

CLINICAL METRIC ERRORS INDUCED BY THE PROPAGATION OF LANDMARKING UNCERTAINTY IN AUTOMATED TKA PLANNING

Ysé Roch (1, 3), Dorian Lozano (3), Nolwenn Fougeron (5), Benoît Gaulin (1,2), Clément Horteur (1,2), Yohan Payan (1), Marek Bucki (3), Antoine Perrier (1,3,4)

1. TIMC, Univ. Grenoble Alpes, CNRS, France; 2. Grenoble-Alpes CHU, France; 3. Twinsight, France; 4. GH Diaconesses Croix St-Simon, France; 5. Univ. Rennes, Inria, CNRS, France

Correspondence: yse.roch@univ-grenoble-alpes.fr

1. Introduction

Automated total knee arthroplasty (TKA) planning relies on anatomical landmarks to compute surgical metrics. These algorithms are typically evaluated on localization accuracy alone, yet a small positioning error can have very different clinical consequences depending on the metric considered. We present a Monte Carlo framework quantifying how landmark errors propagate into TKA planning metrics, illustrated by two contrasting cases: the mechanical hip-knee-ankle angle (mHKA) and the external tibial torsion angle (ETT).

2. Materials and Methods

Thirty osteoarthritic TKA subjects (67.1 ± 8.8 years) underwent preoperative full-leg CT imaging. Landmarks were manually annotated twice by two surgeons on segmented lower-limb meshes, characterizing landmark positioning variability [1]. These distributions were propagated through 14 TKA planning metrics via atlas-based inverse morphing and Monte Carlo simulation (50,000 iterations/metric/patient). Sensitivity was quantified by the 90% confidence interval of metric deviation (CI_{90}) and the Chatterjee coefficient (ξ), which measures statistical dependency between landmark displacement and metric deviation [2]. An automatic landmarking procedure, custom-made, was also applied to the same subjects.

3. Results

Automatic landmarking achieved a mean error of 2.4 mm, comparable to manual inter-operator variability (2.3 mm). Error propagation revealed a 146-fold range in angular variance across metrics. The mHKA was robust ($CI_{90} < 2^\circ$), with heterogeneous but predominantly low landmark-metric dependency (ξ ranging 0–0.29), driven primarily by tibial plateau landmarks along the X-axis. In contrast, the ETT was highly sensitive ($CI_{90} = 16.9^\circ$), with strong statistical

dependency at malleolar landmarks (Fig 1, medial malleolus: $\xi = 0.231$), reflecting four landmarks defining two short transverse axes where in-plane errors compound without geometric averaging.

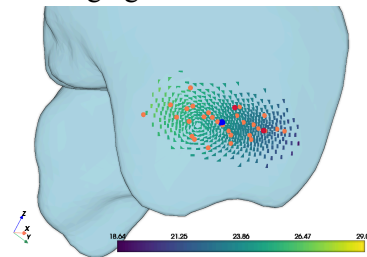


Fig 1: ETT sensitivity to medial malleolus displacement, subject 11 (lateral view). Arrow color encodes the induced angular error magnitude ($^\circ$); orange dots show manual annotation variability; blue dot indicates the reference position.

4. Discussion and Conclusions

Standard landmarking accuracy is an insufficient predictor of clinical metric reliability. The ETT/mHKA contrast shows that geometric configuration, not error magnitude, determines sensitivity: short transverse axes amplify errors multiplicatively, while a long mechanical axis absorbs them. Patient morphology further modulates this effect, making landmark criticality both metric-specific and patient-dependent. This framework directly supports risk-stratified verification and validation of automated TKA planning systems.

5. References

1. Horteur C, et al. Eur J Radiol. 2025 Oct;191:112292. PMID: 40652606.
2. Roch Y et al. How do lower-limb landmarking errors impact morphometrics in total knee arthroplasty planning? (under review), 2025.

Acknowledgements:

Funded by FRM (DIC20161236448), Région AuRA (PIT), and GENCI/IDRIS (grant 2023-A0160313879). IRMaGe facility: ANR-11-INSB-0006. CIFRE thesis (Twinsight/UGA). Clinical data: CHU Grenoble Alpes.