@BoneJointOpen

Interobserver reliability of Coronal Plane Alignment of the Knee (CPAK) phenotype classification

external validation using data from the Osteoarthritis Initiative

From Imperial College, London, UK

Correspondence should be sent to W. Wynell-Mayow william.wynell-mayow@nhs.net

Cite this article:
Bone Jt Open 2025;6(11): 1425–1435.

DOI: 10.1302/2633-1462.611.BJO-2025-0153.R1

W. Wynell-Mayow,[1] T. A. G. Hall,[2] A. Singh,[3] R. J. van Arkel,[2] G. G. Jones[1,4]

[1]Imperial College Healthcare NHS Trust, St Mary's Hospital, London, UK
[2]Biomechanics Group, Department of Mechanical Engineering, Imperial College London, London, UK
[3]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[4]MSk Lab, Department of Surgery and Cancer, Imperial College London, London, UK

Aims
Coronal Plane Alignment of the Knee (CPAK) phenotyping is gaining momentum in research and clinical practice to understand individualized knee alignments and predict knee balance in total knee arthroplasty (TKA). The nine CPAK classes are based on joint line obliquity (JLO) and arithmetic hip-knee-ankle angle (aHKA), which are calculated using the medial proximal tibial angle (MPTA) and lateral distal femoral angle (LDFA). This study aims to assess CPAK classification reproducibility, and analyze what level of angular error is associated with CPAK misclassification.

Methods
Two readers labelled 75 long-leg radiographs (LLRs) from the Osteoarthritis Initiative database for analyses of CPAK inter-reader reproducibility. A single reader then labelled and classified phenotypes for an aggregate total of 1,128 LLRs. Finally, Monte Carlo simulations were run based on 1,128-patient phenotype distribution and the inter-reader reproducibility statistics to understand how CPAK agreement rates were affected by the reproducibility of MPTA and LDFA measurements.

Results
There was excellent reproducibility in MPTA and LDFA measurements (mean absolute error: 0.41°/0.71°; and intraclass correlation coefficient: 0.96°/0.91°, respectively). These small angular deviations led to one-in-five disagreement in CPAK classification (20.0%; 95% CI 10.9% to 29.1%). An aHKA mean absolute error of < 0.1°, which is potentially unattainable, would be required to reduce inter-reader CPAK disagreement to below 95%.

Conclusion
CPAK phenotyping from long-leg radiographs may result in clinically significant rates of misclassification. CT imaging may improve reliability, particularly in cases where aHKA and JLO are near to discriminatory values.

Take home message
- Despite excellent inter-reader reliability in medial proximal tibial angle and lateral distal femoral angle, there was a 20% inter-reader disagreement in Coronal Plane Alignment of the Knee (CPAK) classification.
- Caution should be exercised when using CPAK phenotype measured from long-leg radiographs in either surgical planning or clinical studies.

Bone & Joint Open

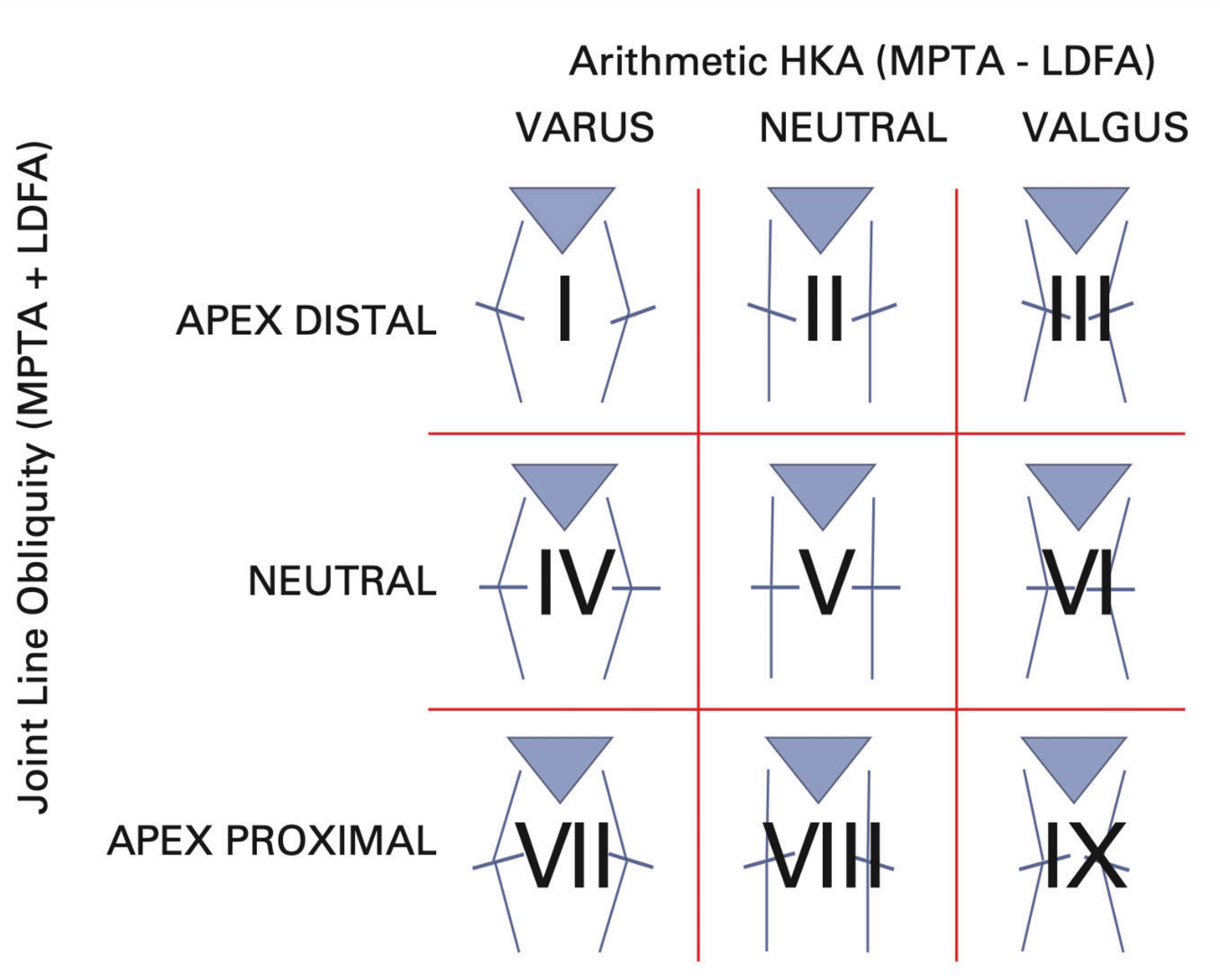

Fig. 1
Schematic of the Coronal Plane Alignment of the Knee (CPAK) classification developed by MacDessi et al.[7] LDFA, lateral distal femoral angle; MPTA, medial proximal tibial angle.

Introduction

In knee joint arthroplasty, recognition of the natural variations in pre-morbid coronal alignment has led to surgeons attempting to recreate patients' constitutional 'kinematic alignment' (KA) during surgery.[1-7] In 2021, MacDessi et al[7] proposed the Coronal Plane Alignment of the Knee (CPAK) as a 'pragmatic and comprehensive' pre- and post-arthritic phenotypic classification. CPAK combined their method for assessing pre-morbid arithmetic hip-knee-ankle alignment (aHKA) with joint line obliquity (JLO).[7-9] Categorizing knees as neutral, valgus, or varus by aHKA and neutral, apex distal, or apex proximal by JLO results in nine CPAK phenotypes (Figure 1), and MacDessi et al[7] demonstrated that CPAK classes I, II, and IV were significantly more likely to be successfully soft-tissue balanced with restricted KA than conventional mechanical alignment (MA).

CPAK is typically calculated using standard long-leg radiographs (LLRs), in which alignment measurements are known to be affected by limb rotation and fixed flexion deformities.[10-13] We hypothesized that these errors, together with challenges in labelling the distal femoral and proximal tibial joint lines would introduce CPAK phenotype variation between readers. Therefore, this study aimed to externally validate CPAK reproducibility on standard LLRs using the prescribed methodology. While previous studies have assessed reproducibility of medial proximal tibial angle (MPTA), lateral distal femoral angle (LDFA), JLO, and aHKA, none have assessed the combined impact of these constitutive measurements on CPAK classification error.[8,14]

Methods

Study design and source data

This is a cross-sectional study of data sourced from the Osteoarthritis Initiative (OAI), a multicentre, longitudinal, prospective, ten-year observational study focused on knee osteoarthritis, with full enrolment criteria in its permanent archive.[15] Broadly, the OAI contains data on patients of both sexes aged 45 to 79 years either with, or at risk of, developing symptomatic knee osteoarthritis. Central image assessments performed under supervision of OAI include standardized HKA values measured from full-limb radiographs.[16]

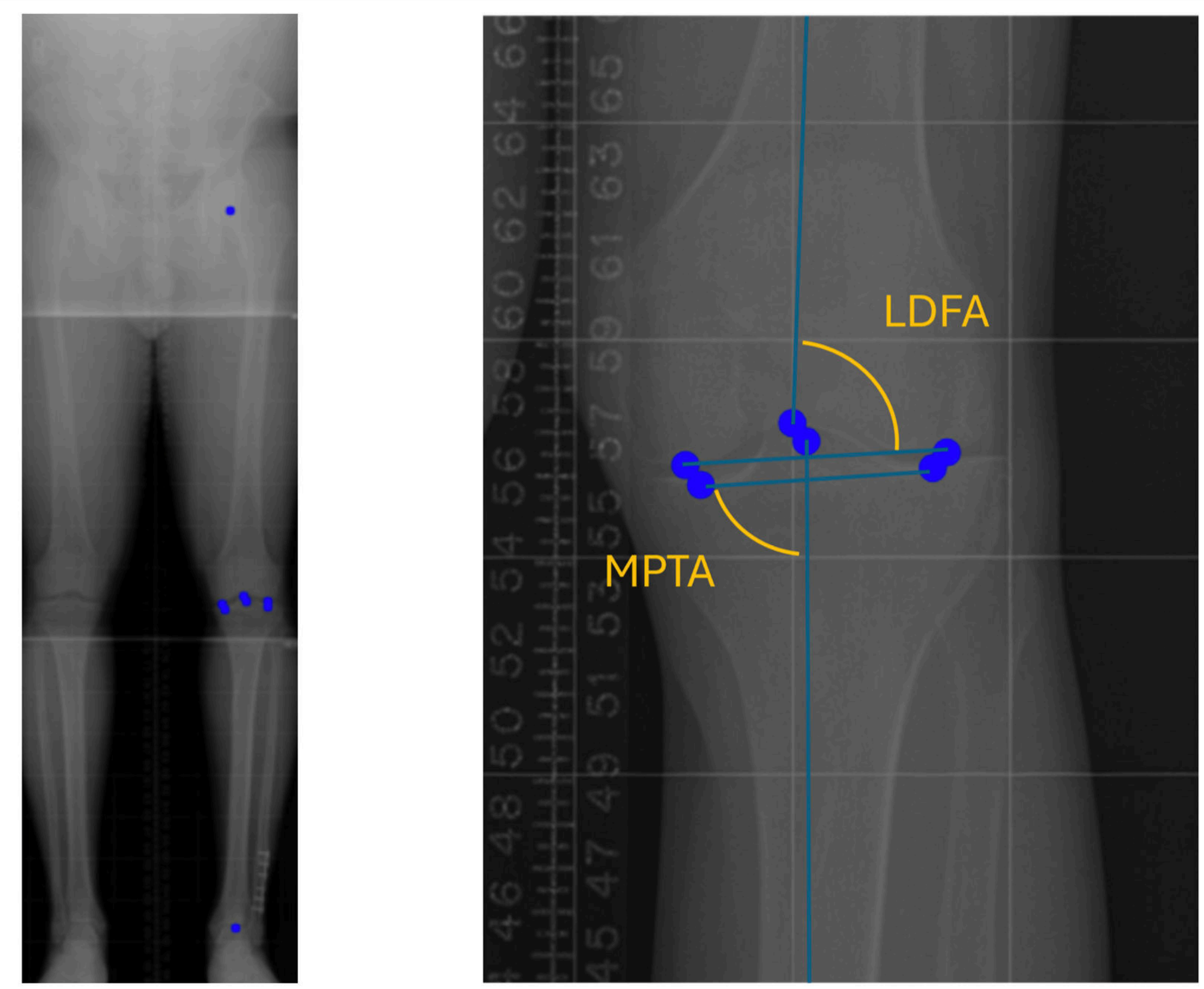

Fig. 2
Labelling of eight anatomical points using RectLabel software[20] and measurement of medial proximal tibial angle (MPTA) and lateral distal femoral angle (LDFA).

Exclusion criteria

Major exclusion criteria in the original study were inflammatory arthritis, contraindication to 3 T MRI, and bilateral end-stage osteoarthritis. The following additional exclusion criteria were applied for this study: poor image quality (low resolution, inaccurate stitching, incomplete review), malrotation (eccentric patella),[17] and evidence of prior surgery (metalwork) or trauma (deformity).

Images meeting exclusion criteria were highlighted by the primary readers and independently reviewed by the senior author (GGJ).

Ethics statement

All patient data in this study were given by informed consent to the OAI.[15] The original OAI study was overseen by an independent Observational Study Monitoring Board (OSMB) appointed by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) at the National Institutes of Health (NIH).[15] No specific ethical approval was required for this study.

Alignment measurement and classification

Two senior orthopaedic trainees (WWM, AS) were the primary readers, having been trained by the senior author, a fellowship-trained consultant knee surgeon. OS1 performed measurements from the left legs of all eligible patients and OS2 independently performed measurements for the first 75 patients which was sufficient for convergence of mean absolute errors (MAE) for constituent angles (Supplementary Material). The pre-hoc sample size calculation to assess reliability between OS1/OS2 was based on a method for hypothesis testing and stated that 55 radiographs would be sufficient.[18] The parameters were two raters, 0.8 expected intraclass correlation (conservative estimate), 0.05 significance level (α), 80% power ($1-\beta$), and 10% dropout rate.

The eight landmarks required for calculating CPAK were labelled using RectLabel, an open-source image annotation software widely used in the field of deep learning.[19,20] Landmarks included: 1 – hip centre; 2 – apex of femoral intercondylar notch; 3 – midpoint of the tibial spines; 4 – midpoint of the tibial plafond; 5 – medial femoral joint surface; 6 – lateral femoral joint surface; 7 – medial tibial joint surface;

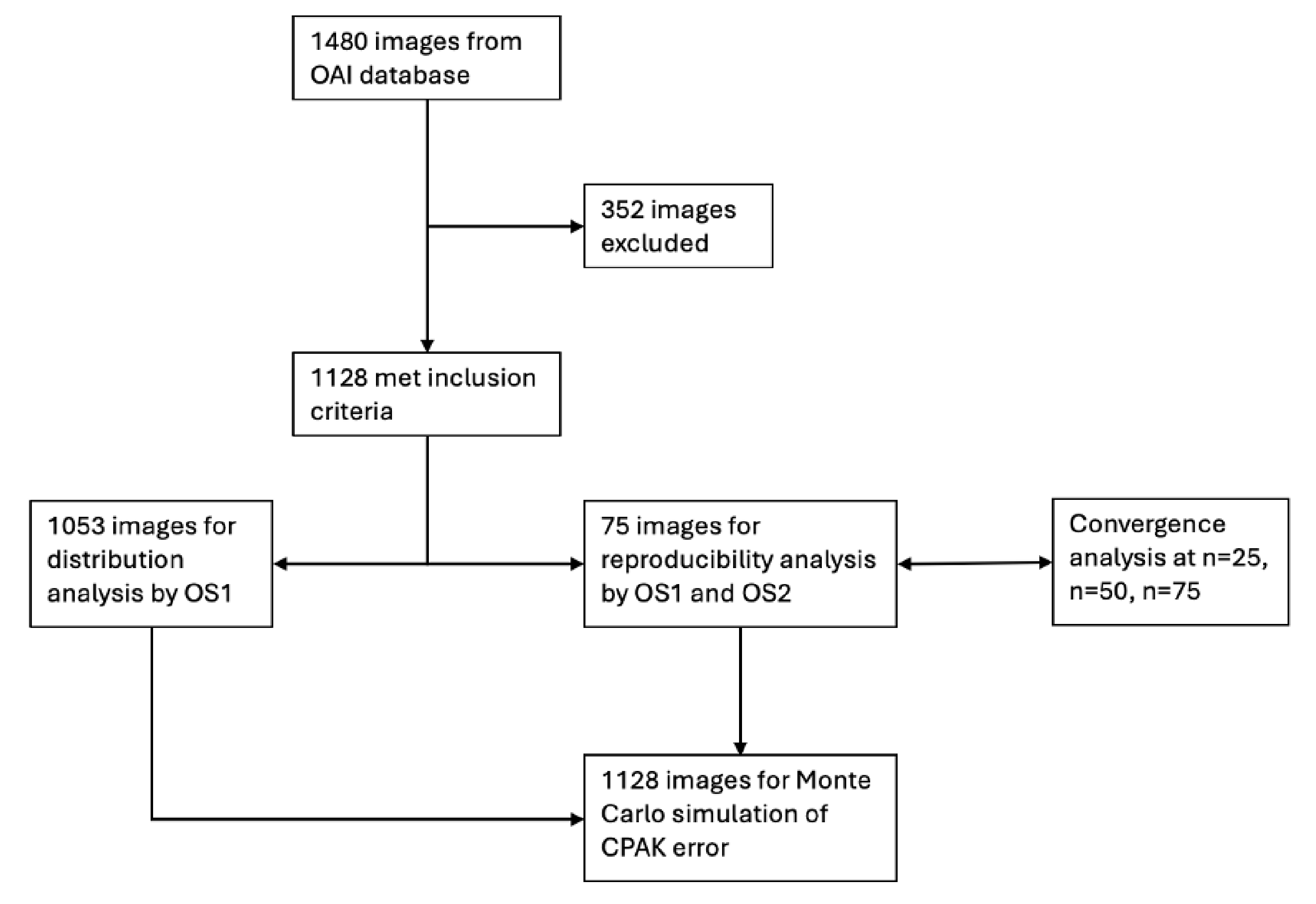

Fig. 3
Flowchart to demonstrate study methodology. CPAK, Coronal Plane Alignment of the Knee; OAI, Osteoarthritis Initiative.

Table I. Inter-reader reliability for mechanical hip-knee-ankle angle between two orthopaedic registrars (OS1/OS2) and the original Osteoarthritis Initiative (OAI) dataset.

Statistic	OS1 vs OS2	OS1 vs OAI	OS2 vs OAI
Mean absolute error	0.23°	0.45°	0.56°
Intraclass correlation coefficient	1.00	0.99	0.99
Mean difference	–0.09°	0.36°	0.49°
95% Limits of agreement	(–0.37° to 0.18°)	(–0.08° to 0.80°)	(0.02° to 0.96°)

8 – lateral tibial joint surface (Figure 2).[7,17] Landmarks were identified according to methods described by Paley.[17] The femoral joint surfaces were the most distal points on the convexity of the medial and lateral distal femur. The tibial joint surfaces were the central points of the subchondral line on the medial and lateral tibial plateau.

The following lines and angles were extracted from the anatomical landmarks (Figure 2): femoral mechanical axis (FMA), tibial mechanical axis (TMA), joint line of the distal femur (JLDF), joint line of the proximal tibia (JLPT), mechanical hip-knee-ankle angle (mHKA), LDFA, MPTA, aHKA defined as MPTA – LDFA; and JLO defined as MPTA + LDFA. All landmarks and angles are available in the Supplementary Material.

As described by MacDessi et al,[7,8] knees with less than –2° aHKA were labelled 'varus', more than +2° 'valgus', and the remaining knees 'neutral'. Knees with less than 177° JLO were labelled 'apex distal', more than 183° 'apex proximal', and the remainder 'neutral'. Classification into the nine CPAK phenotypes followed the original methodology: I – varus, apex distal; II – neutral, apex distal; III – valgus, apex distal; IV – varus, neutral; V – neutral, neutral; VI – valgus, neutral; VII – varus, apex proximal; VIII – neutral, apex proximal; IX – valgus, apex proximal.

Statistical analysis

Inter-reader reproducibility for the 75-patient cohort with measurements from OS1 and OS2 for mHKA, MPTA, LDFA, aHKA, and JLO was assessed using MAE, intraclass correlation coefficient (ICC), mean difference (MD), and 95% limits of agreement (95%LoA). The same statistics were calculated for mHKA versus the original OAI measurements.

CPAK phenotype distribution among the full cohort was compared to healthy and arthritic populations from the original CPAK study using Jensen-Shannon divergence (JSD). This method compares statistical similarity for two datasets, where 0 means the datasets are identical and 1 means they are completely different.[8]

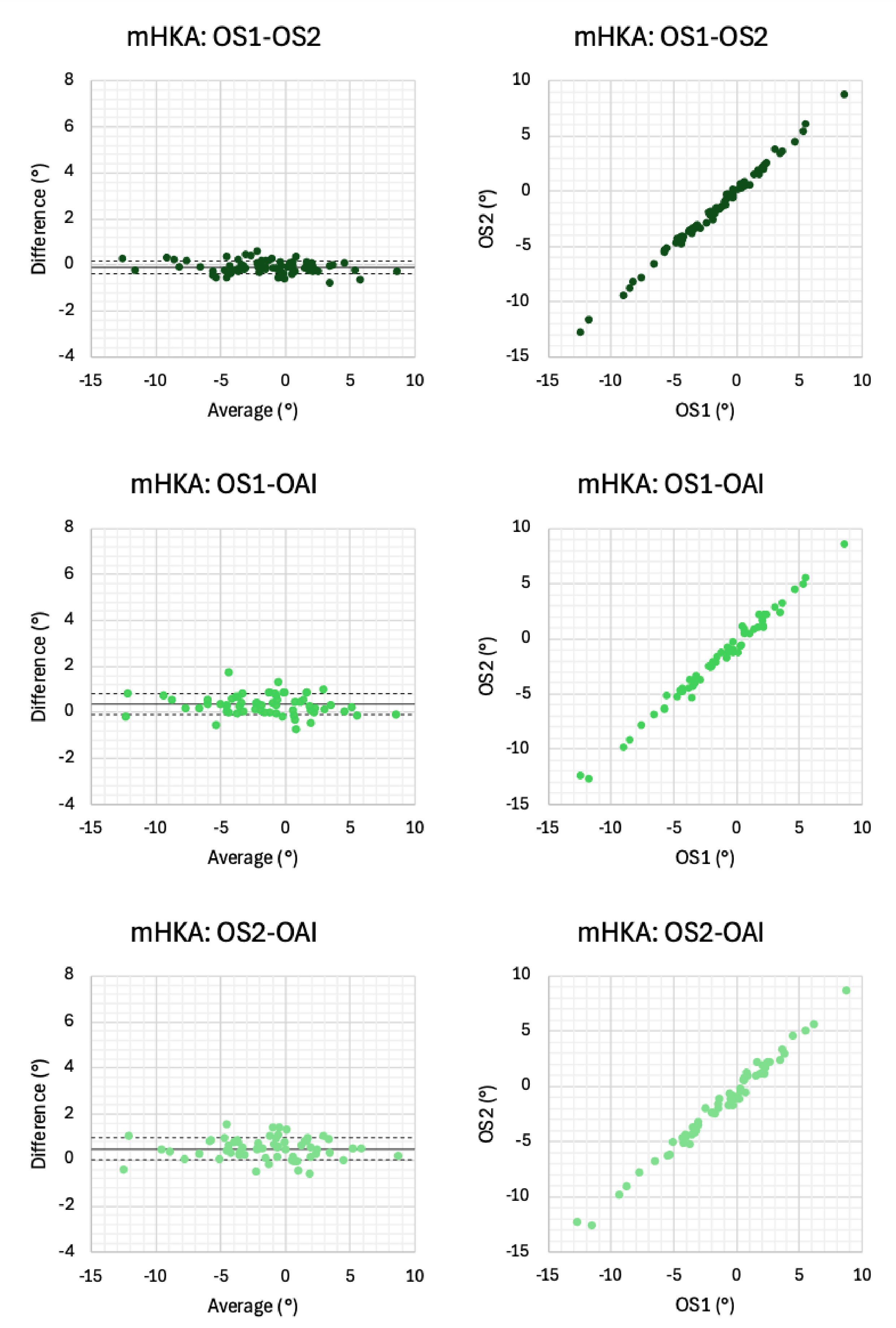

Fig. 4
Inter-reader reliability for mechanical hip-knee-ankle (mHKA) angle between two orthopaedic registrars (top row), the first registrar and the original Osteoarthritis Initiative (OAI) dataset (middle), and the second registrar and the OAI dataset (bottom), shown as Bland-Altman (left column) and scatter plots (right).

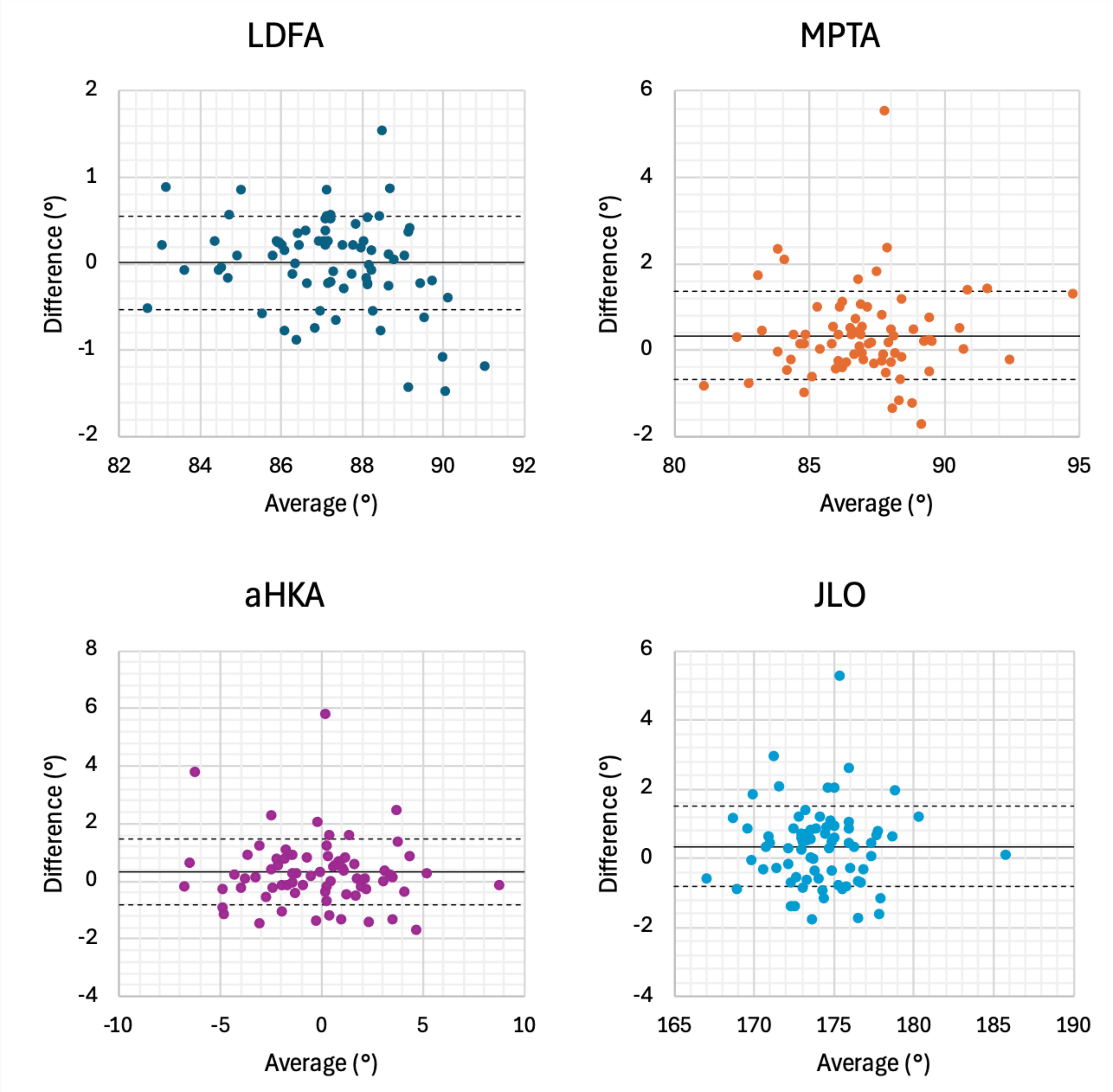

Fig. 5

Inter-reader reliability between two orthopaedic registrars for angles used in Coronal Plane Alignment of the Knee phenotyping, shown as Bland-Altman plots. aHKA, arithmetic hip-knee-ankle alignment; JLO, joint line obliquity; LDFA, lateral distal femoral angle; MPTA, medial proximal tibial angle.

Monte Carlo simulations were used to analyze how agreement rates were affected by the reproducibility of the angular measurements. Monte Carlo simulation is a computational technique that uses random sampling to model the probability of different outcomes based on known uncertainties. In this case, CPAK measurement uncertainty between OS1 and OS2 was learned from the first 75 patients. A total of 1,000 synthetic second-reader datasets were then randomly generated from the 1,128 measurements from OS1 based on the learned error matrix, which was scaled from 5% to 200%. The reproducibility analysis was conducted using built-in functions in Microsoft Excel Version 2506 (Microsoft, USA). The Monte Carlo simulations used to analyse the relationship between angular error and classification were computed using Matlab Version 9.7 (Matlab, USA).

Results

A total of 1,480 LLRs were extracted from the baseline OAI visit with corresponding measures of mHKA. After applying the exclusion criteria, 1,128 radiographs were included for analysis (Figure 3).

Reproducibility

OS1 and OS2 were in near-perfect agreement for mHKA on the 75-patient cohort (Figure 4; Table I), both with each other (MAE: 0.23°; ICC: 1.00) and the original OAI measurements

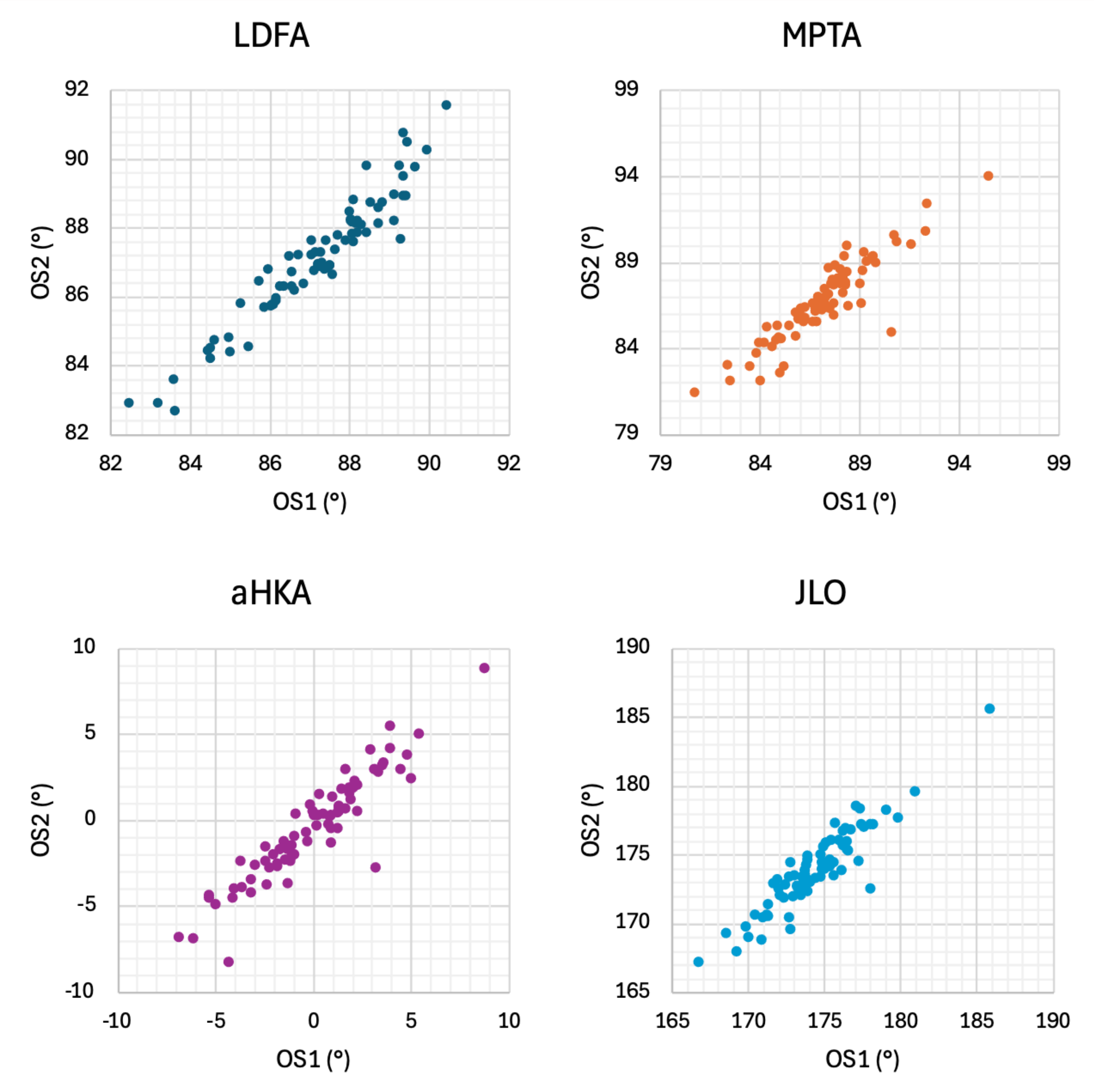

Fig. 6

Correlation between measurements from two orthopaedic registrars for angles used in Coronal Plane Alignment of the Knee phenotyping. aHKA, arithmetic hip-knee-ankle alignment; JLO, joint line obliquity; LDFA, lateral distal femoral angle; MPTA, medial proximal tibial angle.

(MAE: 0.45° and 0.56°; ICC: 0.99 and 0.99, respectively), albeit with a small systematic offset to the latter.

The constituent angles for CPAK (LDFA and MPTA) were less reproducible (Figure 5) but still strongly correlated (Table II). LDFA (MAE: 0.41°; ICC: 0.96) was almost twice as reproducible as MPTA (MAE: 0.71°; ICC: 0.91), corresponding to variation in labelling of the distal femoral and proximal tibial joint line. Mean absolute errors were also higher in aHKA (MAE: 0.79°; ICC: 0.92) and JLO (MAE: 0.91°; ICC: 0.92), the two measurements used to calculate CPAK phenotype (Figure 6).

Discrepancies in aHKA and JLO led to a one in five (20.0%; 95% CI 10.9% to 29.1%) disagreement in CPAK phenotype (Figure 7). CPAK phenotype distribution was highly comparable to both the healthy (JSD: 0.03) and arthritic populations (JSD: 0.04) in the original CPAK study (Table III).[7]

Relationship between angular error and classification

Monte Carlo simulations on the 1,128-patient cohort (Figure 8) indicated that an aHKA mean absolute error of < 0.3° would be required to achieve a CPAK disagreement of less than one in ten (10%). Measurement error of < 0.1° would be required for a CPAK disagreement less than one in 20 (5%).

Discussion

This is the first study to examine CPAK reproducibility on LLRs from an external dataset. CPAK inter-reader disagreement between two trained clinicians was nearly one in five (20%).

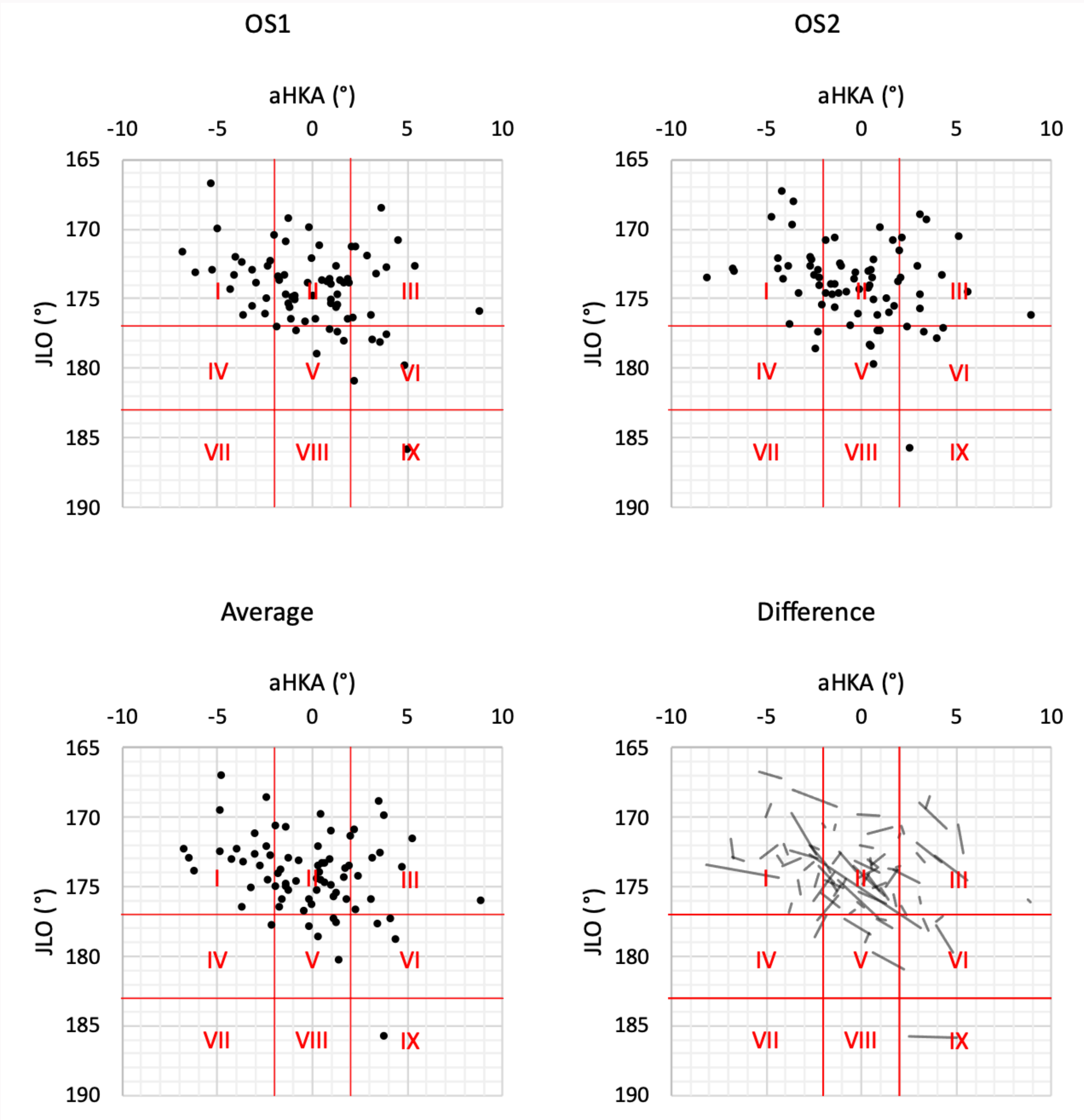

Fig. 7
Arithmetic hip-knee-ankle angle (aHKA) versus joint line obliquity (JLO) from two orthopaedic speciality registrars (OS1/ OS2; top row), with average measurements (bottom left) and measurement differences (bottom right). Red lines indicate Coronal Plane Alignment of the Knee boundaries.

Reducing CPAK disagreement to below one in 20 (5%) would require an eight-fold improvement in aHKA reproducibility to < 0.1°, which may be unachievable using LLRs.

CPAK phenotype distribution was comparable (JSD < 0.04) between the OAI and the original Australian datasets.[7] Similarly, mHKA measurements were in near-perfect agreement between OS1, OS2, and the original OAI measurements (Table I; ICCs > 0.99). Yet, aHKA measurement was four-fold less reproducible than mHKA (MAE: 0.79° vs 0.23°; ICC: 1.00 vs 0.92). This demonstrates the significant reproducibility challenges in a calculated rather than direct method of measuring coronal plane alignment, relying on accurate measurement of MPTA and LDFA. Our ICC for aHKA is similar, albeit slightly less, than that reported by Macdessi et al[8] of 0.97/0.95. When comparing the constituent angle measurements, interobserver reliability was almost two-fold worse for MPTA (MAE: 0.71°; ICC: 0.91) than LDFA (MAE: 0.41°; ICC: 0.96). These MPTA and LDFA values for ICC are near-identical to those reported in a recent multicentre study by Wilhelm et al[14] of 0.97 for LDFA and 0.92 for MPTA.

The observed error in measurement is not surprising, given that avoiding subtle rotational positioning errors when obtaining LLRs is challenging for radiographers. Indeed, previous models have shown that 15° fixed flexion and 10° external rotation together result in 3° LDFA error,[21-26] with similar effects on MPTA.[10,22] The importance of leg positioning

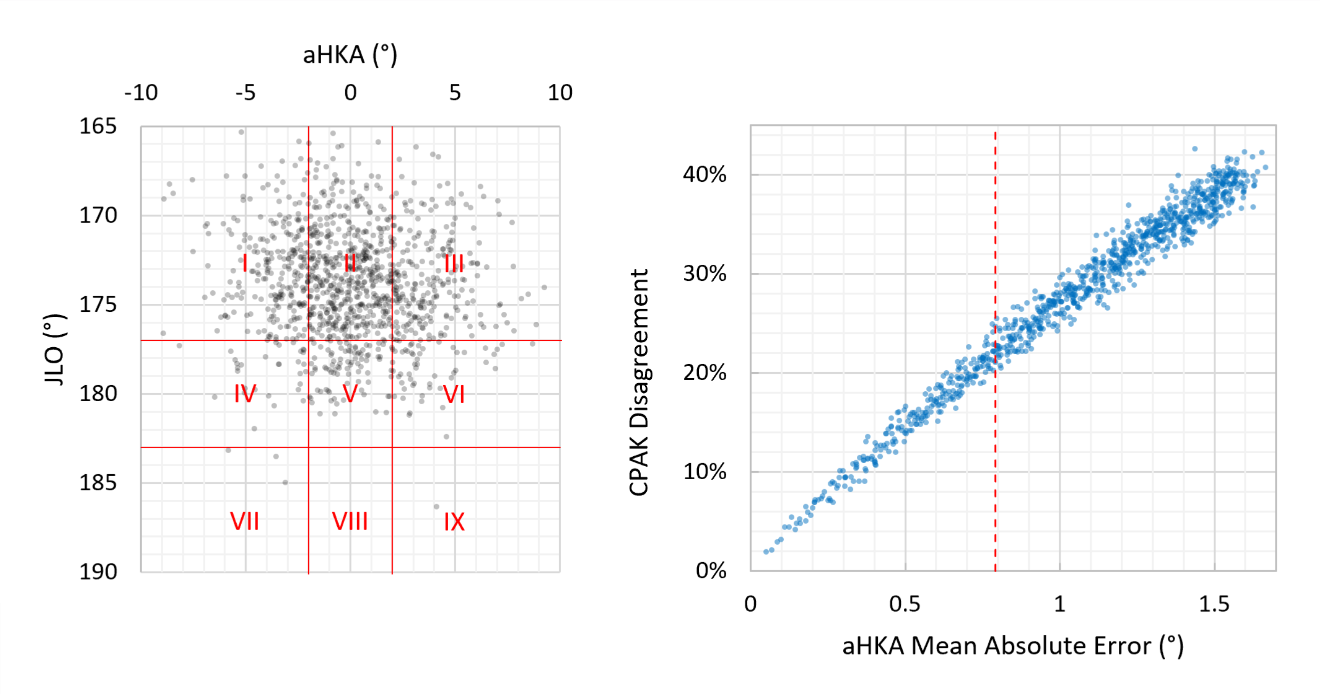

Fig. 8

Arithmetic hip-knee-ankle angle (aHKA) versus joint line obliquity (JLO) for a population of 1,128 patients measured by a single reader (WWM) (left). Coronal Plane Alignment of the Knee (CPAK) disagreement rates for 1,000 Monte Carlo simulations with a synthetic second reader and varying error magnitudes (right). The observed error level is shown as a dashed red line.

Table II. Inter-reader reliability for Coronal Plane Alignment of the Knee constituent angles between two orthopaedic registrars.

Statistic	LDFA	MPTA	aHKA	JLO
Mean absolute error	0.41°	0.71°	0.79°	0.91°
Intraclass correlation coefficient	0.96	0.91	0.92	0.92
Mean difference	0.01°	0.33°	0.32°	0.34°
95% Limits of agreement	(−0.53° to 0.55°)	(−0.69° to 1.36°)	(−0.83° to 1.47°)	(−0.82° to 1.51°)

aHKA, arithmetic hip-knee-ankle alignment; JLO, joint line obliquity; LDFA, lateral distal femoral angle; MPTA, medial proximal tibial angle.

Table III. Distribution of Coronal Plane Alignment of the Knee (CPAK) phenotypes in this study versus healthy and arthritic populations from the original CPAK paper.[7]

CPAK phenotype	This study	Healthy	Arthritic
I – varus, apex distal	20.7%	26.4%	19.4%
II – neutral, apex distal	42.5%	39.2%	32.2%
III – valgus, apex distal	21.2%	9.8%	15.4%
IV – varus, neutral	2.8%	5.4%	9.8%
V – neutral	7.7%	15.4%	14.6%
VI – valgus, neutral	4.8%	3.4%	7.4%
VII – varus, apex proximal	0.3%	0.2%	0.6%
VIII – neutral, apex proximal	0.0%	0.0%	1.6%
VII – valgus, apex proximal	0.1%	0.2%	0.4%

for LLRs in reinforced by Odenbring et al's[27] finding of 2° scan-rescan variation in mHKA measurements across eight patients, and Hall et al's[28] finding of ± 2° for femorotibial angle (FTA) in 2,586 patients with repeated radiographs. CPAK reproducibility error may therefore be significantly higher than acceptable clinical thresholds when considering the potential combined interobserver and scan-rescan errors, although we are not aware of any studies comparing scan-rescan reproducibility of MPTA, LDFA, aHKA, or JLO.

Our findings highlight the challenges in accurately labelling joint lines when measuring MPTA and LDFA. This makes conceptual sense, as is it difficult to identify tangential maxima and minima on convex and concave surfaces. In contrast, the centre of the femoral head was more reliably labelled using Mose circles and the midpoints of the ankle, intercondylar notch, and tibial spines with bisectors.[17] Additionally, the proximity of joint-line landmarks, particularly with loss of joint space in osteoarthritis, means angular measurements are more sensitive to disagreement. Osteoarthritis also makes identification of the 'tibial plateau subchondral line' more challenging due to subchondral sclerosis and potential fixed flexion.[17,29]

Cross-sectional CT is a potential solution to the issues presented in this study, given that rotational and sagittal plane errors can be minimized using 3D frames of reference. Additionally, because aHKA calculations are based on within-bone angular measurements, the lack of weightbearing during CT scanning is not an issue. Furthermore, in institutions using restricted KA strategies or robotics, hip-knee-ankle CT may already be routine during preoperative planning and not represent a deviation from standard practice.[13] CT and LLR have been compared in terms of CPAK phenotyping by Tarassoli et al,[30] who demonstrated LLRs to significantly underestimate proximal tibial varus and JLO compared to CT. However, this study did not compare the interobserver reliability of these methods. Gieroba et al[13] also investigated the reliability of CT for CPAK phenotyping reporting almost two-fold improvements in interobserver coefficient of reliability (COR) of CT scans compared with LLRs for measurement of aHKA, 3.56° versus 2.0°. However, it is important to note that this study used indirect measurements from CT by varying TKA implant positioning on Mako software (Stryker, USA) in 0.5° increments, and reported COR (defined as two standard deviations of the mean) rather than mean absolute error. Of note, a similar two-fold improvement in mean absolute error from 0.79° to 0.40° would be expected to reduce CPAK disagreement to below 10% based on our Monte Carlo simulation.

Compared to CT scans, LLRs are more widely available, cost less, and expose patients to lower radiation doses.[31] Hence, while the authors accept that LLRs will likely remain an accepted standard for assessing CPAK, it is important the implications of a potential 20% misclassification rate are considered when planning and interpreting studies. This particularly applies to studies seeking to correlate CPAK and implant alignment strategies with patient-reported outcome measures (PROMs), where misclassification may mask small differences between groups.[32-34] The findings of this study may also be relevant to alternative coronal knee phenotype classifications that incorporate measures of JLO, such as the functional knee phenotypes proposed by Hirschmann et al.[35] However, mHKA, a measure of coronal alignment typically used in surgical planning, was demonstrated to be four-fold more reliable than arithmetic HKA.

Limitations include the lack of comparative data from CT scans, which is an avenue for further study. Inter-reader reliability could have been performed with more readers and utilizing more comparative images, but the a priori power analysis and post-hoc convergence analysis (Supplementary Material) both determined that the sample sizes were sufficient.

In conclusion, despite excellent inter-reader reliability in MPTA and LDFA, there was a 20% inter-reader disagreement in CPAK classification. Monte Carlo simulation indicated that a mean absolute error for aHKA would need to be < 0.1° to achieve a CPAK disagreement of less than 5%. This degree of accuracy is unlikely to be achievable using LLRs, and therefore caution should be exercised when using LLRs for CPAK phenotyping, surgical planning, or interpreting clinical studies.

Social media

Follow W. Wynell-Mayow on LinkedIn.

Supplementary material

Convergence analysis included which demonstrates 75 patients to be sufficient to achieve convergence of mean absolute errors for constituent angles.

References

1. **Maderbacher G**, **Keshmiri A**, **Krieg B**, **Greimel F**, **Grifka J**, **Baier C**. Kinematic component alignment in total knee arthroplasty leads to better restoration of natural tibiofemoral kinematics compared to mechanic alignment. *Knee Surg Sports Traumatol Arthrosc*. 2019;27(5):1427–1433.
2. **Blakeney W**, **Clément J**, **Desmeules F**, **Hagemeister N**, **Rivière C**, **Vendittoli P-A**. Kinematic alignment in total knee arthroplasty better reproduces normal gait than mechanical alignment. *Knee Surg Sports Traumatol Arthrosc*. 2019;27(5):1410–1417.
3. **Hirschmann MT**, **Hess S**, **Behrend H**, **Amsler F**, **Leclercq V**, **Moser LB**. Phenotyping of hip–knee–ankle angle in young non-osteoarthritic knees provides better understanding of native alignment variability. *Knee Surg Sports Traumatol Arthrosc*. 2019;27(5):1378–1384.
4. **Lin YH**, **Chang FS**, **Chen KH**, **Huang KC**, **Su KC**. Mismatch between femur and tibia coronal alignment in the knee joint: classification of five lower limb types according to femoral and tibial mechanical alignment. *BMC Musculoskelet Disord*. 2018;19(1):411.
5. **Bellemans J**, **Colyn W**, **Vandenneucker H**, **Victor J**. The Chitranjan Ranawat award: is neutral mechanical alignment normal for all patients? The concept of constitutional varus. *Clin Orthop Relat Res*. 2012;470(1):45–53.
6. **Cooke D**, **Scudamore A**, **Li J**, **Wyss U**, **Bryant T**, **Costigan P**. Axial lower-limb alignment: comparison of knee geometry in normal volunteers and osteoarthritis patients. *Osteoarthritis Cartilage*. 1997;5(1):39–47.
7. **MacDessi SJ**, **Griffiths-Jones W**, **Harris IA**, **Bellemans J**, **Chen DB**. Coronal Plane Alignment of the Knee (CPAK) classification a new system for describing knee phenotypes. *Bone Joint J*. 2021;103-B(2).
8. **MacDessi SJ**, **Griffiths-Jones W**, **Harris IA**, **Bellemans J**, **Chen DB**. The arithmetic HKA (aHKA) predicts the constitutional alignment of the arthritic knee compared to the normal contralateral knee: a matched-pairs radiographic study. *Bone Jt Open*. 2020;1(7):339–345.
9. **Toyooka S**, **Osaki Y**, **Masuda H**, **et al**. Distribution of coronal plane alignment of the knee classification in patients with knee osteoarthritis in Japan. *J Knee Surg*. 2023;36(7):738–743.
10. **Brunner J**, **Jörgens M**, **Weigert M**, **Kümpel H**, **Degen N**, **Fuermetz J**. Significant changes in lower limb alignment due to flexion and rotation-a systematic 3D simulation of radiographic measurements. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(4):1483–1490.
11. **Maderbacher G**, **Baier C**, **Benditz A**, **et al**. Presence of rotational errors in long leg radiographs after total knee arthroplasty and impact on measured lower limb and component alignment. *Int Orthop*. 2017;41(8):1553–1560.
12. **Nguyen HC**, **van Egmond N**, **Hevesi M**, **Weinans H**, **Gielis WP**, **Custers RJH**. A new protocol for obtaining whole leg radiographs with excellent reproducibility. *J Cartil Jt Preserv*. 2022;2(1):100042.
13. **Gieroba TJ**, **Marasco S**, **Babazadeh S**, **Di Bella C**, **van Bavel D**. Arithmetic hip knee angle measurement on long leg radiograph versus computed tomography-inter-observer and intra-observer reliability. *Arthroplasty*. 2023;5(1):35.
14. **Wilhelm NJ**, **von Schacky CE**, **Lindner FJ**, **et al**. Multicentric development and validation of a multi-scale and multi-task deep learning model for comprehensive lower extremity alignment analysis. *Artif Intell Med*. 2024;150:102843.
15. **Nevitt MC**, **Felson DT**, **Lester G**. The Osteoarthritis Initiative: protocol for the cohort study (Vol. 1). 2006. https://nda.nih.gov/static/docs/StudyDesignProtocolAndAppendices.pdf (date last accessed 15 October 2025).
16. **Osteoarthritis Initiative (OAI)**. Central Image Assessments. NIMH Data Archive. 2018. https://nda.nih.gov/oai/image-assessments (date last accessed 8 October 2025).
17. **Paley D**. *Principles of Deformity Correction*. Springer, 2002.
18. **Walter SD**, **Eliasziw M**, **Donner A**. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17(1):101–110.
19. **Guy S**, **Jacquet C**, **Tsenkoff D**, **Argenson J-N**, **Ollivier M**. Deep learning for the radiographic diagnosis of proximal femur fractures: limitations and programming issues. *Orthop Traumatol Surg Res*. 2021;107(2):102837.
20. **No authors listed**. RectLabel. https://rectlabel.com/ (date last accessed 8 October 2025).

21. **Ahrend M-D**, **Baumgartner H**, **Ihle C**, **Histing T**, **Schröter S**, **Finger F**. Influence of axial limb rotation on radiographic lower limb alignment: a systematic review. *Arch Orthop Trauma Surg*. 2022;142(11):3349–3366.
22. **Nguyen HC**, **Gielis WP**, **van Egmond N**, **et al**. The need for a standardized whole leg radiograph guideline: the effects of knee flexion, leg rotation, and X-ray beam height. *J Cartil Jt Preserv*. 2021;1(3):100022.
23. **Radtke K**, **Becher C**, **Noll Y**, **Ostermeier S**. Effect of limb rotation on radiographic alignment in total knee arthroplasties. *Arch Orthop Trauma Surg*. 2010;130(4):451–457.
24. **Marques Luís N**, **Varatojo R**. Radiological assessment of lower limb alignment. *EFORT Open Rev*. 2021;6(6):487–494.
25. **Jakma TSC**, **Verhaar JAN**, **Brouwer KH**, **Brouwer RW**. Pitfalls in determining knee alignment. *J Knee Surg*. 2007;20(3):210–215.
26. **Maderbacher G**, **Matussek J**, **Greimel F**, **et al**. Lower limb malrotation is regularly present in long-leg radiographs resulting in significant measurement errors. *J Knee Surg*. 2021;34(1):108–114.
27. **Odenbring S**, **Berggren AM**, **Peil L**. Roentgenographic assessment of the hip-knee-ankle axis in medial gonarthrosis. A study of reproducibility. *Clin Orthop Relat Res*. 1993;289:195–196.
28. **Hall TAG**, **Jones GG**, **van Arkel RJ**. Femorotibial angle scan-rescan reproducibility: a high-precision calculation on a large cohort. *Knee Surg Sports Traumatol Arthrosc*. 2024;32(12):3133–3140.
29. **Braun HJ**, **Gold GE**. Diagnosis of osteoarthritis: imaging. *Bone*. 2012; 51(2):278–288.
30. **Tarassoli P**, **Corban LE**, **Wood JA**, **Sergis A**, **Chen DB**, **MacDessi SJ**. Long leg radiographs underestimate the degree of constitutional varus limb alignment and joint line obliquity in comparison with computed tomography: a radiographic study. *Knee Surg Sports Traumatol Arthrosc*. 2023;31(11):4755–4765.
31. **Henckel J**, **Richards R**, **Lozhkin K**, **et al**. Very low-dose computed tomography for planning and outcome measurement in knee replacement. The imperial knee protocol. *J Bone Joint Surg Br*. 2006;88-B(11):1513–1518.
32. **Konishi T**, **Hamai S**, **Tsushima H**, **et al**. Pre- and postoperative coronal plane alignment of the knee classification and its impact on clinical outcomes in total knee arthroplasty. *Bone Joint J*. 2024;106-B(10):1059–1066.
33. **Corban LE**, **van de Graaf VA**, **Chen DB**, **Wood JA**, **Diwan AD**, **MacDessi SJ**. How often do we alter constitutional limb alignment, joint line obliquity, and Coronal Plane Alignment of the Knee (CPAK) phenotype when performing mechanically aligned TKA? *Bone Jt Open*. 2024;5(2):109–116.
34. **van de Graaf VA**, **Clark GW**, **Collopy D**, **Wood JA**, **Chen DB**, **MacDessi SJ**. Functional alignment minimizes changes to joint line obliquity in robotic-assisted total knee arthroplasty: a CT analysis of functional versus kinematic alignment in 2,116 knees using the Coronal Plane Alignment of the Knee (CPAK) classification. *Bone Jt Open*. 2024;5(12): 1081–1091.
35. **Hirschmann MT**, **Moser LB**, **Amsler F**, **Behrend H**, **Leclerq V**, **Hess S**. Functional knee phenotypes: a novel classification for phenotyping the coronal lower limb alignment based on the native alignment in young non-osteoarthritic patients. *Knee Surg Sports Traumatol Arthrosc*. 2019; 27(5):1394–1402.

Author information

W. Wynell-Mayow, MB BChir, MA (Cantab) MRCS, PGCert (Darzi), Trauma and Orthopaedic Surgeon,
Imperial College Healthcare NHS Trust, St Mary's Hospital, London, UK.

T. A. G. Hall, MEng, PhD, EPSRC Doctoral Prize Fellow (Imperial)
R. J. van Arkel, MEng, PhD, Senior Lecturer
Biomechanics Group, Department of Mechanical Engineering, Imperial College London, London, UK.

A. Singh, BMBS, BSc (hons), MRCS, MSc, NIHR Doctoral Research Fellow, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK.

G. G. Jones, MBBS, BSc (Hons), PhD, FRCS (Tr & Orth), Senior Clinical Lecturer in Orthopaedic Surgery & Trauma, Imperial College Healthcare NHS Trust, St Mary's Hospital, London, UK; MSk Lab, Department of Surgery and Cancer, Imperial College London, London, UK.

Author contributions

W. Wynell-Mayow: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing.
T. A. G. Hall: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Writing – review & editing.
A. Singh: Conceptualization, Data curation, Methodology, Writing – review & editing.
R. J. van Arkel: Writing – review & editing, Writing – original draft, Resources, Software, Supervision.
G. G. Jones: Conceptualization, Methodology, Writing – review & editing, Resources, Supervision.

Funding statement

The authors received no financial or material support for the research, authorship, and/or publication of this article.

ICMJE COI statement

G. G. Jones reports payments for educational work from Smith & Nephew, unrelated to this study. R. van Arkel reports institutional grants/materials for orthopaedic research from J&J MedTech, Zimmer Biomet, and London Devices, consulting fees from InnovateOrtho, and stock or stock options in Embed Biotech, all of which are unrelated to this study.

Data sharing

The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Acknowledgements

This article was prepared using an Osteoarthritis Initiative (OAI) public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners. The authors wish to thank the participants, PIs, co-Is, and hospital staff who have contributed data to the OAI. The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation; GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health.

Ethical review statement

No specific ethical approval was required for this study.

Open access funding

The open access fee for this article was funded by Imperial College London.