Title

Concurrent Validity and Reliability of a Semi-automated Approach to Measuring the Magnetic Resonance Imaging Morphology of the Knee Joint in Active Youth

Authors

Traven C.R. Blaney,[1,2] Janet L. Ronsky,[1,3] Erin M. Macri,[4] Jacob L. Jaremko,[5] Gregor Kuntze,[1,3] Amir Pakdel,[5] Jackie L. Whittaker,[1,6,7,8] Carolyn A. Emery.[1,8,9]

Affiliations

[1]McCaig Institute of Bone and Joint Health, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Faculty of Engineering, McGill University, Montreal, Canada

[3]Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

[4]Department of General Practice, Department of Orthopedics and Sports Medicine, Erasmus MC, Rotterdam, Amsterdam

[5]Department of Radiology & Diagnostic Imaging, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada.

[6]Deparment of Physical Therapy, University of British Columbia

[7]Arthritis Research Canada

[8]Sport Injury Prevention Research Centre, Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada

[9]The Alberta Children's Hospital Research Institute for Child and Maternal Health, Faculty of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author Contact:

Dr. Janet Ronsky, jlronsky@ucalgary.ca

ABSTRACT

Post-traumatic knee osteoarthritis is attributed to alterations in joint morphology, alignment, and biomechanics triggered by injury. While magnetic resonance (MR) imaging-based measures of joint morphology and alignment are relevant to understanding osteoarthritis risk, time consuming manual data extraction and measurement limit the number of outcomes that can be considered and deter widespread use. This paper describes the development and evaluation of a semi-automated software for measuring tibiofemoral and patellofemoral joint architecture using MR images from youth with and without a previous sport-related knee injury. After prompting users to identify and select key anatomical landmarks, the software can calculate 37 (14 tibiofemoral, 23 patellofemoral) relevant geometric features (morphology and alignment) based on established methods. To assess validity and reliability, 11 common geometric features were calculated from the knee MR images (proton density and proton density fat saturation sequences; 1.5 Tesla) of 76 individuals with a 3–10-year history of youth sport-related knee injury and 76 uninjured controls. Spearman's or Pearson's correlation coefficients (95% CI) and Bland-Altman plots were used to assess the concurrent validity of the semi-automated software (novice rater) versus expert manual measurements, while intra-class correlation coefficients ($ICC_{2,1}$; 95%CI), standard error of measurement (95%CI), 95% minimal detectable change, and Bland-Altman plots were used to assess the inter-rater reliability of the semi-automated software (novice versus resident radiologist rater). Correlation coefficients ranged between 0.89 (0.84,0.92; Lateral Trochlear Inclination) and 0.97 (0.96,0.98; Patellar Tilt Angle). ICC estimates ranged between 0.79 (0.63,0.88; Lateral Patellar Tilt Angle) and 0.98 (0.95,0.99; Bisect Offset). Bland-Altman plots did not reveal systematic bias. These measurement properties estimates are equal, if not better than previously reported methods suggesting that this novel semi-automated software is an accurate, reliable, and efficient alternative method for measuring large numbers of geometric features of the tibiofemoral and patellofemoral joints from MR studies.

Keywords: Alignment, Biomechanics, Morphology, MRI, Patellofemoral, Tibiofemoral, MATLAB

INTRODUCTION

Youth who suffer a sport-related knee injury are at increased risk of developing symptomatic radiographic knee osteoarthritis.[1, 2] Alongside the inflammatory response to injury, it is hypothesized that one of the contributing mechanisms underlying the elevated risk of post-traumatic osteoarthritis are alterations in tibiofemoral and patellofemoral joint bone shape (i.e., morphology), alignment and subsequent biomechanics precipitated by injury.[3]

Knee joint geometry are commonly quantified as distances, angles and ratios measured from Magnetic Resonance (MR) images.[4] Traditionally, geometric values are manually generated by an experienced rater with extensive knowledge of anatomy, radiology, and data manipulation, using a Digital Imaging and Communications in Medicine (DICOM) image file viewing application with measurement abilities.[5, 6] This resource intensive and time-consuming approach restricts the types of geometric measurements that can be assessed (i.e., DICOM applications typically only measure distances and angles), is prone to data transcription errors when data are extracted for post-processing, and limits the number of measurements that can be assessed in large MR datasets.

A purpose built, user-friendly, semi-automated software for quantifying knee joint geometries, that houses both measurement and post-processing functionalities could overcome the limitations of the traditional manual approach to obtaining knee joint geometric measurements. A semi-automated approach could also contribute to efforts to standardize how knee joint geometric outcomes are named and quantified across studies, and homogenize MR image slice selection and anatomical landmarks.[7, 8, 9] On top of being faster, less prone to error due to data manipulation, and standardizing measurements to enable data synthesis, a semi-automated software could also enable less-experienced raters to perform accurate and reliable measurements if it incorporated clear instructions. The objective of this research was to develop a semi-automated software to quantify relevant tibiofemoral and patellofemoral geometric measurements from MR images, and to assess the software's inter-rater reliability and concurrent validity to a traditional manual approach using MR images of youth with and without a previous sport-related knee injury.

METHODS

Identification of Geometric Parameters

All knee joint geometric features potentially relevant to knee injury and osteoarthritis were identified from existing systematic reviews and relevant studies.[6, 7, 8, 9, 10, 11, 12, 13, 14, 15] This list was narrowed to 37 parameters (14 tibiofemoral, 23 patellofemoral) based on discussions with an experienced rater (EM) and a musculoskeletal fellowship trained radiologist with more than 15 years of imaging experience (JJ), who informed the design of the custom-built software. The selected measures have all been previously reported and used to quantify MRI knee joint geometries in studies that assess the relationship between knee joint geometry and MRI features (e.g., cartilage damage or bone marrow lesions) or clinical outcomes (e.g., injury, osteoarthritis, pain). As we narrowed down the list of candidate measures for this study, we placed emphasis on measurements that have been shown to statistically differ between groups or conditions, as well as commonly used measurements that have appeared across multiple studies. As part of the selection process, the established methods for calculating the geometric features (exclusively distances, angles, and ratios) were extracted. Table 1 and 2 summarizes the tibiofemoral and patellofemoral features, respectively.

Table 1. Tibiofemoral Geometric Features Calculated by the Semi-automated Software

Outcome	Description	Reference points on Figures 2-4	MRI Slice (if applicable)
Tibiofemoral Angle	Angle formed by line bisecting femoral shaft and line bisecting tibial shaft	Angle between C17 → C18 and C20 → C19	Midline coronal slice[1]
Tibial Slope	Angle between a line perpendicular to the bisecting tibial axis and a line connecting the lateral and medial tibial plateau *(A negative value indicates the medial side of the plateau is more proximal than the lateral).*	Angle between C9 → C10 and C9 → C21	Midline coronal slice[1]
Medial Tibial Slope	Angle between the medial tibial slope and line perpendicular to the tibial bisecting axis *(Negative value = the anterior point of the medial tibial slope is more proximal than the posterior).*	Medial: Angle between B10 → B11 and B10 → B20	Midline coronal slice[1]
Lateral Tibial Slope	Angle between the lateral tibial slope and line perpendicular to tibial bisecting axis *(Negative value = the anterior point of the lateral tibial slope is more proximal than the posterior)*	Lateral: Angle between B10 → B11 and B10 → B20*	Midline coronal slice[1]
Intercondylar Notch Width	Distance from the most medial point of the intercondylar notch to the most lateral	C15 → C16	Slice with the most-pronounced popliteal groove[2]

Bicondylar Width	Distance from the most lateral point of the lateral femoral condyle to the most medial point of the medial femoral condyle	C11 → C12	Slice with the most-pronounced popliteal groove[2]
Medial Condylar Width	Distance from the most medial/lateral point of the medial femoral condyle to the medial side of the intercondylar notch	C12 → C18	Slice with the most-pronounced popliteal groove[2]
Lateral Condylar Width	Distance from the most medial/lateral point of the lateral femoral condyle to the lateral side of the intercondylar notch	C11 → C15	Slice with the most-pronounced popliteal groove[2]
Medial Compartment Ratio	Ratio of medial femoral condylar width and medial tibial plateau width	Medial: (B13 → B12)/(B10 → B11)	Slice with the most-pronounced popliteal groove[2]
Lateral Compartment Ratio	Ratio of lateral femoral condylar width and lateral tibial plateau width	Lateral: (B13 → B12)/(B10 → B11)	Slice with the most-pronounced popliteal groove[2]
Medial Plateau Concavity	Ratio of medial tibial plateau depth *(line perpendicular to a line that connects the anterior and posterior points of the medial tibial plateau and runs through the deepest point of the medial tibial plateau)* and the medial tibial plateau length	Medial: (B14 → B18)/(B10 → B11)	Midline coronal slice[1]
Lateral Plateau Convexity	Ratio of lateral tibial plateau height *(line perpendicular to a line that connects the anterior and posterior points of the lateral tibial plateau and runs through the highest point of the lateral tibial plateau)* and the lateral tibial plateau length	Lateral: (B14 → B18)/(B10 → B11)	Midline coronal slice[1]
Medial Condylar Convexity	Ratio of the medial condylar width and a line perpendicular to the line that runs through the most anterior and posterior points of the medial femoral condyle and runs through the deepest point of the medial femoral condyle	Medial: (B15 → B19)/(B12 → B13)	Midline coronal slice[1]
Lateral Condylar Convexity	Ratio of the lateral condylar width and a line perpendicular to the line that runs through the most anterior and posterior points of the lateral femoral condyle and runs through the deepest point of the lateral femoral condyle	Lateral: (B15 → B19)/(B12 → B13)	Midline coronal slice[1]

[1]MRI slice located halfway between the most posterior slice containing the tibia and femur, and the slice containing the most posterior aspect of the patella.
[2] The most-pronounced popliteal slice is the MRI slice where the patellar groove has the greatest area and extends the furthest medially.

Table 2. Patellofemoral Geometric Features Calculated by the Semi-automated Software

Outcome	Description	Reference points on Figures 2-4	MRI Slice (as applicable)
Insall-Salvati Ratio	Ratio of the distance from the patellar tendon tibial attachment to the most distal point of the patella, and the distance line connecting the most proximal point of the patella and the most distal.	(B6 → B7)/ (B5 → B6)	Midline patellar sagittal slice[1]
Modified Insall-Salvati Ratio	Ratio of the distance from the patellar tendon tibial attachment to the most distal point of the patellar articular surface, and the distance line connecting	(B9 → B7)/ (B8 → B9)	Midline patellar sagittal slice[1]

	the most proximal point of the patellar articular surface and the most distal.		
Blackburne-Peel Index	Ratio of the distance between the distal and proximal points of the patellar articular surface, and the distance between the distal point of the patellar articular surface and a point created from a line that is perpendicular to the medial tibial slope line, that also runs through the distal point of the patellar articular surface.	(B8 → B9)/ (B9 → B21)	Midline patellar sagittal slice[1]
Caton-Deschamps Index	Ratio of the distance between the distal and proximal points of the patellar articular surface, and the distance between the distal point of the patellar articular surface and the anterior point of the medial tibial plateau	(B8 → B9)/ (B9 → B10)	Midline patellar sagittal slice[1]
Lateral Displacement 1	Percent of patellar length that lies lateral to the anterior point of the lateral femoral trochlea	(A6 → A18)/ (A6 → A7)*100	Slice with greatest mediolateral patellar diameter[2]
Lateral Displacement 2	Distance between a line perpendicular to the posterior condylar line and a line that runs through the most anterior point of the medial femoral trochlea and the most medial point of the patella	A7 → A20	Slice with greatest mediolateral patellar diameter[2]
Bisect Offset	Percent of the patella length that lies lateral to the line perpendicular to the posterior condylar line that runs through the deepest point of the trochlea	(A6 → A14)/ (A7 → A14)*100	Slice with greatest mediolateral patellar diameter[2]
Medial Trochlear Inclination Angle	Angle between the posterior condylar line and the medial trochlea	Angle between A3 → A4 and A5 → A1	Femoral condylar slice[3]
Lateral Trochlear Inclination Angle 1	Angle between the posterior condylar line and the lateral trochlea	Angle between A3 → A4 and A5 → A2	Femoral condylar slice[3]
Lateral Trochlear Inclination Angle 2	Angle between the posterior condylar line and the lateral trochlea	Angle between A3 → A4 and A5 → A2	Most proximal axial slice showing complete cartilage coverage of the trochlea
Lateral Patellofemoral Angle	Angle between the anterior condylar line and the lateral articular surface of the patella *(Negative = the most lateral point of the patellar articular surface is posterior to the most medial point of the patellar articular surface)*	Angle between A1 → A2 and A9 → A8	Slice with greatest mediolateral patellar diameter[2]
Sulcus Angle	Angle between the medial and lateral trochlea	Angle between A5 → A1 and A5 → A1	Femoral condylar slice[3]
Trochlear Angle	Angle between the posterior condylar line and anterior condylar line *(Negative = the most anterior point of the medial trochlea is anterior to the most anterior point of the lateral trochlea)*	Angle between A3 → A4 and A1 → A2	Femoral condylar slice[3]

Lateral Patellar Tilt Angle	Angle between the posterior condylar line and the lateral articular surface of the patella *(Negative = the most lateral point of the patellar articular surface is posterior to the most medial point of the patellar articular surface)*	Angle between A3 → A4 and A9 → A8	Femoral condylar slice[3]
Patellar Tilt Angle	Angle between the posterior condylar line and a line that runs through the widest part of the patella *(Negative = the lateral point of the patella is anterior to the medial point)*	Angle between A3 → A4 and A6 → A7	Slice with greatest mediolateral patellar diameter[2]
Tibial Tuberosity to Trochlear Groove Distance (TT-TG)	Distance between the tibial tubercle and the line that is perpendicular to the posterior condylar line that runs through the deepest part of the trochlea	A10 → A15	(1) Femoral condylar slice[3] (2) Most proximal axial slice including the tibial patellar tendon attachment
Trochlear Depth	Average of the distance from the posterior condylar line to the medial and lateral condyles, minus the distance from the deepest point of the trochlea to the posterior condylar line	((A1 → A16) + (A2 → A17))/2) – (A5 → A18)	Slice with greatest mediolateral patellar diameter[2]
Congruence Angle	Angle between the line that bisects the sulcus angle and a line that connects the deepest part of the trochlea with the most posterior point of the patella	Angle between A3 → A21* and A5 → A11	Femoral condylar slice[3]
Patellar Angle	Angle between the medial and lateral patellar articular surfaces	Angle between A11 → A6 and A11 → A6	Slice with greatest mediolateral patellar diameter[2]
Trochlear Width	Length of the anterior condylar line	A1 → A2	Femoral condylar slice[3]
Axial Engagement Index	Ratio of the patellar width (greatest width) and the trochlear width (greatest width)	(A6 → A7)/ (A1 → A2)	(1) Femoral condylar slice[3] (2) Slice with greatest mediolateral patellar diameter[2]
Patellar Facet Asymmetry	Ratio of the medial and lateral patellar articular surface lengths	(A8 → A9)/ (A12 → A13)	Slice with greatest mediolateral patellar diameter[2]
Trochlear Facet Asymmetry	Ratio of the medial and lateral trochlear lengths	(A5 → A1)/ (A5 → A2)	Femoral condylar slice[3]

Note: Measurements in ***Bold font*** *were used to assess the validity and reliability of the software*

[1]The middle MRI slice containing the patella (i.e., total number of MRI slices containing the patella divided by two)

[2]The axial slice showing the greatest mediolateral patellar diameter

[3]The axial slice with the most-prominent femoral condyles – the slice showing the greatest area of femoral condyles.

Software Development

Matlab (Matlab v. 2019b, MathWorks, USA) was used to develop a custom semi-automated software (https://bit.ly/KneeMorphSAM)[16] with a custom graphical user interface (GUI) for use in conjunction with axial, coronal and sagittal proton density and proton density fat saturation sequences, and balanced steady-state gradient echo pulse sequences.

When the user initiates our software, they are prompted to select which of the 37 geometric features they want to measure (i.e., all features, a subset of features, or an individual feature). As geometric measurements of the knee are taken from individual MR image slices extracted from MR sequences based on the presence of unique anatomical landmarks, users are prompted through an anatomical point selection process of key morphological landmarks in the MR sequences. For example, to select the appropriate midline axial slice (a common slice used across multiple measurements), the user is prompted to identify both the most posterior slice containing the tibia and femur, and the slice containing the most posterior aspect of the patella. Using this information, the software then calculates and displays the exact slice representative of the halfway point. While the software recommends the use of the displayed slice, the user has the option to override the software if they believe the slice is not an accurate representation of the slice in question, if there are resolution errors, or if slice mal alignment is present due to technologist errors and variability. The anatomical point selection process is customized to include only points and relevant MR sequences (e.g., frontal, sagittal and axial) required to calculate the relevant geometric features (Tables 1 and 2, Figures 2, 3 and 4). To assist novice users and facilitate standardization, the GUI provides detailed instructions to guide users as the identify anatomical points. These instructions were developed with access to the source publications. To help orientate the user, the main GUI window includes composite images which overlay the selected MR sequence for conceptual visualization of the knee joint (Figure 1). The GUI also has tools to adjust MR image brightness and contrast, add overlaying gridlines, and zoom features. If the MR study does not include an axial sequence the software has an option to generate an axial sequence based on a reformat of available high resolution minimal slice thickness sagittal sequences.

PLACE FIGURE 1 HERE

Once anatomical point selection is complete, the software calculates the selected distance, angle and ratio geometric features using cartesian plane methods (Tables 1 and 2).[6, 7, 8, 9, 10, 11, 12, 13, 14, 15] The features and their values are then summarized and can be exported as an Excel spreadsheet (Microsoft Corporation, Redmond, WA). A detailed standard operating procedure (https://bit.ly/KneeMorphSAM)[16] was also developed to guide users and describe all functionalities of the software application.

PLACE FIGURE 2 HERE

PLACE FIGURE 3 HERE

PLACE FIGURE 4 HERE

Software Evaluation

Participants: Participants included a sub-sample of the Alberta Youth Prevention of Early-Osteoarthritis (PrE-OA) cohort study with sagittal 3D gradient echo FIESTA MR sequences (repetition time 10.5 ms, echo time 4.2 ms, slice thickness 1.0 mm, flip angle 55°, matrix 512x512; 1.5 Tesla). This included 76 youth who experienced a sport-related intra-articular knee injury 3-10 years prior to undergoing MR images, and 76 uninjured controls of similar age, sex and sporting history.[17] Balanced steady-state gradient echo pulse sequences (FIESTA MR sequences) were used for the present study because the slice thickness was smaller in comparison to other sequences obtained in the parent PrE-OA study, and the pixel resolution was higher and isotropic, allowing for high quality axial image reformatting. Information regarding participant recruitment and eligibility for the PrE-OA cohort study is detailed elsewhere.[18] Ethics approval was granted by the local Conjoint Health Research Ethics Board (CHREB, ETHICS ID # E-25075).

Procedures: A sub-set of 11 of the 37 features was selected to assess the software performance due to time and resource restraints associated with manual measurements. These 11 features (Table 2) encompass distances, angles, and ratios, and were selected, based on their expected relevance to the current cohort (i.e., youth with and without a knee injury).[4, 5, 7, 15] Given that all 37 measures are performed and calculated with similar procedures for identifying anatomical landmarks on MR images, there is no reason to believe the performance of the software would differ for the remaining 26 features.

Knee geometric measurements were assessed by three unique blinded raters. First, a novice rater (Rater 1) used the semi-automated software to measure the subset of 11 geometric features across all 152 participants. Second, a rater with nine years' experience (Rater 2) used an established manual method in conjunction with OsiriX Lite 10.0 (Pixmeo SARL, Switzerland) to estimate the same 11 geometric features across all 152 participants. Finally, a radiologist resident (Rater 3) used the semi-automated software to estimate the 11 geometric features in a subset of 30 participants. These measurements were used to estimate concurrent validity (semi-automated software, Rater 1 vs. manual measurement, Rater 2) and inter-rater reliability (Rater 1 vs. Rater 3, semi-automated software). Rater 1 developed the software, and embedded instructions based on source publications for the established manual method used by Rater 2. Rater 1 had no previous experience using the manual approach to perform the measurements. For this reason, it can be assumed that Raters 1 and 2 differed on level experience but not access to the source publications. Rater 3 only used the instructions embedded in the software and did not have access to the source publications. Rater 3 is a trained radiologist, with experience performing standard measurements on DICOM image viewing applications.

Statistical Analysis: Data analyses were performed using Stata Version 14.2 (StataCorp LP, College Station, TX). After assessing the distribution of all measurements by visual inspection and Shapiro-Wilk normality tests, the mean, standard deviation (SD) and minimum and maximum values for all 11 geometric features were calculated and summarized by rater. Levene's test was evaluated to confirm homoscedasticity of samples.[19]

Concurrent validity can be assessed in the absence of a criterion measure or 'gold standard' and is evaluated by comparing two measures on the same construct by different systems at relatively similar times (i.e., concurrently). In the case of the current study we, evaluate the concurrent validity of the newly developed semi-automated in comparison to a traditional manual measurement approach. Similar performance, of the two measurement systems give confidence that this novel method could be used in place of the traditional method, with the additional benefit of being more standardized and taking less time.[20] The concurrent validity between the semi-automated software and a traditional manual approach (criterion standard) was assessed with Spearman's or Pearson's correlation coefficients as appropriate, and 95% limits of agreement (mean difference between measurements ± 1.96 SD).[20, 21] Bland-Altman plots were also used to assess for systematic bias, outliers, and relationships between the difference in values of parameters between methods.[22] Further, these plots were visually assessed to ensure homoscedasticity (i.e., heteroscedasticity was considered present if the scatter of values changed progressively with increasing average values).[24] Bland-Altman plots for all reported measures are presented in Appendix A.

Reliability is the extent to which a measurement system is consistent and free from error.[23] Inter-rater reliability refers to the agreement in measurements between two or more raters.[23] A subset of 30 studies were used to assess reliability. Agreement between novice and experienced raters (inter-rater reliability) using the semi-automated software was assessed with Intra-class Correlation Coefficients ($ICC_{2,1}$, 95% CI),[23] and measurement precision (Standard Error of Measurement; SEm; $SD * \sqrt{1 - ICC}$).[23] Model $ICC_{2,1}$ was selected as the three raters in this study are expected to be representative of larger populations of raters (raters with minimal radiological experience, and raters with practicing radiologic history).[20] Finally, 95% Minimal Detectable Change (MDC_{95}), which represents the minimal change (in units of parameter) that must occur to be 95% confident that a true change has occurred, was calculated for each outcome as $1.96 * SEm * \sqrt{ICC}$ to assist in interpretation.[25]

RESULTS

The median participant age was 23 years (minimum-maximum 14-27) and 40% were female. Amongst previously injured participants the median age of injury was 16 years (11-19). The mean, SD and minimum and maximum values of all 11 geometric features by rater are summarized in Tables 3 and 4.

Table 3. Geometric Feature Measurements by Rater for Concurrent Validity Comparison (N = 152).

Outcome	Rater 1[1] Mean±SD (min-max) (N = 152)	Rater 2[2] Mean±SD (min-max) (N = 152)
Insall-Salvati Ratio	1.1±0.2 (0.7-1.5)	1.1±0.2 (0.7-1.5)
Bisect Offset (%)	54.0±7.2 (37.0-84.4)	54.3±7.1 (37.1-86.4)
Medial Trochlear Inclination Angle (^{o})	28.6±5.4 (15.3-42.7)	29.6±5.7 (16.4-47.3)

Lateral Trochlear Inclination Angle 1 (°)	26.2±5.2 (14.7-43.4)	27.0±5.6 (12.2-44.4)
Lateral Trochlear Inclination Angle 2 (°)	19.5±5.9 (4.1-32.1)	18.9±6.0 (1.1-33.0)
Sulcus Angle (°)	126.7±9.0 (99.0-149.8)	126.0±9.6 (96.0-153.4)
Trochlear Angle (°)	1.7±2.3 (-4.5-8.7)	1.9±2.5 (-4.1-10.0)
Lateral Patellar Tilt Angle (°)	7.9±5.4 (-8.5-22.8)	8.5±5.6 (-6.3-24.8)
Patellar Tilt Angle (°)	9.3±4.5 (-2.0-19.9)	10.8±5.1 (-2.9-23.0)
TT-TG (mm)	8.6±4.2 (1.0-23.2)	8.0±3.9 (0.5-20.0)
Trochlear Depth (mm)	3.9±1.2 (1.4-8.9)	3.8±1.2 (0.9-8.1)

mm = millimeters, SD = standard deviation. TT-TG = Tibial Tuberosity to Trochlear Groove Distance.
[1]Novice rater using semi-automatic software method.
[2]Experienced rater using manual method.

Table 4. Geometric Feature Measurements by Rater for Reliability Comparison (N = 30).

Outcome	Rater 1[1] Mean±SD (min-max) (N = 30)	Rater 3[3] Mean±SD (min-max) (N = 30)
Insall-Salvati Ratio	1.1±0.2 (0.9-1.4)	1.1±0.2 (0.8-1.5)
Bisect Offset (%)	53.0±7.0 (37.0-68.8)	53.3±7.8 (35.7-71.1)
Medial Trochlear Inclination Angle (°)	25.1±5.4 (15.3-37.5)	23.4±5.4 (14.3-35.8)
Lateral Trochlear Inclination Angle 1 (°)	25.4±3.5 (18.9-31.9)	25.1±4.1 (17.3-32.0)
Lateral Trochlear Inclination Angle 2 (°)	18.8±5.4 (7.2-29.7)	18.9±5.9 (4.9-29.0)
Sulcus Angle (°)	129.5±8.5 (115.0-145.5)	130.0±8.4 (114.2-144.3)
Trochlear Angle (°)	2.5±2.3 (-3.8-6.3)	2.4±2.5 (-2.9-9.0)
Lateral Patellar Tilt Angle (°)	8.6±5.1 (-1.7-21.2)	7.2±4.9 (-3.5-16.0)
Patellar Tilt Angle (°)	8.1±5.5 (-2.0-19.9)	7.4±4.8 (-1.2-16.1)
TT-TG (mm)	8.5±3.7 (2.9-15.7)	8.1±4.2 (0.1-15.5)
Trochlear Depth (mm)	4.0±1.1 (1.7-6.1)	4.3±1.2 (2.1-6.9)

mm = millimeters, SD = standard deviation. TT-TG = Tibial Tuberosity to Trochlear Groove Distance.
[1]Novice rater using semi-automatic software method.
[2]Radiologist rater using semi-automated software.

Concurrent Validity

The correlation and limits of agreement of between values generated by rater 1 using the novel semi-automated software (target test) and rater 2 using a traditional manual approach are summarized in Table 4. Pearson Product-Moment Correlation Coefficient (95% CI) was estimated for normally distributed measurements (i.e., Insall-Salvati Ratio, Medial Trochlear Inclination, Lateral Trochlear Inclination 2, Sulcus Angle, Trochlear Angle, Lateral Patellar Tilt, Patellar Tilt), and Spearman Rank Correlation Coefficient for non-normally distributed measurements (i.e., Bisect Offset, Lateral Trochlear Inclination 1,Tibial Tuberosity to Trochlear

Groove Distance (TT-TG), and Trochlear Depth).[21] Correlation coefficients ranged from 0.89 (Lateral Trochlear Inclination 1) to 0.97 (Patellar Tilt Angle) for angled measurement measurements, from 0.92 (Insall-Salvati Ratio) to 0.95 (Bisect Offset) for ratio measurements, and from 0.95 (TT-TG) to 0.96 (Trochlear Depth) for linear distance measurements.

Mean differences ± 2 SD (95% limits of agreement) between rater 1 and 2 ranged from 0.17° (Trochlear Angle) to 1.44° (Patellar Tilt Angle) for angled measurement measurements, from 0.003 (Bisect Offset) to 0.04 (Insall-Salvati Ratio) for ratio measurements, and from 0.295mm (Trochlear Depth) to 0.56mm (TT-TG) for linear distance measurements. Bland-Altman plots did not reveal any major systematic bias or relationships between the difference in magnitude between rater 1 and 2 and were reviewed to ensure homoscedasticity.

Table 5. Correlation and Agreement between Semi-automatic Software and Manual Approach (Rater 1 vs. 2) to Estimate Concurrent Validity (N = 152).

Outcome	Pearson, or Spearman Correlation Coefficient (95% CI)	p-value	Mean Difference (95% CI), (LOA)
Insall-Salvati Ratio	0.92[a] (0.89, 0.94)	< 0.001	0.04 (0.03 0.05), (-0.09, 0.17)
Bisect Offset (%)	0.95[b] (0.92, 0.96)	< 0.001	0.28% (-0.59 0.04), (-4.21, 3.65)
Medial Trochlear Inclination Angle (°)	0.91[a] (0.87, 0.93)	< 0.001	-1.02° (-1.41 -0.64), (-5.84, 3.79)
Lateral Trochlear Inclination Angle 1 (°)	0.89[b] (0.84, 0.92)	< 0.001	0.75° (-1.12 -0.38), (-3.53, 4.30)
Lateral Trochlear Inclination Angle 2 (°)	0.91[a] (0.88, 0.95)	< 0.001	0.64° (0.24 1.03), (-4.33, 5.60)
Sulcus Angle (°)	0.95.[a] (0.94, 0.97)	< 0.001	0.33° (-0.13 0.79), (-5.46, 6.12)
Trochlear Angle (°)	0.96[a] (0.95, 0.98)	< 0.001	-0.17° (-0.27 -0.07), (-1.42, 1.08)
Lateral Patellar Tilt Angle (°)	0.95[a] (0.94, 0.97)	< 0.001	-0.64° (-0.91 -0.37), (-4.04, 2.76)
Patellar Tilt Angle (°)	0.97[a] (0.96, 0.98)	< 0.001	-1.44° (-1.65 -1.23), (-4.06, 1.18)
TT-TG (mm)	0.95[b] (0.93, 0.97)	< 0.001	0.56mm (0.36 0.75), (-1.87, 2.93)
Trochlear Depth (mm)	0.96[b] (0.93, 0.97)	< 0.001	-0.14mm (0.09 0.19), (-0.54, 0.82)

LOA = limits of agreement, mm = millimeters, SEM = Standard error of measurement, TT-TG = Tibial Tuberosity to Trochlear Groove Distance.
[a]Pearson Correlation Coefficient
[b]Spearman Correlation Coefficient

Inter-rater Reliability

Inter-rater reliability ($ICC_{2,1}$ (95% CI)), SEM, MDC_{95}, mean differences and 95% limits of agreement (± 1.96 SD) between rater 1 and 3 using the semi-automated software are summarized in Table 5. $ICC_{2,1}$ (95% CI) values ranged between 0.79 (95% CI 0.63, 0.88) for Lateral Patellar

Tilt Angle and 0.98 (0.95, 0.99) for Bisect Offset. SEM values ranged from 0.783° (Trochlear Angle) to 3.382° (Sulcus Angle) for angled measurement measurements, from 0.010 (Bisect Offset) to 0.054 (Insall-Salvati Ratio) for ratio measurements, and from 0.348mm (Trochlear Depth) to 1.109mm (TT-TG) for linear distance measurements. Bland-Altman plots showed homoscedasticity and did not reveal any significant systematic bias or relationships between the differences in magnitude between measurements.

Table 6. Correlation and Agreement between Novel Rater and Radiologist (Rater 1 vs. 3) to Estimate Reliability (N = 30).

Outcome	$ICC_{2,1}$ (95%CI)	SEM	MDC_{95}	Mean Difference (95% CI), (LOA)
Insall-Salvati Ratio	0.89 (0.80, 0.94)	0.05	0.15	0.07 (-0.02 0.04), (-0.15, 0.17)
Bisect Offset (%)	0.98 (0.95, 0.99)	1.02%	2.82%	-0.27% (-0.88 0.33), (-3.52, 2.97)
Medial Trochlear Inclination Angle (°)	0.92 (0.86, 0.96)	1.52°	4.20°	1.77° (0.96 2.58), (-2.59, 6.13)
Lateral Trochlear Inclination Angle 1 (°)	0.86 (0.76, 0.93)	1.94°	5.37°	0.38° (-0.35 1.11), (-3.54, 4.30)
Lateral Trochlear Inclination Angle 2 (°)	0.89 (0.81, 0.94)	1.97°	5.46°	-0.04° (-1.02 0.93), (-5.27, 5.18)
Sulcus Angle (°)	0.86 (0.75, 0.92)	3.38°	9.37°	-0.50° (-2.16 1.16), (-9.37, 8.38)
Trochlear Angle (°)	0.90 (0.82, 0.95)	0.78°	2.05°	0.09° (-0.30 0.49), (-2.03, 2.21)
Lateral Patellar Tilt Angle (°)	0.81 (0.66, 0.89)	2.35°	6.53°	1.44° (0.30 2.58), (-4.68, 7.55)
Patellar Tilt Angle (°)	0.79 (0.63, 0.88)	2.08°	5.77°	0.68° (-0.57 1.92), (-5.99, 7.34)
TT-TG (mm)	0.93 (0.87, 0.96)	1.11mm	3.08mm	0.35mm (-0.21 0.90), (-2.61, 3.30)
Trochlear Depth (mm)	0.92 (0.86, 0.96)	0.35mm	0.97mm	-0.30mm (-0.46 -0.13), (-1.18, 0.59)

Abbreviations: LOA = limits of agreement, MDC_{95} = Minimal detectable change, mm = millimeters, SEM = Standard error of measurement, TT-TG = Tibial Tuberosity to Trochlear Groove Distance.

DISCUSSION

The traditional manual approach to quantify the geometry of the knee joint from MR images is resource intensive, time-consuming, and can be prone to data transcription errors. This study reports the development and preliminary evaluation (validity and inter-reliability) of a semi-automated approach for measuring tibiofemoral and patellofemoral joint geometry using MR images from youth with and without a previous sport-related knee injury. The novel semi-automated software, which guides users to select anatomical landmarks associated with desired geometrics features demonstrates strong concurrent validity and inter-rater reliability for measuring linear distance, angle, and ratio measurements suggesting that it is an accurate, reproducible, and efficient alternative method for measuring large numbers of tibiofemoral and patellofemoral geometric features from MR studies. Although the current study only assessed the validity and reliability of a sub-set of varied geometric features, there is no reason to suggest that the performance of the custom software would yield inferior results for the remaining features as the measurement process is consistent.

To understand the relationship between measurements made with the target and criterion standard test we estimated the monotonic association between measurements from the two methods with a correlation coefficient,[21] calculated mean differences (95% limits of agreements) and employed Bland-Altman plots to assess for systematic bias.[22] All monotonic associations assessing concurrent validity exceeded 0.89 which suggests a good to excellent correlation between the measurements made with the semi-automated software and traditional manual approach.[21] Mean differences less than 1.5° for angular measurements, 1mm for linear distance measurements, and 5% for ratio measurements, as well as absence of systematic bias provide further evidence of the promising accuracy of the semi-automated software. To provide a brief context for these reported mean difference values, several studies were identified which produced statistically significant differences, and compared knee joint geometry with MRI features (e.g., cartilage damage or bone marrow lesions) or clinical outcomes (e.g., injury, osteoarthritis, pain). Crossley et al. 2009[26] reported the following mean differences between a population with patellofemoral joint OA and an uninjured control group: 19.86% for Bisect Offset, -1.31 for Lateral Patellar Tilt Angle. Ali et al. 2010[27] reported the following mean differences between an under 40-year-old population with severe cartilage defects and an uninjured control group: -8.1° for Lateral Trochlear Inclination 1, 22° for Sulcus Angle, -3.39mm for Trochlear Depth, 0.09 for Insall-Salvati Ratio. Stefanik et al. 2012[28] reported the following ranges between quartiles with a low and high prevalence of patellofemoral joint cartilage damage and bone marrow legions, respectively: <4.96° and <0.8° for the Trochlear Angle. Macri et al. 2018[5] reported mean differences for the following: 3.6% for Bisect Offset, and 4.9° for Sulcus Angle.

To evaluate inter-rater reliability between a novice and experienced rater using the semi-automated software we assessed the association between rater measurements (ICC), measurement precision (SEm) and examined Bland-Altman plots. All ICC estimates exceeded 0.79 which suggests good to excellent reliability.[29] SEm estimates less than 3.5° for angular measurements, 1.2mm for linear measurements, and 0.5% for ratio measurements, as well as absence of systemic bias suggest raters of all levels can generate relatively consistent measurements of angular, linear and ratio measurements with the semi-automated software. MDC and SEm values are somewhat appreciable, provided the subtlety of quantifiable morphological changes precipitated by a knee injury, which infers differences between groups may go overlooked by the current iteration of the software, if there are multiple raters. Similarly, many clinical measures (e.g., strength testing, range of motion testing) have relatively high MDCs, so it is typically recommended that measures on an individual patient are made by the same rater to enhance interpretation of the values. Our results suggest that interpretation of some of the measures made with the custom software will also benefit from a measurement system that includes the same rater particularly if longitudinal comparison is desired. That withstanding, when evaluating group differences, in research settings, it is recommend that MDC is calculated as $\frac{MDC}{\sqrt{N}}$, which would result in a much smaller group MDC value that supports the usefulness of these measures.[25]

The error associated with appreciable MDC values is likely attributed to rater experience and skill. To ameliorate this potential source of error, instructions to the software could be supplemented by source publications with more elaborate measurements descriptions. This source of error may also be compounded by pixel resolution. The pixel resolution for the sagittal MR images used in this study was 0.293mm in both the x and y axis. However, as the software created the axial reformats for measures in requiring an axial viewing plane, this distance increased to 1mm in the y direction which may have led to greater between rater differences and margin of error in terms of real-world distances. Furthermore, the custom software only allows for selection of the center of the pixel. DICOM image viewing applications typically make use of interpolation formulas for this reason, and future iterations of the custom software should apply interpolation formulas to increase the resolution in this direction, thereby minimizing this error of measurement. Future studies should clarify these discrepancies with a more robust study design, where intra-rater reliability is assessed, and resultant values between injured and uninjured groups are compared between the custom software and established traditional approaches with standardized experience levels.

We were unable to identify previous studies that assessed geometric features of the knee joint from MRI measurements with, or assessed the concurrent validity of, semi-automated methods. One of the most crucial criteria for establishing concurrent validity is the validity of the criterion standard. For this study, the criterion or “gold” standard was a manual method employing an image processing application for DICOM images (DICOM viewing applications) to extract distances and angles, and Microsoft Excel® (Microsoft, USA) for data manipulation and analysis. Although measurements of this manual approach have not been compared to real-world, in-situ bone geometry measurements, it has been used on numerous occasions to measure and assess the relationship of knee geometrics to clinical features.[5, 15, 30, 31] The novel semi-automated software developed in this study produced virtually identical measurements to this established manual method. It is interesting to note that Rater 2 (EM) estimated that the process of measuring and calculating the 11 geometric measurements using the established manual method took on an average 20 minutes per participant compared to approximately 10 minutes per participants with the semi-automated approach. Given that the semi-automatic approach allows investigators to measure up to 37 geometric features in approximately half of the time, while eliminating opportunities for data transcription errors, suggests that it is a promising alternative.

Although we were unable to identify any studies that have assessed the reliability of semi-automated approaches that measure geometric features of the knee, there are studies that examine the reliability of traditional manual measurement approaches. An example subset of six such studies have estimated ICC values for inter and or intra-rater reliability for the traditional manual approach.[4, 11, 15, 30, 31, 32] Using the picture archiving software Centricity® (General Electric Healthcare, USA), Sebro et al. 2017[11] reported intra-rater ICC estimates ranging from 0.54-0.93, while Mundy et al. 2016[32] reported intra-rater ICCs between (0.78-0.90), and inter-rater ICCs between (0.81-0.96), both in pediatric populations. Stefanik et al. 2010[31] and 2012[30], using OsiriX (Pixmeo, SARL, Switzerland), reported ranges of intra and inter-rater ICCs values across several geometric features of the knee in the MOST cohort (adults with or at increased risk of osteoarthritis) greater than 0.90, and 0.70, respectively. Macri et al. 2017[4] and 2018[15] reported estimates of reliability across several geometric features of the knee in both the Framingham Community cohort and in a population 1 year after anterior cruciate ligament reconstruction surgery - presenting intra-rater ICCs between 0.89-0.99, and inter-rater ICCs between 0.85-0.98. Given the susceptibility of ICCs to be inflated with samples containing large variance, stability and agreement of a measurement system should also be considered.[20] In relation to the reliability of established methods, the ICC estimates of inter-rater reliability of the semi-automated software presented in this paper are directly comparable, if not superior to, previously established methods.

Limitations: The findings of this study demonstrate that a novel semi-automated software for measuring geometric features of the knee is valid and reliable for measuring 11 features including linear, angular and ratio measurements. Similarly, only MR images from a subset of 30 participants were used to estimate inter-rater reliability, which could mask variability – although Tables 3 and 4 present similar magnitudes of descriptive statistics between different sample sizes, suggesting that the subset of 30 participants could be extrapolated. A larger sample size may result in different estimates of reliability and should be evaluated in future studies to confirm reproducibility. Despite being small, the clinical relevance, if any, of the mean difference in measurements generated with the traditional manual and novel semi-automated software has yet to be determined officially. To further assess clinical relevance, reported measurements using the custom software between an injured and uninjured control population could be evaluated for statistical relationships between study groups, and then compared to the results of the same measurements using an established manual method - thus ensuring the mean differences do not have a substantial influence when assessing measurements between injured and uninjured populations. Further, it is important to note that the imaging processing software OsiriX may prove to be more sophisticated and precise in interpreting data from high resolution images than the novel semi-automated software developed in this study, which is limited by the fact that the user is only able to select the center of a pixel, where each pixel represents a real-world dimension that is provided in the DICOM (MRI) file.

It is important to recognize that both measurement systems assessed in this study (software and manual) included a rater, and that it is impossible to compare these systems without considering the level of rater experience. Although there would be less potential for variability in scores if we had compared the two measurement systems using the same experienced rater, we would not have been able to inform the real-world in-situ use of the software and understand if it could walk an inexperienced user through the measurement process and successfully generate knee joint measurements.

The cohort with which MRI sequences were drawn for this study was well-standardized in terms of imaging protocol, meaning all MR images used to perform measurements were produced on the same equipment at the same site. While this is a benefit for validating the software, the reproducibility of these results has not yet been tested on other equipment, at different testing sites, at different time points, upon patient repositioning, or with other patient populations. In its current iteration, and without the use of interpolation formulas to increase pixel resolution, the software measurements would likely change with larger slice thicknesses, or poorer MR image resolutions. With that said, the primary source of information required by the software to perform measurements is the pixel-spacing attribute found in DICOM image files, which suggests that the software could be used for all MR image studies using the same file format and regardless of equipment used – although fluctuations based on resolution could still occur.

CONCLUSIONS

This study describes the development and preliminary evaluation of a novel semi-automated method for characterizing knee joint geometry using anatomical landmark features from MRI. We found the software showed concurrent validity and inter-rater reliability equivalent or better than previously reported methods for measuring geometric features, with accuracy, reproducibility, and efficiency potentially suitable for measuring large numbers of tibiofemoral and patellofemoral features from MRI. This method provides a viable alternative measurement approach to the resource intensive, time consuming traditional manual approach, and may be of interest to investigators interested in considering multiple geometric features of the knee joint in large MR datasets. Although the results are promising, further studies are still required to officially validate the clinical capabilities of the software.

Future Directions:

Future evaluation of the novel semi-automated software developed in this study could include further feasibility and acceptability measurements including the time saved in using it compared to traditional manual approaches, and feedback from both naïve and expert users. Future functionality related to use with other sagittal and frontal MR sequences and measurements that exploit the three-dimensional nature of MR sequence data vs individual two-dimensional slices. Finally, future studies are required to assess differences in the geometric features of the knee as well as the relationship between these features and clinical measurements in youth with and without a past intra-articular knee injury, as well as other clinical populations (e.g., knee osteoarthritis).

Authorship: Conceptualization, C.E., J.J., J.W., and J.R.; methodology, E.M., J.J., J.R.; software, T.B., G.K., and J.J.; validation, T.B., E.M., and A.P.; formal analysis, T.B.; investigation, T.B., E.M., and A.P.; resources, C.E., J.R.; data curation, T.B. and A.P.; writing—original draft preparation, T.B.; writing—review and editing, J.W., G.K., E.M., J.J., J.R., C.E.; supervision, C.E., J.R., J.J., G.K.; project administration, C.E. All authors have read and agreed to the published version of the manuscript.

Funding: The Alberta PrE-OA cohort is funded by the Canadian Institutes of Health Research (MOP 133597), the Alberta Team Osteoarthritis Team supported by Alberta Innovates, and the Alberta Children’s Hospital Research Institute Chair in Pediatric Rehabilitation (Alberta Children’s Hospital Foundation). The Sport Injury Prevention Research Centre (SIPRC) is one of ten International Olympic Committee Research Centers focused on Injury and Illness Prevention in Sport.

Declaration of Conflicting Interests: JLW is an Associate Editor of the British Journal of Sports Medicine. All other authors declare no conflict of interest.

REFERENCES

1. Whittaker JL, Toomey CM, Nettel-Aguirre A, et al. Health-related Outcomes after a Youth Sport-related Knee Injury. *Med Sci Sports Exerc* 2019; 51: 255-263. DOI: 10.1249/MSS.0000000000001787.

2. Richmond SA, Fukuchi RK, Ezzat A, et al. Are joint injury, sport activity, physical activity, obesity, or occupational activities predictors for osteoarthritis? A systematic review. *J Orthop Sports Phys Ther* 2013; 43: 515-B519. 20130611. DOI: 10.2519/jospt.2013.4796.

3. Baker-LePain JC and Lane NE. Role of bone architecture and anatomy in osteoarthritis. *Bone* 2012; 51: 197-203. 20120124. DOI: 10.1016/j.bone.2012.01.008.

4. Macri EM, Felson DT, Zhang Y, et al. Patellofemoral morphology and alignment: reference values and dose-response patterns for the relation to MRI features of patellofemoral osteoarthritis. *Osteoarthritis Cartilage* 2017; 25: 1690-1697. 20170623. DOI: 10.1016/j.joca.2017.06.005.

5. Macri EM, Culvenor AG, Morris HG, et al. Lateral displacement, sulcus angle and trochlear angle are associated with early patellofemoral osteoarthritis following anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc* 2018; 26: 2622-2629. 20170509. DOI: 10.1007/s00167-017-4571-1.

6. Haj-Mirzaian A, Thawait GK, Tanaka MJ, et al. Diagnosis and Characterization of Patellofemoral Instability: Review of Available Imaging Modalities. *Sports Med Arthrosc Rev* 2017; 25: 64-71. DOI: 10.1097/JSA.0000000000000148.

7. Macri EM, Stefanik JJ, Khan KK, et al. Is Tibiofemoral or Patellofemoral Alignment or Trochlear Morphology Associated With Patellofemoral Osteoarthritis? A Systematic Review. *Arthritis Care Res (Hoboken)* 2016; 68: 1453-1470. DOI: 10.1002/acr.22842.

8. Diederichs G, Issever AS and Scheffler S. MR imaging of patellar instability: injury patterns and assessment of risk factors. *Radiographics* 2010; 30: 961-981. DOI: 10.1148/rg.304095755.

9. van Diek FM, Wolf MR, Murawski CD, et al. Knee morphology and risk factors for developing an anterior cruciate ligament rupture: an MRI comparison between ACL-ruptured and non-injured knees. *Knee Surg Sports Traumatol Arthrosc* 2014; 22: 987-994. 20130706. DOI: 10.1007/s00167-013-2588-7.

10. Paiva M, Blond L, Holmich P, et al. Quality assessment of radiological measurements of trochlear dysplasia; a literature review. *Knee Surg Sports Traumatol Arthrosc* 2018; 26: 746-755. 20170318. DOI: 10.1007/s00167-017-4520-z.

11. Sebro R and Weintraub S. Knee morphometric and alignment measurements with MR imaging in young adults with central cartilage lesions of the patella and trochlea. *Diagn Interv Imaging* 2017; 98: 429-440. 20170223. DOI: 10.1016/j.diii.2017.01.010.

12. Schneider A, Si-Mohamed S, Magnussen RA, et al. Tibiofemoral joint congruence is lower in females with ACL injuries than males with ACL injuries. *Knee Surg Sports Traumatol Arthrosc* 2018; 26: 1375-1383. 20171020. DOI: 10.1007/s00167-017-4756-7.

13. Fucentese SF, von Roll A, Koch PP, et al. The patella morphology in trochlear dysplasia--a comparative MRI study. *Knee* 2006; 13: 145-150. 20060209. DOI: 10.1016/j.knee.2005.12.005.

14. Lykissas MG, Li T, Eismann EA, et al. Does medial patellofemoral ligament reconstruction decrease patellar height? A preliminary report. *J Pediatr Orthop* 2014; 34: 78-85. DOI: 10.1097/BPO.0b013e3182a12102.

15. Macri EM, Culvenor AG, Morris HG, et al. Lateral displacement, sulcus angle and trochlear angle are associated with early patellofemoral osteoarthritis following anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc* 2018; 26: 2622-2629. 20170509. DOI: 10.1007/s00167-017-4571-1.

16. Whittaker JL, Blaney T. Semi-automated Approach to Measuring MRI Knee Morphology. osf.io/mknp2 (2021).

17. Whittaker JL, Toomey CM, Woodhouse LJ, et al. Association between MRI-defined osteoarthritis, pain, function and strength 3-10 years following knee joint injury in youth sport. *Br J Sports Med* 2018; 52: 934-939. 20171010. DOI: 10.1136/bjsports-2017-097576.

18. Steffen K, Meeuwisse WH, Romiti M, et al. Evaluation of how different implementation strategies of an injury prevention programme (FIFA 11+) impact team adherence and injury risk in Canadian female youth football players: a cluster-randomised trial. *Br J Sports Med* 2013; 47: 480-487. 20130313. DOI: 10.1136/bjsports-2012-091887.

19. Brown MB and Forsythe AB. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association* 1974; 69: 364-367. DOI: 10.2307/2285659.

20. Portney LGWMP. *Foundations of clinical research : applications to practice*. 2015.

21. Schober P, Boer C and Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg* 2018; 126: 1763-1768. DOI: 10.1213/ane.0000000000002864.

22. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310.

23. Shrout PE and Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-428. DOI: 10.1037//0033-2909.86.2.420.

24. Ludbrook J. Confidence in Altman-Bland plots: a critical review of the method of differences. *Clin Exp Pharmacol Physiol* 2010; 37: 143-149. 20090828. DOI: 10.1111/j.1440-1681.2009.05288.x.

25. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34-42. 20060824. DOI: 10.1016/j.jclinepi.2006.03.012.

26. Crossley KM, Marino GP, Macilquham MD, et al. Can patellar tape reduce the patellar malalignment and pain associated with patellofemoral osteoarthritis?1. *Arthritis Care & Research* 2009; 61: 1719-1725. DOI: 10.1002/art.24872.

27. Ali SA, Helmer R and Terk MR. Analysis of the Patellofemoral Region on MRI: Association of Abnormal Trochlear Morphology With Severe Cartilage Defects. *American Journal of Roentgenology* 2010; 194: 721-727. DOI: 10.2214/AJR.09.3008.

28. Stefanik JJ, Roemer FW, Zumwalt AC, et al. Association between measures of trochlear morphology and structural features of patellofemoral joint osteoarthritis on MRI: the MOST study. *J Orthop Res* 2012; 30: 1-8. 20110624. DOI: 10.1002/jor.21486.

29. Koo TK and Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15: 155-163. 20160331. DOI: 10.1016/j.jcm.2016.02.012.

30. Stefanik JJ, Roemer FW, Zumwalt AC, et al. Association between measures of trochlear morphology and structural features of patellofemoral joint osteoarthritis on MRI: the MOST study. *J Orthop Res* 2012; 30: 1-8. 20110624. DOI: 10.1002/jor.21486.

31. Stefanik JJ, Zhu Y, Zumwalt AC, et al. Association between patella alta and the prevalence and worsening of structural features of patellofemoral joint osteoarthritis: the multicenter osteoarthritis study. *Arthritis Care Res (Hoboken)* 2010; 62: 1258-1265. DOI: 10.1002/acr.20214.

32. Mundy A, Ravindra A, Yang J, et al. Standardization of patellofemoral morphology in the pediatric knee. *Pediatr Radiol* 2016; 46: 255-262. 20150917. DOI: 10.1007/s00247-015-3459-9

FIGURE CAPTIONS

Figure 1. Illustration of anatomical points included in the calculation of geometric outcomes, for the axial viewing plane. Points labeled with an asterisk are not included in the point selection process of the custom software, as outlined in the Software Development section of this report and are therefore interpreted by the software to calculate outcomes. All other points are included in the point selection process. Points were labeled for reference for Tables 1 and 2, that describe the outcomes included in the custom software.

Figure 2. Illustration of anatomical points included in the calculation of architectural outcomes, for the sagittal viewing plane. Points labeled with an asterisk are not included in the point selection process of the custom software, as outlined in the Software Development section of this report and are therefore interpreted by the software to calculate outcomes. All other points are included in the point selection process. Points were labeled for reference for Tables 1 and 2, that describe the outcomes included in the custom software.

Figure 3. Illustration of anatomical points included in the calculation of geometric outcomes, for the coronal viewing plane. Points labeled with an asterisk are not included in the point selection process of the custom software, as outlined in the Software Development section of this report and are therefore interpreted by the software to calculate outcomes. All other points are included in the point selection process. Points were labeled for reference for Tables 1 and 2, that describe the outcomes included in the custom software.

Figure 4. Composite renderings created by the custom software used to assist the rater with orientation and visualizations. The user has the option to view one of the three renderings at a time. From left to right: a simple Composite Overlay of each MR image in the sequence, an Average Intensity Projection (AIP) is an image of the average intensity values of pixels between all MR images in the sequence, and a Maximal Intensity Projection (MIP) is an image of the maximum intensity value of pixels between all MR images in the sequence.

FIGURES

Figure 1

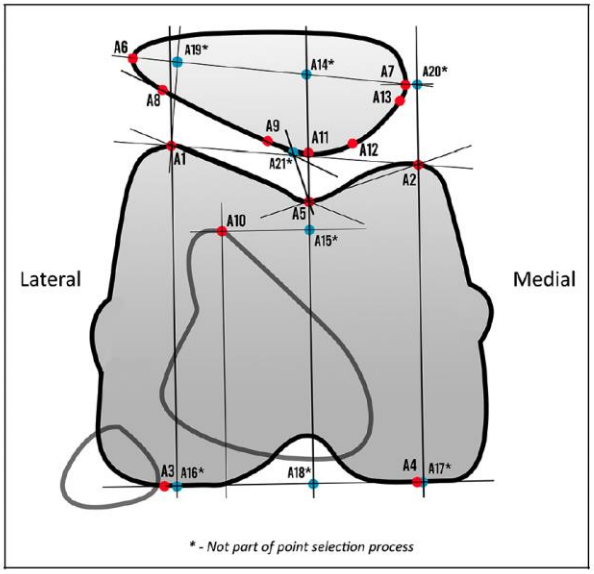

Figure 2

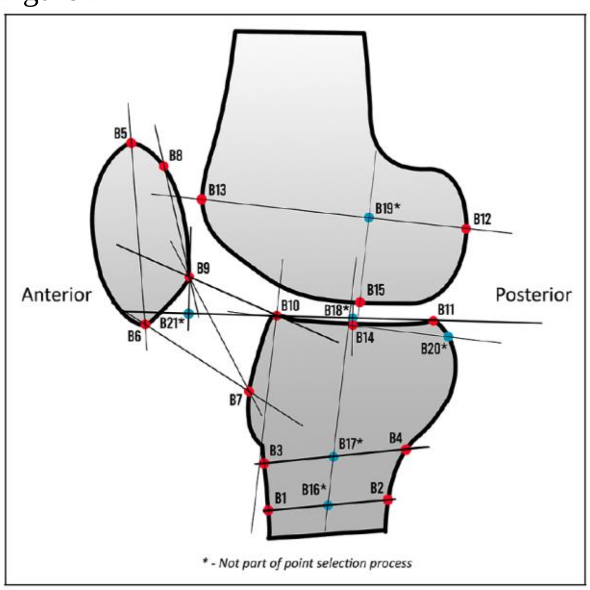

Figure 3

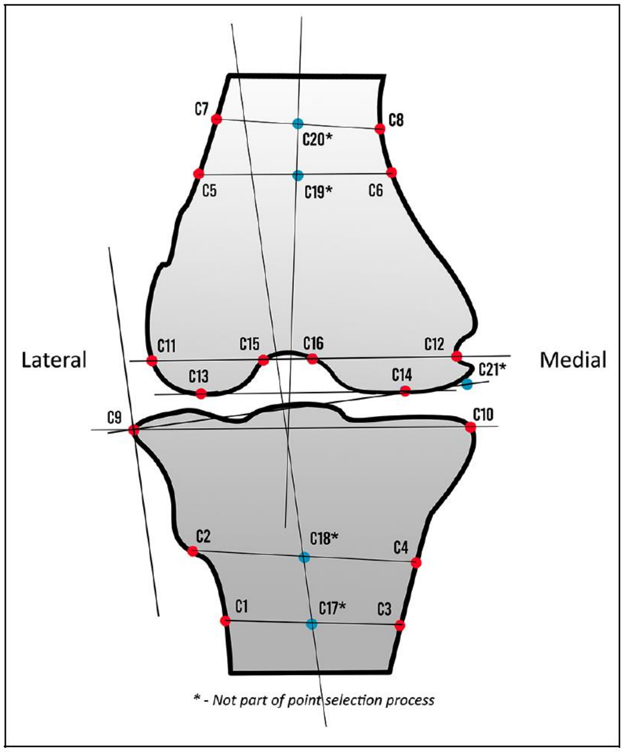

Figure 4