



Research article

How do lower-limb landmarking errors impact morphometrics in total knee arthroplasty planning?



Ysé Roch^{a,c,*}, Dorian Lozano^c, Nolwenn Fougeron^e, Benoît Gaulin^{a,b}, Clément Horteur^{a,b}, Yohan Payan^a, Marek Bucki^c, Antoine Perrier^{a,c,d}

^a Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

^b Department of Orthopedics and Sport Surgery, Grenoble-Alpes CHU, South Teaching Hospital, Kimberley Avenue, BP 338, 38434 Échirolles Cedex, France

^c Twinsight, Grenoble, France

^d Department of Orthopedics and Traumatology, Groupe Hospitalier Diaconesses Croix St-Simon, Paris, France

^e Univ. Rennes, Inria, CNRS, IRISA-UMR 6074, Rennes, France

ARTICLE INFO

Keywords:

Total knee arthroplasty

Landmark positioning

Sensitivity analysis

Monte Carlo simulation

Morphometric measurements

ABSTRACT

Background: Total knee arthroplasty (TKA) planning relies on accurate anatomical landmark identification to compute morpho-functional metrics guiding implant positioning. While landmark positioning errors inevitably propagate into surgical measurements, current validation practices apply uniform tolerance thresholds regardless of clinical impact. Previous sensitivity analyses identified categorical distinctions and metric-dependent patterns, yet no framework exists for quantifying individual landmark-axis contributions or translating these findings into risk-stratified acceptance criteria.

Purpose: To establish an evidence-based hierarchy of anatomical landmark criticality for TKA planning through comprehensive sensitivity analysis, and to propose risk-stratified acceptance criteria for quality control protocols.

Methods: Monte Carlo-based sensitivity analysis (50,000 iterations per metric) was performed across 30 osteoarthritic patients (67.1 ± 8.8 years) for 14 morpho-functional metrics (11 angles, 3 distances) derived from 34 anatomical landmarks. Manual annotations by two experienced orthopedic surgeons provided empirical positioning error distributions. Atlas-based inverse morphing enabled cross-patient statistical aggregation. Two complementary sensitivity indices were employed: the novel Morphometrical Impact Rate (MIR), quantifying metric deviation per millimeter of landmark displacement, and the Chatterjee coefficient (ξ), capturing non-linear statistical dependencies including reference frame contamination effects.

Results: Metric sensitivity exhibited 12-fold variability (CI_{90} : 1.37° – 16.87° for angular metrics). The inferior medial femoral condyle Z-coordinate emerged as the most critical landmark coordinate (MIR = 0.854 deg/mm for angles, 0.792 mm/mm for distances). Malleolar landmarks exhibited maximal sensitivity for rotational metrics (medial malleolus X-axis: MIR = 1.395 deg/mm). Z-axis perturbations dominated coronal measurements, with complete reversal for axial rotational metrics. Plane-defining landmarks demonstrated elevated Chatterjee coefficients despite moderate MIR values, confirming reference frame contamination as a distinct error propagation mechanism.

Conclusion: This study provides quantitative evidence that positioning accuracy and clinical impact are not synonymous. We propose a hierarchical validation framework in which tolerance thresholds scale inversely with quantified clinical impact, prioritizing plane-defining landmarks and high-impact landmark-axis combinations identified through MIR and Chatterjee coefficient analysis. Plane-defining landmarks warrant prioritized validation regardless of individual MIR values, as reference frame errors cascade into multiple geometrically independent metrics.

* Corresponding author at: Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France.

E-mail address: yse.roch@univ-grenoble-alpes.fr (Y. Roch).

<https://doi.org/10.1016/j.ejrad.2026.112894>

Received 30 December 2025; Received in revised form 6 April 2026; Accepted 28 April 2026

Available online 3 May 2026

0720-048X/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Total Knee Arthroplasty (TKA) is indicated for end-stage knee osteoarthritis. Although this procedure provides effective pain relief and generally good functional outcomes, a relatively high rate of patient dissatisfaction [1] has led to a recent paradigm shift [2] from a systematic “one-fits-all” technique toward a personalized surgical approach. Several strategies have been proposed, most of which focus on restoring the patient’s native anatomy (e.g., kinematic [3] and restricted kinematic alignment [4]) and/or achieving symmetric flexion and extension gaps (functional alignment [5]). Traditional 2D radiographs are primarily limited to coronal plane planning (e.g., Hip-Knee-Ankle (HKA) angle, mechanical Lateral Distal Femoral Angle (mLDFA), or mechanical Medial Proximal Tibial Angle (mMPTA). Although modern 3D Computed-Tomography (CT)-based planning such as robotic and navigation systems enable the unrestricted multi-planar positioning required for personalized alignment strategies (slope, torsions...), they focus mainly on coronal assessment. Axial and sagittal plans from 3D CT still lack advanced sensitivity analysis.

The growing demand for personalized surgery has driven the development of advanced technologies, including robotic-assisted [6] and navigation systems [7,8], which provide surgeons with increased precision [9] while allowing unrestricted implant positioning in all three dimensions of space. These technologies rely on accurate acquisition of the patient’s anatomy, either through intraoperative landmark registration or in combination with preoperative CT-scan data. The accuracy of bone resections performed using these systems, as well as the reliability of 3D-CT-based planning, strongly depends on the quality of landmark identification performed either *in vivo* or on CT images.

Therefore, the landmarking process represents a cornerstone in preoperative planning to ensure surgical accuracy and reduce the risk of complications associated with implant malpositioning.

Anatomical landmark identification constitutes a fundamental challenge in orthopedic imaging, stemming primarily from inherent definitional ambiguity [10,11]: no consensus exists on which landmarks or axes represent the most reliable standard [12,13], and many conventional landmarks correspond to ill-defined anatomical regions, including femoral epicondyles and tibial plateau margins, rather than geometrically distinct points [14–17]. Consequently, radiologically identified features often reflect variable projections of 3D bone, with single-plane CT measurements producing 12.7% outlier values and systematic errors exceeding 2° compared to multiplanar approaches [18,19].

A single landmark often contributes to multiple calculations, including reference frame definitions, projection planes, angular measurements, and distance computations. For instance, the femoral head center is used for frontal plane definition and HKA measurements, and the surgical transepicondylar axis serves both as a rotational reference and as a projected angle on axial planes. The reuse of landmarks across coordinate systems and measurement contexts introduces method-dependent variability, as systematic reviews document high heterogeneity in reported axes depending on coordinate construction choices [10,12]. Pathological changes further complicate landmark identification, as osteophytes and erosions alter contours and induce uncertainty regarding placement on pathological versus native anatomy. While baseline landmarking positioning accuracy on native anatomy can achieve submillimeter accuracy, arthritic knees demonstrate substantially higher interobserver variability [11]. Automated landmark detection reduces operator variability, with recent deep learning models achieving angular accuracy comparable to experts [20] on 2D X-rays. However, algorithms inherit training biases: model performance and generalization depend critically on the training dataset, while studies demonstrate that models trained on limited pathological cases may not generalize across disease stages [21]. Other teams demonstrated accuracy on the pathological population of interest using mathematically based algorithms [22]. When combining the limited precision of ill-

defined landmarks, automated detection errors, the reuse of these landmarks across multiple calculations, and pathological complexity, error propagation becomes inevitable.

While the precision of anatomical landmark identification has been extensively documented, the literature reveals high methodological variability, with positioning errors ranging from sub-millimetric precision to over 20 mm depending on the anatomical site and soft tissue coverage [17,23]. Both manual annotations and automated detection methods exhibit these inherent positioning uncertainties [24]. More importantly, the clinical impact of such variability is not uniform: previous sensitivity analyses indicate that specific landmarks—particularly those closer to the knee joint or used for epicondylar registration—possess highly restricted safe zones (<2 mm) and disproportionately dictate overall registration accuracy and surgical success [25,26].

Schlatterer et al. introduced a Monte Carlo simulation framework [27] demonstrating geometric amplification whereby short epiphyseal axes exhibit 10-fold greater sensitivity than long mechanical axes. Their simulation revealed maximal angular errors of 3.84° for tibial axial rotation versus only 0.98° for varus-valgus alignment. More recently, Luo et al. employed Jacobian-based sensitivity analysis concluding that “landmarks used for the origin of reference frames have the greatest effect on system output” and establishing categorical prioritization of plane-defining landmarks [25].

While these foundational studies established important principles, several opportunities for extension remain. First, both Luo et al. and Schlatterer et al. [25,27] applied uniform uncertainty parameters across all landmarks simultaneously—while appropriate for their objectives—precluding quantification of individual landmark contributions to specific metrics. Second, analyses focused primarily on implant positioning angles rather than the broader spectrum of morpho-functional metrics (alignment angles, joint geometry, limb lengths) employed in contemporary personalized TKA planning. Third, no framework exists for translating sensitivity findings into differentiated acceptance thresholds: current validation paradigms remain anchored to uniform criteria, most notably the $\pm 3^\circ$ mechanical axis tolerance, applied indiscriminately regardless of landmark function or metric sensitivity.

Indeed, Schelker et al. highlighted that the universal use of the “safe zone” of $\pm 3^\circ$ derived from mechanical alignment is hardly applicable to modern personalized strategies, emphasizing an “urgent need for evidence-based thresholds” [28]. As Reinke et al. recently demonstrated in image analysis validation [29], metric selection without consideration of task-specific requirements leads to systematic pitfalls and misleading conclusions. Such uniform approaches are inadequate for risk-stratified validation, as they fail to distinguish high-impact landmarks requiring stringent control from those with minimal downstream clinical consequence. In the context of modern TKA planning, uncontrolled errors in critical landmarks propagate directly into unintended gap asymmetries, altered joint line restoration, and failure to achieve the targeted personalized alignment.

In this context, the present study aims to establish, for the first time, an evidence-based hierarchy of anatomical landmark criticality for TKA planning. To achieve this, we developed a hybrid sensitivity analysis framework combining:

- Monte Carlo uncertainty propagation (50,000 iterations across 30 patients) to quantify the intrinsic sensitivity of 14 key morpho-functional metrics.
- Multidimensional sensitivity analysis introducing the Morphometrical Impact Rate (MIR) to measure error magnitude, alongside the Chatterjee coefficient (ξ) to evaluate statistical functional dependency.

The ultimate goal of this research is to propose a risk-stratified classification of landmarks, enabling the definition of differentiated acceptance criteria to secure automated planning systems and optimize

quality assurance protocols in orthopedic surgery.

2. Materials and methods

2.1. Study cohort and data acquisition

2.1.1. Patient population

Thirty subjects (11 men, 65.6 $\pm$ 9.8 years old, and 19 women, 71.2 $\pm$ 7.8 years old) were included in this study following ethical committee approval (2019-A02124-53) after providing their informed consent. All subjects underwent Total Knee Arthroplasty (TKA) at the CHU Grenoble Alpes hospital, France. Subjects who underwent previous geometry-altering surgeries of the ipsilateral limb, including the femur, tibia, fibula, and patella bones, were excluded. Demographic characteristics and preoperative clinical parameters are summarized in Table 1. For all subjects, presurgical pseudonymized full-leg CT scans were performed. Preoperative mechanical axis alignment was measured using the HKA angle from full-leg anteroposterior radiographs.

2.1.2. CT imaging protocol

Full-leg CT scans were acquired preoperatively using a Siemens multidetector scanner (140 kV, 49–266 mA auto-modulation, 1 mm slice thickness, 512x512 matrix). Following pseudonymization, all DICOM images were exported according to the left-posterior-superior (LPS) coordinate system convention: x-axis = medio-lateral (positive = lateral), y-axis = antero-posterior (positive = posterior), z-axis = distal-proximal (positive = proximal). This standardization ensures consistent spatial referencing across all subsequent processing steps. For all subjects, presurgical full-leg CT scans were used to segment and reconstruct the 3D geometries of the femur, tibia, fibula, and patella bones of the operated limb. Segmentation was performed automatically for all 30 subjects using a validated in-house deep learning pipeline based on a patch-based U-Net architecture (256x256x256 voxels at 0.8 mm isotropic resolution), trained on 192 lower limb CT scans. Pipeline performance was assessed on an independent test dataset of 91 subjects, yielding a mean Dice similarity coefficient (DSC) of 0.97 for the four bones of interest. This pipeline is implemented within a surgical planning software that has received FDA 510(k) clearance (K250290, January 31, 2025) as a Medical Image Management and Processing System (Class II, 21 CFR 892.2050, Product Codes LLZ/QIH), providing regulatory-grade external validation of the image processing workflow.

2.2. Landmark positioning data

This study aims to quantify how anatomical landmark positioning errors propagate into morphometric (distances and angles) parameters used for TKA surgical planning. Rather than focusing on landmarking algorithm performance, we investigate the differential clinical impact of positioning uncertainties across landmarks and coordinate axes to establish evidence-based criticality hierarchies for quality control protocols.

Our methodology builds upon a previously validated landmarking framework [22], from which we derive empirical positioning error distributions. For this study, we analyzed 34 anatomical landmarks (out of 87 initially validated) necessary to compute 14 morpho-functional metrics across 30 subjects.

Table 1

Demographic details of the study participants and preoperative mechanical alignment.

Parameters	Men (n = 11)	Women (n = 19)	All (n = 30)
Age (years)	65.6 $\pm$ 9.8	71.2 $\pm$ 7.8	67.1 $\pm$ 8.8
Height (mm)	178.3 $\pm$ 6.7	160.8 $\pm$ 4.9	167.2 $\pm$ 10.2
BMI (kg/m ²)	28.1 $\pm$ 4.7	28.7 $\pm$ 6.0	28.5 $\pm$ 5.1
HKA angle (°)	176.5 $\pm$ 4.7	179.4 $\pm$ 4.1	178.3 $\pm$ 4.5

Manual landmark annotations were performed independently by two experienced orthopedic surgeons on 30 subjects randomly divided into two equal groups. Each operator performed landmarking twice on their assigned group, with a seven-day interval in-between sessions, and once on the other group, enabling analysis of both intra-operator repeatability and inter-operator reproducibility. Repeatability analysis demonstrated a mean median error of 1.83 mm and reproducibility of 2.34 mm. These three expert annotations per landmark constitute the empirical basis for modeling positioning uncertainty in the sensitivity analysis detailed in Section §2.4.1.

2.3. Morphometric parameters

2.3.1. Metric calculations methodology

Morphometric parameters were selected based on clinical utility in TKA planning. Specifically, the 34 landmarks enabled calculation of 11 angles and 3 distances. The ISB (International Society of Biomechanics) planes define femoral and tibial anatomical reference frames for morphometry. Metrics calculations employ standardized coordinate geometry: (1) define anatomical planes: coronal, sagittal, and axial planes using landmark-defined reference frames (e.g., transepicondylar axis for coronal plane orientation), (2) project 3D landmark coordinates onto appropriate measurement planes (e.g., femoral mechanical axis from hip center to notch center), and (3) compute angles and distances with sign conventions following ISB standards (e.g., varus = negative HKA deviation). Metrics (angles and distances) descriptions are documented in Annex A.

Detailed formulas, landmark dependencies for each metric and reference frames definitions are provided in Supplementary Table S1. Visual illustrations of landmark configurations and vector constructions for representative metrics (mLDFA and mMPTA) are provided in Fig. 1.

2.4. Sensitivity statistical analysis

Statistical analysis was performed using Python 3.12 (Python Software Foundation) with common scientific computing libraries.

The complete analytical pipeline is composed of four steps (described below); Fig. 2 summarizes this pipeline.

2.4.1. Cross-patient error distribution modeling

A fundamental challenge in aggregating landmark positioning errors across patients lies in anatomical variability: positioning uncertainty observed in one patient's coordinate system cannot be directly compared with uncertainties from geometrically distinct anatomies. To enable inter-subject comparability and aggregation of landmark positioning errors across patient-specific anatomies, we employed an atlas-based inverse morphing methodology. This approach uses a fixed anatomical atlas ($Atlas_0$) as a common coordinate framework.

For each patient p , a spatial transformation $\phi_{0 \rightarrow p}$ maps the atlas geometry to patient-specific anatomy using a scaled Iterative Closest Point (ICP) algorithm for initial global alignment, followed by a non-linear warping. The inverse transformation $\phi_{0 \rightarrow p}^{-1}$ enables projection of patient landmarks back into the unified atlas coordinate system.

The methodology follows a systematic workflow: (1): for each landmark, compute barycenter from manual annotations ($n = 3$) and project it onto patient bone surfaces, as the centroid of multiple annotations may fall outside the bone surface; (2): apply inverse transformation $\phi_{0 \rightarrow p}^{-1}$ to map patient landmarks to $Atlas_0$, (3) build statistical distributions in $Atlas_0$, (4) forward-transform landmarks back to patient-specific coordinate systems via $\phi_{0 \rightarrow p}$ and (5) build statistical distributions on the patients and perform patient-specific sensitivity analysis. Steps (3) and (5) require statistical characterization of positioning uncertainty. For each landmark k , we modeled positional uncertainty using a multivariate normal distribution $X_k \sim N(\mu_k, \Sigma_k)$, where μ_k represents the empirical mean position and Σ_k denotes the 3×3 covariance matrix

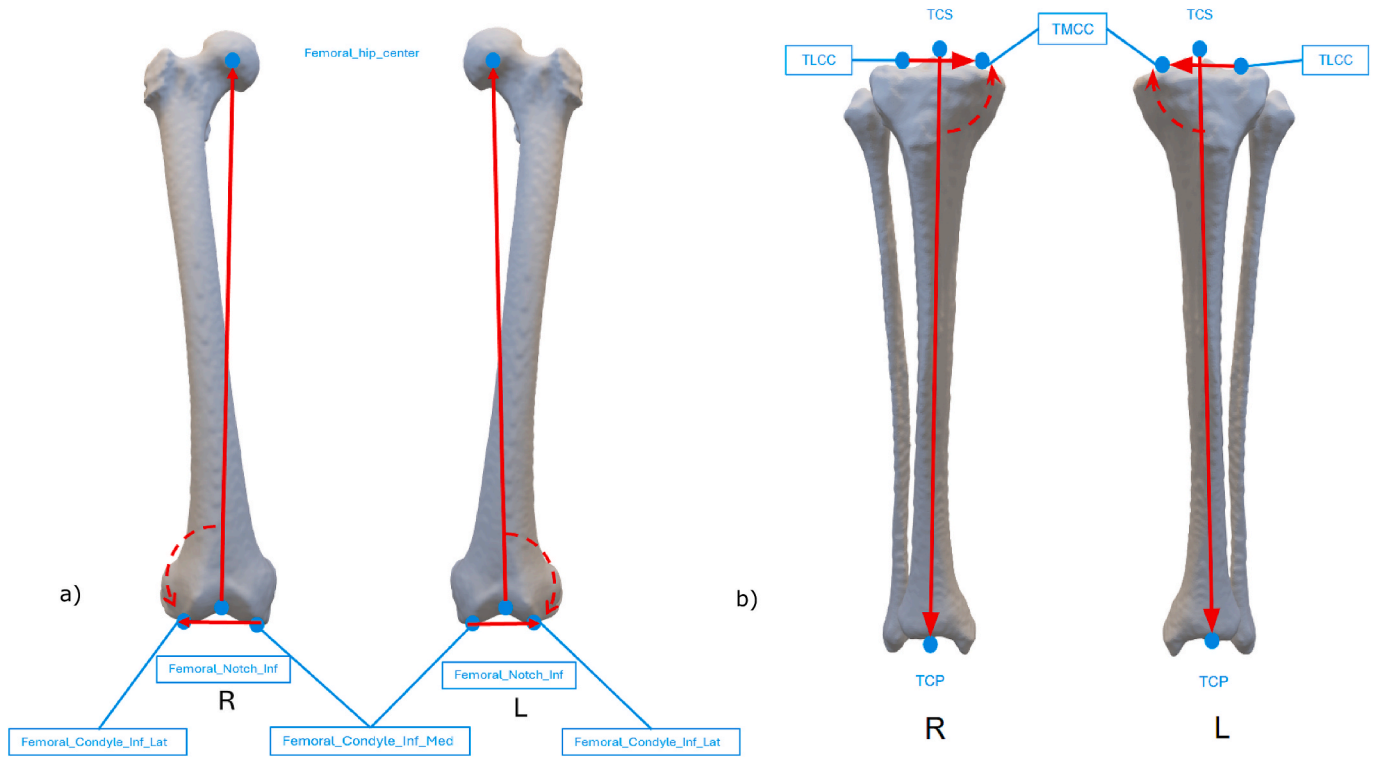


Fig. 1. Anatomical landmark configurations and geometric vector constructions used to compute morpho-functional metrics in TKA planning. (a) Femoral landmarks defining the mechanical Lateral Distal Femoral Angle (mLDFA): the mechanical axis vector (proximal-to-distal red arrow) and the joint line vector (transverse red arrow) are shown for both right (R) and left (L) femora; the dashed red arc illustrates the resulting angle. (b) Tibial landmarks defining the mechanical Medial Proximal Tibial Angle (mMPTA): the tibial mechanical axis vector and the joint line vector are shown for both right (R) and left (L) tibiae. Cyan spheres represent anatomical landmarks.

capturing both error magnitude and directional dependencies. A rejection sampling algorithm with orthogonal surface projection ensured anatomically valid landmark positions (acceptance threshold: projection distance < 1 mm). We generated 50,000 valid samples per landmark, a number determined through empirical convergence analysis demonstrating sensitivity measure stabilization (maximum deviation < 0.02 between runs) beyond 40,000 samples. This sampling procedure was repeated independently for each of the 30 patients, yielding patient-specific sensitivity distributions that account for inter-subject morphological variability.

Complete mathematical formulations, including barycentric coordinate transformations and surface projection algorithms, are provided in [Supplementary Materials S2](#).

2.4.2. Sensitivity analysis framework

To quantify how landmark positioning errors propagate into morpho-functional metrics, we developed a comprehensive analysis pipeline combining one established method with several novel contributions. The width of the 90% confidence interval (CI_{90}), defined as the empirical interval $[Q_5, Q_{95}]$ of the 50,000 Monte Carlo output values, was used as the primary measure of metric sensitivity. Two complementary sensitivity indices were then employed. The Morphometrical Impact Rate (MIR), introduced in this study, quantifies the amplitude of metric deviation per millimeter of landmark displacement: it captures how much a given landmark-axis combination can move a metric value through direct geometric participation. The Chatterjee coefficient [30] quantifies statistical dependency between landmark coordinate perturbations and metric deviations, detecting non-linear and reference frame contamination effects invisible to amplitude-based measures alone: a landmark may exhibit low MIR yet high ξ if its errors systematically corrupt multiple metrics through reference frame distortion rather than direct lever-arm effects. Together, MIR and ξ characterize two

mechanistically distinct error propagation pathways. Three complementary visualization approaches support their interpretation: Hanabi plots characterizing how metric values respond to landmark coordinate perturbations, sensitivity heatmaps displaying ellipsoidal uncertainty regions overlaid on 3D bone surfaces to visualize spatial error distributions and critical positioning zones, and criticality maps plotting MIR against the Chatterjee coefficient to visualize the dual-index sensitivity distribution landmark-axis combination. Together, this multi-faceted framework enables identification of critical landmarks requiring prioritized validation.

2.4.2.1. Chatterjee coefficients. The Chatterjee rank coefficient $\xi(X, Y)$ quantifies statistical dependency between landmark coordinate X and metric Y , ranging from 0 (independence) to 1 (perfect functional dependence) [30]. Unlike Pearson or Spearman correlations, Chatterjee coefficients detect non-linear, non-monotonic relationships, handle correlated, non-independent variables, and do not require distributional assumptions.

Our implementation divided the 50,000 projected samples into 500 consecutive packets of 100 samples each. For each packet, mean landmark coordinate displacement and mean metric deviation were computed. This approach was applied independently to each anatomical axis (X , Y , and Z), enabling characterization of axis-specific dependencies between landmark positioning errors and metric deviations.

2.4.2.2. Morphometrical impact Rate (MIR). We introduce the Morphometrical Impact Rate (MIR), a novel metric developed in this study to quantify the rate of change of metric values with respect to landmark coordinate perturbations, expressed in millimeters of absolute displacement along each anatomical axis. The MIR is computed as a range-based sensitivity measure across the empirical error distribution:

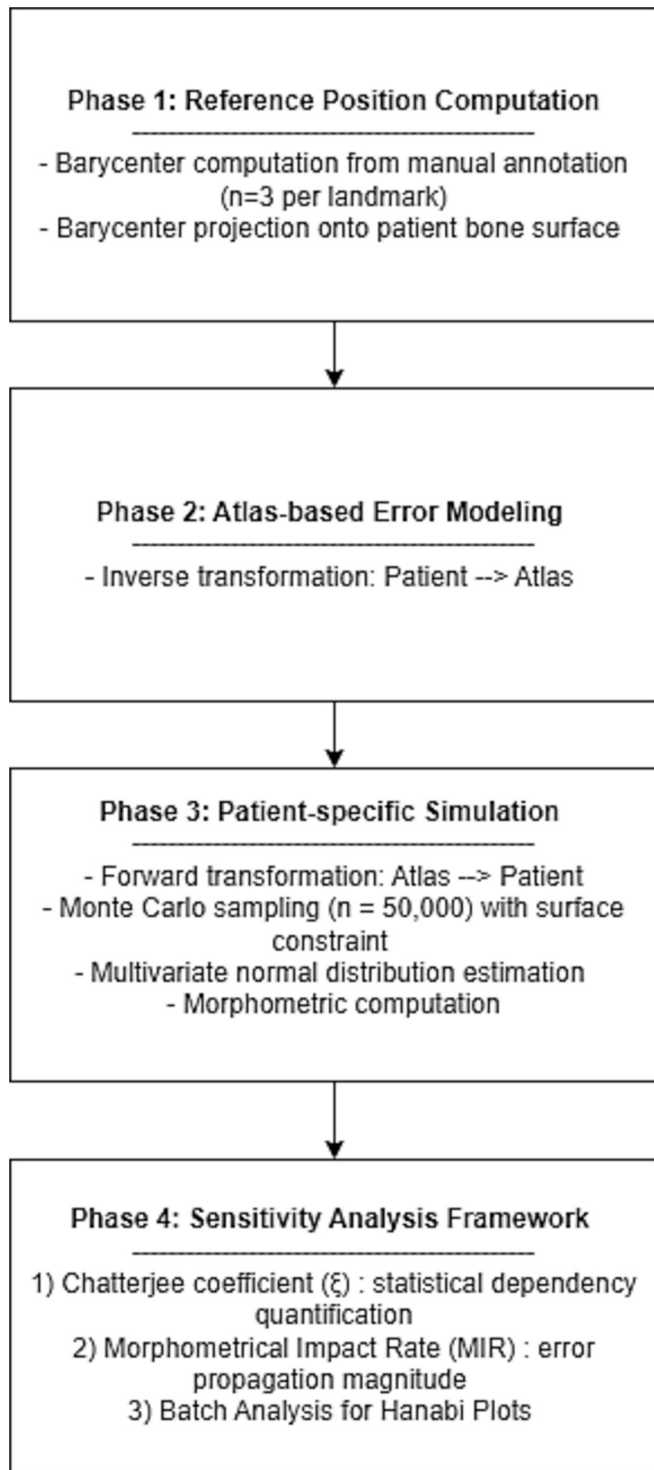


Fig. 2. Methodological workflow for Monte Carlo-based sensitivity analysis of anatomical landmark positioning errors in TKA surgical planning metrics.

$$MIR(L_k^{axis}, M) = \frac{\max(M) - \min(M)}{\max(L_k^{axis}) - \min(L_k^{axis})}$$

where M and L_k^{axis} represent, respectively, the metric and landmark coordinate values observed across the 500 packets.

For angles, MIR is expressed in deg/mm, representing the average angular deviation induced per millimeter of landmark displacement. For distance metrics, MIR is dimensionless (mm/mm), representing the proportional change in measured distance per unit landmark

displacement.

2.4.2.3. Hanabi plots: A novel visualization method. We introduce Hanabi plots (Fig. 3), a new visualization approach (named after the Japanese word for fireworks), reflecting the characteristic scatter pattern that emerges when displaying the relationship between landmark coordinates and perturbations and resulting metric variations.

For each landmark-metric pair, Hanabi plots display 50,000 samples organized as 500 batches of 100 points. This representation reveals relationships between dependent landmark coordinates and morphometric outcomes, enabling identification of linear versus non-linear dependencies and scatter patterns indicative of measurement stability.

The MIR corresponds to the slope of the envelope connecting extreme points in the Hanabi plots. While MIR quantifies the amplitude of metric sensitivity, the Chatterjee coefficient ξ quantifies how tightly metric values follow landmark coordinate variations.

The complete Hanabi plot generation methodology is provided in [Supplementary Materials](#) Section S3.

2.4.2.4. Integrated criticality interpretation. The combination of MIR and the Chatterjee coefficient enables a two-dimensional characterization of landmark sensitivity. To illustrate this framework, we present examples from the mLDFA (mechanical Lateral Distal Femoral Angle) metric analysis:

- High MIR + High ξ : the Inferior Lateral Femoral Condyle Z-coordinate (MIR = 0.866 deg/mm, ξ = 0.282) exhibits both large metric range and tight functional relationship, representing a dual-threat landmark requiring primary validation attention
- High MIR + Low ξ : The Inferior Medial Femoral Condyle Z-coordinate (MIR = 0.630 deg/mm, ξ = 0.055) demonstrates large metric range but high scatter, indicating geometric amplification with unpredictable error propagation despite substantial lever-arm effects.

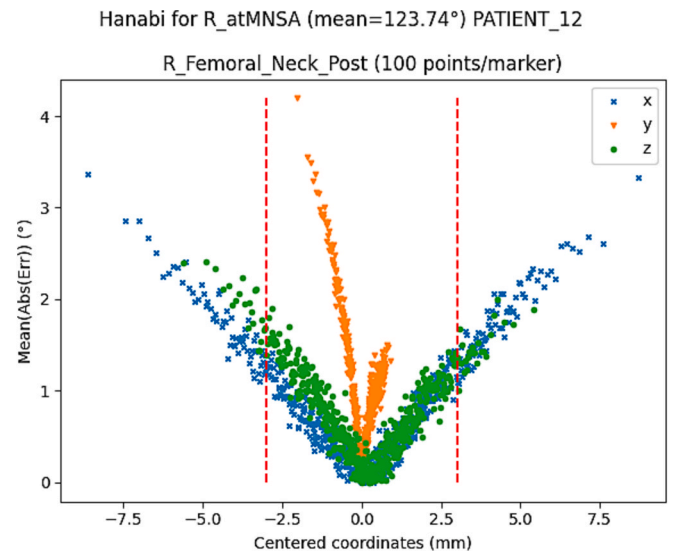


Fig. 3. Hanabi plot for the atMNSA metric at the Femoral Neck Posterior landmark (Patient 12). Each marker represents the mean absolute metric deviation (°) for a batch of 100 Monte Carlo samples, plotted against the centered landmark coordinate displacement (mm) along the X-axis (blue crosses), Y-axis (orange triangles), and Z-axis (green circles). The characteristic “firework” scatter pattern reveals axis-specific functional dependencies: the Y-coordinate exhibits a pronounced V-shaped relationship indicating strong metric sensitivity (high MIR), while X and Z coordinates demonstrate broader dispersion with weaker dependency. Vertical dashed red lines indicate a ± 3 mm, a positioning range representative of typical manual annotation variability reported in the literature.

- Low MIR + high ξ : The Inferior Lateral Femoral Condyle Y-coordinate (MIR = 0.053 deg/mm, $\xi = 0.356$) shows a small metric range but a tight relationship, representing a naturally stable landmark that nonetheless introduces systematic, potentially correctable biases.
- Low MIR + low ξ : The Inferior Femoral Notch Y-coordinate (MIR = 0.053 deg/mm, $\xi < 0.005$) exhibits both a small metric range and high scatter, indicating functional independence where positioning errors contribute negligibly to measurement uncertainty.

This two-dimensional representation provides a systematic basis for quality control decision-making. Landmarks with elevated values on both indices should be prioritize for manual review, while those combining high MIR with low ξ or high ξ with low MIR warrant conditional review based on the specific error propagation mechanism. Landmarks with low values on the indices contribute negligibly to measurement uncertainty. The boundaries between regions are not discrete thresholds but reflect continuous gradients in sensitivity, as illustrated in Figs. 8 and 9.

2.4.2.5. Sensitivity heatmap. To facilitate results interpretation of landmark-specific sensitivity, sensitivity heatmaps (Fig. 4) display ellipsoidal uncertainty regions overlaid on 3D bone surfaces, visualizing error distributions and critical positioning zones for individual patients and landmarks.

3. Results

3.1. Metric sensitivity analysis

3.1.1. Monte Carlo uncertainty propagation

Monte Carlo uncertainty propagation across 50,000 iterations per metric quantified how positioning uncertainties propagate through morphometric calculations to affect surgical planning measurements. Analysis across 30 patients revealed a heterogeneous sensitivity landscape, quantified by Monte Carlo output variance, spanning 0.13 mm^2 (femoral length variance) to 26.30° (external tibial torsion variance, equivalent to a standard deviation of 5.13°). Mean CI_{90} width across angular metrics was $6.44 \pm 0.69^\circ$ (median 6.09°), and $2.35 \pm 0.69 \text{ mm}$ (median 2.65 mm) for distance metrics.

Fig. 5a) and b) display the complete morphometric sensitivity spectrum, with metrics ranked by mean CI_{90} width across the 30-patient cohort. Error bars represent inter-patient variability (standard deviation of CI_{90} values across subjects), indicating whether sensitivity is consistent across anatomical morphologies or subject to individual anatomical variations.

3.1.2. Metric sensitivity spectrum

3.1.2.1. Angular metrics. The ranked CI_{90} distribution (Fig. 5a) reveals a continuous sensitivity spectrum spanning from mHKA ($\text{CI}_{90} = 1.37^\circ \pm 0.22^\circ$) to ETT ($16.87^\circ \pm 0.90$). The two lowest-sensitivity metrics are mHKA and mL DFA ($\text{CI}_{90} = 1.91^\circ \pm 0.51^\circ$). mJLCA ($\text{CI}_{90} = 3.85^\circ \pm 0.48^\circ$), mMPTA ($\text{CI}_{90} = 3.46^\circ \pm 0.37^\circ$), and tibial slope ($\text{CI}_{90} = 4.82^\circ \pm$

atMNSA on patient 12

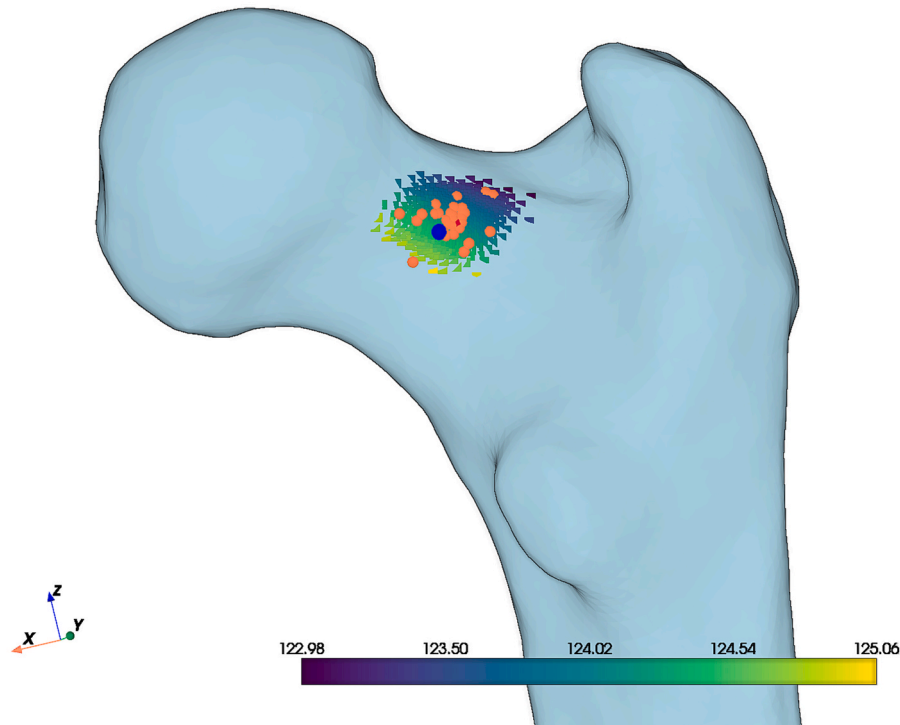


Fig. 4. Landmark positioning variability and metric sensitivity at the Posterior Femoral Neck landmark (Patient 12, atMNSA metric). Orange markers (1–30): barycenters of manual annotations from each patient, inverse-morphed to Atlas_0 then forward-morphed onto Patient 12's anatomy. Color scale: resulting atMNSA angle value ($^\circ$) for each landmark position.

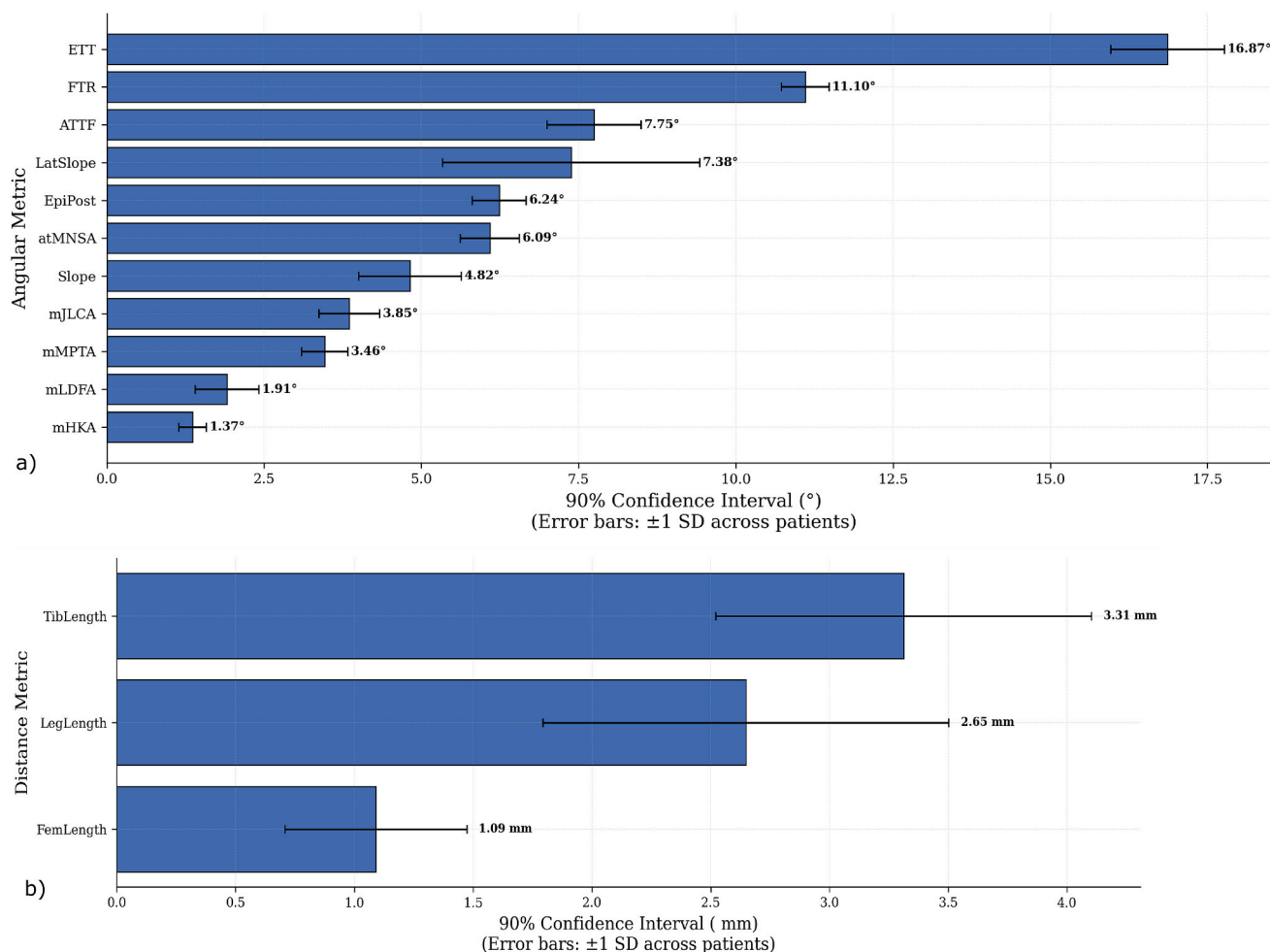


Fig. 5. Morphometric sensitivity spectrum ranked by mean 90% confidence interval (CI₉₀) width across 30 patients for a) angular metrics and b) distance metrics. Error bars represent inter-patient variability (SD of CI₉₀ values across subjects), illustrating the range of individual anatomical responses to landmark positioning uncertainty.

0.81°) exhibited moderate sensitivity. atMNSA (CI₉₀ = 6.09° ± 0.47°), EpiPost (CI₉₀ = 6.24° ± 0.43°) demonstrated low inter-patient variability (SD = ± 0.43°-0.47°), indicating consistent sensitivity across anatomical configurations. LatSlope (CI₉₀ = 7.38° ± 2.04°) exhibits the highest inter-patient variability of all angular metrics (SD = ± 2.04°). ETT (CI₉₀ = 16.87° ± 0.90°) and FTR (CI₉₀ = 11.10° ± 0.38°) occupy the upper end of the spectrum, both with low inter-patient variability. The low standard of deviation (SD = ± 0.38°-0.90°) indicates systematic

sensitivity independent of anatomical configuration.

3.1.2.2. Distance metrics. The three distance metrics span a 3-fold range (CI₉₀: 1.09–3.31 mm). Femoral length shows the lowest sensitivity (CI₉₀ = 1.09 ± 0.38 mm, variance = 0.13 mm²); however, this must be interpreted with caution, as the femoral head center was treated as fixed, meaning landmark positioning uncertainty was not propagated through this anatomical point. Consequently, the reported low variance

Table 2

Top 15 landmark-axis combinations ranked by Morphometrical Impact Rate (MIR) for angular metrics.

#	Landmark	Axis	Mean MIR (deg/mm)	Mean Chatterjee (ξ)	Projection Plane	Primary Affected Metrics
1	Tibial_Mal_Med	X	1.395	0.150	Axial	ETT
2	Fibula_Mal_Lat	X	1.295	0.068	Axial	ETT
3	Femoral_Epicondyle_Med	X	1.138	0.077	Axial, Coronal	EpiPost, mHKA
4	Femoral_Neck_Post	Y	1.056	0.198	Axial, Coronal	ATTF, atMNSA
5	Femoral_Condyle_Inf_Med	Z	0.854	0.115	Coronal	mJLCA, mLDFA
6	Femoral_Condyle_Inf_Lat	Z	0.833	0.180	Coronal	mJLCA, mLDFA
7	Femoral_Condyle_Post_Med	Y	0.777	0.041	Axial	EpiPost, FTR
8	Tibial_Plate_Med	Z	0.748	0.071	Axial, Coronal	ETT, FTR, mJLCA
9	Femoral_Condyle_Post_Lat	Y	0.673	0.042	Axial	ATTF, EpiPost
10	Femoral_Epicondyl_Lat	X	0.657	0.106	Axial, Coronal	EpiPost, mHKA
11	Tibial_Plate_Lat	Z	0.635	0.049	Axial, Coronal	ETT, FTR, mJLCA
12	Tibial_Plate_Lat_Sup_Post	Z	0.566	0.125	Sagittal, Coronal	LatSlope, Slope, mJLCA
13	Fibula_Mal_Lat	Y	0.561	0.110	Axial	ETT
14	Tibial_Mal_Med	Y	0.551	0.232	Axial	ETT
15	Tibial_Plate_Med	X	0.532	0.046	Axial, Coronal	ETT, FTR, mJLCA

represents only the contribution from the distal landmark (medial condyle inferior), and a more comprehensive error model incorporating femoral head variability would likely yield higher estimates for this metric. Tibial length and leg length exhibited higher sensitivity ($CI_{90} = 3.31 \pm 0.79$ mm and 2.65 ± 0.85 mm, respectively).

3.1.3. Dual-index sensitivity rankings

3.1.3.1. Angular metrics. Table 2 presents the 15 landmark-axis combinations ranked by MIR for angular metrics, with corresponding Chatterjee coefficients. The complete rankings by Chatterjee coefficient are provided in [Supplementary Table S4](#).

Malleolar landmarks dominated mechanical sensitivity: the medial tibial malleolus (X-axis: $MIR = 1.395$ deg/mm, Chatterjee $\xi = 0.151$) and the lateral fibular malleolus (X-axis: $MIR = 1.295$ deg/mm, $\xi = 0.068$) exhibited the highest MIR values, serving as distal references for ETT. Their moderate-to-low ξ values indicate that, despite large metric ranges, the input–output relationship exhibits substantial scatter. Examples of malleolar and tibial plateau landmark distribution range for ETT angle calculation are given in [Fig. 6](#).

Femoral epicondyles demonstrated substantial impact despite serving primarily to define the reference plane rather than being direct measurement endpoints (medial epicondyle X-axis: $MIR = 1.138$ deg/mm, $\xi = 0.077$; lateral epicondyle X-axis: $MIR = 0.657$ deg/mm, $\xi = 0.106$) ([Fig. 7\(a\)](#)). The medial epicondyle's high MIR reflects its X-axis role in transepicondylar axis definition; the lateral epicondyle's elevated ξ on the Y-axis reflects anteroposterior sensitivity to coronal plane orientation. Femoral condylar landmarks showed consistently high MIR values across multiple metrics: the inferior medial femoral condyle (Z-axis: $MIR = 0.854$ deg/mm, $\xi = 0.115$) and the inferior lateral femoral condyle (Z-axis: $MIR = 0.833$ deg/mm, $\xi = 0.180$) contribute to three geometrically distinct calculations (mLDFA, mJLCA, and femoral length ([Fig. 7\(b\)](#))).

Posterior femoral condyle landmarks exhibited elevated Y-axis sensitivity (posterior medial: $MIR = 0.777$ deg/mm, $\xi = 0.041$; posterior lateral: $MIR = 0.673$ deg/mm, $\xi = 0.042$). The morphometrical impact hierarchy revealed fundamental asymmetry wherein distal landmarks consistently exhibited higher MIR values than proximal landmarks despite comparable absolute positioning errors.

Across the top 15 angular combinations, a systematic dissociation between MIR and ξ is evident: landmarks with the highest MIR (malleoli, $MIR > 1.2$ deg/mm) exhibit $\xi < 0.20$, while landmarks with the highest ξ (posterior condyles, femoral neck) exhibit $MIR < 1.1$ deg/mm. This dissociation, visible in [Fig. 8](#), reflects two mechanistically distinct error propagation pathways: geometric amplification (high MIR, low ξ) versus reference frame contamination (moderate MIR, high ξ).

3.1.3.2. Distance metrics. Table 3 presents MIR rankings for distance metrics with corresponding ξ values. The complete Chatterjee rankings are provided in [Supplementary Table S5](#).

The inferior medial femoral condyle Z-coordinate dominated both indices simultaneously ($MIR = 0.792$ mm/mm, $\xi = 0.533$), the only landmark-axis combination to achieve top rank by both criteria. This co-elevation reflects its role as the distal terminus for femoral length calculation, where vertical positioning directly determines total femoral dimension without intermediate geometric transformations or averaging effects. Its X- and Y-axes also maintained detectable ξ values (0.439 and 0.342, respectively) despite near-perpendicular orientation to the measurement axis, reflecting non-idealized 3D geometry.

Distal tibial landmarks showed moderate but consistent dual sensitivity: distal medial tibia (Z-axis: $MIR = 0.340$ mm/mm, $\xi = 0.262$) and the distal lateral tibia (Z-axis: $MIR = 0.316$ mm/mm, $\xi = 0.391$) exhibited higher ξ values relative to their MIR than tibial spine landmarks, indicating that distal tibial positioning errors maintain more structured functional relationships with length metrics than plateau-

level contributions. The lower sensitivity compared to femoral measurements reflects the tibial length calculation methodology involving centroid computations (tibial center of shaft, tibial center of plateau) that partially average positioning errors across multiple contributing landmarks. Tibial spine landmarks demonstrated lower ξ despite comparable MIR values (medial and lateral, Z-axis: $MIR \approx 0.316$ – 0.340 mm/mm, $\xi \approx 0.110$ – 0.157), consistent with centroid-based averaging attenuating individual landmark influence on output consistency. [Fig. 9](#) illustrates the spatial distribution of sensitivity ellipsoids for this posterior uncertainty at the lateral tibial plateau: ellipsoids extend onto osteophytic formations, with the automated landmark position (blue sphere) localizes to native bone surface anterior to the osteophyte margin, whereas manual expert annotations (represented by ellipsoid distributions) exhibit greater posterior dispersion, a pattern consistent with the interpretive ambiguity between pathological bone overgrowth and native anatomical boundaries in degenerative joints.

The distance metric landscape is more compact than angular metrics ([Fig. 10](#)): no landmark-axis combination exceeds $MIR = 1.0$ mm/mm, and the ξ range (0.014–0.533) is dominated by a single outlier (Femoral Condyle Inf Med Z). This compactness reflects the simpler geometric structure of length calculations compared to projected angular measurements.

3.1.4. Axis-specific sensitivity patterns

To identify which coordinate perturbations dominate metric deviations, we decomposed landmark positioning errors along three orthogonal anatomical axes: X (mediolateral), Y (anteroposterior), and Z (proximodistal). This directional analysis enables axis-prioritized quality control protocols.

Aggregated sensitivity analysis across all landmark-axis combinations contributing to the 14 metrics ($n = 69$ for angular metrics, $n = 15$ for distance metrics) revealed distinct directional patterns ([Table 4](#) and [Fig. 11](#)).

Ten out of 14 metrics exhibited the highest sensitivity to Z-axis positioning errors, reflecting lower limb morphometry, where surgical planning metrics predominantly quantify alignment or dimensions along the limb's long axis. For distance measurements, Z-axis perturbations generated a median MIR of 0.33 mm/mm compared to 0.15 mm/mm for the X-axis and 0.07 mm/mm for the Y-axis. Coronal plane angular metrics (mHKA, mLDFA, mMPPTA, mJLCA) demonstrated even stronger Z-axis dominance, with a median MIR of 0.22 deg/mm versus 0.08 deg/mm (X) and 0.06 deg/mm (Y)—a 2.8-fold difference. This pattern reflects that proximodistal positioning errors directly alter the projected vertical coordinates of landmark pairs defining mechanical axes, while mediolateral and anteroposterior errors primarily affect perpendicular coordinates, contributing minimally to coronal angle calculations.

In contrast, rotational metrics measured within the transverse plane (ETT, FTR, EpiPost, ATTF) exhibited a completely inverted sensitivity hierarchy. Medirolateral (X-axis) errors dominated with a median MIR of 0.62 deg/mm, followed by anteroposterior (Y-axis) at 0.44 deg/mm, with proximodistal (Z-axis) exhibiting the lowest impact at 0.29 deg/mm. This reversal arises because rotational metrics quantify angular relationships between landmark pairs lying within the transverse (XY) plane: for instance, external tibial torsion compares the transmalleolar axis to the tibial plateau axis, both defined by landmarks separated by only 60–80 mm horizontally. Positioning errors occurring within this measurement plane (X and Y axes) directly alter the apparent orientation of these short axes, whereas Z-axis errors displace landmarks perpendicular to the measurement plane, preserving angular relationships. With such short inter-landmark distances, even modest in-plane errors (1–2 mm) translate into substantial angular deviations (0.7–1.4°).

Sagittal plane metrics (Slope, LatSlope, ATTF) demonstrated an intermediate pattern where Z-axis sensitivity remained dominant (median $MIR = 0.42$ deg/mm) but Y-axis contribution increased substantially (0.32 deg/mm) compared to coronal metrics (0.06 deg/mm). The Z-to-Y

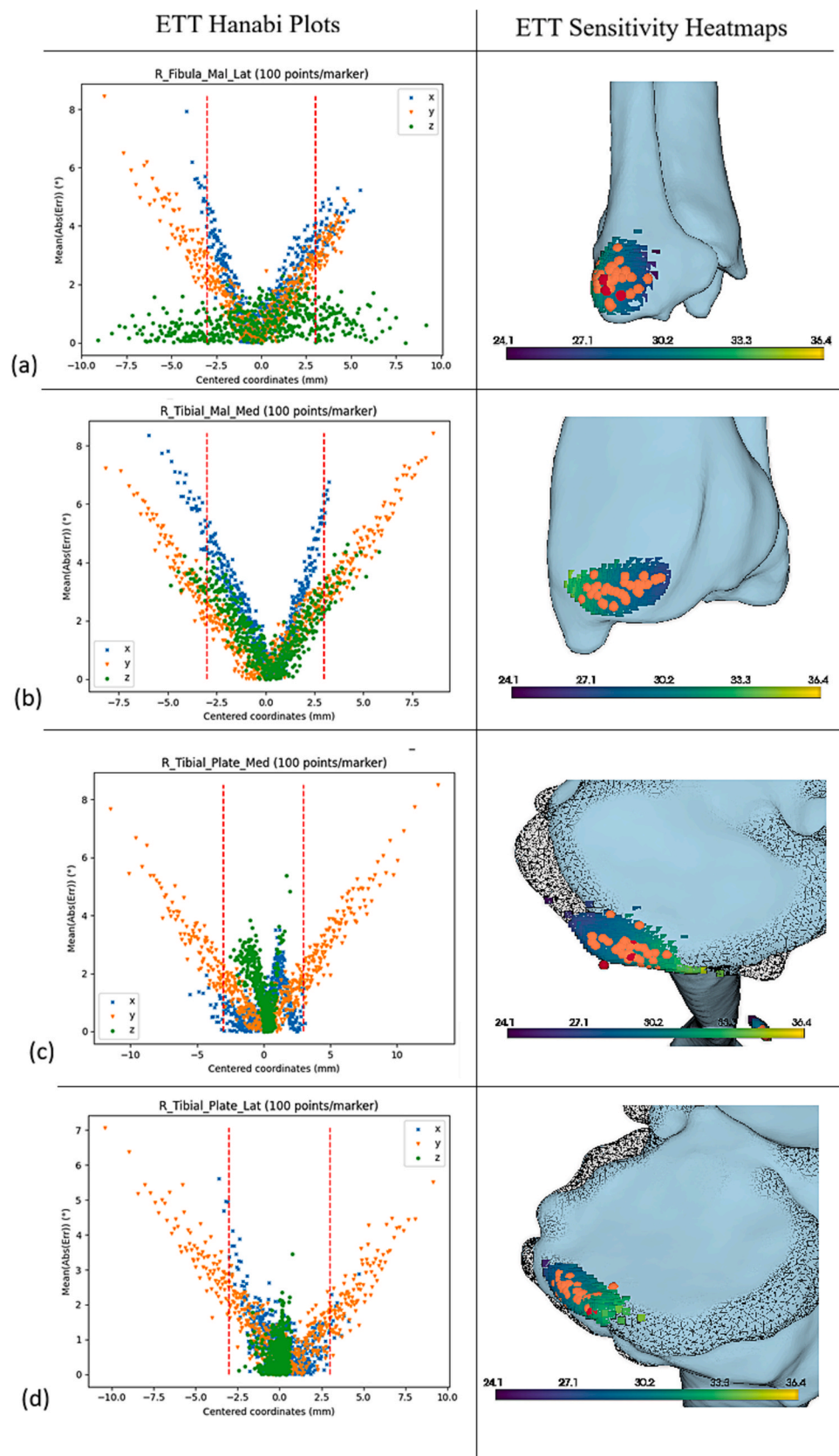


Fig. 6. Hanabi plots and sensitivity maps for ETT for patient 27. Represented landmarks are (a) lateral malleolar, (b) medial malleolar, (c) medial tibial plateau and (d) lateral tibial plateau. Each point represents a batch of 100 samples (500 batches total).

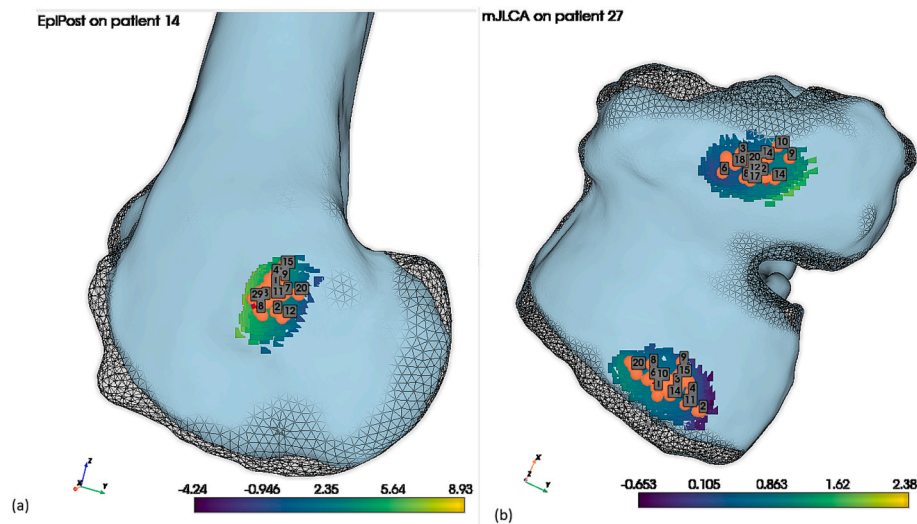


Fig. 7. Sensitivity ellipsoids for high-impact landmarks. (a) Medial femoral epicondyle demonstrating mediolateral positioning uncertainty affecting trans-epicondylar axis definition. (b) Inferior femoral condyles showing proximodistal uncertainty propagating into mLDFA and mJLCA calculations.

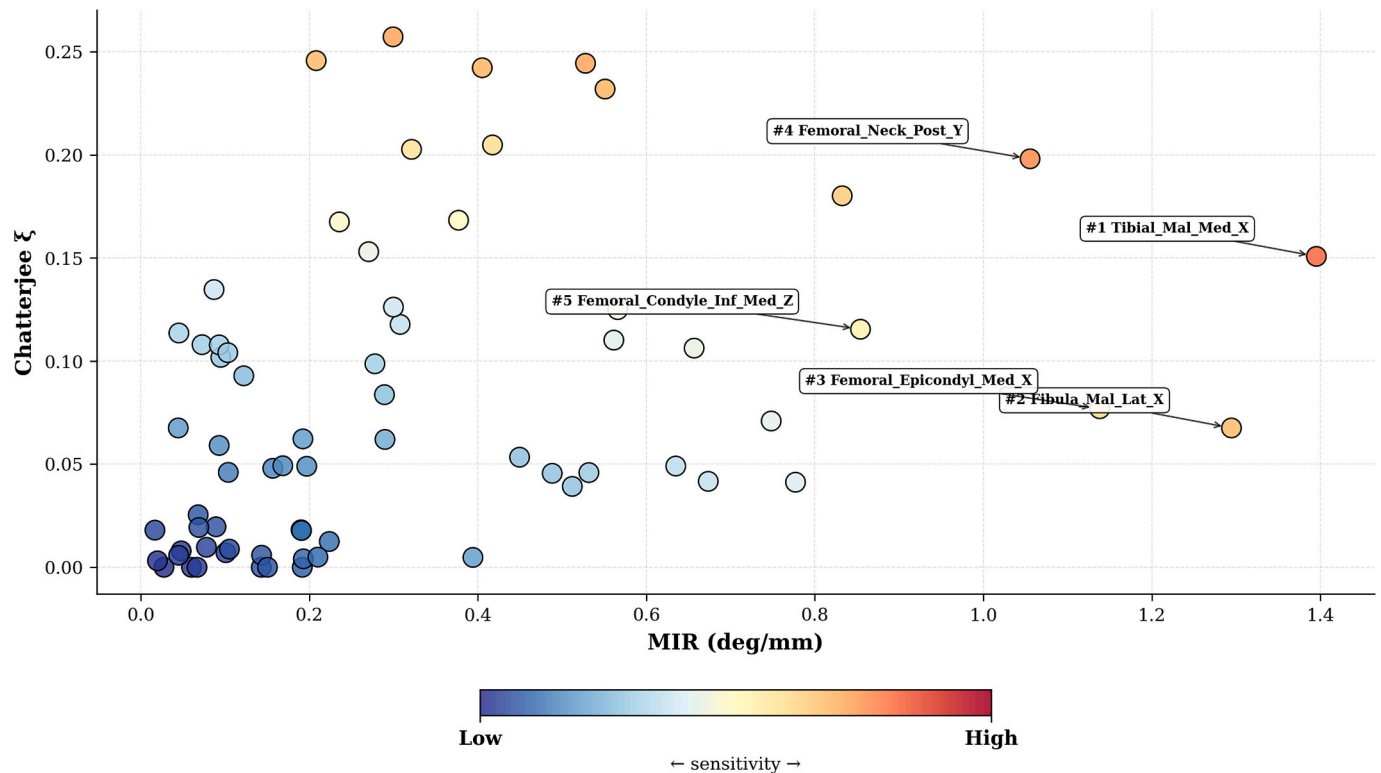


Fig. 8. Criticality map for angular metrics. Each point represents one landmark-axis combination (n = 69). The x-axis displays the MIR (deg/mm); the y-axis displays the Chatterjee coefficient (ξ). Point colour encodes the Euclidean distance from the origin in the min-max normalized (MIR, ξ) space, defined as $\sqrt{(\text{MIR}_{\text{norm}}^2 + \xi_{\text{norm}}^2)} / \sqrt{2}$, providing a visual summary of combined dual-index magnitude (0 = low sensitivity, 1 = high). This color scale is intended as a visual aid only; MIR and ξ should be interpreted independently as described in the text. Labels identify the five highest-impact combinations. Two mechanistically distinct clusters are visible: high-MIR/low- ξ combinations (malleolar landmarks) reflecting geometric amplification, and moderate-MIR/high- ξ combinations (posterior condyles, femoral neck) reflecting reference frame contamination.

ratio of only 1.3-fold indicates meaningful sensitivity to both proximodistal and anteroposterior positioning errors, reflecting dual geometric dependencies of sagittal angle calculations. Femoral anteversion (ATF) exhibited similar dual sensitivity but with reversed magnitudes: Y-axis MIR (0.52 deg/mm) exceeded Z-axis MIR (0.36 deg/mm), as anteversion primarily measures anteroposterior orientation of the femoral neck axis relative to posterior condylar reference.

4. Discussion

This study presents a comprehensive quantitative sensitivity analysis of landmark positioning errors on morphometric metrics used for TKA planning. Through atlas-based inverse morphing, Chatterjee coefficient analysis, and Monte Carlo simulation across 30 osteoarthritic patients with 50,000 iterations per metric, we established criticality hierarchies

LatSlope on patient 14

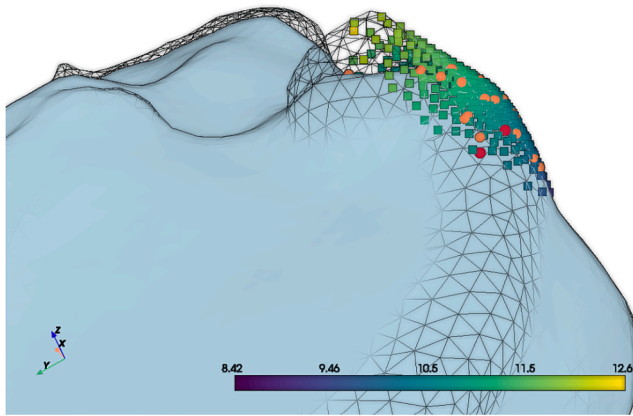


Fig. 9. Sensitivity ellipsoids for lateral tibial plateau superior anterior landmark demonstrating posterior positioning uncertainty. Ellipsoid distributions onto osteophytic formations present in osteoarthritic knees reflecting interpretive ambiguity between pathological bone overgrowth and native anatomical boundaries.

Table 3

Top 15 landmark-axis combinations ranked by Morphometrical Impact Rate (MIR) for distance metrics.

#	Landmark	Axis	Mean MIR (mm/mm)	Mean Chatterjee (ξ)	Primary Affected Metrics
1	Femoral_Condyle_Inf_Med	Z	0.792	0.533	FemLength
2	Tibial_Dist_Med	Z	0.340	0.262	LegLength, TibLength
3	Tibial_Plate_Spine_Lat	Z	0.340	0.157	TibLength
4	Tibial_Dist_Lat	Z	0.316	0.391	TibLength
5	Tibial_Plate_Spine_Med	Z	0.316	0.110	LegLength, TibLength
6	Tibial_Dist_Med	X	0.224	0.182	TibLength
7	Tibial_Dist_Lat	X	0.156	0.243	LegLength, TibLength
8	Tibial_Plate_Spine_Med	X	0.152	0.031	LegLength, TibLength
9	Tibial_Plate_Spine_Lat	X	0.120	0.049	TibLength
10	Tibial_Plate_Spine_Lat	X	0.112	0.078	TibLength
11	Femoral_Condyle_Inf_Med	X	0.110	0.439	TibLength
12	Tibial_Plate_Spine_Med	Y	0.091	0.033	FemLength
13	Tibial_Dist_Lat	Y	0.074	0.014	TibLength
14	Femoral_Condyle_Inf_Med	Y	0.070	0.342	LegLength, TibLength
15	Tibial_Dist_Med	Y	0.058	0.017	FemLength

for both angular and distance metrics.

This study builds upon foundational sensitivity analyses by Schlatterer et al. [27] and Luo et al. [25]. Schlatterer's Monte Carlo framework first demonstrated geometric amplification whereby short epiphyseal axes (≈ 100 mm) exhibit 10-fold greater sensitivity than long mechanical axes (≈ 400 mm), with rotational errors reaching 3.84° versus $\leq 0.98^\circ$ for coronal alignment—a principle our findings strongly corroborate. Luo et al. extended this work using Jacobian-based analysis, establishing the critical insight that “landmarks used for the origin of reference frames have the greatest effect on system output.”

The present study extends this framework in four directions: (1) systematic evaluation of 14 morpho-functional metrics spanning alignment, joint geometry, and limb dimensions; (2) introduction of the Morphometrical Impact Rate (MIR) providing landmark-specific, axis-

resolved sensitivity quantification in clinically interpretable units (deg/mm, mm/mm); (3) application of Chatterjee coefficients to detect non-linear dependencies and reference frame contamination effects not captured by variance-based methods alone; and (4) propagation of patient-specific empirical error distributions across 30 anatomies rather than uniform uncertainty parameters.

Notably, our error model treated the femoral head center as geometrically fixed (sphere-fitting eliminates positioning ambiguity); consequently, the inferior medial femoral condyle Z-coordinate emerged as the most critical landmark (MIR = 0.854 deg/mm for angular metrics, 0.792 mm/mm for distances). This finding complements rather than contradicts Luo et al.'s conclusion regarding reference frame origins: both studies identify proximal landmarks as high-impact sites, but our analysis further resolves which specific coordinates within these anatomical regions drive measurement uncertainty. The clinical implication is that validation protocols should prioritize not merely categorical landmark types (reference frame vs. endpoint) but specific landmark-axis combinations based on quantified downstream impact.

Our results provide independent corroboration of Schlatterer's geometric amplification principle: mHKA ($CI_{90} = 1.37^\circ$, lever arm 800–1000 mm) demonstrated substantially lower sensitivity than mJLCA ($CI_{90} = 3.85^\circ$, lever arm 60–80 mm). The broader sensitivity range observed in our study (1.37° to 16.87° , 12-fold) compared to Schlatterer's analysis (0.98° to 3.84° , 4-fold) reflects primarily our inclusion of axial rotational metrics—external tibial torsion (ETT, $CI_{90} = 16.87^\circ$), femoro-tibial rotation (FTR, $CI_{90} = 11.10^\circ$), femoral neck anteversion (ATTF, $CI_{90} = 7.75^\circ$), and the epicondylar-posterior condylar angle (EpiPost, $CI_{90} = 6.24^\circ$)—all dependent on particularly short transverse axes (60–80 mm spans). This extended metric scope reveals that geometric amplification effects are most pronounced for rotational measurements in the axial plane, where lever-arm distances are minimal and in-plane positioning errors (X and Y axes) translate directly into angular deviations.

The distinction between plane-defining and endpoint landmarks, established by Luo et al. through Jacobian analysis, receives quantitative support from our dual-index framework: plane-defining landmarks exhibited higher Chatterjee coefficients (median $\xi = 0.082$) than endpoint landmarks ($\xi = 0.063$) despite comparable or lower MIR values. This pattern—moderate individual impact but elevated statistical coupling—reflects the reference frame contamination mechanism whereby errors at plane-defining sites propagate into multiple geometrically independent metrics simultaneously. Our contribution lies in quantifying this effect and demonstrating that Chatterjee coefficients, rather than variance-based measures alone, are necessary to detect such multivariate error propagation.

4.1. Landmark criticality and validation implications

Our results challenge the assumption that landmark positioning error magnitude correlates directly with clinical impact. The superior posterior femoral shaft landmark demonstrated a high reproducibility positioning error in the Z-axis in the cohort (median 2.67 mm) and yet contributed marginally to TKA-relevant metrics (MIR < 0.05 deg/mm, Chatterjee ξ < 0.02). Conversely, the inferior medial femoral condyle exhibited low positioning errors in the Z-axis (median 0.50 mm) and yet appeared as the most critical landmark for both angular (MIR = 0.854 deg/mm, $\xi = 0.309$) and distance (MIR = 0.792 mm/mm, $\xi = 0.533$) metrics. This landmark's influence extends across multiple metrics simultaneously: a positioning error propagates into coronal alignment (mL DFA), joint geometry (mJLCA), and femoral length quantification, compounding its clinical relevance despite modest absolute positioning uncertainty.

Current quality-assurance practices and tools for landmark-based planning most often rely on uniform prediction-to-true landmark distance thresholds (typically 2–4 mm) applied identically across all landmarks. Such heuristics are used both in regulatory submissions and

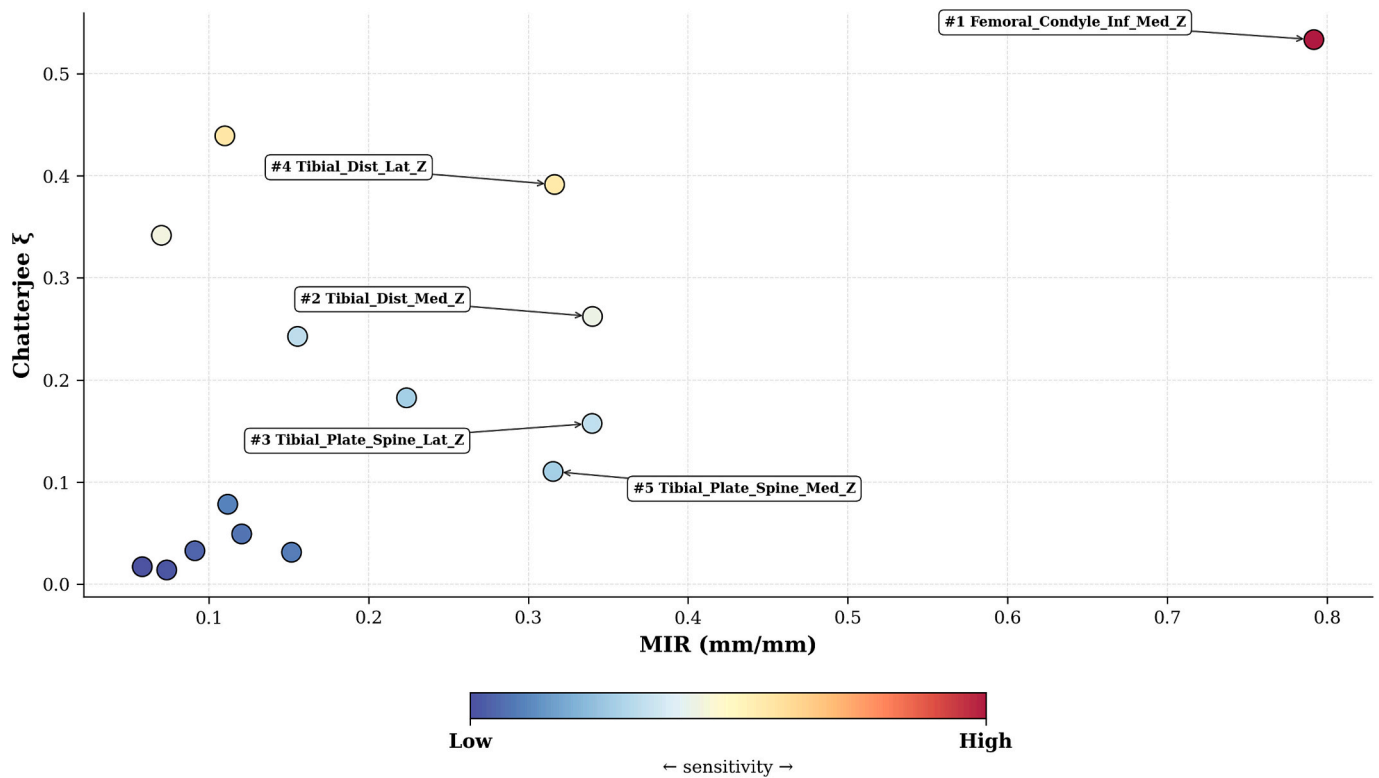


Fig. 10. Criticality map for distance metrics ($n = 15$). Point color encodes the same continuous sensitivity score as Fig. 8: $\sqrt{(\text{MIR}_{\text{norm}}^2 + \xi_{\text{norm}}^2)} / \sqrt{2}$, with MIR and ξ independently min–max normalized within the distance metric subset. Labels identify the five highest-impact combinations ranked by MIR. The Femoral Condyle Inferior Medial Z-coordinate (#1) dominates both indices (MIR = 0.79 mm/mm, $\xi = 0.53$); distal tibial landmarks (#2, #4) show elevated ξ relative to their MIR, consistent with structured error propagation through centroid-based calculations, as well as tibial spine landmarks (#3, #5).

Table 4

Axis-Specific Sensitivity by Metric Type. Values represent median MIR (deg/mm or mm/mm) computed across all landmark-metric pairs within each anatomical plane category.

Metric Type	n Metrics	X-axis (ML) Median MIR	Y-axis (AP) Median MIR	Z-axis (PD) Median MIR	Dominant Axis
Distances	3	0.15	0.07	0.33	Z
Coronal Angles	4	0.08	0.06	0.22	Z
Sagittal Angles	3	0.12	0.32	0.42	Z
Axial Angles	4	0.62	0.44	0.29	X

in validation studies [24,31–33]. This one-size-fits-all strategy allows large errors at high-impact sites while demanding unnecessary precision at stable locations. A landmarking algorithm achieving 3.0 mm error at the Femoral Condyle Inferior Medial (passing ± 3 mm validation) induces an average 2.6° mechanical axis deviation and a 1.8° joint line convergence error, clinically significant deviations that escape current detection protocols.

These findings call for a sensitivity-stratified approach to validation effort: landmarks occupying the upper end of both the MIR and ξ distributions, including malleolar landmarks, femoral epicondyles, posterior femoral neck, and the inferior medial femoral condyles, warrant the most stringent positioning requirements, while those at the lower end contribute minimally and may tolerate relaxed thresholds without compromising surgical planning accuracy. This has direct implications for automated planning systems: current validation workflows, whether robot-assisted, navigation-based, or 2D radiograph software, typically apply uniform accuracy thresholds across all landmarks, whereas our

findings suggest that validation effort should concentrate on landmarks critical to the specific metrics of surgical interest, plus plane-defining landmarks regardless of the procedure. Automated systems inheriting bias from training data would particularly benefit from targeted uncertainty quantification at high-impact anatomical sites.

4.2. The invisible threat: Geometric mechanisms of error amplification

Monte Carlo uncertainty propagation revealed heterogeneous sensitivity spanning 146-fold range for angular metrics (variance range: $0.18\text{--}26.30^\circ$) and 8-fold range (variance range: $0.13\text{--}1.08\text{ mm}^2$) for distance metrics. Three distinct geometric mechanisms explain this variability.

4.2.1. Lever-arm effects

Angular metrics sensitivity to landmark positioning errors scales inversely with the distance separating the landmarks defining each axis. This geometric relationship creates opposing effects depending on measurement configuration (e.g., projection planes).

For long-span measurements, large inter-landmark distances attenuate the angular impact of positioning errors. The mHKA angle, defined by hip-to-ankle landmarks spanning 800–1000 mm (depending on patients' morphology) exhibited minimal sensitivity. This protection scales systematically with measurement span: mHKA outperforms mL DFA, which outperforms mJLCA, reflecting progressively shorter lever arms. Previous studies reporting scan–rescan and inter-observer errors of $1\text{--}2^\circ$ for HKA measurements on long-leg radiographs [34–36] are consistent with our Monte Carlo results.

Conversely, short-span measurements amplify positioning errors into large angular deviations. External tibial torsion quantifies the angular offset between two short transverse axes: the transmalleolar axis (60–70 mm) and the tibial plateau axis (70–80 mm). With four independent

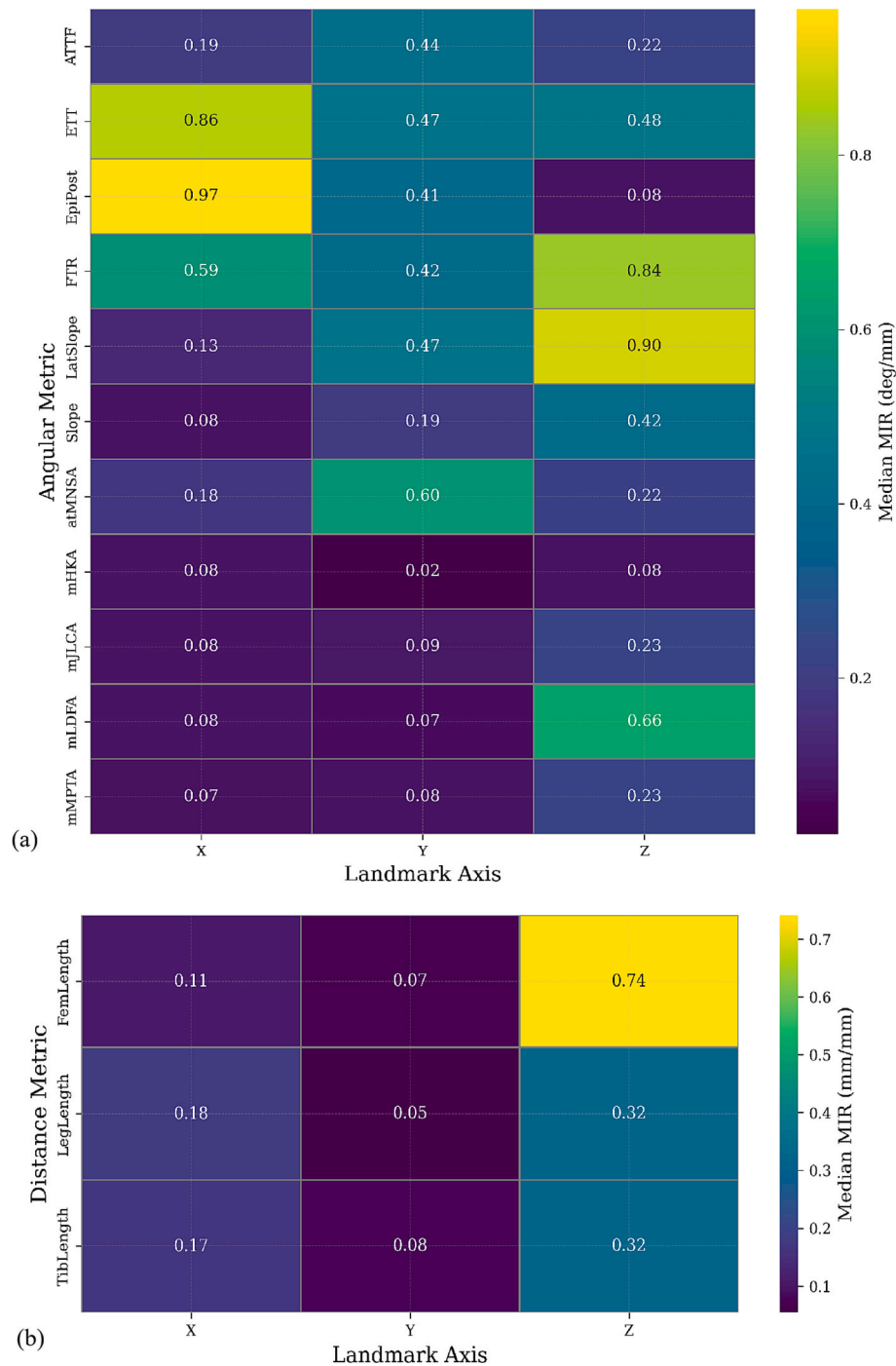


Fig. 11. Metric-axis interaction heatmap across a) angular metrics and b) distance metrics.

landmarks contributing errors and no geometric averaging, uncertainty compounds substantially to produce $CI_{90} = 16.87^\circ$. Recent cadaveric studies confirm only moderate inter-observer reliability for tibial torsion ($ICC \approx 0.49\text{--}0.70$) with absolute errors of $4\text{--}8^\circ$ despite standardized training [37], in agreement with our simulated uncertainty.

4.2.2. Reference frame contamination

Beyond direct geometric effects, errors in plane-defining landmarks cascade into geometrically independent metrics through reference frame distortion. The transepicondylar axis (TEA), connecting medial and lateral epicondyles separated by approximately 80 mm, defines the orientation of the coronal plane for multiple metrics. Berger et al. established the surgical transepicondylar axis as the most reliable

reference for femoral component rotation in total knee arthroplasty [38].

Consider a scenario where the lateral femoral epicondyle landmark experiences a 2.0 mm lateral displacement error. The transepicondylar axis rotates by $\arctan(2.0/80) \approx 1.43^\circ$ in the coronal plane. All subsequent calculations performed in the “coronal plane” now occur in a plane tilted 1.43° from true anatomical orientation. Since mHKA, mLDFA, mMPTA, mJLCA, and atMNSA all rely on this transepicondylar axis to define their coronal reference frame, this frame distortion corrupts all five parameters simultaneously, inducing false varus/valgus interpretation with approximately 1.4° bias, spurious joint line asymmetry, and systematic cross-domain bias affecting component sizing, alignment strategy, and soft tissue balancing predictions.

This mechanism explains why plane-defining landmarks exhibited higher Chatterjee coefficients (median $\xi = 0.082$) than endpoint landmarks ($\xi = 0.063$) despite lower MIR values. The distinction lies between univariate amplitude and multivariate correlation structure: plane-defining landmarks induce moderate deviations (MIR ≈ 0.3 – 0.5 deg/mm) in five metrics simultaneously through reference frame contamination, whereas endpoint landmarks induce large deviations (MIR ≈ 0.8 – 1.2 deg/mm) in only one or two metrics through direct geometric participation.

Critically, contaminated metrics may individually be considered valid because absolute values remain within plausible physiological ranges, a mildly varus mHKA value would not raise concern in isolation. The underlying reference frame error becomes detectable only through cross-metric consistency analysis: if mHKA, mL DFA, mMPTA, mJLCA, and atMNSA all exhibit systematic bias in the same direction (e.g., all shifted toward varus), this correlated pattern reveals reference plane distortion that would escape single-metric validation.

4.2.3. Sensitivity amplification in unstable metrics

The sensitivity of rotational metrics carries direct clinical consequences. Multiple CT-based case–control studies have shown that even modest internal component malrotation (≈ 4 – 6°) is strongly associated with anterior knee pain and patellofemoral complications after TKA, and that revision of malrotated components improves outcomes [39–41]. Patients with combined internal femoral–tibial rotation are more than five times as likely to experience anterior knee pain as those with combined external rotation. Given that ETT measurement uncertainty ($CI_{90} = 16.87^\circ$) substantially exceeds this clinically significant threshold, automated rotational measurements require mandatory manual validation before informing surgical decisions.

Translational positioning errors in TKA also carry significant clinical consequences. Joint line elevation—directly influenced by Z-axis landmark positioning of tibial plateau and femoral condyle landmarks—demonstrates a significant negative correlation with functional outcomes (KSS function score: $\rho = -0.496$, $p < 0.001$). Elevation exceeding 4–5 mm induces mid-flexion instability, patellar impingement, decreased range of motion, and quadriceps weakness [42]. Patients within ± 5 mm of native joint line height are 3.88 times more likely to achieve substantial clinical benefit [43]. Given that our analysis identified the inferior medial femoral condyle Z-coordinate as the highest-impact distance landmark (MIR = 0.792 mm/mm), Z-axis positioning accuracy at this site directly influences joint line restoration and subsequent clinical outcomes.

Tibial length ($CI_{90} = 3.31 \pm 0.79$ mm) and leg length ($CI_{90} = 2.65 \pm 0.85$ mm) exhibited substantially higher uncertainty than femoral length, reflecting cumulative error propagation through multi-stage centroid calculations. The tibial center of shaft and tibial center of plateau each aggregate positioning errors from contributing landmarks. The leg length measurement, combining femoral and tibial segments through summation, exhibited the highest variance (1.08 mm^2).

Distance measurement uncertainty carries direct clinical relevance in other arthroplasty contexts. In total hip arthroplasty, leg length discrepancy (LLD) remains one of the most frequent complications (prevalence 3–32%) and a leading cause of orthopedic litigation [44]. LLD exceeding 10 mm is associated with gait disturbances, back pain, sciatica, and patient dissatisfaction, with studies reporting 18–27% reduction in Oxford Hip Score at one-year follow-up [45]. While our TKA-focused analysis found leg length $CI_{90} = 2.65$ mm—within acceptable margins when considered in isolation—cumulative uncertainties from multiple surgical steps may approach clinically significant thresholds, underscoring the importance of landmark accuracy throughout the measurement chain.

4.2.4. Morphology-dependent sensitivity

Lateral tibial slope exhibited the most dramatic morphology-dependent pattern with $CI_{90} = 7.38^\circ \pm 2.04^\circ$. This inter-patient

variability substantially exceeds that of systematically unstable metrics like ETT and FTR. This pattern indicates that some patients achieve lateral slope $CI_{90} < 5^\circ$ (acceptable for surgical planning where guidelines recommend 3– 7° posterior inclination), while others exceed 10° . The morphology explanation likely involves variations in lateral tibial plateau anatomy. Patients with relatively flat plateaus, exhibiting large anteroposterior dimensions, provide longer measurement baselines and thus create geometric configurations resistant to angular amplification. Conversely, patients with curved plateaus, exhibiting shorter anteroposterior dimensions, create sensitivity to short lever-arm effects. Additionally, cortical definition quality in CT images varies substantially based on bone density and osteoarthritic remodeling patterns.

4.3. A novel hierarchical validation protocol strategy

Based on our findings from CT-based landmarking in osteoarthritic patients undergoing TKA planning, we propose a three-phase hierarchical validation protocol (HVP) addressing both direct geometric effects and reference frame contamination. This protocol is designed for validation of automated landmarking algorithms [22] against manual expert annotations and applies to CT-based surgical planning workflows where inter- and intra-operator variability and algorithmic positioning errors must be quantified prior to clinical deployment.

- Phase 1 prioritize plane-defining landmark validation. All landmarks serving reference plane definition roles (including the femoral epicondyles for coronal plane definition, posterior femoral condyles for axial rotation reference, and tibial plateau boundaries for sagittal plane orientation) require validation before accepting morphometric calculations. These landmarks demand stringent positioning control (± 1.5 mm) regardless of their individual MIR values. If any plane-defining landmark exceeds this tolerance, the entire anatomical plane should be flagged for comprehensive manual review.
- Phase 2 conditionally validate the endpoint landmarks after confirming reference plane integrity. Validation thresholds scale to each landmark's position in the MIR and Chatterjee coefficient distributions: higher-impact combinations require more stringent tolerances, while lower-impact combinations may tolerate relaxed requirements without meaningful consequences for surgical accuracy. This conditional approach prevents unnecessary efforts to validate endpoints when their underlying reference frames already contain systematic errors.
- Phase 3 implements cross-metric consistency checking through automated correlation analysis. If three or more metrics sharing a common reference plane exhibit correlated deviations exceeding 1.5° in the same direction, plane-defining landmarks should be reinspected regardless of their Phase 1 validation status, providing redundant protection against subtle reference frame errors that might escape individual landmark inspection.

Translating these findings into clinical practice requires addressing three implementation-agnostic challenges. First, consistency in landmark definition requires distinguishing geometrically stable landmarks from ill-defined regions susceptible to pathological remodeling; for the latter, consensus guidelines tailored to arthritic anatomy would reduce a primary source of inter-observer variability (as illustrated by the posterior dispersion of annotations at the lateral tibial plateau (Fig. 9), where the absence of consensus on whether to target native bone boundaries or osteophytic overgrowth introduces systematic uncertainty regardless of the landmarking method). Second, automated algorithms can improve reproducibility provided they are trained on such consensus definitions. Third, for manual or hybrid workflows, implementing the cross-metric consistency checks proposed in Phase 3 could enable real-time feedback flagging systematic reference frame distortions, independently of the underlying landmark solution.

4.4. Methodological advances and broader applicability

The atlas-based inverse morphing methodology enables aggregation of landmark positioning errors across patient-specific anatomies into a unified statistical framework, facilitating population-level sensitivity quantification while preserving individual anatomical variability. Previous landmarking validation studies typically reported mean errors and standard deviations for individual landmarks without quantifying downstream impact on surgical morphometrics. Our approach connects positioning accuracy directly to clinical metrics through Monte Carlo uncertainty propagation.

The combination of MIR (amplitude-based) and Chatterjee coefficient (dependency-based) values provides complementary perspectives essential for comprehensive sensitivity characterization. MIR identifies mechanical leverage through direct geometric relationships, revealing which landmarks participate most strongly in metric calculations. Chatterjee coefficients reveal functional coupling through statistical dependencies, detecting multivariate correlation patterns including reference frame contamination effects invisible to univariate analyses.

The framework extends beyond TKA to other landmark-dependent surgical planning applications. Hip arthroplasty relies on acetabular rim landmarks for cup positioning and femoral head-neck landmarks for stem alignment. Spinal surgery employs pedicle landmarks for screw trajectory planning. Craniofacial reconstruction utilizes cephalometric landmarks for symmetry assessment and implant positioning. The core methodological principles—geometric leverage quantification, correlation structure analysis, and Monte Carlo uncertainty propagation—apply wherever anatomical landmark positioning uncertainties propagate into surgical planning measurements. Each application would require domain-specific adaptation of reference frames, clinically relevant metrics, and acceptance thresholds, but the fundamental sensitivity analysis framework remains generalizable.

4.5. Limitations

Several limitations warrant consideration. First, our cohort consisted exclusively of osteoarthritic patients undergoing TKA evaluation; osteoarthritic remodeling, osteophyte formation, and cartilage loss may alter landmark positioning uncertainty patterns compared to healthy joints, and generalizability to non-pathological anatomy or other procedures requires validation. Second, positioning error distributions were derived from manual annotations by two experienced orthopedic surgeons; inter- and intra-operator variability patterns may differ with other expertise levels, annotation protocols, or automated landmarking algorithms employing deep learning approaches. Third, Monte Carlo simulations assumed multivariate normal distributions; alternative distribution models including heavy-tailed distributions or mixture models were not explored. Fourth, the study focused on static measurement uncertainties without considering dynamic factors such as patient positioning variability during CT acquisition, weight-bearing versus supine imaging differences, or surgical workflow constraints. Additionally, no explicit pose normalization beyond the LPS coordinate convention was enforced; residual variability in limb orientation may introduce noise in the reported axis-specific sensitivity patterns, which should therefore be interpreted as population-level tendencies rather than perfectly patient-invariant directional decompositions. Fifth, the analysis assumed accurate bone segmentation. Although the pipeline achieved mean DSC = 0.97 on an independent test set, residual segmentation errors were not modeled; the femoral head center was treated as geometrically fixed to isolate landmark-specific contributions; its sensitivity to upstream segmentation inaccuracies is therefore not captured here. In workflows with lower segmentation quality, cumulative uncertainty from both stages may exceed the values reported. Sixth, the proposed criticality hierarchies were derived from a single-center cohort acquired under a homogeneous CT protocol, using a specific atlas-based morphing pipeline. Generalizability to other imaging

systems, acquisition parameters, osteoarthritis severity profiles, and anatomical morphotypes warrants prospective validation, though the sensitivity analysis framework itself remains protocol-agnostic.

4.6. Future works

Future research should investigate whether patient-specific characteristics—particularly osteoarthritis severity and morphological variants—systematically modulate error propagation patterns. Our cohort comprised exclusively osteoarthritic patients (Ahlbäck grade 2–4), yet preliminary observations suggest that advanced joint degradation, especially lateral compartment involvement, may amplify positioning uncertainty at specific landmarks, raising the question of whether disease severity necessitates adapted validation requirements.

Beyond patient factors, extending this framework to other orthopedic procedures—hip arthroplasty, shoulder reconstruction, spinal fusion—where landmark-based planning similarly governs surgical outcomes would test the generalizability of geometric criticality hierarchies. Finally, integration of uncertainty quantification into surgical planning interfaces, displaying measurement confidence intervals alongside nominal values, could fundamentally change how surgeons interpret borderline measurements in clinical practice.

5. Conclusion

This study reveals a fundamental paradigm shift in how landmark positioning errors should be evaluated for surgical planning: positioning accuracy and clinical impact are not synonymous. Current validation practices, which apply uniform tolerance thresholds across all landmarks, fail to capture the true propagation of measurement uncertainty into surgical metrics. Our findings demonstrate that seemingly acceptable positioning errors at geometrically critical sites can induce clinically significant deviations in alignment and joint geometry measurements, while larger errors at non-critical landmarks remain inconsequential.

To address this gap, we developed novel sensitivity indicators: the Morphometrical Impact Rate (MIR) quantifies error amplitude per millimeter of landmark displacement, while the Chatterjee coefficient captures statistical dependency patterns including reference frame contamination effects. Complementary visualization tools—Hanabi plots, sensitivity heatmaps, and criticality maps—translate these quantitative findings into intuitive representations supporting clinical decision-making and quality control implementation.

We propose a hierarchical risk-stratified validation framework comprising three priorities: first, ensuring the reliability of plane-defining landmarks that establish anatomical reference frames, as errors at these sites cascade into multiple geometrically independent metrics; second, validating endpoint landmarks according to their demonstrated mechanical impact on specific surgical measurements; third, implementing cross-metric consistency checks to detect subtle reference frame distortions that would escape individual landmark inspection. This framework, developed here for CT-based TKA planning, must be adapted to the imaging modality and clinical context—notably, lower-resolution intraoperative imaging (e.g., robotic systems with limited field-of-view) may exhibit different error distributions requiring modality-specific validation requirements. Looking forward, key questions remain on whether patient-specific morphological characteristics can enable adaptive validation protocols, and how uncertainty-aware interfaces could translate these findings into routine clinical practice.

CRediT authorship contribution statement

Ysé Roch: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Dorian Lozano:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal

analysis, Data curation, Conceptualization. **Nolwenn Fougeron:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Benoît Gaulin:** Writing – review & editing, Investigation, Data curation, Conceptualization. **Clément Horteur:** Writing – review & editing, Investigation, Data curation. **Yohan Payan:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Marek Bucki:** Writing – review & editing, Validation, Supervision, Software, Methodology, Conceptualization. **Antoine Perrier:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yohan Payan reports financial support was provided by Fondation pour la Recherche Médicale (FRM DIC20161236448).

Antoine Perrier reports financial support was provided by Fondation pour la Recherche Médicale (FRM DIC20161236448).

Nolwenn Fougeron reports financial support was provided by Programme Innovation Technologique of the Région Auvergne-Rhône-Alpes.

Ysé Roch reports equipment, drugs, or supplies was provided by

Annex.

Table 1, Table 2 and Table 3 present the angles and distances with their corresponding descriptions and the centroid's definition. Each metric is computed from 2 vectors (U and V), each defined by 2 landmarks (U1 and U2; V1 and V2).

Table 1

Angular metric names with their description, 2D projection plane and landmarks defining U and V.

Angular metrics	Description	2D Projection Plane	Landmarks (* are detailed in Table 3)
mHKA	Mechanical Hip Knee Ankle Angle	Coronal	U1: TCS* U2: Femoral Hip Center V1: TCS* V2: TCP*
mLDFA	Lateral Distal Femoral Angle	Coronal	U1: Femoral Notch Inf U2: Femoral Hip Center V1: Femoral Condyle Inf. Med. V2: Femoral Condyle Inf Lat.
mMPTA	Medial Proximal Tibial Angle	Coronal	U1: TCS* U2: TCP* V1: TLCC* V2: TMCC*
mJLCA	Joint Line Convergence Angle	Coronal	U1: Femoral Condyle Inf. Med. U2: Femoral Condyle Inf Lat. V1: TLCC* V2: TMCC*
atMNSA	Coronal Cervico-Diaphyseal Angle	Coronal	U1: Femoral Shaft Sup Sup Center U2: Femoral Shaft Inf Inf Center V1: Femoral Neck Center V2: Femoral Hip Center
Slope	Posterior inclination of the tibial plateau	Sagittal	U1: TCCsag* U2: TCPsag* V1: TAC* V2: TPC*
LatSlope	Posterior inclination of the lateral tibial plateau	Sagittal	U1: TCCsag* U2: TCPsag* V1: Tibial Plate Lat. Sup. Ant. V2: Tibial Plate Lat. Sup. Post
ATTf – condyl post	Femoral Neck Anteversion Posterior Condyles	Axial	U1: Femoral Neck Center U2: Femoral Hip Center V1: Femoral Condyle Post Lat. V2: Femoral Condyle Post Med.
EpiPost	Epicondyle—Condyle Post angle Positive means internal rotation of the posterior condyles relative to the epicondyle's axis	Axial	U1: Femoral Epicondyle Lat. U2: Femoral Epicondyle Med. V1: Femoral Condyle Post Lat. V2: Femoral Condyle Post Med.

(continued on next page)

Grand Équipement National De Calcul Intensif (GENCI-IDRIS grant 2023-A0160313879).

Antoine Perrier reports a relationship with Twinsight that includes: board membership, employment, and equity or stocks.

Ysé Roch reports a relationship with Twinsight that includes: employment, equity or stocks.

Marek Bucki reports a relationship with Twinsight that includes: board membership, employment, and equity or stocks.

Dorian Lozano reports a relationship with Twinsight that includes: employment.

If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by Fondation pour la Recherche Médicale under the project FRM DIC20161236448.

This work received the financial support of the *Programme Innovation Technologique* of the Région Auvergne-Rhône-Alpes.

This project was provided with computing HPC and storage resources by GENCI at IDRIS thanks to the grant 2023-A0160313879 on the supercomputer Jean Zay's A100 partition .

Table 1 (continued)

Angular metrics	Description	2D Projection Plane	Landmarks (* are detailed in Table 3)
ETT	External tibial torsion angle	Axial	U1: Tibial Plate Med. U2: Tibial Plate Lat. V1: Tibial Mal. Med. V2: Tibial Mal. Lat.
FTR	Rotation between Femur and Tibia	Axial	U1: Femoral Condyle Post Med. U2: Femoral Condyle Post Lat. V1: Tibial Plate Med. V2: Tibial Plate Lat.

Table 2

Distance metric names with their description, 2D projection plane and landmarks defining U and V.

Distance Metrics	Description	Landmarks (* are detailed in Table 3)
Femur Length	3D femoral length from femoral head to medial condyle	U1: Femoral Hip Center U2: Femoral Condyle Inf. Med.
Tibia Length	3D tibial length	U1: TCS* U2: TCP*
Leg Length	3D leg length	U1: Femoral Hip Center U2: TCP*

Table 3

Tibial Sparse Centroids definitions.

	Description	Landmarks
TPC	Tibial Posterior Interplateau Center	1/2 Tibial Plate Med. Sup. Post. 1/2 Tibial Plate Lat. Sup. Post
TAC	Tibial Anterior Interplateau Center	1/2 Tibial Plate Med. Sup. Ant. 1/2 Tibial Plate Lat. Sup. Ant.
TCP	Tibial Plafond Center	1/2 Tibial Distal Med. 1/2 Tibial Distal Lat.
TCS	Tibial Spine Centroid	1/2 Tibial Plate Spine Med 1/2 Tibial Plate Spine Lat
TCCsag	Tibial Median Interplateau Center	1/5 TPC 4/5 TAC
TCPsag	Tibial Plafond Sagittal Center	1/2 Tibial Distal Post. 1/2 Tibial Distal Ant.
TMCC	Tibial Medial Plateau Center	1/4 Tibial Plate Med. Sup. Post. 1/4 Tibial Plate Med. Sup. Ant. 1/4 Tibial Plate Spine Med. 1/4 Tibial Plate Med.
TLCC	Tibial Lateral Plateau Center	1/4 Tibial Plate Lat. Sup. Post. 1/4 Tibial Plate Lat. Sup. Ant. 1/4 Tibial Plate Spine Lat. 1/4 Tibial Plate Lat.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejrad.2026.112894>.

References

- [1] M.J. DeFrance, G.R. Scuder, Are 20% of patients actually dissatisfied following total knee arthroplasty? A systematic review of the literature, *J. Arthroplasty* 38 (3) (2023 Mar) 594–599, <https://doi.org/10.1016/j.arth.2022.10.011>. Epub 2022 Oct 14 PMID: 36252743.
- [2] C. Rivière, J.M. Vigdorchik, P.A. Vendittoli, Mechanical alignment: the end of an era!, *Orthop. Traumatol. Surg. Res.* 105 (7) (2019 Nov) 1223–1226, <https://doi.org/10.1016/j.otsr.2019.07.005>. Epub 2019 Aug 1 PMID: 31378698.
- [3] S.M. Howell, M.L. Hull, A.J. Nedopil, C. Rivière, Caliper-verified kinematically aligned total knee arthroplasty: rationale, targets, accuracy, balancing, implant survival, and outcomes, *Instr. Course Lect.* 72 (2023) 241–259. PMID: 36534860.
- [4] Morcos MW, Beckers G, Salvi AG, Bennani M, Massé V, Vendittoli PA. Excellent results of restricted kinematic alignment total knee arthroplasty at a minimum of 10 years of follow-up. *Knee Surg Sports Traumatol Arthrosc.* 2025 Feb;33(2):654–665. doi: 10.1002/ksa.12452. Epub 2024 Sep 9. PMID: 39248213; PMCID: PMC11792112.
- [5] Shatrov J, Battelier C, Sappey-Mariner E, Gunst S, Servien E, Lustig S. Functional Alignment Philosophy in Total Knee Arthroplasty - Rationale and technique for the varus morphotype using a CT based robotic platform and individualized planning. *SICOT J.* 2022;8:11. doi: 10.1051/sicotj/2022010. Epub 2022 Apr 1. Erratum in: *SICOT J.* 2022;8:18. doi: 10.1051/sicotj/2022017. PMID: 35363136; PMCID: PMC8973302.
- [6] R. Pailhé, Total knee arthroplasty: latest robotics implantation techniques, *Orthop. Traumatol. Surg. Res.* 107 (1S) (2021 Feb) 102780, <https://doi.org/10.1016/j.otsr.2020.102780>. Epub 2021 Jan 11. PMID: 33333275.
- [7] D. Saragaglia, L. Sigwalt, J. Gaillot, V. Morin, B. Rubens-Duval, R. Pailhé, Results with eight and a half years average follow-up on two hundred and eight e-Motion FP® knee prostheses, fitted using computer navigation for knee osteoarthritis in patients with over ten degrees genu varum, *Int. Orthop.* 42 (4) (2018 Apr)

- 799–804, <https://doi.org/10.1007/s00264-017-3618-8>. Epub 2017 Aug 22 PMID: 28831536.
- [8] B.J. Keyes, D.C. Markel, R.M. Meneghini, Evaluation of limb alignment, component positioning, and function in primary total knee arthroplasty using a pinless navigation technique compared with conventional methods, *J. Knee Surg.* 26 (2) (2013 Apr) 127–132, <https://doi.org/10.1055/s-0032-1319788>. Epub 2012 Jul 13 PMID: 23288760.
 - [9] Horteur C, Gaulin B, Pascal P, Leroy C, Gaii J, Murgier J, Barth J, Pailhé R. Computed Tomography-Based Analysis of Implant Positioning after Total Knee Arthroplasty: A Randomized Controlled Trial Comparing Conventional and Robotic Arm-Assisted Procedures. *J Knee Surg.* 2025 Aug 6. doi: 10.1055/a-2664-7551. Epub ahead of print. PMID: 40769204.
 - [10] Veerman QWT, Ten Heggeler RM, Tuijthof GJM, de Graaff F, Fluit R, Hoogeslag RAG. High variability exists in 3D leg alignment analysis, but underlying principles that might lead to agreement on a universal framework could be identified: A systematic review. *Knee Surg Sports Traumatol Arthrosc.* 2025 Jun;33(6):2063–2077. doi: 10.1002/ksa.12512. Epub 2024 Oct 26. PMID: 39460613; PMCID: PMC12104779.
 - [11] Park A, Nam D, Friedman MV, Duncan ST, Hillen TJ, Barrack RL. Inter-observer precision and physiologic variability of mri landmarks used to determine rotational alignment in conventional and patient-specific TKA. *J Arthroplasty.* 2015 Feb;30 (2):290–5. doi: 10.1016/j.arth.2014.08.015. Epub 2014 Sep 6. PMID: 25267537; PMCID: PMC4323956.
 - [12] M. Saffarini, L. Nover, R. Tandogan, R. Becker, L.B. Moser, M.T. Hirschmann, P. F. Indelli, The original Akagi line is the most reliable: a systematic review of landmarks for rotational alignment of the tibial component in TKA, *Knee Surg. Sports Traumatol. Arthrosc.* 27 (4) (2019 Apr) 1018–1027, <https://doi.org/10.1007/s00167-018-5131-z>. Epub 2018 Sep 10 PMID: 30203197.
 - [13] E.J. De Valk, J.C. Noorduyt, E.L. Mutsaerts, How to assess femoral and tibial component rotation after total knee arthroplasty with computed tomography: a systematic review, *Knee Surg. Sports Traumatol. Arthrosc.* 24 (11) (2016 Nov) 3517–3528, <https://doi.org/10.1007/s00167-016-4325-5>. Epub 2016 Sep 21 PMID: 27655141.
 - [14] J. Victor, D. Van Doninck, L. Labey, F. Van Glabbeek, P. Parizel, J. Bellemans, A common reference frame for describing rotation of the distal femur: a ct-based kinematic study using cadavers, *J. Bone Joint Surg. Br.* 91 (5) (2009 May) 683–690, <https://doi.org/10.1302/0301-620X.91B5.21827>. PMID: 19407308.
 - [15] J.G. Seo, S.H. Choi, S. Shin, B.H. Lee, M.J. Chang, Relationship between mechanical axis-derived and anatomic landmark-derived femoral rotation in TKA: a three-dimensional CT study, *J. Arthroplasty* 29 (12) (2014 Dec) 2314–2318, <https://doi.org/10.1016/j.arth.2014.07.017>. Epub 2014 Jul 24 PMID: 25138615.
 - [16] K. Lei, L.M. Liu, J.M. Luo, C. Ma, Q. Feng, L. Yang, L. Guo, Could surgical transepicondylar axis be identified accurately in preoperative 3D planning for total knee arthroplasty? a reproducibility study based on 3D-CT, *Arthroplasty* 4 (1) (2022 Oct 17) 46, <https://doi.org/10.1186/s42836-022-00147-2>. PMID: 36244969; PMCID: PMC9575283.
 - [17] J. Victor, D. Van Doninck, L. Labey, B. Innocenti, P.M. Parizel, J. Bellemans, How precise can bony landmarks be determined on a CT scan of the knee? *Knee* 16 (5) (2009 Oct) 358–365, <https://doi.org/10.1016/j.knee.2009.01.001>. Epub 2009 Feb 5 PMID: 19195896.
 - [18] K. Liu, X. Liu, Y. Guan, H. Ma, D. Fu, Z. Fan, Accuracy and reproducibility analysis of different reference axes for femoral prosthesis rotation alignment in TKA based on 3D CT femoral model, *BMC Musculoskelet. Disord.* 24 (1) (2023 Aug 18) 660, <https://doi.org/10.1186/s12891-023-06781-4>. PMID: 37596664; PMCID: PMC10439596.
 - [19] C.L. Wai Hung, Y. Wai Pan, C. Kwong Yuen, L. Hon Bong, L.W. Lei Sha, S.W. Ho Man, Interobserver and intraobserver error in distal femur transepicondylar axis measurement with computed tomography, *J. Arthroplasty* 24 (1) (2009 Jan) 96–100, <https://doi.org/10.1016/j.arth.2007.11.014>. Epub 2008 Mar 28 PMID: 18534429.
 - [20] N.J. Wilhelm, C.E. von Schacky, F.J. Lindner, M.J. Feucht, Y. Ehmann, J. Pogorzelski, S. Haddadin, J. Neumann, F. Hinterwimmer, R. von Eisenhart-Rothe, M. Jung, M.F. Russe, K. Izadpanah, S. Siebenlist, R. Burgkart, M.C. Rupp, Multicentric development and validation of a multi-scale and multi-task deep learning model for comprehensive lower extremity alignment analysis, *Artif. Intell. Med.* 150 (2024 Apr) 102843, <https://doi.org/10.1016/j.artmed.2024.102843>. Epub 2024 Mar 14 PMID: 38553152.
 - [21] A. Tiulpin, I. Melekchov and S. Saarakkala, “KNEEL: Knee Anatomical Landmark Localization Using Hourglass Networks,” 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 352–361, doi: 10.1109/ICCVW.2019.00046.
 - [22] C. Horteur, N. Fougeron, B. Gaulin, Y. Roch, M. Bucki, E. Palluel, Y. Payan, A. Perrier, Accuracy evaluation of an automatic landmarking method based on 3D segmented CT scans of full lower limbs with knee osteoarthritis, *Eur. J. Radiol.* 191 (2025 Oct) 112292, <https://doi.org/10.1016/j.ejrad.2025.112292>. Epub 2025 Jul 8 PMID: 40652606.
 - [23] U. Della Croce, A. Cappozzo, D.C. Kerrigan, Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles, *Med. Biol. Eng. Comput.* 37 (2) (1999 Mar) 155–161, <https://doi.org/10.1007/BF02513282>. PMID: 10396818.
 - [24] J. Grammens, A. Van Haver, I. Lumban-Gaol, F. Danckaers, P. Verdonk, J. Sijbers, Automated landmark annotation for morphometric analysis of distal femur and proximal tibia, *J. Imaging* 10 (4) (2024) 90, <https://doi.org/10.3390/jimaging10040090>.
 - [25] T. Luo, N. Martensson, J. Howard, D. Stevens, K. McIsaac, B. Lanting, Identifying sources of error in computer-navigated total knee arthroplasty using sensitivity analyses in knee models, *J. Arthroplasty* 40 (2025), <https://doi.org/10.1016/j.arth.2025.02.061>.
 - [26] D.F. Amanatullah, P.E. Di Cesare, P.A. Meere, G.C. Pereira, Identification of the landmark registration safe zones during total knee arthroplasty using an imageless navigation system, *J. Arthroplasty* 28 (6) (2013 Jun) 938–942, <https://doi.org/10.1016/j.arth.2012.12.013>. Epub 2013 Apr 6 PMID: 23566700.
 - [27] B. Schlatterer, J.M. Linares, P. Chabrand, J.M. Sprael, J.N. Argenson, Influence of the optical system and anatomic points on computer-assisted total knee arthroplasty, *Orthop. Traumatol. Surg. Res.* 100 (4) (2014 Jun) 395–402, <https://doi.org/10.1016/j.otsr.2013.12.029>. Epub 2014 May 14 PMID: 24835533.
 - [28] Schelker BL, Nowakowski AM, Hirschmann MT. What is the “safe zone” for transition of coronal alignment from systematic to a more personalised one in total knee arthroplasty? A systematic review. *Knee Surg Sports Traumatol Arthrosc.* 2022 Feb;30(2):419–427. doi: 10.1007/s00167-021-06811-5. Epub 2022 Jan 1. PMID: 34973095; PMCID: PMC8866271.
 - [29] Reinke A, Tizabi MD, Baumgartner M, Eisenmann M, Heckmann-Nötzl D, Kavur AE, Rädtsch T, Sudre CH, Acion L, Antonelli M, Arbel T, Bakas S, Benis A, Buettner F, Cardoso MJ, Cheplygina V, Chen J, Christodoulou E, Cimini BA, Farahani K, Ferrer L, Galdran A, van Ginneken B, Glocker B, Godau P, Hashimoto DA, Hoffman MM, Huisman M, Isensee F, Jannin P, Kahn CE, Kainmueller D, Kainz B, Karargyris A, Kleesiek J, Kofler F, Kooi T, Kopp-Schneider A, Kozubek M, Kreshuk A, Kurc T, Landman BA, Litjens G, Madani A, Maier-Hein K, Martel AL, Meijering E, Menze B, Moons KGM, Müller H, Nychyporuk B, Nickel F, Petersen J, Rafelski SM, Rajpoot N, Reyes M, Riegler M, Rieke N, Saez-Rodriguez J, Sánchez CI, Shetty S, Summers RM, Taha AA, Tiulpin A, Tsaftaris SA, Van Calster B, Varoquaux G, Yaniv ZR, Jäger PF, Maier-Hein L. Understanding metric-related pitfalls in image analysis validation. *Nat Methods.* 2024 Feb;21(2):182–194. doi: 10.1038/s41592-023-02150-0. Epub 2024 Feb 12. PMID: 38347140; PMCID: PMC1181963.
 - [30] S. Chatterjee, A new coefficient of correlation, *J. Amer. Statist. Assoc.* 116 (536) (2021) 2009–2022.
 - [31] Schulze M, Juergensen L, Rischen R, Toennemann M, Reischle G, Puetzler J, Gosheger G, Hasselmann J. Quality assurance of 3D-printed patient specific anatomical models: a systematic review. *3D Print Med.* 2024 Mar 27;10(1):9. doi: 10.1186/s41205-024-00210-5. PMID: 38536566; PMCID: PMC10967057.
 - [32] H. Jacinto, S. Valette, R. Prost, Multi-atlas automatic positioning of anatomical landmarks, *J. Vis. Commun. Image Represent.* 50 (2017), <https://doi.org/10.1016/j.jvcir.2017.11.015>.
 - [33] R.J.A. Kuiper, P.R. Seevinck, M.A. Viergever, H. Weinans, R.J.B. Sakkars, Automatic assessment of lower-limb alignment from computed tomography, *J. Bone Joint Surg. Am.* 105 (9) (2023 May 3) 700–712, <https://doi.org/10.2106/JBJS.22.00890>. Epub 2023 Mar 22 PMID: 36947661.
 - [34] E. Skyttä, V. Haapamäki, M. Koivikko, H. Huhtala, V. Remes, Reliability of the hip-to-ankle radiograph in determining the knee and implant alignment after total knee arthroplasty, *Acta Orthop. Belg.* 77 (2011) 329–335.
 - [35] Y. Saiki, T. Kabata, T. Ojima, et al., Reliability and validity of OpenPose for measuring hip-knee-ankle angle in patients with knee osteoarthritis, *Sci. Rep.* 13 (2023) 3297, <https://doi.org/10.1038/s41598-023-30352-1>.
 - [36] S.H. Kim, Y.B. Park, M.K. Song, J.W. Lim, H.J. Lee, Reliability and validity of the femorotibial mechanical axis angle in primary total knee arthroplasty: navigation versus weight bearing or supine whole leg radiographs, *Knee Surg Relat Res.* 30 (4) (2018 Dec 1) 326–333, <https://doi.org/10.5792/ksrr.18.028>. PMID: 30466253; PMCID: PMC6254869.
 - [37] E.W. Edmonds, K.C. Parvareh, M.J. Price, C.L. Farnsworth, J.D. Bomar, J. L. Hughes, V.V. Upasani, The reliability of measurements for tibial torsion: a comparison of CT, MRI, biplanar radiography, and 3D reconstructions with and without standardized measurement training, *J Pediatr Soc North Am.* 5 (3) (2024 Feb 12) 661, <https://doi.org/10.55275/JPOSNA-2023-661>. PMID: 40433330; PMCID: PMC12088203.
 - [38] R.A. Berger, H.E. Rubash, M.J. Seel, W.H. Thompson, L.S. Crossett, Determining the rotational alignment of the femoral component in total knee arthroplasty using the epicondylar axis, *Clin. Orthop. Relat. Res.* 286 (1993 Jan) 40–47. PMID: 8425366.
 - [39] R.L. Barrack, T. Schrader, A.J. Bertot, M.W. Wolfe, L. Myers, Component rotation and anterior knee pain after total knee arthroplasty, *Clin. Orthop. Relat. Res.* 392 (2001 Nov) 46–55, <https://doi.org/10.1097/00003086-200111000-00006>. PMID: 11716424.
 - [40] R.A. Berger, L.S. Crossett, J.J. Jacobs, H.E. Rubash, Malrotation causing patellofemoral complications after total knee arthroplasty, *Clin. Orthop. Relat. Res.* 356 (1998 Nov) 144–153, <https://doi.org/10.1097/00003086-199811000-00021>. PMID: 9917679.
 - [41] Valkering KP, Breugem SJ, van den Bekerom MP, Tuinebreijer WE, van Geenen RC. Effect of rotational alignment on outcome of total knee arthroplasty. *Acta Orthop.* 2015;86(4):432–9. doi: 10.3109/17453674.2015.1022438. Epub 2015 Feb 24. PMID: 25708694; PMCID: PMC4513597.
 - [42] van Lieshout WAM, Valkering KP, Koenraadt KLM, van Etten-Jamaludin FS, Kerkhoffs GMMJ, van Geenen RCI. The negative effect of joint line elevation after total knee arthroplasty on outcome. *Knee Surg Sports Traumatol Arthrosc.* 2019 May;27(5):1477–1486. doi: 10.1007/s00167-018-5099-8. Epub 2018 Aug 14. PMID: 30109369; PMCID: PMC6527530.
 - [43] L.T. Buller, C.M. Metzger, E.R. Deckard, R.M. Meneghini, The effect of joint line elevation on patient-reported outcomes after contemporary revision total knee

- arthroplasty, *J. Arthroplasty* 37 (6) (2022 Jun) 1146–1152, <https://doi.org/10.1016/j.arth.2022.02.062>. Epub 2022 Feb 21 PMID: 35202760.
- [44] A.S. Desai, A. Dramis, T.N. Board, Leg length discrepancy after total hip arthroplasty: a review of literature, *Curr. Rev. Musculoskelet. Med.* 6 (4) (2013 Dec) 336–341, <https://doi.org/10.1007/s12178-013-9180-0>. PMID: 23900834; PMCID: PMC4094096.
- [45] M.R. Whitehouse, N.S. Stefanovich-Lawbuary, L.R. Brunton, A.W. Blom, The impact of leg length discrepancy on patient satisfaction and functional outcome following total hip arthroplasty, *J. Arthroplasty* 28 (8) (2013 Sep) 1408–1414, <https://doi.org/10.1016/j.arth.2012.12.009>. Epub 2013 Mar 16 PMID: 23507069.