. No (serious) adverse events were reported by the enrolled patients.

Feasibility

Ease of device positioning, strap attachment, and application of moment were rated as easy in most cases, with some difficulty noted in patients with high leg circumference. Mean moment applied in valgus was 18.9 Nm (SD: 2.92) and 19.0 Nm (SD: 2.60) in varus. Target moment of 20 Nm was reached in 62 (86.1%) cases in the valgus direction and 59 (80.6%) of cases in the varus direction. Pain due to impingement of the straps into the skin was cited as the reason for not achieving the target moment in most cases. Median time spent in CT-room room was 9 minutes (IQR: 8.00-11.00).

Reliability

Good to excellent interrater correlation coefficients were found for rotation about the screw axis, mTRE and MTPM with medians 0.98 (IQR: 0.97-0.99), 0.89 (IQR:0.84-0.93) and 0.93 (IQR: 0.90-0.96) respectively. Intra-rater correlation coefficients were excellent (ICC: 0.91 [IQR: 0.64-0.98]) for rotation about the screw axis, excellent (ICC: 0.96 [IQR: 0.84-0.99]) for mTRE and moderate (ICC: 0.74 [IQR 0.02-0.94]) for MTPM.

KOOS scores and magnitude of displacement

Analysis of spearman correlations between rotation about the screw-axis, mTRE, MTPM and KOOS sub scores resulted in almost all statistically significant negative correlations between the parameters and all KOOS sub scores (Table 2). Magnitude of displacement is shown in Table 1, and Figure 4.

	KOOS-Pain	KOOS- Other Symptoms	KOOS-ADL	KOOS-Sport/Rec	KOOS-QoL
MTPM (mm)	-0.36 (p=0.0023)	-0.29 (p=0.013)	-0.36 (p=0.0022)	-0.26 (p=0.029)	-0.34 (p=0.0041)
mTRE (mm)	-0.35 (p=0.0031)	-0.30 (p=0.012)	-0.34 (p=0.0033)	-0.23 (p=0.059)*	-0.30 (p=0.012)
Rotation about screw-axis (deg)	-0.28 (p=0.018)	-0.23 (p=0.053)*	-0.29 (p=0.013)	-0.26 (p=0.03)	-0.32 (p=0.0061)

Table 2: Spearman Correlations between Movement parameters and KOOS sub scores for all included patients. MTPM; Maximum Total Point Motion; mTRE: Mean Target Registration Error. KOOS; Knee injury and Osteoarthritis Outcome Score; ADL: Function in daily living; Sport/Rec; Function in Sport/Recreation; QoL; Knee Related Quality of Life.

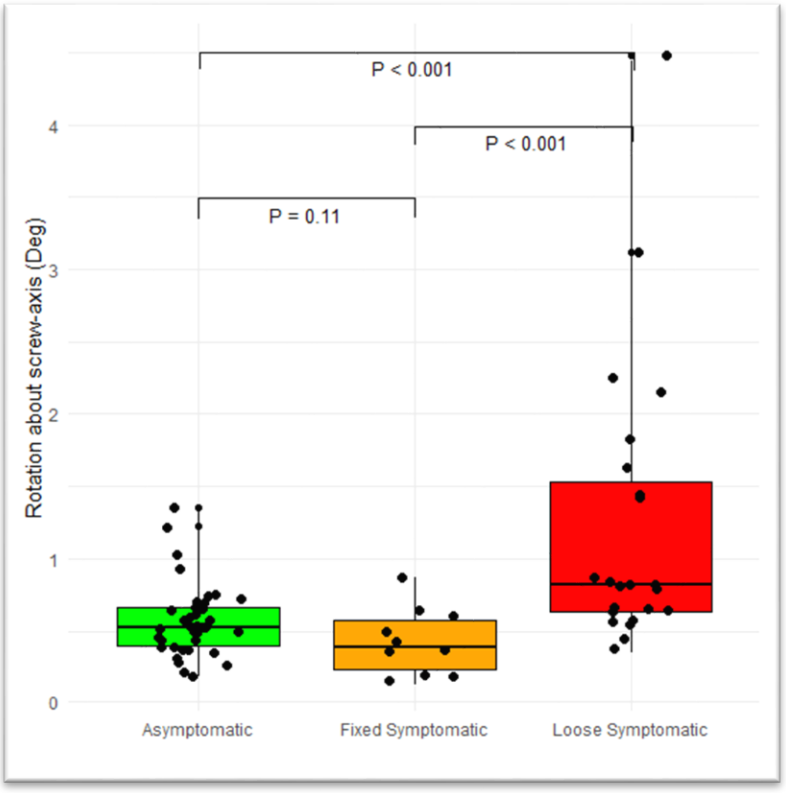


Figure 4 A; Box- and Scatterplot for Rotation about screw-axis in degrees (deg) per group. Median rotation about the screw axis was 0.52 (IQR: 0.40-0.67) degrees, was 0.39 (IQR: 0.23-0.58) degrees and 0.82 (IQR: 0.64-1.54) for asymptomatic, symptomatic fixed and symptomatic loose knees, respectively.

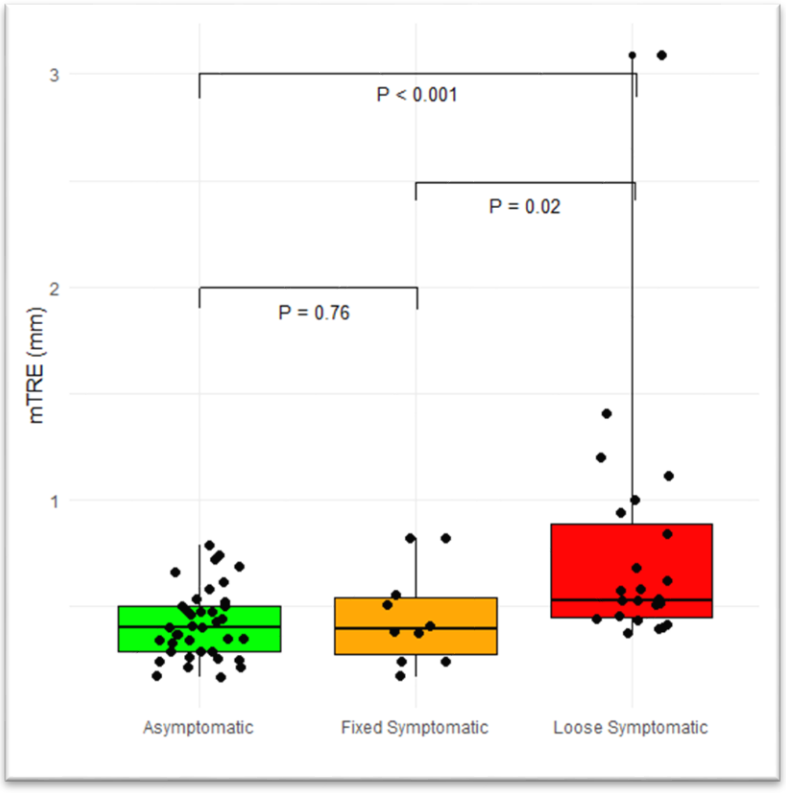


Figure 4 B; Box- and Scatterplot for Mean Target Registration Error (mTRE) millimeter (mm) per group. Median mTRE was 0.40 (IQR: 0.29-0.50) mm, 0.40 (IQR: 0.28-0.54) mm and 0.53 (IQR: 0.45-0.89) mm for asymptomatic, symptomatic fixed and symptomatic loose knees, respectively.

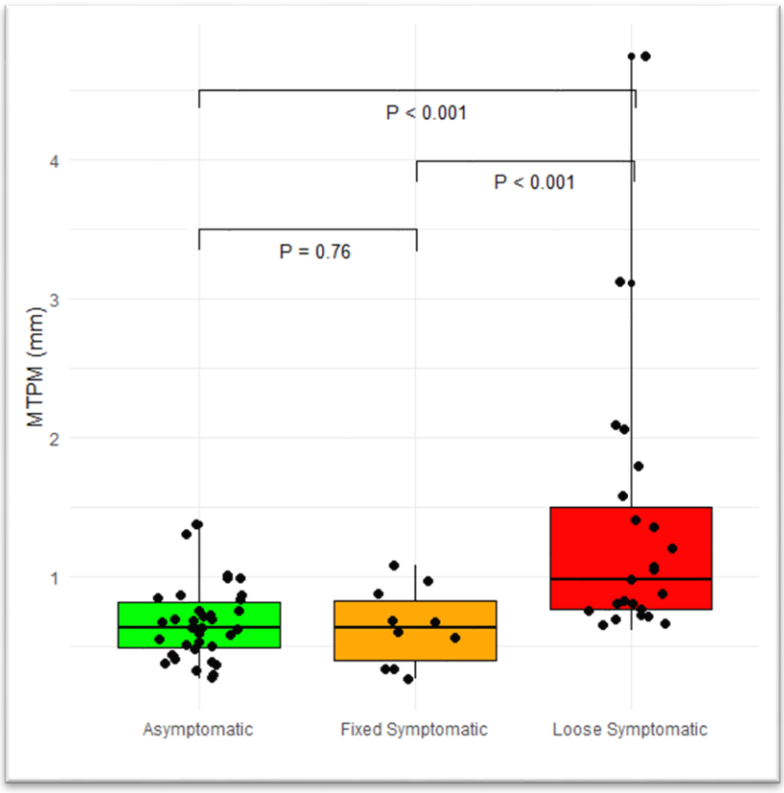


Figure 4 C; Box- and Scatterplot for Maximum Total Point Motion (MTPM) in millimeter (mm) per group. Median MTPM was 0.63 (IQR: 0.49-0.82) mm, 0.64 (IQR: 0.40-0.83) mm and 0.98 (IQR: 0.77-1.51) mm for asymptomatic, symptomatic fixed and symptomatic loose knees, respectively.

Diagnostic accuracy

Based on separate ROC curves, optimal thresholds to differentiate between a loose and fixed tibial component were 0.53 degrees of rotation about the screw axis, a mTRE of 0.42 mm and a MTPM of 0.70 mm, respectively. Consequent diagnostic accuracy metrics and ROC curves are displayed in Figure 5 and Table 3.

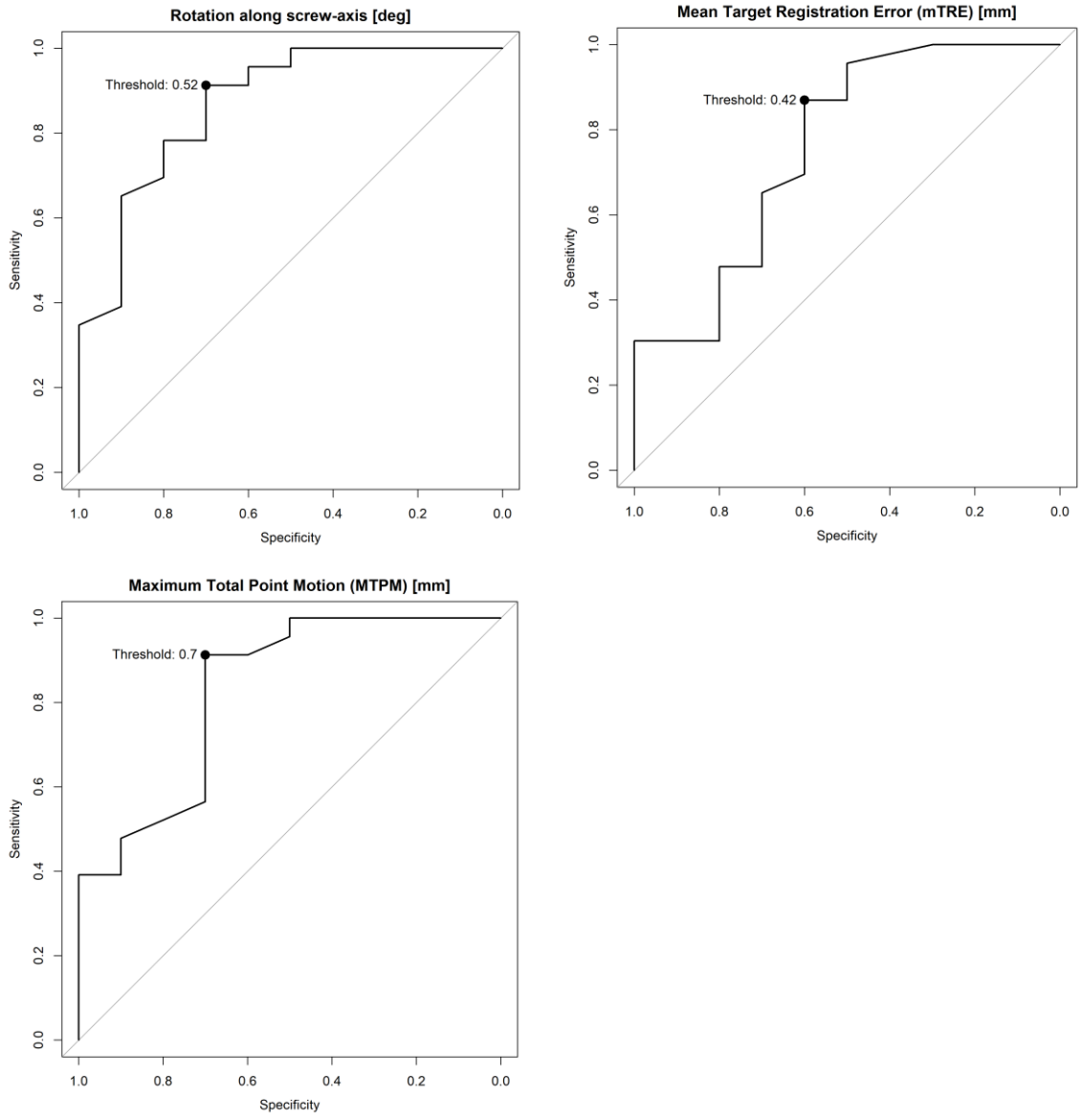


Figure 5 A/B/C: ROC curves for rotation about the screw axis in degrees, Mean Target Registration Error (mTRE) in mm, and Maximum Total Point Motion (MTPM) in mm, with optimal thresholds highlighted.

Combining all three parameters in a multivariate logistic regression model resulted in VIF of 2.03 for rotation about the screw-axis, 5.60 for mTRE and 7.79 for MTPM (Figure 6, Table 3). The multivariate ROC analysis yielded an AUC of 0.90 indicating excellent discriminatory ability. An optimal probability threshold of 0.54 was determined via the ROC analysis, maximizing sensitivity and specificity trade-off (Figure 6). At this threshold, the model exhibited a sensitivity of 0.91 (95% CI: 0.72-0.97), specificity of 0.72 (95% CI: 0.43-0.90), PPV of 0.87 (95% CI: 0.68-0.95), and NPV of 0.80 (95% CI: 0.49-0.95) (Table 3).

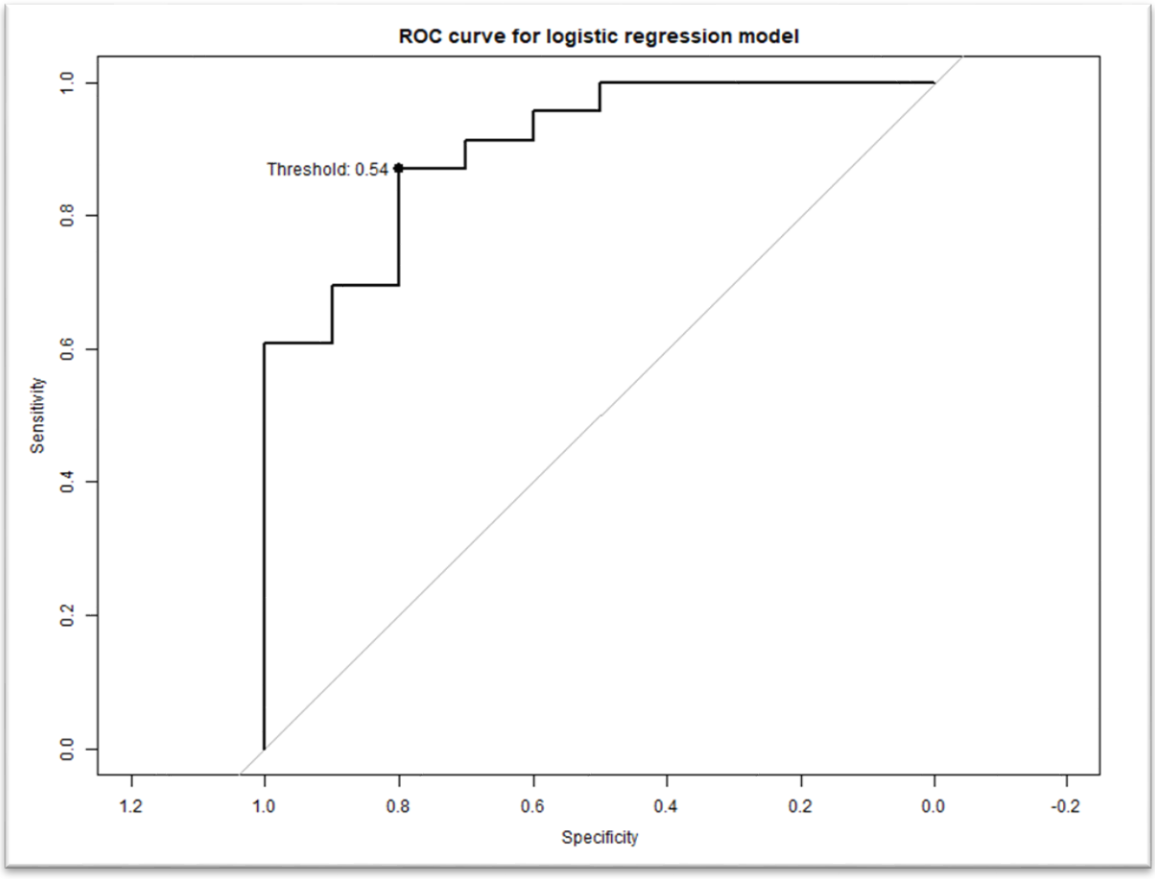


Figure 6: ROC curve of multivariate logistic regression model with optimal probability threshold highlighted.

	Optimal Threshold	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Rotation about the screw-axis	0.53 deg	0.91 (0.52-0.91)	0.70 (0.30-0.70)	0.88 (0.68-0.97)	0.78 (0.40-0.97)
mTRE	0.42 mm	0.87 (0.35-0.87)	0.60 (0.20-0.60)	0.83 (0.63-0.95)	0.67 (0.30-0.93)
MTPM	0.70 mm	0.91 (0.35-0.87)	0.70 (0.30-0.70)	0.88 (0.68-0.97)	0.78 (0.40-0.97)
Multivariate regression model	0.57 *(probability)*	0.91 (0.72-0.97)	0.72 (0.43-0.90)	0.87 (0.68-0.95)	0.80 (0.49-0.95)

Table 3: Results of Uni- and multivariate accuracy metrics. CI; Confidence Interval; mTRE; Mean Target Registration Error, MTPM; Maximum Total Point Motion, deg; degrees, mm; millimeter.

Discussion

The most important finding of this study is that the proposed non-invasive modality is safe, feasible and reliable with a diagnostic accuracy that is better or at least as good as current diagnostic imaging modalities such as CT, MRI, and nuclear scans.

The induced displacement method demonstrated favorable safety, with tolerable pain and discomfort levels and no serious adverse events, highlighting its suitability for clinical use. It was found to be practical and easy for medical staff to use, even with the challenge of large leg circumferences in some patients. The method demonstrated high diagnostic accuracy, with 91% sensitivity and 72% specificity, underscoring its potential as a non-invasive alternative to current modalities.

The outcomes of the asymptomatic patients were not statistically significantly different from the patients that were evaluated as fixed at the time of revision surgery. Referring to the KOOS outcomes, the asymptomatic patients did not have complaints, but it cannot be ruled out that these patients had a loose tibial TKA component. After all, ten out of 38 asymptomatic knees would have been diagnosed as loose if applying the thresholds for rotation about the screw-axis, mTRE and MTPM. Nevertheless, it is not yet established that a loose TKA may go undetected because of the absences of symptoms and additionally, the reported thresholds should be externally validated before clinical use.

Testing this technique against greater sets of databases enabling separate analysis for different types of implants could potentially increase and individualize the diagnostic accuracy of this method. It would be beneficial for future research to include loosening of the femoral component and long-term follow-up of patients diagnosed with stable implants to confirm the absence of progressive loosening and to ensure that the method not only has diagnostic accuracy but also predictive validity. As added value, TKA component migration over time can be measured and might serve as a less invasive alternative for model and/or marker based radiostereometric analysis.

Despite the positive findings, this study is not without limitations. First and foremost, the sample size needs to be expanded in future research to improve the robustness of the calculated thresholds. The current study evaluated the diagnostic accuracy in a diverse patient group with prostheses of different brands, materials, and sizes, and reported promising accuracy results. The current study did not differentiate between different TKA designs or types of fixations. To provide more detailed information on the extent of loosening for these different types, future studies should include a detailed evaluation of various implant designs, alignment, and fixation techniques. Secondly, our current study involved 3D image analysis. During this process, we segmented and registered the whole tibia. The recent cadaveric study by Ter Wee

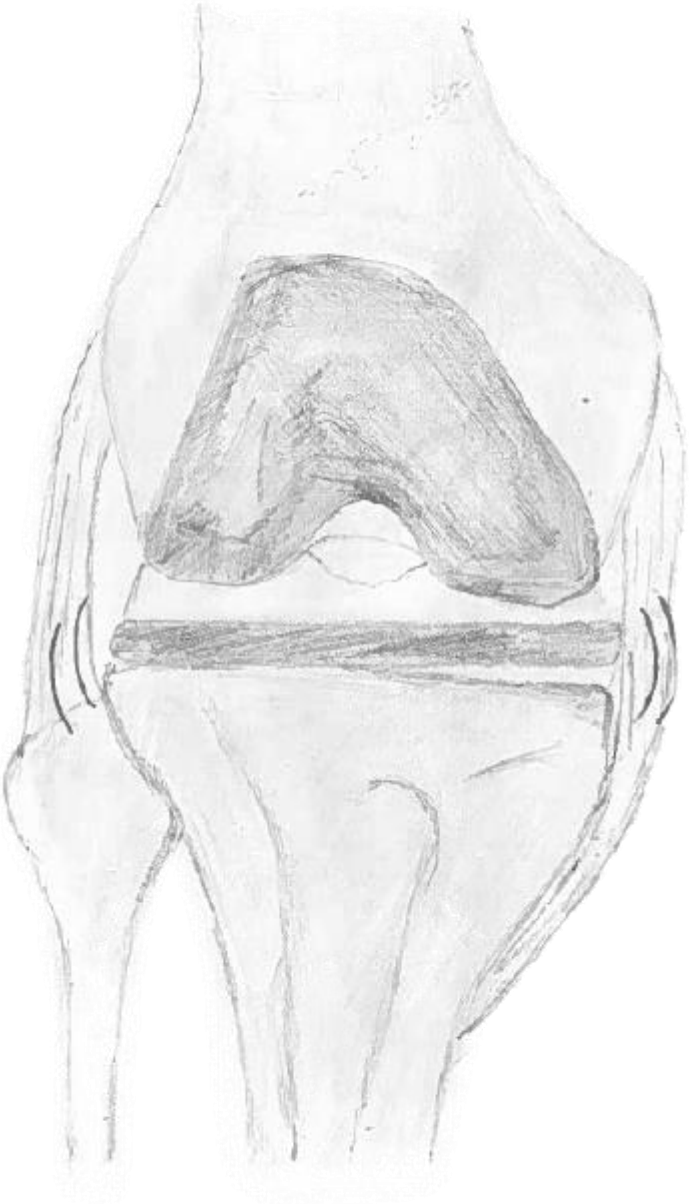
et al. indicates that because of tibial deformation under loading, tibial component displacement is overestimated if the entire tibia is segmented [22]. Thirdly, the comparison to the intraoperative “gold standard” does not fully address the potential for surgeon bias or variation in intraoperative assessment of the fixation status of the tibial component. Lastly, while patients with either tibial or femur component loosening report with the same complaints (e.g., weight bearing pain), this study only addresses potential loosening of the tibial component (20.4% of the reasons for revision) yet neglects possible loosening of the femoral component (8.8% of reasons for rTKA) [10, 11].

Conclusion

This non-invasive method, which measures induced displacement of the tibial TKA component has potential as a safe, feasible, reliable, and accurate diagnostic tool to aid the diagnosis of aseptic TKA loosening.

References

1. Barnsley L, Barnsley L (2019) Detection of aseptic loosening in total knee replacements: a systematic review and meta-analysis. Skeletal Radiology 48:1565-1572
2. Bartlett JW, Frost C (2008) Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in Obstetrics & Gynecology 31:466-475
3. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310
4. Buijs GS, Kievit AJ, Ter Wee MA, Magg C, Dobbe JGG, Streekstra GJ, et al. (2024) Non-invasive quantitative assessment of induced component displacement can safely and accurately diagnose tibial component loosening in patients: A prospective diagnostic study. Knee Surg Sports Traumatol Arthrosc;10.1002/ksa.12299
5. Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis.10.1002/ksa.12206. Knee Surg Sports Traumatol Arthrosc2024.
6. Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis.10.1002/ksa.12206. Knee Surg Sports Traumatol Arthrosc 2024.
7. Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284
8. Hext RA, Kaptein BL, Howard JL, Lanting BA, Teeter MG (2024) Inducible displacement of cementless total knee arthroplasty components with conventional and weight-bearing CT-based radiostereometric analysis. J Orthop Res;10.1002/jor.26017
9. ICMJE. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. https://www.icmje.org/icmje-recommendations.pdf.
10. Kievit AJ, Buijs GS, Dobbe JGG, Ter Wee A, Kerkhoffs G, Streekstra GJ, et al. (2023) Promising results of a non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon) 104:105930
11. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine 15:155-163
12. Magg C, ter Wee M, Buijs G, Kievit A, Krap D, Dobbe JG, et al. *Towards automation in non-invasive measurement of knee implant displacement*. Vol 12927: SPIE; 2024.
13. Mokkink LB, de Vet H, Diemeer S, Eekhout I (2023) Sample size recommendations for studies on reliability and measurement error: an online application based on simulation studies. Health Services and Outcomes Research Methodology 23:241-265
14. Mokkink LB, Eekhout I, Boers M, van der Vleuten CPM, de Vet HCW (2023) Studies on Reliability and Measurement Error of Measurements in Medicine - From Design to Statistics Explained for Medical Researchers. Patient Relat Outcome Meas 14:193-212
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 63:737-745
16. Niesen AE, Garverick AL, Hull ML (2021) Maximum Total Point Motion of Five Points Versus All Points in Assessing Tibial Baseplate Stability. J Biomech Eng 143:
17. Olivecrona H, Olivecrona L, Weidenhielm L, Noz ME, Kardos J, Maguire GQ, Jr., et al. (2008) A new technique for diagnosis of acetabular cup loosening using computed tomography: preliminary experience in 10 patients. Acta Orthop 79:346-353
18. Sandberg O, Carlsson S, Harbom E, Cappelen V, Tholén S, Olivecrona H, Wretenberg P (2022) Inducible displacement CT increases the diagnostic accuracy of aseptic loosening in primary total hip arthroplasty. Acta Orthop 93:831-836
19. Sandberg OH, Karrholm J, Olivecrona H, Rohrl SM, Skoldenberg OG, Broden C (2023) Computed tomography-based radiostereometric analysis in orthopedic research: practical guidelines. Acta Orthop 94:373-378
20. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420-428
21. Ter Wee MA, Dobbe JGG, Buijs GS, Kievit AJ, Schafroth MU, Maas M, et al. (2023) Load-induced deformation of the tibia and its effect on implant loosening detection. Sci Rep 13:21769
22. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, et al. (2021) Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 30:2729-2754
23. Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454
24. Wretenberg P, Tholen S, Carlsson S, Olivecrona H (2021) Implant Movement Analysis (IMA), A New CT Based Technique for Diagnosis of Aseptic Loosening of Total Knee Arthroplasty. Orthopedic Research Online Journal



Chapter IX

General discussion and future prospects

General discussion

The aim of this thesis was twofold: first, to evaluate the current state of diagnostics for aseptic knee arthroplasty loosening; and second, to propose and validate a new modality that measures induced micromovement of the tibial component as a potential diagnostic aid for aseptic knee arthroplasty loosening.

Based on a systematic review and meta-analysis, MRI and SPECT/CT currently appear to be the best imaging modalities to aid the diagnosis of knee arthroplasty loosening **(Chapter II)**. MRI primarily provides detailed soft tissue contrast and can detect periprosthetic bone marrow edema, synovitis, and fluid collections, while SPECT/CT combines metabolic and anatomical imaging to identify increased periprosthetic bone turnover. The perceived superiority of SPECT/CT and MRI should be interpreted with caution due to significant heterogeneity across the included studies, broad confidence intervals, and a high risk of bias in most of the included studies, leading to a generally low certainty of evidence [2, 6]. It is therefore very likely that an updated study in a few years might provide different results, as indicated by the 95% prediction region (**Chapter II, Figure 4)**.

When comparing diagnostic modalities for knee arthroplasty loosening one should not solely focus on diagnostic accuracy, but should also consider factors such as radiation exposure, availability, and scan duration. SPECT/CT involves ionizing radiation with an effective dose of 6-8 mSv, which is comparable to several years of natural background radiation [1, 16]. In contrast, MRI does not involve radiation exposure and typically takes only 30-45 minutes, compared to the 1.5 to 3 hours required for a SPECT/CT scan. MRI is more widely available, especially in high-income countries. SPECT/CT scanners are less common and primarily found in larger hospitals due to high costs and the need for specialized equipment and radiopharmaceuticals. These factors, along with the shorter scan time, often make MRI the preferred choice, despite the availability challenges of the necessary technology.

Between the available additional imaging modalities, CT, however, remains the most widely available and low-cost imaging modality. It is less expensive, more accessible, and faster than both SPECT/CT and MRI, with a typical radiation exposure of approximately 1 mSv for a standard knee scan. Despite these advantages, CT was not included in the review in **Chapter II** because the search identified only a single study assessing its diagnostic accuracy for aseptic knee arthroplasty loosening. This study by Foti et al. was excluded from the meta-analysis because it was not possible to extract true positive, true negative, false positive, and false negative values [8].

The limited identification of CT studies in this review is surprising, especially considering that **Chapter IV** highlighted a consensus on four radiological signs indicative of aseptic knee

arthroplasty loosening—namely, obvious implant migration, subsidence, progressive radiolucency, and radiolucency greater than 2 mm on radiographs or CT [4].

In recent history, bone scintigraphy was the most used additional diagnostic modality in our academic hospital. In **Chapter III,** its in-practice diagnostic performance was evaluated. The retrospective evaluation revealed that bone scintigraphy has a diagnostic accuracy of 84% for detection of aseptic loosening when evaluated by an orthopedic surgeon, compared to 69% when evaluated by a nuclear physicist. There was moderate agreement between the orthopedic surgeons' and nuclear physicist's pre-operative prediction of TKA component looseness. However, these diagnostic accuracy results should be interpreted with caution due to the unblinded, retrospective study design, which inherently includes all aspects of the complete diagnostic process, along with the associated biases (e.g. selection bias, observer bias, recall bias, and information bias).

The results of this study showed poorer specificity when the nuclear physicist interpreted the bone scan, compared to the orthopedic surgeon. This implies that bone scintigraphy, as interpreted by the nuclear physicist, is more likely to identify patients who truly have aseptic loosening but also incorrectly labels some patients without aseptic loosening as having an aseptic loosened prosthesis (false positives). As bone scintigraphy merely displays indirect signs of aseptic loosening (e.g., bone remodeling), a tendency to interpret the results in a way that aligns with the expected diagnosis is likely. Such false positives were less frequent with the orthopedic surgeon's interpretation. It is very likely that superior accuracy arises from their ability to access a more extensive range of clinical details.

The results of this practical evaluation of bone scintigraphy align with the findings of the studies included in the meta-analysis of **Chapter II**, although direct comparison to these studies is challenging [3, 6]. This challenge stems from differences in study design and insufficiently detailed methodologies. For example, in the study by Claassen et al., it is unclear who interpreted the bone scans and whether this was done in a blinded manner [7]. Similarly, in the study by Smith et al., it is noted that a single experienced radiologist evaluated the bone scintigraphy. However, the authors also acknowledge that ensuring the evaluation was objective proved challenging, as quantification was deemed to offer little additional value [17]. As a result, the radiologist made a broad distinction between "completely normal" and "increased uptake". While this approach is methodologically sound and practical, it may limit the real-world applicability of the results, where diagnostic tools like bone scintigraphy often lead to incomplete or uncertain diagnosis. The lack of standardized interpretation and the subjective nature of these assessments limit the generalizability of the findings to everyday clinical scenarios, particularly because increased uptake around the prosthesis can either still be physiological or be associated with various alternative diagnoses in patients with a painful knee post-TKA, such as (low-grade) infection.

Due to insufficient power, the study in **Chapter III** was unable to adequately assess the value of adding SPECT/CT to the diagnostic pathway to correctly diagnose aseptic loosening [3]. Among the three studies included in **Chapter II** that evaluated the diagnostic performance of bone scintigraphy, the study by Mandegaran et al. is likely the most insightful and relevant for addressing this question [15]. Their research directly compared the diagnostic performance of bone scintigraphy with SPECT/CT. Specifically for confirming or ruling out aseptic loosening, they found that SPECT/CT had better sensitivity (100%% vs. 89%) and specificity (75% vs. 30%) compared to bone scintigraphy [6, 15]. However, considering its high risk of bias for the patient population, reference test and flow and timing, one should interpret these results with caution [6]. The conflicting results, further highlighted by the wide 95% probability intervals in **Chapter II, Figure 4**, and the varying outcomes reported by different types of evaluators of these modalities, particularly emphasize that there is no clear answer to the question of how effective these modalities are in detecting or ruling out aseptic knee arthroplasty loosening.

Building on the methodological and data limitations identified in **Chapters II and III**, the studies presented in **Chapters IV and V** were conducted to assess the extent of the observed variability in diagnostic criteria and thresholds for the most used reference test, and to attempt to reduce this variability by finding consensus. **Chapter IV** objectified the perceived variability of clinical and radiological criteria fitting the diagnosis of aseptic loosening. **Chapter V** attempted to address one of the critical limitations highlighted in **Chapter II**, the inconsistent use of reference standards in diagnostic accuracy studies. As **Chapter II, Appendix 2** shows, many included studies lacked or employed various thresholds for the reference test, leading to a high risk of bias and questionable reliability and comparability of the results. Using a Delphi consensus method, visible fluid motion at the interface observed during specific movements such as flexion, extension, varus, or valgus stress or when gently applying direct force with a finger or instrument was identified as the threshold that received consensus. However, it is important to note that the method used to establish this definition is not intended to determine a definitive clinical judgment on when a prosthesis is clinically and functionally loose. Instead, it provides a standardized definition that can serve as a universal threshold, improving the comparability of future studies.

Contemplating on the findings described in **Chapter I - V**, it can be concluded that most modalities used to enhance the diagnosis of knee arthroplasty loosening merely show indirect signs of loosening. This conclusion is reinforced by the American College of Radiology, which highlights that the results of currently available tests rely on indirect indicators. These tests are costly, insufficiently sensitive and specific, time-consuming, and can only suggest an increased likelihood of suspected implant loosening [10, 20]. To improve the diagnostic pathway for knee arthroplasty loosening, the studies described in **Part II** of this thesis were based on the hypothesis that demonstrating and quantifying the mobility of the prosthesis relative to the bone can be a direct measure of loosening if done in a clinically appropriate, precise, and

reproducible manner. With this aim, **Part II** of this thesis proposes and investigates the AtMoves Knee System.

The AtMoves Knee System is non-invasive, as precise as model- marker based RSA (**Chapter IV**), proven feasible, safe for clinical use and non-implant specific [12, 19]. Besides the AtMoves Knee Systems loading device and software, all it requires is a CT-scanner, which is usually available at hospitals and clinics where knee arthroplasties are performed.

The studies presented in **Part II** of this thesis, together with other studies conducted by our research group [14, 18], provide a very thorough validation of this novel diagnostic technique. When comparing the AtMoves Knee System to a similar modality, like, for example Implant Movement Analysis (IMA) by Sectra B.V. [21], the pitfalls of the AtMoves Knee System are well known and accounted for. The AtMoves Knee System uses a patented loading device with a four-point bending mechanism **(Chapter VI)**, which ensures a 20Nm load over the index knee with negligible operator variance when applied by different trained users (**Chapter VII**) [13]. Sectra's IMA uses only cushions and tension traps to load the leg in two different directions, in each of which a CT scan is performed [21]. Segmentation and registration produce two superimposed images that are presented to a radiologist. This system likely does not or cannot quantify the induced implant movement, primarily because of its undiscussed yet well-known limitations, such as unknown and untested reproducibility. As a result, the final diagnosis remains subjective, unsupported, and difficult to validate thoroughly. We are particularly proud that we have been able to provide and share this thorough validation in leading orthopedic scientific journals.

AtMoves Knee System is the first technique employed to evaluate induced implant motion in both symptomatic patients awaiting revision surgery to diagnose aseptic loosening and asymptomatic patients (**Chapter VIII**) [5]. The results were notably insightful. Applying all the thresholds reported in **Chapter VIII** to the three different clinical parameters of induced implant motion (i.e., rotation about the screw-axis, mTRE, and MTPM), ten out of 38 asymptomatic knees would have been diagnosed as loose, while at the same time, worse patient reported outcome measures showed a statistically significant correlation with increased induced implant motion [5]. As this is the first study to report induced movement in a standardized and reproducible manner, one can only speculate on the reasons for these findings. It may be explained by the fact that no standardized threshold was used for the reference test, as provided in **Chapter V** of this thesis. In addition, the study reported in **Chapter VIII** did not differentiate between different TKA designs or types of fixations. To provide more detailed information on the extent of loosening for these different types, future studies should include larger cohorts and a detailed evaluation of different implant designs, alignment, and fixation techniques to see if implant type or fixation specific thresholds can be identified. Artificial intelligence and machine learning models can play a significant role in addressing this issue.

Another question that remains partially unanswered is the most optimal segmentation strategy. A recent study by Ter Wee et al. suggests that the tibia deforms under the applied load with the AtMoves Knee System's Loading device, resulting in generally larger displacement measurements, which may result in an overestimation of the actual induced implant displacement, as it introduces a registration error [18]. This registration error is the result of the valgus deformed segmented geometry of the tibia being registered to a varus deformed segmented tibia. They propose that segmenting only 20% of the tibia, instead of 100%, reduces this effect with only a minimal increase in methodological error.

Both protocols are also investigated in **Chapter VII** of this thesis, where the potential variance introduced by the loading device operator was examined. This study also found that the standard error of measurement decreases when the 20% segmentation protocol is applied. Segmenting the tibial bone as a reference is necessary, but it remains unclear what the most ideal cutoff value or balance is between a greater systematic error or a greater influence of tibial deformation. Future studies will need to address this question, using finite element analyses and advanced simulations to provide a definitive answer. Furthermore, it is crucial to assess whether this updated segmentation protocol also improves the diagnostic accuracy of the AtMoves Knee System. It is possible that tibial deformation and the resulting measurements could incorporate predictive factors such as bone density, that might be neglected when excluding tibial deformation. It is imperative that for such a study a standardized reference test as proposed in **Chapter V** is employed.

Future prospects

Although the studies presented in this thesis offer a comprehensive evaluation of the current state of diagnostics for knee arthroplasty loosening and the diagnostic performance of the AtMoves Knee System, several challenges persist, and promising directions for future research remain.

The diagnostic accuracy of the AtMoves Knee System was evaluated and presented in **Chapter VIII** of this thesis. In the study detailed in this chapter, diagnostic accuracy was derived by dichotomizing three different continuous variables into categories of "loose" or "fixed." The thresholds used were optimized for sensitivity and specificity based on the cohort included in this study. However, before these thresholds can be extrapolated to other populations or applied in clinical practice, they must be validated in an external study cohort.

A significant issue with current diagnostic modalities is their high cost, coupled with persistent uncertainty in achieving an accurate diagnosis. This uncertainty can lead to an increased risk of delayed or incorrect diagnoses, potentially resulting in unnecessary revision surgeries. The costs related to these diagnostic modalities and unwarranted revision surgeries are hypothesized to be substantial. The introduction of the AtMoves Knee System into the diagnostic pathway for patients suspected of component loosening is proposed to reduce these costs. First, by reducing the costs of the diagnostic pathway, as employing a diagnostic AtMoves Knee System scan is cheaper than other employed advanced imaging modalities. Second, by providing a more insightful and direct evidence for aseptic loosening, if present, thereby potentially reducing the number of false positives and false negatives. To validate this hypothesis, an open label randomized controlled trial should be conducted, comparing one arm using the AtMoves Knee System with another arm that does not. The primary outcome should be the frequency of correct treatment outcomes in each arm. Secondly, it should be assessed whether the availability of the AtMoves Knee System effectively reduces the reliance on other more invasive and expensive diagnostic modalities. Additionally, it is important to evaluate whether its use shortens the time interval from the initial presentation of symptoms to an accurate diagnosis.

The reported incidence of tibial component loosening is higher than that of femoral component loosening (20.4% vs. 8.8% of the reasons for revision surgery) [9]. The consensus study **(Chapter IV)** concluded that weight-bearing pain is a clinical criterion for diagnosing loosening of both the femoral and tibial components, indicating that patients with loosening of the tibial, femoral, or both components present with the same complaint. This thesis specifically investigated the diagnostic accuracy and precision of the tibial component but did not assess the femoral component, despite the possibility that patients may have presented with femoral component loosening. Besides this being a limitation of the studies included in this thesis, it is a remaining

objective for our research team to repeat those studies for the femoral component and for the AtMoves Knee System to be able to provide a full analysis of potentially loose knee components.

With increased life expectancy, both the number of people requiring joint replacement and the number of joint replacements—per joint as well as per individual—pose a significant challenge. As the number of primary procedures increases, the demand for diagnostics will also rise. It is therefore important to expand this potentially cost-effective and efficient technique to other joints, prioritized by prevalence and feasibility. In this context, the development of a loading device for the hip joint has been initiated and the applicability of the AtMoves Knee System for unicondylar knee arthroplasties is currently under evaluation. Furthermore, building on the findings of this thesis, efforts have begun to establish, identify, and describe a universally recognized gold standard for intraoperatively determined hip prothesis loosening.

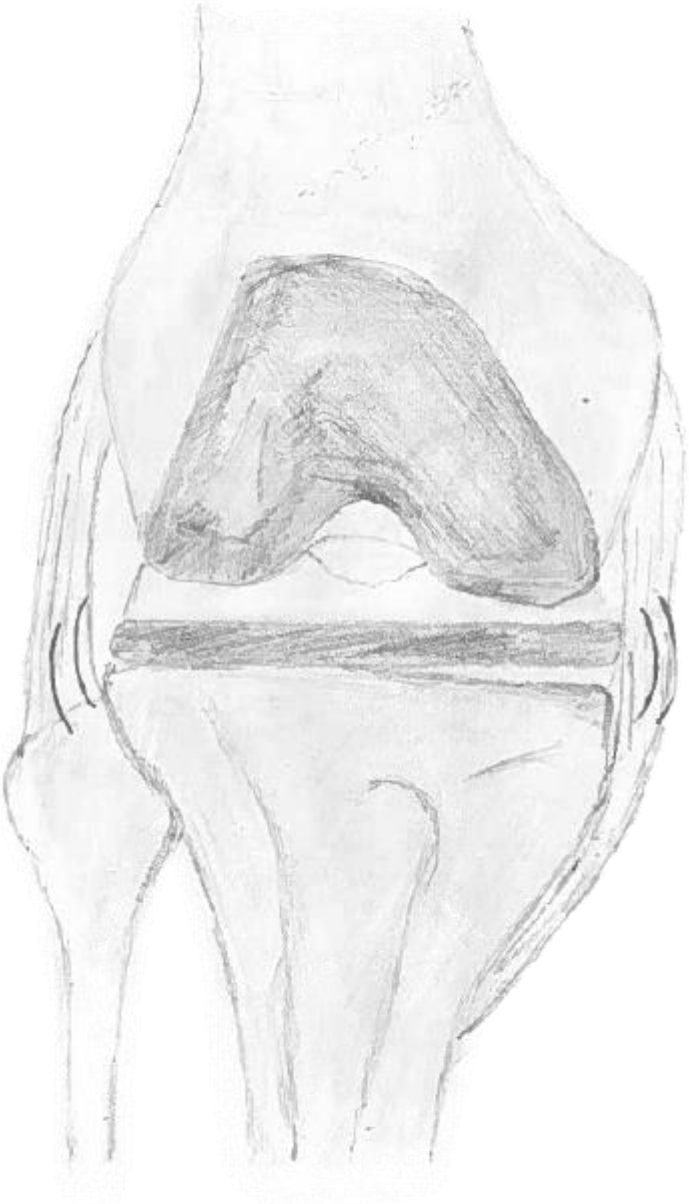
Given the good availability of CT scanners, low invasiveness, and high precision, the AtMoves Knee System has the potential to replace model- or marker-based RSA as the gold standard for migration measurement as employed in current day pre- and post-market implant follow-up studies [11]. A relatively simple measurement of induced implant movement, followed by a subsequent measurement after a set period, may offer significantly greater insight into the degree of fixation than repeated static measurements.

Building on this concept, we are currently conducting studies within a cross-sectional cohort of patients. Interestingly, even among asymptomatic patients, we observe a considerable degree of implant mobility. These findings suggest that, in the future, standard follow-up using conventional X-rays could potentially be replaced by an AtMoves scan. By establishing baseline measurements and monitoring changes over time, this approach could serve as an early indicator of potential prosthesis loosening. Early detection of loosening could significantly shorten the diagnostic process, preserve the remaining bone stock before revision surgery, and, most importantly, help maintain the quality of life for patients with knee prostheses.

While the AtMoves Knee System presents a promising alternative to the less sensitive, costly, and often debated options currently available, it is important not to focus solely on this solution. We should also explore other potentially more affordable and accessible innovations, such as AI and machine learning integrated with patient data—provided they are validated against a reproducible, widely accepted reference standard. These emerging technologies can be effectively integrated with insights gained from AtMoves scans, offering a more comprehensive approach to monitoring implant performance. However, it is crucial to recognize that more research is not always synonymous with better outcomes. Therefore, optimizing and standardizing the use of existing data should also be a key priority to maximize its value and ensure meaningful improvements in patient care.

References

1. *Alrehily F, Alshamrani AF (2023) Estimation of radiation dose associated with bone SPECT/CT and establishing local diagnostic reference levels using size-specific dose estimate. Journal of Radiation Research and Applied Sciences 16:100527*
2. *Brozek JL, Canelo-Aybar C, Akl EA, Bowen JM, Bucher J, Chiu WA, et al. (2021) GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence-An overview in the context of health decision-making. J Clin Epidemiol 129:138-150*
3. *Buijs GS, Kievit AJ, Schafroth MU, Blankevoort L (2024) An evaluation of the diagnostic performance of the triphasic bone scintigraphy in patients suspected of aseptic total knee arthroplasty loosening. Journal of Orthopaedics;https://doi.org/10.1016/j.jor.2024.08.007*
4. *Buijs GS, Kievit AJ, Schafroth MU, Hirschmann MT, Blankevoort L. Weight-bearing pain and implant migration, progressive radiolucencies, radiolucency more than 2 mm and subsidence on radiographs and CT are generally accepted criteria for knee arthroplasty loosening: An international Delphi consensus study2024, Knee Surg Sports Traumatol Arthrosc.*
5. *Buijs GS, Kievit AJ, Ter Wee MA, Magg C, Dobbe JGG, Streekstra GJ, et al. (2024) Non-invasive quantitative assessment of induced component displacement can safely and accurately diagnose tibial component loosening in patients: A prospective diagnostic study. Knee Surg Sports Traumatol Arthrosc;10.1002/ksa.12299*
6. *Buijs GS, Kooijenga AC, Rikken RGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening; a systematic comparative diagnostic test review and meta-analysis.10.1002/ksa.12206. Knee Surg Sports Traumatol Arthrosc 2024.*
7. *Claassen L, Ettinger M, Plaass C, Daniilidis K, Calliess T, Ezechieli M (2014) Diagnostic value of bone scintigraphy for aseptic loosening after total knee arthroplasty. Technol Health Care 22:767-773*
8. *Foti G, Longo C, D'Onofrio M, Natali S, Piovan G, Oliboni E, et al. (2023) Dual-Energy CT for Detecting Painful Knee Prosthesis Loosening. Radiology 306:e211818*
9. *Gademan MGJ, Van Steenbergen LN, Cannegieter SC, Nelissen R, Marang-Van De Mheen PJ (2021) Population-based 10-year cumulative revision risks after hip and knee arthroplasty for osteoarthritis to inform patients in clinical practice: a competing risk analysis from the Dutch Arthroplasty Register. Acta Orthop 92:280-284*
10. *Hochman MG, Melenevsky YV, Metter DF, Roberts CC, Bencardino JT, Cassidy RC, et al. (2017) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty. Journal of the American College of Radiology 14:S421-S448*
11. *Kaptein BL, Pijls B, Koster L, Kärrholm J, Hull M, Niesen A, et al. (2024) Guideline for RSA and CT-RSA implant migration measurements: an update of standardizations and recommendations. Acta Orthop 95:256-267*
12. *Kärrholm J (1989) Roentgen stereophotogrammetry. Review of orthopedic applications. Acta Orthop Scand 60:491-503*
13. *Kievit AJ, Buijs GS, Dobbe JGG, Ter Wee A, Kerkhoffs G, Streekstra GJ, et al. (2023) Promising results of an non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon) 104:105930*
14. *Magg C, ter Wee M, Buijs G, Kievit A, Krap D, Dobbe JG, et al. Towards automation in non-invasive measurement of knee implant displacement. Vol 12927: SPIE; 2024.*
15. *Mandegaran R, Agrawal K, Vijayanathan S, Gnanasegaran G (2018) The value of 99mTc-MDP bone SPECT/CT in evaluation of patients with painful knee prosthesis. Nucl Med Commun 39:397-404*
16. *Mathis DT, Tschudi S, Amsler F, Hauser A, Rasch H, Hirschmann MT (2022) Correlations of typical pain patterns with SPECT/CT findings in unhappy patients after total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 30:3007-3023*
17. *Smith SL, Wastie ML, Forster I (2001) Radionuclide Bone Scintigraphy in the Detection of Significant Complications after Total Knee Joint Replacement. Clinical Radiology 56:221-224*
18. *Ter Wee MA, Dobbe JGG, Buijs GS, Kievit AJ, Schafroth MU, Maas M, et al. (2023) Load-induced deformation of the tibia and its effect on implant loosening detection. Sci Rep 13:21769*
19. *Valstar ER, Vrooman HA, Toksvig-Larsen S, Ryd L, Nelissen RG (2000) Digital automated RSA compared to manually operated RSA. J Biomech 33:1593-1599*
20. *Walker EA, Fox MG, Blankenbaker DG, French CN, Frick MA, Hanna TN, et al. (2023) ACR Appropriateness Criteria® Imaging After Total Knee Arthroplasty: 2023 Update. J Am Coll Radiol 20:S433-s454*
21. *Wretenberg P, Tholen S, Carlsson S, Olivecrona H. Implant Movement Analysis (IMA), A New CT Based Technique for Diagnosis of Aseptic Loosening of Total Knee Arthroplasty.10.31031/OPROJ.2021.08.000690. Orthopedic Research Online Journal: Crimson Publishers; 2021.*



Appendices

Summary

Nederlandse samenvatting

List of publications

PhD Portfolio

Dankwoord

Summary

Knee arthroplasty is a commonly used surgical procedure to relieve pain and restore mobility in patients with advanced osteoarthritis. However, the need for revision surgeries remains a concern, with 13% of patients requiring a revision within 10 years. The most common causes for revision are aseptic loosening, periprosthetic infection, and mechanical instability. Identifying and improving diagnostic modalities for aseptic loosening is crucial to prevent unnecessary revisions and ensure timely and accurate treatment for patients. To this end, the aim of this thesis was twofold: first, to evaluate the current state of diagnostics for aseptic knee arthroplasty loosening; and second, to propose and validate a new modality that measures induced micromovement of the tibial component as a diagnostic aid for aseptic knee arthroplasty loosening.

Part I: Current State of Diagnostics

In **Chapter II**, the diagnostic accuracy of various modalities for diagnosing aseptic loosening of knee arthroplasties is evaluated and compared. A systematic review and meta-analysis following Cochrane and PRISMA guidelines was conducted, including 14 studies. These studies examined diagnostic tools such as bone scintigraphy, FDG-PET-CT, SPECT/CT, radionuclide arthrogram, and MRI. The results showed that MRI and SPECT/CT had the highest diagnostic accuracy, although the overall certainty of the evidence was low due to bias, particularly in patient selection and reference tests. The findings highlight the need for more reliable and standardized diagnostic methods.

In **Chapter III**, the diagnostic accuracy of bone scintigraphy for detecting aseptic loosening in total knee arthroplasties is evaluated. The study involved 59 patients who underwent revision surgery for suspected aseptic loosening. Bone scintigraphy evaluations were performed by a nuclear physicist and an orthopedic surgeon. The orthopedic surgeon had higher diagnostic accuracy (84%) compared to the nuclear physicist (69%), with moderate kappa agreement between the two. This emphasizes the challenges and variability in interpreting bone scintigraphy results and suggests that having a comprehensive view of the patient and outcomes from other modalities significantly impacts diagnostic accuracy.

In **Chapter IV**, the variability in clinical and radiological criteria for diagnosing aseptic loosening of knee arthroplasties is investigated. Through a consensus method, an international panel of specialized knee revision surgeons was asked to identify criteria they consider appropriate for diagnosing aseptic loosening. High variability was found in the criteria, with only a few reaching consensus. Weight-bearing pain was recognized as a crucial clinical criterion, while radiological criteria such as implant migration, progressive radiolucencies, prosthesis subsidence, and

radiolucencies greater than 2 mm on X-ray or CT were deemed important. This underscores the need for standardized diagnostic protocols.

In **Chapter V**, the lack of a uniform definition for intraoperative assessment of knee arthroplasty component loosening is addressed. The absence of such a definition complicates the interpretation of diagnostic studies where this is used as a reference test. A Delphi consensus method was employed, where a panel of specialized knee revision surgeons reached consensus on the minimum criteria for defining intraoperative component loosening. The panel agreed that visible fluid movement at the interface during specific movements or when gentle force is applied should define a component as loose. This new consensus provides a standardized reference for future diagnostic research.

Part II: Evaluation of Induced Implant Movement

In **part II**, a new non-invasive method for detecting aseptic loosening is introduced and evaluated, the AtMoves Knee System.

The AtMoves Knee System is an advanced diagnostic tool designed to accurately and non-invasively detect aseptic loosening of the tibial component in knee arthroplasties. By applying standardized load on the knee during a CT scan, the system measures and analyzes the induced micromovement of the tibial component using specialized software. This software quantifies the degree of implant displacement and rotation, using parameters such as rotation about the screw-axis, mean target registration error (mTRE), and maximum total point motion (MTPM).

In **Chapter VI**, the reproducibility and reliability of implant displacement measurements using the AtMoves Knee System are evaluated. Ten cadaveric knees were implanted with loose tibial components and scanned using the AtMoves Knee System's loading device. The study quantified displacement from different positions within the CT scanner and rotation differences between loose and fixed conditions, demonstrating that the method is both reproducible and reliable. The findings suggest that this non-invasive technique could be a valuable tool for diagnosing tibial component loosening in clinical settings.

In **Chapter VII**, the inter-operator reliability of the AtMoves Knee System is assessed. The study involved 16 patients and showed that the system exhibited good reliability between different operators, with minimal measurement errors. The results indicate that the system can be reliably operated by various trained operators, making it a viable option for routine clinical use.

In **Chapter VIII**, the clinical safety, feasibility, and accuracy of the AtMoves Knee System are evaluated. The study involved 72 patients, both symptomatic and asymptomatic, and assessed the discomfort, side effects, and diagnostic accuracy of the system. The results showed that the system is safe, with a median discomfort score of 6/10 in symptomatic patients and 2/10 in

asymptomatic patients, with no reported side effects. The diagnostic accuracy was high, with a sensitivity of 0.91 and specificity of 0.72 when comparing the AtMoves Knee System outcomes with intraoperative findings during revision surgery. The study concludes that the AtMoves Knee System is a promising tool for accurately diagnosing aseptic loosening in knee arthroplasty patients.

In **Chapter IX**, the main findings of the thesis are discussed, highlighting the limitations of current diagnostic methods for aseptic loosening of knee arthroplasties and the potential of the AtMoves Knee System as a new diagnostic tool. This chapter emphasizes the potential of the AtMoves Knee System, but also discusses the need for further research to refine and validate this method in larger clinical populations and to explore its applicability to other components of the knee prosthesis.

Nederlandse samenvatting

Knieprothese (arthroplastiek) is een veelgebruikte chirurgische ingreep om pijn te verlichten en mobiliteit te herstellen bij patiënten met vergevorderde artrose. Echter, de noodzaak voor revisie-operaties blijft een belangrijk aandachtspunt, aangezien 13% van de patiënten binnen 10 jaar een revisie nodig heeft. De meest voorkomende oorzaken voor revisie zijn aseptische loslating, peri prothetische infectie, en mechanische instabiliteit. Het identificeren en verbeteren van diagnostische modaliteiten voor aseptische loslating is essentieel om onnodige revisies te voorkomen en patiënten tijdig en juist te behandelen. Om deze reden was het doel van deze thesis tweeledig: ten eerste om de huidige stand van diagnostiek voor aseptische loslating van knieprothesen te evalueren; en ten tweede om een nieuwe modaliteit voor te stellen en te valideren die geïnduceerde micromobiliteit van de tibia component meet als een potentieel diagnostisch hulpmiddel voor aseptische loslating van knieprothesen.

Deel I: Huidige stand van de diagnostiek

In **hoofdstuk II** wordt de diagnostische nauwkeurigheid van verschillende modaliteiten voor de diagnose van aseptische loslating van knieprothesen geëvalueerd en vergeleken. Een systematische review en meta-analyse volgens Cochrane- en PRISMA-richtlijnen werd uitgevoerd, waarbij 14 studies werden geïncludeerd. Deze studies onderzochten diagnostische hulpmiddelen zoals botscintigrafie, FDG-PET-CT, SPECT/CT, radionuclide arthrogram en MRI. De resultaten toonden aan dat MRI en SPECT/CT de hoogste diagnostische nauwkeurigheid hadden, hoewel de algehele zekerheid van het bewijs laag was door bias, vooral in de patiëntselectie en referentietests. De bevindingen onderstrepen de noodzaak voor betrouwbaardere en gestandaardiseerde diagnostische methoden.

In **hoofdstuk III** wordt de diagnostische nauwkeurigheid van botscintigrafie voor de detectie van aseptische loslating bij totale knieprothesen geëvalueerd. Het betrof een studie met 59 patiënten die een revisieoperatie ondergingen voor vermoedelijke aseptische loslating. Botscintigrafie-beoordelingen werden uitgevoerd door een nucleair fysicus en een orthopedisch chirurg. De orthopedisch chirurg had een hogere diagnostische nauwkeurigheid (84%) vergeleken met de nucleair fysicus (69%), met een matige kappa-overeenkomst tussen de twee. Dit benadrukt de uitdagingen en variabiliteit bij de interpretatie van botscintigrafie-resultaten en suggereert dat het hebben van een volledig beeld van de patiënt en de uitkomsten van overige modaliteiten een aanzienlijke impact hebben op de diagnostische nauwkeurigheid.

In **hoofdstuk IV** wordt de variabiliteit in klinische en radiologische criteria voor de diagnose van aseptische loslating van knieprothesen onderzocht. Middels een consensus methode werd een internationaal panel van gespecialiseerde knie revisie chirurgen gevraagd naar criteria die zij

passend achten bij de diagnose aseptische loslating. Er werd een hoge variabiliteit gevonden in de criteria, met slechts enkele criteria die consensus bereikten. Gewichtdragende pijn werd als klinisch criterium erkend. Radiologische criteria zoals implantaatmigratie, progressieve radiolucenties, verzakking van de prothese en radiolucenties groter dan 2 mm op röntgenfoto of CT werden ook als belangrijk beschouwd. Dit onderstreept de noodzaak van gestandaardiseerde diagnostische protocollen.

In **hoofdstuk V** wordt het gebrek aan een uniforme definitie voor intra-operatieve beoordeling van loslating van knieprothese-componenten aangekaart. Het gebrek hieraan compliceert de interpretatie van diagnostische studies waarin dit als referentietest wordt toegepast. Een Delphi-consensusmethode werd gebruikt, waarbij een panel van gespecialiseerde knie-revisiechirurgen consensus bereikte over minimale criteria voor het definiëren van intra-operatieve componentloslating. Het panel was het eens dat zichtbare vloeistofbeweging op de interface tijdens specifieke bewegingen of bij het voorzichtig aanbrengen van kracht, een component als los zou moeten definiëren. Deze nieuwe consensus biedt een gestandaardiseerde referentie voor toekomstig diagnostisch onderzoek.

Deel II: Evaluatie van geïnduceerde implantaatbeweging

In **deel II** wordt een nieuwe niet-invasieve methode voor de detectie van aseptische loslating gepresenteerd en geëvalueerd, het AtMoves Knee System.

Het AtMoves Knee System is een geavanceerd diagnostisch hulpmiddel voor het nauwkeurig en niet-invasief detecteren van aseptische loslating van de tibiale component in knieprothesen. Door tijdens een gestandaardiseerde belasting op de knie toe te passen tijdens het verrichten van een CT-scan, worden verplaatsingen van de tibiale component gemeten en geanalyseerd met behulp van gespecialiseerde software. Deze software kwantificeert de mate van geïnduceerde implantaat verplaatsing en rotatie, gebruikmakend van parameters zoals rotatie om de schroefas, gemiddelde puntverplaatsing (mTRE), en maximale puntverplaatsing (MTPM).

In **hoofdstuk VI** wordt de reproduceerbaarheid en betrouwbaarheid van implantaatverplaatsingsmetingen middels het AtMoves Knee System, geëvalueerd. Tien kadaverknieën werden geïmplanteerd met loszittende tibia componenten en gescand met behulp van het AtMoves Knee System's belastingsapparaat. De studie kwantificeerde verplaatsing van een vaste component vanuit verschillende posities in de CT-scanner en rotatieverschillen tussen losse en vaste condities, wat aantoonde dat de methode zowel reproduceerbaar als betrouwbaar is. De bevindingen suggereren dat deze niet-invasieve techniek een waardevol hulpmiddel kan zijn voor het diagnosticeren van tibia componentloslating in klinische settings.

In **hoofdstuk VII** wordt de inter-operator betrouwbaarheid van het AtMoves Knee System geëvalueerd. De studie omvatte 16 patiënten en toonde aan dat het systeem een goede betrouwbaarheid vertoonde tussen verschillende operators, met slechts lage meetfouten. De resultaten geven aan dat het systeem betrouwbaar kan worden bediend door verschillende getrainde operators, wat het een haalbare optie maakt voor routinematig klinisch gebruik.

In **hoofdstuk VIII** wordt de klinische veiligheid, haalbaarheid en nauwkeurigheid van het AtMoves Knee System beoordeeld. De studie betrof 72 patiënten, zowel symptomatisch als asymptomatisch, en evalueerde het ongemak, de bijwerkingen en de diagnostische nauwkeurigheid van het systeem. De resultaten toonden aan dat het systeem veilig is, met een mediane ongemaksscore van 6/10 bij symptomatische patiënten en 2/10 bij asymptomatische patiënten, zonder gerapporteerde bijwerkingen. De diagnostische nauwkeurigheid was hoog, met een sensitiviteit van 0,91 en specificiteit van 0,72 wanneer de uitkomsten van het AtMoves Knee System vergeleken werden met de intra-operatieve bevindingen tijdens revisiechirurgie. De studie concludeert dat het AtMoves Knee System een veelbelovend hulpmiddel is voor het nauwkeurig diagnosticeren van aseptische loslating bij knieprothese patiënten.

In **hoofdstuk IX** worden de belangrijkste bevindingen van de thesis besproken en bediscussieerd, waarbij de beperkingen van huidige diagnostische methoden voor aseptische loslating van knieprothesen en het potentieel van het AtMoves Knee System als nieuw diagnostisch hulpmiddel worden besproken. Dit hoofdstuk benadrukt de potentie van het AtMoves Knee System, maar bespreekt ook de noodzaak voor verder onderzoek om deze methode te verfijnen en te valideren in grotere klinische populaties en om de toepasbaarheid ervan op andere componenten van de knieprothese te onderzoeken.

List of publications

Peer-reviewed publications

Within the scope of this thesis

Kievit AJ, **Buijs GS**, Dobbe JGG, Ter Wee A, Kerkhoffs GMMJ, Streekstra GJ, Schafroth MU, Blankevoort L. Promising results of a non-invasive measurement of knee implant loosening using a loading device, CT-scans and 3D image analysis. Clin Biomech (Bristol, Avon). 2023 Apr;104:105930. doi: 10.1016/j.clinbiomech.2023.105930. Epub 2023 Mar 3.

Buijs GS, Kooijenga AC, Rikken QGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening: A systematic comparative diagnostic test review and meta-analysis. Knee Surg Sports Traumatol Arthrosc. 2024 Aug;32(8):2061-2074. doi: 10.1002/ksa.12206.

Buijs GS, Kievit AJ, Ter Wee MA, Magg C, Dobbe JGG, Streekstra GJ, Schafroth MU, Blankevoort L. Non-invasive quantitative assessment of induced component displacement can safely and accurately diagnose tibial component loosening in patients: A prospective diagnostic study. Knee Surg Sports Traumatol Arthrosc. 2024 May 31. doi: 10.1002/ksa.12299.

Buijs GS, Kievit AJ, Walinga AB, Schafroth MU, Hirschmann MT, Blankevoort L; International Consensus Panel. Visible fluid motion on manipulation as the new threshold for intraoperatively determined knee arthroplasty component loosening: A Delphi study. Knee Surg Sports Traumatol Arthrosc. 2024 Jul 15. doi: 10.1002/ksa.12357.

Buijs GS, Kievit AJ, Schafroth MU, Hirschmann MT, Blankevoort L; International Consensus Panel. Weight-bearing pain and implant migration, progressive radiolucencies, radiolucency more than 2 mm and subsidence on radiographs and CT are generally accepted criteria for knee arthroplasty loosening: An international Delphi consensus study. Knee Surg Sports Traumatol Arthrosc. 2024 Aug 10. doi: 10.1002/ksa.12419.

Caroline Magg, Maaike A. ter Wee, **George S. Buijs**, Arthur J. Kievit, Dennis A. Krap, Johannes G. G. Dobbe, Geert J. Streekstra, Leendert Blankevoort, Clara I. Sánchez, "Towards automation in non-invasive measurement of knee implant displacement," Proc. SPIE 12927, Medical Imaging 2024: Computer-Aided Diagnosis, 129270R (3 April 2024); https://doi.org/10.1117/12.3008090

Caroline Magg, Lukas P.E. Verweij, Maaike A. ter Wee, **George S. Buijs**, Johannes G.G. Dobbe, Geert J. Streekstra, Leendert Blankevoort, Clara I. Sánchez. Training-free Prompt Placement by

Propagation for SAM Predictions in Bone CT Scans. Published: 06 Jun 2024, Last Modified: 06 Jun 2024. MIDL 2024.

Ter Wee MA, Dobbe JGG, **Buijs GS,** Kievit AJ, Schafroth MU, Maas M, Blankevoort L, Streekstra GJ. Load-induced deformation of the tibia and its effect on implant loosening detection. Sci Rep. 2023 Dec 8;13(1):21769. doi: 10.1038/s41598-023-49177-z.

Outside the scope of this thesis

Te Velde JP, **Buijs GS**, Schafroth MU, Saouti R, Kerkhoffs GMMJ, Kievit AJ. Total Hip Arthroplasty in Teenagers: A Systematic Literature Review. J Pediatr Orthop. 2024 Feb 1;44(2):e115-e123. doi: 10.1097/BPO.0000000000002578. Epub 2023 Nov 29.

Kroese TE, **Buijs GS**, Burger MDL, Ruurda JP, Mook S, Brosens LAA, van Rossum PSN, van Hillegersberg R. Metastasectomy or Stereotactic Body Radiation Therapy With or Without Systemic Therapy for Oligometastatic Esophagogastric Cancer. Ann Surg Oncol. 2022 Aug;29(8):4848-4857. doi: 10.1245/s10434-022-11541-0.

Kroese TE, Christ SM, van Rossum PSN, Burger MDL, **Buijs GS**, Mühlematter U, Andratschke N, Ruurda JP, Hüllner M, Gutschow CA, van Hillegersberg R, Guckenberger M. Incidence and survival of patients with oligometastatic esophagogastric cancer: A multicenter cohort study. Radiother Oncol. 2022 Aug;173:269-276. doi: 10.1016/j.radonc.2022.06.012. Epub 2022 Jun 24.

Buijs GS, de Gendt EEA, Sadiqi S, Post MW, Muijs SPJ, Oner FC. Long-Term Reliability and Validity of the AO Spine PROST (Patient-Reported Outcome Spine Trauma). Spine (Phila Pa 1976). 2022 Sep 1;47(17):E562-E569. doi: 10.1097/BRS.0000000000004379.

PhD Portfolio

Name:	George Samuel Buijs
PhD Period:	January 2022 – August 2024
Supervisors:	Dr. ir. Leendert Blankevoort (promotor) Dr. Matthias Ulrich Schafroth (co-promotor) Dr. Arthur Johan Kievit
Research location:	Amsterdam UMC, location AMC

1. PhD Training

General courses	Year	ECTS
The Amsterdam UMC World of Science	2020	0.7
Basic Course for clinical investigators (BROK) V2.0, NFU	2022	1.5
Clinical epidemiology; randomized controlled trials	2022	0.6

Seminars, workshops and masterclasses	Year	ECTS
Weekly Journal Clubs (Amsterdam UMC)	2023 - 2024	1.0
Wetenschapsbesprekingen Orthopedie en Sportgeneeskunde, Amsterdam UMC	2022 - 2024	1.0
Amsterdam Movement Sciences Annual Meeting	2022	0.6
Amsterdam Movement Sciences Annual Meeting	2023	0.6
Amsterdam Movement Sciences Annual Meeting	2024	0.6
Amsterdam Movement Sciences PhD day, Amsterdam Movement Sciences	2022	0.25
Amsterdam Movement Sciences PhD day, Amsterdam Movement Sciences	2023	0.25
Amsterdam Movement Sciences PhD day, Amsterdam Movement Sciences	2024	0.25
Mini symposium; Skeletal Muscle, from patient to cell, Amsterdam Movement Sciences	2024	0.25
Mini symposium; Osteosarcoma, Amsterdam Movement Sciences	2023	0.25
Mini symposium: Ankle Cartilage lesions, Amsterdam Movement Sciences	2023	0.25

(Inter)national congresses	Year	ECTS
ESSKA, Paris, France	2022	1.0
NOV jaarcongres 2022, Utrecht, The Netherlands	2022	1.0
NOV najaarscongres 2022, Alkmaar, The Netherlands	2022	1.0
SEOHS 2022, Amsterdam, The Netherlands	2022	0.5
ISAKOS 2023, Boston, USA	2023	1.0
EHS 2023, Bern, Switserland	2023	1.0
ESSKA Speciality days 2022, Warsaw, Poland	2023	1.0
AAOS Annual Meeting 2024, San Fransisco, USA	2024	1.0

Word Arthroplasty Congress 2024, Madrid, Spain	2024	1.0
ESSKA, Milan, Italy	2024	1.0
NOV/NOF, Rotterdam, The Netherlands	2024	1.0

Presentations	Year	ECTS
Wetenschapsbespreking Orthopedie en Sportgeneeskunde, Amsterdam UMC (1)	2023	1.0
Wetenschapsbespreking Orthopedie en Sportgeneeskunde, Amsterdam UMC (2)	2024	1.0
NOV najaarscongres 2022, Alkmaar, The Netherlands (Oral presentation)	2022	1.0
SEOHS 2022, Amsterdam, The Netherlands (Oral presentation)	2022	0.5
ISAKOS 2023, Boston, USA (Oral presentation)	2023	1.0
ESSKA Speciality days 2022, Warsaw, Poland (Oral presentation)	2023	1.0
Word Arthroplasty Congres 2024, Madrid, Spain (oral presentation #1)	2024	1.0
Word Arthroplasty Congres 2024, Madrid, Spain (oral presentation #2)	2024	1.0
Word Arthroplasty Congres 2024, Madrid, Spain (ePoster #1)	2024	0.5
Word Arthroplasty Congres 2024, Madrid, Spain (ePoster #2)	2024	0.5
NOV/NOF, Rotterdam, The Netherlands (oral presentation #1)	2024	1.0
NOV/NOF, Rotterdam, The Netherlands (oral presentation #2)	2024	1.0

Peer reviews	Year	ECTS
Knee Surgery, Sports Traumatology, Arthroscopy (3)	2024	1.5
Cartilage (1)	2024	0.5
Journal of Experimental Orthopedics (1)	2024	0.5

2. Teaching

Teaching activities	Year	ECTS
Supervison BSc (3)	2022 - 2025	3
Supervision MSc (2)	2022 - 2024	4
Teaching at Arthroscopic courses (2)	2022 - 2023	1
Teaching at fracture fixation courses (2)	2022 - 2023	1
Knee course (knie cursus) ROGO Amsterdam UMC (2)	2022 - 2023	2
Arts & Topsport (2)	2022 - 2023	6
Lecture Personalized Medicine	2024	0,25

3. Parameters of esteem

Grants	Year
Unrestricted research grant by Zimmerbiomet	2023

Awards	Year
SEOHS 2022, nominated for best abstract	2022

Dankwoord

Als ik terugkijk op de periode die heeft geleid tot dit proefschrift, herinner ik mij vooral het moment waarop ik begon aan dit promotietraject. Ik had de ambitie om onderzoek te doen naar een onderwerp met wezenlijke potentie, iets wat meer zou kunnen worden dan slechts een boek in de kast. Drie jaar later kan ik concluderen dat die potentie er nog steeds is en dat ik deze met heel veel plezier heb mogen verifiëren. Dat stemt mij dankbaar en had niet zo kunnen zijn zonder eenieder die een bijdrage heeft geleverd aan dit proefschrift en de onderzoeken die hierin worden beschreven. Naast alle deelnemers en proefpersonen aan de beschreven studies, wil ik graag een aantal mensen specifiek bedanken.

Allereerst wil ik graag de opponenten bedanken voor de tijd en aandacht die zij hebben genomen om dit proefschrift te beoordelen.

Beste Leendert, Arthur en Matthias, bedankt dat jullie mij de kans hebben gegeven om van het Utrechtse naar het Amsterdamse af te reizen om samen met jullie dit loslatingsavontuur aan te gaan. Ik wil jullie bedanken voor het vertrouwen dat jullie mij hebben gegeven, voor de dagelijkse begeleiding, de discussies, de gezellige momenten op de donderdagochtenden, maar ook op de vele gelegenheden in steden ver weg en dichtbij. Samen hebben we knieprothese-loslating terug op de (inter)nationale agenda gezet en jullie oplossing voor de loslatingsdiagnostiek gevalideerd en aan de orthopedische wereld gepresenteerd. Ik zie er naar uit om deze oplossing straks gemeengoed te zien worden. Jullie hebben mij jullie uitvinding toevertrouwd en daar ben ik jullie heel dankbaar voor. Ik had mij geen betere eerste baan kunnen wensen.

Beste Annemiek en Caroline, drie achtergronden en één doel: knieprothese-loslating beter begrijpen en de diagnostiek verbeteren. Dank dat ik te pas en te onpas een beroep kon doen op alle kennis en talenten die jullie meebrachten naar het team dat we samen zijn geworden.

Beste Iwan en Geert, bedankt voor jullie kritische doch altijd opbouwende begeleiding en ondersteuning van de onderzoeken in dit proefschrift.

Beste Dennis, jouw betrokkenheid bij dit project opende voor mij de deuren naar een wereld waar ik anders nooit mee in aanraking zou komen. Bedankt voor je vertrouwen, alle gesprekken, discussies, koffie en koekjes.

Beste Milo en Chiel, met het afronden van dit proefschrift rust de dagelijkse verantwoordelijkheid van het beantwoorden van alle resterende vragen op jullie schouders. Ik zie uit naar de antwoorden die we samen gaan vinden.

Beste Gino, jouw leiderschap creëert een sfeer waarin collega's vrienden worden. Bedankt hiervoor.

Beste Joep, Kaj, Gitte, Julian, Myelene, Alon, Jari, Jason, Kenny, Anne, Noortje, Jacobien, Liam, Tobias, Pim, Priscilla, Milo, Kishan, Lukas, Milad, Quinten en Alex, samen hebben we topklinisch en basaal wetenschappelijk onderzoek verricht, de wereld bereisd en op en naast het podium laten zien hoe leuk het is om wetenschappelijk onderzoek te doen in het Amsterdam UMC. Ik wil jullie bedanken voor al jullie steun, uitleg, potjes tafeltennis, skireizen, diners, feestavonden, het collectief zeuren over de medische ethische toetsingscommissie, het steppen over de Champs-Élysées, het linedansen in Boston en jullie support telkens wanneer ik weer hetzelfde loslatingsverhaal stond te vertellen.

Beste Frits, Hammie, Koen, Koos, Heru, Tim en Ruben, waar jullie je vastketenden op bruggen voor een betere toekomst, vloog ik de wereld kapot om te vertellen over loslating. Dank voor jullie begrip, humor en onvoorwaardelijke vriendschap. Af en toe een klein bubbeltje op de dinsdagavond gaf de burger moed.

Beste Chiem, Julius, Mark, Martijn, en Matthias, en Casper, Tijmen en Bas, en Max, Joris, Moos en Roy, ik kijk met heel veel plezier terug op de avonden, weekenden en weken die we de afgelopen jaren samen in binnen- en buitenland hebben doorgebracht. Deze momenten waren een zeer gewaardeerde afwisseling op de inhoud van dit proefschrift. Bedankt voor jullie vriendschap.

Beste paranimfen, Alex, als mijn wetenschappelijke overbuurman kon ik altijd rekenen op jouw kennis en relativeringsvermogen. Bedankt dat je jouw wetensschappelijke ervaring, Rotterdamse nuchterheid en gezelligheid de afgelopen jaren met mij wilde delen. Arthur, er zijn weinig promovendi die de eer hebben om hun copromotor ook hun paranimf te mogen noemen. Een geluk bij een ongeluk. Bedankt dat ik ook in deze rol weer op je kan rekenen!

Laura, bedankt voor je hulp bij het maken van de cover van dit boek!

Lieve Tessa en Olivier, ik prijs mijzelf gelukkig met zulke schoonouders. Bedankt voor alle steun en ruggespraak.

Lieve Arjan en Marleen, lieve papa en mama, bedankt voor alles wat jullie in de jaren voorafgaand aan het afronden van dit proefschrift voor mij hebben gedaan en betekend. Het is evident dat dit zonder jullie fijne opvoeding en ondersteuning nooit zo goed was gelukt.

Lieve Celine, in de afgelopen jaren schreven we samen niet één maar twee proefschriften en tegelijkertijd bouwden we samen een geweldig fijn huis. Ik had de afgelopen jaren niet zonder jou willen doorbrengen. Ik ben trots op ons en ongekend blij met jou aan mijn zijde. Bedankt voor je onvoorwaardelijke geduld, enthousiasme en liefde.