



Inter- and intra-rater reliability of knee flexion angle measurements on X-ray and MRI

Harry Kyle Campbell Summers^{1^}, Stephen Picken¹, Oday Al-Dadah^{1,2}

¹Department of Trauma and Orthopaedic Surgery, South Tyneside District Hospital, South Shields, UK; ²Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Framlington Place, Newcastle-upon-Tyne, UK

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: O Al-Dadah; (IV) Collection and assembly of data: HKC Summers, S Picken; (V) Data analysis and interpretation: O Al-Dadah; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Harry Kyle Campbell Summers, MBBS, Foundation Doctor. Department of Trauma and Orthopaedic Surgery, South Tyneside District Hospital, Harton Lane, South Shields, UK. Email: harry.summers3@nhs.net.

Background: Range of motion (ROM) is an important aspect of orthopaedic patient assessment. It can be measured at the knee joint by determining the knee flexion angle (KFA) a patient can achieve at extremes of flexion and extension. As with any measurement, the accuracy and reliability of the method used determine its validity. The consistency of magnetic resonance imaging (MRI) scans as compared to the current gold standard of X-ray remains unknown in terms of KFA evaluation. The aim of this study was to assess and compare the reliability of measuring KFA between X-ray and MRI scans.

Methods: This study included 80 patients (94 knees) who had attended a specialist knee clinic due to varying knee pathologies and undergone both X-ray and MRI scans. Lateral and T1-weighted sagittal imaging views (respectively) were used to measure KFA by two trained observers independently at two separate time points, 8 weeks apart. The data was then statistically analysed and intra- and inter-observer reliability calculated using the intraclass correlation coefficient (ICC).

Results: The intra-observer reliability for X-ray was 0.96 ($P < 0.001$) and that for MRI was 0.83 ($P < 0.001$). The inter-observer reliability for X-ray was 0.99 ($P < 0.001$) and that for MRI was 0.81 ($P < 0.001$). All the intra-class correlation coefficients were graded as excellent in both the intra- and inter-observer reliability analysis. Overall, the mean KFA was notably higher on X-ray measurements than that on MRI scans. There was a statistically significant difference between Time 1 and Time 2 measurements (17.7° vs. 16.8°) for MRI data ($P = 0.022$). No significant difference was found for X-ray measurements (46.4° vs. 45.6°) in this regard ($P = 0.182$).

Conclusions: Both X-ray and MRI allow KFA to be measured with an excellent degree of reliability. However, X-ray measurements were overall superior to that of MRI mainly due to the larger field of view of the visible on-screen image which more readily identifies the anatomical landmarks required to measure KFA.

Keywords: Knee; X-ray; magnetic resonance imaging (MRI); reliability; range of motion

Received: 17 January 2022; Accepted: 07 June 2022; Published: 15 October 2022.

doi: [10.21037/aoj-22-2](https://doi.org/10.21037/aoj-22-2)

View this article at: <https://dx.doi.org/10.21037/aoj-22-2>

[^] ORCID: [0000-0002-4375-0181](https://orcid.org/0000-0002-4375-0181).

Introduction

Orthopaedic surgery often requires a comprehensive range of motion (ROM) assessment both preceding and following an operation to determine if treatment has facilitated improvement. Pre-operative ROM can be an indicator as to the results that can be achieved following surgery and post-operative ROM can be an indicator of function and of patient satisfaction (1). Therefore, it is important to measure these values reliably and accurately.

Reliability and accuracy are indicators of the validity of a measurement. An unreliable method of measurement is therefore unlikely to be valid (2). While accuracy of ROM measurements is widely discussed in orthopaedic literature (3,4), very little research has investigated the reliability of ROM measurements (5). A reliable measurement is one that can be repeated and a similar or identical result obtained. Reliable measurements of ROM allow different clinicians to arrive at the same conclusion, which is particularly important in the case of multiple referrals or poor documentation. Additionally, as patients with chronic musculoskeletal conditions are often followed-up over a long period of time, clinicians must be satisfied that any changes in ROM are due to improvement in the patient's condition and not unreliable measurements. Intra-rater reliability evaluates the results measured more than once by the same observer (i.e., re-test reliability) whilst inter-rater reliability assesses the consistency of results measured by a number of different observers.

One way of assessing ROM in the lower limb is by measuring the knee flexion angle (KFA) a patient can achieve at their extremes of flexion and extension. This can be physically measured using a goniometer or assessed with appropriate medical imaging technology, the most commonly used being X-ray or magnetic resonance imaging (MRI). Other more experimental imaging techniques which may be appropriate such as electrical linkage, sound-based or robotic measurements (6) are not readily available to clinicians. KFA measurements derived from imaging taken in greater extension have previously been proven to be more reliable than those taken from images with a greater degree of flexion (4).

The aim of this study was to assess and compare the intra- and inter-rater reliability of measuring KFA between X-ray and MRI. We present the following article in accordance with the MDAR reporting checklist (available at <https://aoj.amegroups.com/article/view/10.21037/aoj-22-2/rc>).

Table 1 Demographics of subjects

Demographic	Value
Mean age (yrs) (SD)	44 (17.3)
Gender (Male:Female)	37:43
Laterality (Right:Left:Bilateral)	33:33:14
Mean height (m) (SD)	1.70 (10.0)
Mean weight (kg) (SD)	81.3 (20.6)
Mean BMI (kg/m ²) (SD)	28.1 (6.0)

N=80 patients (n=94 knees). BMI, body mass index; SD, standard deviation.

Methods

This study was exempt from Institutional Review Board (IRB)/Ethics Committee approval at South Tyneside District Hospital as it was a pragmatic study evaluating the existing routine clinical practice of the senior author (consultant orthopaedic surgeon). This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective observational study was part of a larger radiological study that investigated 80 patients (94 knees) who had undergone both X-ray and MRI imaging for varying clinical indications. Patients included in this study were under the care of a single consultant orthopaedic surgeon with a special interest in knee surgery, at a district general hospital. They underwent arthroscopic knee surgery after attending a specialist knee clinic and receiving clinical assessment and radiological investigations. Informed consent was obtained from all patients. Their demographics are detailed in *Table 1*. The lateral X-ray view and the T1-weighted sagittal MRI view images of each patient were used to measure KFA by two observers independently at two separate time points. The exact same X-ray image and MRI image was evaluated on both the first and second data collection periods to allow for re-test reliability to be calculated. At the time the study was conducted, Observer A (HS) was a fourth-year medical student and Observer B (SP) was a clinical research fellow (junior doctor). Both observers received formal training in calculating KFA from the senior author (consultant orthopaedic surgeon) and were provided with an identical measurement pro forma detailing the steps required to calculate KFA.

To reduce the potential for recall bias, each round of data collection was undertaken 8 weeks apart and both observers were blinded to each other's results and also from their

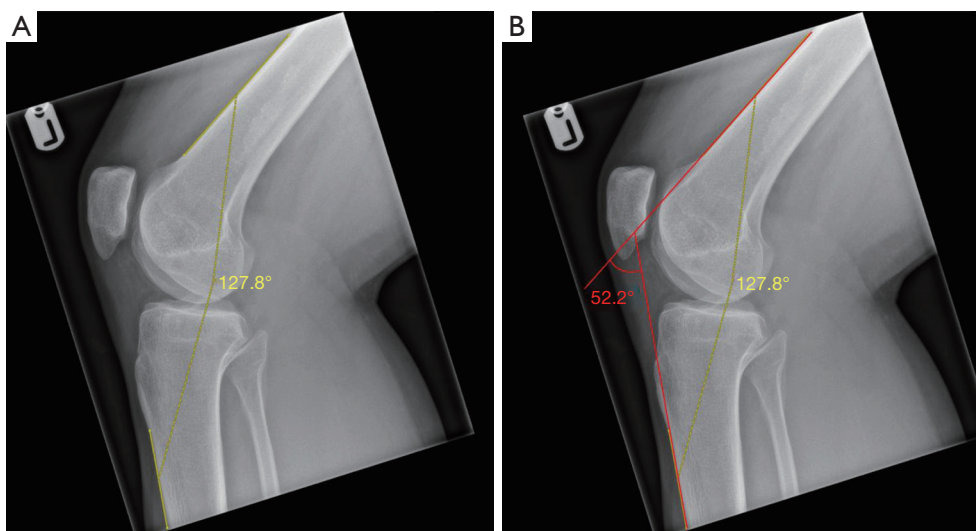


Figure 1 Lateral X-ray images of knee joint. (A) Annotated using PACS angle tool (yellow measurements). (B) Illustrated to show calculation of true KFA (red measurements). KFA, knee flexion angle.

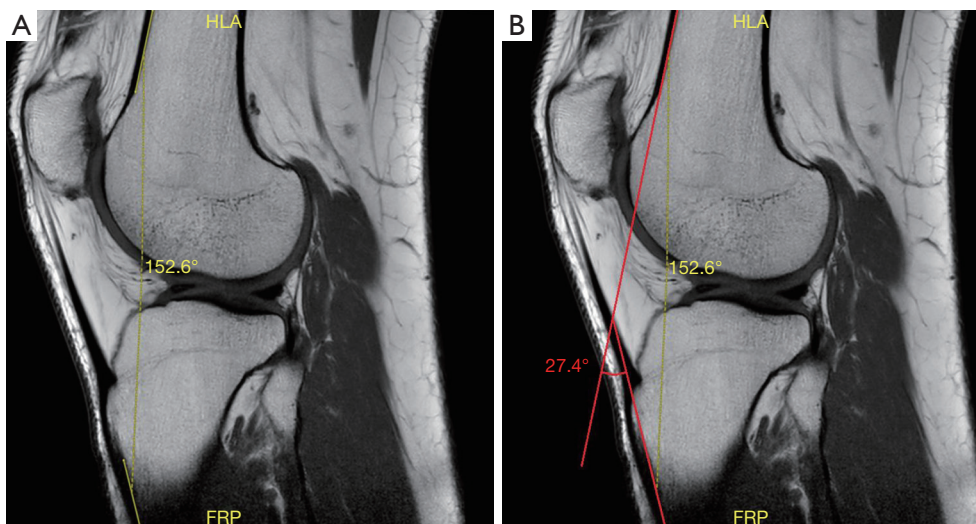


Figure 2 T1-weighted sagittal MRI images of knee joint. (A) Annotated using PACS angle tool (yellow measurements). (B) Illustrated to show calculation of true KFA (red measurements). MRI, magnetic resonance imaging; PACS, Picture Archiving and Communication System; KFA, knee flexion angle.

own first round of data collection. All radiological images were evaluated on the medical imaging platform Picture Archiving and Communication System (PACS) (Centricity version 6, GE Healthcare, Chicago) and the angle tool selected, which allows two lines to be drawn and an angle between them measured. If multiple images of the same patient's knee were available, the image with the least degree of flexion (i.e., the more extended knee) was used, as this

has been proven to produce greater accuracy and reliability when measuring the KFA (4). Lines were drawn along the distal anterior cortex of the femur and the proximal anterior cortex of the tibia, a method validated for determining the true KFA within 2 degrees (4). Because the PACS angle tool measures the angle between the two lines drawn, obtained values were subtracted from 180° to obtain the true KFA value (Figures 1,2).

Table 2 Knee flexion angle (Observer A)

Variable	Time 1 (n=94) mean (range)	Time 2 (n=94) mean (range)	Mean difference	P value ¹	95% CI
X-ray flexion angle (degrees°)	46.4 (2 to 94)	45.6 (2 to 94)	0.8	0.182	-0.4–2.0
MRI flexion angle (degrees°)	17.7 (3 to 35)	16.8 (5 to 31)	0.9	0.022*	0.1–1.7

¹, Paired Student's *t*-test; *, Statistically significant at <0.05 level; MRI, magnetic resonance imaging.

Table 3 Intra-observer reliability (Observer A)

Variable	ICC	Grade ¹	P value	95% CI
X-ray flexion angle	0.96	Excellent	<0.001*	0.94–0.97
MRI flexion angle	0.83	Excellent	<0.001*	0.75–0.88

¹, ICC grading system (Shrout & Fleiss). *, Statistically significant at <0.05 level. MRI, magnetic resonance imaging; ICC, intraclass correlation coefficient.

Table 4 Inter-observer reliability (Observer A vs. Observer B)

Variable	ICC	Grade ¹	P value	95% CI
X-ray flexion angle	0.99	Excellent	<0.001*	0.98–0.99
MRI flexion angle	0.81	Excellent	<0.001*	0.71–0.87

¹, ICC grading system (Shrout & Fleiss). *, Statistically significant at <0.05 level. MRI, magnetic resonance imaging; ICC, intraclass correlation coefficient.

An MRI consists of multiple images arranged as slices. While the true KFA is constant throughout the scan, the landmarks used to measure the angle are not visible on every slice and their inclination can vary across different slices. To obtain consistency from the outset of the study in MRI evaluations, both observers agreed to begin by finding the sagittal slice in which the patella appeared largest (longitudinal dimension) and from there moving to the nearest slice where both the anterior cortex of the proximal tibia and distal femur were visible. Both observers also used the same bony landmarks on the femur and tibia for X-ray evaluations. This method was introduced to reduce error and variability by using a measurement system that was easily reproducible.

Statistical analysis

Plotted histograms with fitted curve lines, box-plots, normal Q-Q plots and the Kolmogorov-Smirnov statistic were used to confirm that a normal distribution was an appropriate assumption for all the continuous variables in the study. The paired Student's *t*-test was used for the within group analysis. The inter- and intra-observer reliability of the KFA (continuous data) were determined using intra-class correlation coefficients (ICC). The ICC analysis was based on a consistency type two-way mixed model. The ICC results were further interpreted and categorised on the basis of the values proposed by Shrout and Fleiss (7) with a score of 0–0.4 indicating *poor* reliability, 0.4–0.75 indicating *moderate* reliability and a score of more than 0.75 indicating *excellent* reliability. The level of statistical significance was set at a two-

sided $P < 0.05$. Statistical analysis was performed using SPSS for Windows version 26.0 (IBM Corp., Armonk, New York).

Results

Table 2 shows the KFA analysis for Observer A. Overall, the mean KFA was notably higher on X-ray measurements than that on MRI as was the range of KFA values. There was a statistically (but not clinically) significant difference between Time 1 and Time 2 measurements for MRI data. No significant difference was found for X-ray measurements in this regard.

The intra-observer reliability for Observer A (*Table 3*) was notably higher for X-ray (0.96) than MRI measurements (0.83). Both groups achieved an excellent reliability grade (>0.75).

The measurements (mean (range)) obtained by Observer B (n=94) for X-ray KFA =47.1° (3° to 87°) and for MRI KFA =17.0° (3° to 37°). The inter-observer reliability between Observer A and Observer B (*Table 4*) was superior for X-ray (0.99) as compared to that of MRI measurements (0.81). Both groups achieved an excellent reliability grade (>0.75).

Discussion

Given the physical space limitations of an MRI coil, flexion is naturally limited, therefore the range of KFA measurements and the mean KFA was expectedly smaller on MRI than X-ray (*Table 2*). While measurements taken in greater degrees of extension have previously been proven to be more reliable (4), in this study a greater degree of

reliability (Tables 3,4) was seen on X-ray measurements, which had a greater degree of flexion (Table 2). This implies that the method of imaging has a greater impact on reliability than the degree of flexion the image is taken in. Measuring the KFA from MRI is therefore notably less reliable than measuring it from X-ray, to the point that the benefit of most of the images being in greater extension is negated.

The difference in reliability can be explained by the nature of the imaging methods. Where X-ray is composed of a single image, MRI is composed of multiple slices. Although theoretically the KFA should be the same throughout all slices of an MRI, measurements require visualisation of certain bony landmarks on the femur and tibia (as described in the materials and methods section). Due to the three-dimensional nature of these structures, the inclination of the measured landmarks may vary on the specific slice used for measurement. Another major advantage of X-ray is its larger field of view of the image as compared to that of MRI which more readily identifies the anatomical landmarks required to measure the KFA. The excellent reliability grades seen in our results (Tables 3,4) are a positive reflection of the reproducibility of the measurement techniques used in this study.

In comparison to other studies, our results for X-ray demonstrate its superior reliability to physical goniometry (6,7) but roughly equivalent reliability to digital goniometry techniques (8). The benefit of physical goniometry is that it is inexpensive and can be conducted quickly and does not require radiological intervention (or exposure to radiation), however, it is rather user dependent. Most contemporary studies agree that X-ray is the gold standard method of KFA measurement (3,9), exhibiting a high degree of reliability, which concurs with our results.

The limitations of this study; although both observers followed a system when measuring the KFA on MRI scans, it is still possible that different slices could have been used for each measurement as the specific slice to be used was not agreed on. This could have caused the angle of inclination of the bony landmarks used to vary and therefore the KFA measured to differ. Although the effect of using different slices was intended to be minimised for this research study, it nonetheless simulates real-time clinical circumstances where clinicians would use different slices and therefore measure different angles. This would make measuring the KFA using MRI less reliable, which is reflected in our results (Tables 3,4). Although Table 2 shows a statistically significant difference for the MRI

KFA comparison between Time 1 and Time 2, it is noted that the mean difference (0.9°) is not clinically significant. Furthermore, the mean difference of the X-ray KFA (0.8°) was only 0.1° less than that of MRI and was found not to be statistically significant. This raises the possibility of a Type I error for the MRI analysis. Another limitation of this study is the relative inexperience of the observers. In real-time clinical practice, it is usual for doctors of a more senior level to interpret radiological imaging in the outpatient clinics. However, both observers received formal training at the start of the study on calculating KFA from X-ray and MRI. Consequently, it is expected that the effect of this particular factor on the final results was minimised.

The clinical relevance of this study is summated by the finding that calculating KFA using X-ray is of greater reliability than using MRI. Although an MRI would never be ordered for the sole purpose of calculating KFA due to cost (among other limitations), a clinician ordering multiple imaging investigations may have chosen to measure the KFA at the limit of extension on MRI as a secondary function, if it had been more reliable to do so.

Conclusions

Both X-ray and MRI allow KFA to be measured with excellent reliability. However, it is notably more reliable to measure KFA using X-ray. The superior reliability of X-ray measurements were due to a larger field of imaging view and more readily identifiable bony landmarks which allowed for more reproducible results.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at <https://aoj.amegroups.com/article/view/10.21037/aoj-22-2/rc>

Data Sharing Statement: Available at <https://aoj.amegroups.com/article/view/10.21037/aoj-22-2/dss>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://aoj.amegroups.com/article/view/10.21037/aoj-22-2/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was exempt from Institutional Review Board (IRB)/Ethics Committee approval at South Tyneside District Hospital as it was a pragmatic study evaluating the existing routine clinical practice of the senior author (consultant orthopaedic surgeon). This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Informed consent was obtained from all patients.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Schurman DJ, Parker JN, Ornstein D. Total condylar knee replacement. A study of factors influencing range of motion as late as two years after arthroplasty. *J Bone Joint Surg Am* 1985;67:1006-14.
2. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 1998;7:301-17.
3. Lavernia C, D'Apuzzo M, Rossi MD, et al. Accuracy of knee range of motion assessment after total knee arthroplasty. *J Arthroplasty* 2008;23:85-91.
4. Manó S, Pálincás J, Kiss L, et al. The Influence of Lateral Knee X-Ray Positioning on the Accuracy of Full Extension Level Measurements: An In Vitro Study. *Eur J Orthop Surg Traumatol* 2012;22:245-50.
5. Jordan K, Dziedzic K, Mullis R, et al. The development of three-dimensional range of motion measurement systems for clinical practice. *Rheumatology (Oxford)* 2001;40:1081-4.
6. Bull AM, Amis AA. Knee joint motion: description and measurement. *Proc Inst Mech Eng H* 1998;212:357-72.
7. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
8. Hancock GE, Hepworth T, Wembridge K. Accuracy and reliability of knee goniometry methods. *J Exp Orthop* 2018;5:46.
9. Phillips A, Goubran A, Naim S, et al. Reliability of radiographic measurements of knee motion following knee arthroplasty for use in a virtual knee clinic. *Ann R Coll Surg Engl* 2012;94:506-12.

doi: 10.21037/aoj-22-2

Cite this article as: Summers HKC, Picken S, Al-Dadah O. Inter- and intra-rater reliability of knee flexion angle measurements on X-ray and MRI. *Ann Joint* 2022;7:34.