Journal of Biomechanics 95 (2019) 109309

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

Variability, agreement and reliability of MRI knee landmarks

Johan van der Merwe [a,*], Dawie J. van den Heever [a], Pieter Erasmus [b]

[a] Biomedical Engineering Research Group, Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Stellenbosch 7600, South Africa
[b] MediClinic, Stellenbosch 7600, South Africa

ARTICLE INFO

Article history:
Accepted 2 August 2019

Keywords:
MRI
Knee
Landmarks
Variability
Agreement

ABSTRACT

Surface mesh reconstructions of bones are often required to define landmark-based coordinate systems, regions of interest and morphological features when studying the soft tissues of the knee from MRI scans. This study reports the variability, agreement and reliability of osseous landmarks to better understand their downstream effects. Fifteen landmarks were defined on the distal femur and twelve on the proximal tibia. Surface meshes were created from twenty right knee MRI scans with a mean subject age of 30.9 years. A single observer identified landmarks on all twenty knees, while three observers repeated the observations three times on a subset of eight knees. All observations were aligned to the Procrustes mean shapes. Principal component analysis was used to study inter-subject variability and two-way ANOVA for inter- and intra-observer agreement and reliability. Inter-subject landmark variation ranged from 0.6 to 5.26 mm, while inter- and intra-observer agreement were at most 5.1 and 5.69 mm respectively. Between-observer reliability ranged from 0.07 to 0.98 while within-observer values were between 0.51 and 0.98. Landmarks derived from fitted spheres or circles often performed well, while most others had their poorest agreement or greatest variation limited to only one or two cardinal directions.

1. Introduction

Magnetic resonance imaging (MRI) plays an important role in the study of the knee and is the preferred modality for investigating soft tissues (Braun and Gold, 2012; Choi and Gold, 2011; Eckstein et al., 2014). However, MRI may also be used to accurately reconstruct surface meshes of the osseous geometries through image segmentation (Van den Broeck et al., 2014). These bone meshes are important for defining joint coordinate systems, regions of interest, and identifying morphological features (Grood and Suntay, 1983; Van der Merwe et al., 2013; Victor et al., 2009). Applications of MRI derived knee bone meshes in the literature include quantitative analyses (Bowes et al., 2015; Bredbenner et al., 2010), anatomical models (DeFrate et al., 2004; Pena et al., 2006; Yue et al., 2011) and design and evaluation of patient-specific instrumentation and implants (Frye et al., 2015; Howell et al., 2008; Van den Heever et al., 2012).

Any such in silico applications must be related back to the intra-operative situation via homologous, surgically relevant and accessible landmarks (Victor et al., 2009). Examples include definition of the aforementioned coordinate systems, such as that of Grood and Suntay (1983) which used the most posterior points on the medial and lateral condyles of the femur as well as the tibial spine center. Landmarks like the posterior aspect of the trochlear groove have been used to delineate regions of interest on the articular surface of the distal femur (Williams et al., 2006). The morphology of the groove itself has also been studied by placing a series of landmarks along its deepest portion (Iranpour et al., 2010). It stands to reason then that any uncertainties associated with the definition of landmarks are liable to affect subsequent measurements and procedures (Della Croce et al., 1999; Kedgley et al., 2015). However, there appears to be a sparsity of information regarding which landmarks are most appropriate for use on MRI derived knee bone meshes.

So the question remains: which landmarks are best when using MRI-derived surface meshes of knee bones? To provide information for an informed decision, we investigated the inter- and intra- observer agreement and reliability as well as inter-subject variability for a large set of landmarks defined on 3D surface meshes reconstructed from 20 MRI scans. All specimens were aligned to a mean reference shape via generalised Procrustes analyses in order to avoid contaminating the results with a landmark-based coordinate system's variation.

* Corresponding author at: Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Cnr of Banhoek and Joubert Str, Stellenbosch 7600, South Africa.
E-mail address: jovdmerwe@sun.ac.za (J. van der Merwe).

https://doi.org/10.1016/j.jbiomech.2019.08.003
0021-9290/

2. Methods

2.1. Data

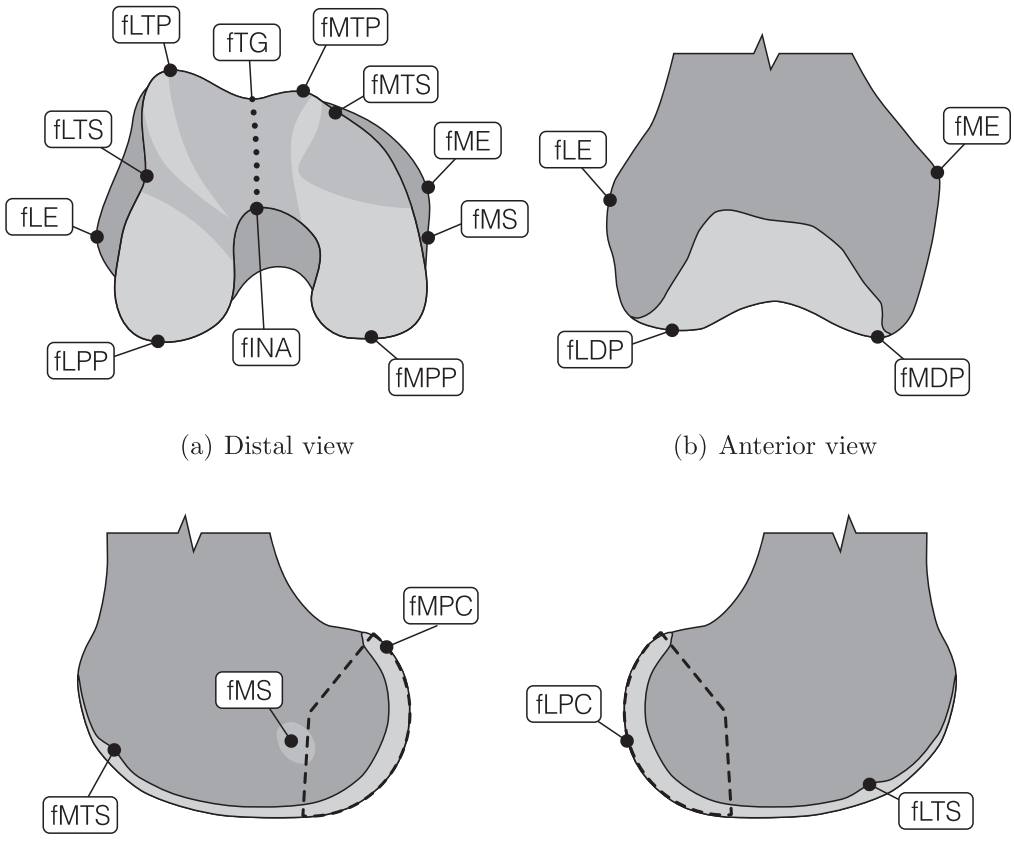
The Faculty of Health Sciences (Stellenbosch University, South Africa) approved the collection of 20 Caucasian (eight male and 12 female) subjects' right knee MRI scans, ranging in age from 16 to 57 with a mean of 30.9 years. Inclusion criteria specified that subjects be skeletally mature with no identifiable osseous and cartilaginous pathology. The thicknesses of the slices were 1.5 mm with a resolution of 512×512 pixels. Pixel sizes ranged between 0.38 and 0.47 mm with an average of 0.4 mm. A single expert segmented the distal femur and proximal tibia of each scan in order to reduce bias. Segmentation was done in the sagittal plane using Mimics software (Materialise, Leuven, Belgium). Based on a visual inspection of the gray values, a threshold was first applied in order to best isolate the bony geometries. Unwanted artefacts, structures and holes were then manually cleaned from every second slice. These edited slices were subsequently interpolated, and the interpolated slices again inspected and cleaned if necessary to form a complete mask from which a 3D surface mesh was created. The mesh was further smoothed to remove any remaining artefacts, and visually validated against the original raw scan. Finally, the mesh was exported as a stereolithography face-vertex data structure.

2.2. Observations

A Matlab (MathWorks, Massachusetts, USA) program with a user interface was written to place landmarks directly on the surface meshes. The program presented each mesh to an observer in random order and random orientation so as to prevent memorisation. Once a mesh was shown, the observer was free to align it as desired. Landmarks were requested in random order for each new mesh and the observer was allowed to make adjustments until they were satisfied. The landmarks were recorded relative to the native orientation of the mesh as segmented from the raw scan data. Landmarks were classified as one of three types:

Point (P): The three dimensional Cartesian coordinates of the intersection between a ray that is perpendicular to the screen and originates on the cursor of the program, and the mesh.
Point Array (PA): A list containing a predetermined number of points, each individual point as defined previously.
Face Array (FA): A list of all mesh faces that lied at least partially within a sphere with a user specified radius and centred around intersection points as defined previously. Instances of the sphere were repeatedly defined as the observer dragged the cursor along the surface of the mesh to 'paint' a portion of it.

The landmarks used here were derived from similar studies in the literature as well as discussions with the collaborating orthopaedic surgeon (Cobb et al., 2008; Esfandiarpour et al., 2009; Victor, 2009; Victor et al., 2009). Fig. 1 and Table 1 describe the full set of landmarks defined on the distal femur, while Fig. 2 and Table 2 does so for the proximal tibia. Each observer was provided with similar tables and figures for reference. Computed landmarks were calculated afterwards from point or face arrays, and therefore only indirectly identified by an observer.



(a) Distal view (b) Anterior view

(c) Medial view (d) Lateral view

Fig. 1. Locations of the femoral landmarks.

Table 1
Femoral landmarks.

ID	Landmark	Description	Type
		Anatomic landmarks	
fME	Medial epicondyle	Most prominent bony protrusion over the medial condyle	P
fMS	Medial sulcus	Indentation on the medial condyle's bone surface	P
fLE	Lateral epicondyle	Most prominent bony protrusion over the lateral condyle	P
fMTP	Medial trochlea peak	The most anterior point of the medial trochlear ridge	P
fLTP	Lateral trochlea peak	The most anterior point of the lateral trochlear ridge	P
fINA	Intercondylar notch apex	The most anterior and distal termination of the intercondylar notch	P
fTG	Trochlear groove	9 equidistant points placed along the deepest line of the trochlear groove	PA
fMTS	Medial terminal sulcus	Medial end of the medial condyle's groove	P
fLTS	Lateral terminal sulcus	Lateral end of the lateral condyle's groove	P
fMPP	Medial posterior point	Most posterior point on the medial condyle	P
fLPP	Lateral posterior point	Most posterior point on the lateral condyle	P
fMDP	Medial distal point	Most distal point on the medial condyle	P
fLDP	Lateral distal point	Most distal point on the lateral condyle	P
fMPC	Medial posterior condyle	The spherical part of the medial condyle's posterior articular surface	FA
fLPC	Lateral posterior condyle	The spherical part of the lateral condyle's posterior articular surface	FA
		Computed landmarks	
fTGc	Trochlear groove center	Center of a circle fitted to the trochlear groove point array (fTG)	P
fMPCc	Medial posterior condyle center	Center of a sphere fitted to the medial posterior condyle face array's vertices (fMPC)	P
fLPCc	Lateral posterior condyle center	Center of a sphere fitted to the lateral posterior condyle face array's vertices (fLPC)	P

2.3. Alignment

One observer identified the femoral and tibial anatomical landmarks on all n_s (20) meshes prior to analysis. The computed landmarks were calculated and included subsequently, replacing their respective point or face arrays as part of the observation. This formed $\mathbf{v}_c$, a set of n_s observations each containing a homologous list of landmark coordinates. Generalised Procrustes analysis was then used to remove variation due to rotation and translation among $\mathbf{v}_c$. During the analysis, the observation with median scale was selected as the initial reference. The least squares ordinary Procrustes problem was then solved by aligning all $\mathbf{v}_c$ to the initial reference to produce a set of n_s aligned point sets, $\mathbf{v}_a$. The Procrustes mean $\mathbf{v}_m$ could then be computed by averaging the coordinates of the individual $\mathbf{v}_a$. The process was repeated iteratively until convergence occurred at an average root mean squared distance of 3.68 mm between the aligned observations $\mathbf{v}_a$ and resulting mean $\mathbf{v}_m$. The final Procrustes mean $\mathbf{v}_m$ served as the mean reference observation for subsequent analyses investigating landmark variation about it (Rohlf and Slice, 1990).

In order to describe landmark variation in terms of anatomically meaningful directions, $\mathbf{v}_m$ was aligned to a coordinate system based on that of Grood and Suntay (1983). To estimate the mean anatomic axis associated with $\mathbf{v}_m$, the anatomic axis of each individual subject was first estimated as the principal axis of a single sheet hyperboloid fitted to mesh vertices along the shaft of the bone. The shaft was isolated beyond 60% of the largest cross sectional area (Andrews and Séquin, 2014). The mean anatomic axis associated with $\mathbf{v}_m$ was then taken as the average of that of the individual subjects aligned into the same space as their respective $\mathbf{v}_a$ point sets.

The medio-lateral x-axis of the mean femoral reference observation was defined through the landmarks contained within $\mathbf{v}_m$ that correspond to the centers of the posterior condyle spheres (Kurosawa et al., 1985). Afterwards, the coronal plane was constrained parallel to the mean anatomic axis and contained the x-axis and its orthogonal, the cranio-caudal z-axis. Finally, the y-axis was orthogonal to both the x- and z-axes and the origin was placed at the intercondylar notch apex of $\mathbf{v}_m$. The origin of the mean tibial reference observation was at the landmark from $\mathbf{v}_m$ corresponding to the spine center, with the z-axis parallel to its anatomic axis. The coronal plane was then set parallel to the z-axis as well as a line passing through the landmarks contained within $\mathbf{v}_m$ that correspond to the medial and lateral tibial condyle centers. The x-axis was contained within the coronal plane and orthogonal to the z-axis, while the y-axis was again orthogonal to both. For a right knee, lateral was taken as positive along the x-axis, cranial along the z-axis and anterior along the y-axis.

The coordinate system described here was derived only once, directly from the homologous landmark coordinates of the mean reference observation $\mathbf{v}_m$. In contrast, additional coordinate systems were not defined on the individual observations themselves. Instead, in order to maintain scatter about the mean, each $\mathbf{v}_a$ was transformed into the same coordinate space as the reference observation $\mathbf{v}_m$. Furthermore, the transformation required to transform each initial observation $\mathbf{v}_c$ into its final aligned observation $\mathbf{v}_a$ was recorded. This transformation was then applied directly to

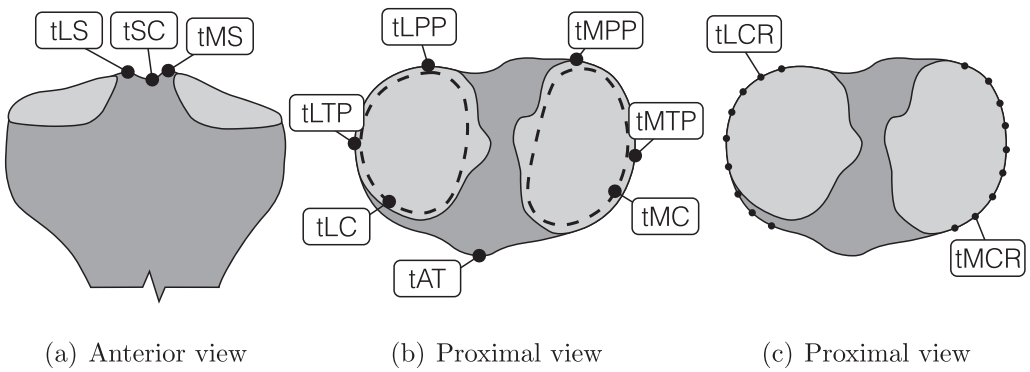


Fig. 2. Locations of the tibial landmarks.

Table 2
Tibial landmarks.

ID	Landmark	Description	Type
		Anatomic landmarks	
tMCR	Medial cortical rim	9 equidistant points placed along the medial half of the medial plateau's cortical rim	PA
tLCR	Lateral cortical rim	9 equidistant points placed along the lateral half of the lateral plateau's cortical rim	PA
tMC	Medial condyle	The articular surface of the medial plateau	FA
tLC	Lateral condyle	The articular surface of the lateral plateau	FA
tMS	Medial spine	Most proximal point on the medial spine	P
tLS	Lateral spine	Most proximal point on the lateral spine	P
tSC	Spine center	Center point of the saddle region between the spines	P
tMTP	Medial transverse point	Most medial point on the medial cortical rim	P
tLTP	Lateral transverse point	Most lateral point on the lateral cortical rim	P
tMPP	Medial posterior point	Most posterior point on the medial cortical rim	P
tLPP	Lateral posterior point	Most posterior point on the lateral cortical rim	P
tAT	Anterior tubercle	Most prominent point on the tibial tubercle	P
		Computed landmarks	
tMCRc	Medial cortical rim center	Center of a circle fitted to the medial cortical rim point array (tMCR)	P
tLCRc	Lateral cortical rim center	Center of a circle fitted to the lateral cortical rim point array (tLCR)	P
tMCc	Medial condyle center	Geometric mean of the medial articular surface face array's vertices (tMC)	P
tLCc	Lateral condyle center	Geometric mean of the lateral articular surface face array's vertices (tLC)	P

subsequent observations made on the same subject in order to maintain inter- and intra-observer variation.

2.4. Analysis

The standard deviation σ of the distance from each Procrustes mean to its 20 individual landmark observations was computed to investigate inter-specimen variability. Furthermore, principal component analysis was performed to find the set of orthogonal basis vectors that best describe the variation of each landmark cluster about its mean. Confidence ellipsoids based on the 95% χ^2 distribution along each principal axis were then determined.

The repeatability and reproducibility of a random sample of observers was investigated by having three observers each repeat three sets of observations on a subset of eight femurs and tibias. Two-way random effects Analysis of Variance (ANOVA) was applied to each of the aligned landmarks' coordinates along the x-, y- and z-axes as well as the Euclidean distance about the mean $\|\mathbb{R}^3\|$. The ANOVA allowed estimation of the inter-observer and intra-observer variances and hence, their standard deviations, σ_o and σ_w, to serve as measures of agreement (Bartlett and Frost, 2008). In addition, inter- and intra-observer reliability for the distance from the mean was defined as the Intraclass Correlation Coefficients ICC_o and ICC_w:

$$ICC_o = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2 + \sigma_w^2} \quad (1)$$

$$ICC_w = \frac{\sigma_b^2 + \sigma_o^2}{\sigma_b^2 + \sigma_o^2 + \sigma_w^2} \quad (2)$$

where σ_b was the inter-subject standard deviation also derived from ANOVA. Negative estimates were treated as evidence that the true value of the variance was zero (Searle et al., 2009).

3. Results

The inter-subject landmark scatter about the Procrustes mean is shown in Fig. 3. The femoral landmarks which exhibited the least inter-subject variation were the medial, lateral and trochlear groove centers as well as the intercondylar notch apex. The remaining landmarks all showed greater variation, with the medial terminal sulcus the most. For the tibia, the medial and lateral spines as well as the spine center had the least variation.

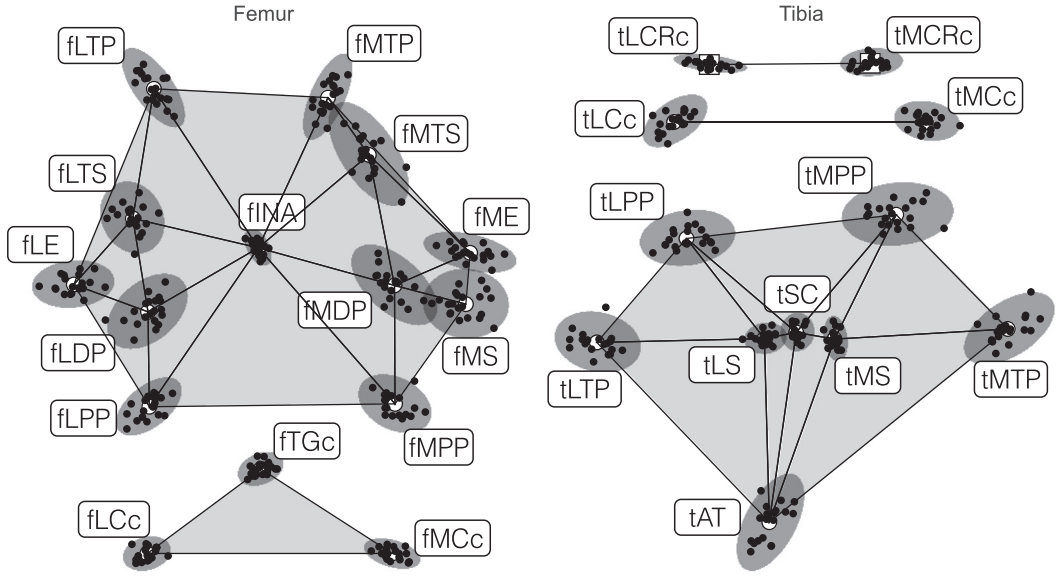


Fig. 3. Inter-subject landmark scatter about the Procrustes mean superimposed upon wire-frames of the femoral and tibial landmarks. The femur is viewed distally, while the tibia is shown proximally. The computed landmarks were offset posteriorly for the purposes of illustration. The mean landmarks are indicated by white markers, the individual observations are black, and the gray ellipsoids are the confidence intervals about the means.

Table 3
Landmark variation.

	Inter-specimen σ [mm]				Inter-observer σ_o [mm], ICC_o					Intra-observer σ_w [mm], ICC_w				
ID	x	y	z	$\|\|\mathbb{R}^3\|\|$	x	y	z	$\|\|\mathbb{R}^3\|\|$	r	x	y	z	$\|\|\mathbb{R}^3\|\|$	r
fME	3.39	1.49	2.11	4.26	0.18	2.20	5.10	1.91	0.58	0.36	1.80	2.17	0.99	0.91
fMS	3.14	2.54	2.10	4.55	0.34	2.55	1.44	0.00	0.90	0.42	3.28	2.62	0.92	0.90
fLE	3.01	1.74	1.91	3.96	0.28	0.67	2.21	0.87	0.79	0.47	1.05	2.64	1.23	0.86
fMTP	1.81	3.14	1.50	3.92	0.00	0.00	0.52	0.18	0.98	0.70	0.15	1.14	0.39	0.98
fLTP	2.31	2.90	1.53	4.01	0.00	0.05	1.26	0.54	0.94	0.51	0.18	1.09	0.52	0.97
fINA	1.00	1.36	1.42	2.20	0.03	0.18	0.00	0.20	0.50	0.60	0.85	0.71	0.65	0.54
fMTS	2.93	3.78	2.15	5.24	3.12	4.03	2.70	0.75	0.76	1.77	2.41	1.70	0.88	0.86
fLTS	2.35	2.63	1.42	3.80	0.25	1.10	0.00	0.00	0.90	0.64	1.51	0.60	0.67	0.90
fMPP	2.67	2.14	1.37	3.68	0.48	0.01	0.53	0.38	0.85	1.17	0.10	1.14	0.75	0.88
fLPP	2.40	2.12	1.61	3.59	1.26	0.12	0.53	0.87	0.86	1.03	0.10	1.32	0.62	0.95
fMDP	3.14	2.94	1.65	4.60	0.00	1.56	0.20	0.00	0.80	1.25	2.07	0.28	1.27	0.80
fLDP	2.94	2.74	1.22	4.20	0.83	3.10	0.28	1.42	0.42	5.69	2.41	0.40	1.70	0.66
fTGc	1.67	1.24	1.98	2.87	1.17	0.40	0.68	0.69	0.58	2.46	0.99	1.08	0.92	0.73
fMCc	2.05	1.11	0.63	2.42	0.00	0.48	0.07	0.18	0.97	0.16	0.41	0.09	0.19	0.98
fLCc	1.75	1.26	0.70	2.27	0.08	0.16	0.00	0.16	0.96	0.26	0.40	0.18	0.24	0.97
tMS	0.77	1.45	1.27	2.08	0.10	0.00	0.02	0.08	0.85	0.30	0.79	0.06	0.34	0.85
tLS	1.35	1.01	1.14	2.03	0.20	0.14	0.02	0.14	0.79	0.54	0.57	0.10	0.47	0.81
tSC	1.09	1.22	0.88	1.86	0.48	1.00	0.34	0.73	0.07	0.61	1.27	0.36	0.78	0.51
tMTP	2.88	2.30	1.00	3.83	0.36	0.60	1.28	0.64	0.88	0.38	2.69	1.26	0.65	0.94
tLTP	2.74	1.95	1.07	3.54	0.42	1.05	0.70	0.66	0.93	0.34	2.40	0.53	0.41	0.98
tMPP	3.56	2.02	1.06	4.23	0.69	1.09	1.31	0.84	0.75	2.37	0.68	0.89	1.42	0.81
tLPP	3.04	1.79	1.12	3.70	0.37	1.10	1.87	0.94	0.78	1.44	0.51	1.06	1.13	0.87
tAT	2.16	3.09	3.67	5.26	0.84	0.00	1.45	1.19	0.79	1.48	0.19	2.26	1.67	0.86
tMCRc	2.00	0.90	0.71	2.31	0.22	0.05	1.15	0.23	0.89	0.91	0.32	0.89	0.73	0.90
tLCRc	2.35	0.60	1.42	2.81	0.69	0.00	2.27	0.76	0.80	1.00	0.51	1.55	0.88	0.89
tMCc	2.10	1.28	0.63	2.54	0.76	0.54	0.15	0.75	0.42	0.87	1.24	0.21	0.91	0.65
tLCc	2.05	1.68	0.89	2.80	1.63	2.71	0.39	1.58	0.46	0.87	1.18	0.25	0.86	0.88

The anterior tubercle along with any landmarks defined on the cortical rim of the tibia varied the most.

Table 3 lists the inter-specimen variation as well as the inter and intra-observer agreement and reliability along each of the anatomic axes as well as the Euclidean distance $\|\|\mathbb{R}^3\|\|$ from the mean. The inter-subject deviation quantifies the results from Fig. 3. In the case of the femur, inter-observer agreement for the computed landmarks ranged between 0 and 1.17 mm, and intra-observer agreement between 0.09 and 2.46 mm. The highest inter- and intra-observer agreements were observed on anatomic landmarks, the medial epicondyle in the z direction and the lateral distal point along the x-axis at 5.1 and 5.69 mm respectively. Inter-observer reliability ranged between 0.42 and 0.98 and intra-observer reliability between 0.54 and 0.98. The tibial landmarks exhibited inter-observer agreement between 0 and 2.71 mm and reliability between 0.07 and 0.93. Intra-observer agreement ranged from 0.06 to 2.69 mm and reliability from 0.51 to 0.98.

4. Discussion

Variability in coordinate systems and the landmarks upon which they are based may have a detrimental effect on downstream measurements (Kedgley et al., 2015; Osis et al., 2015). Therefore, in order to investigate inter-specimen variability and inter- and intra-observer reliability and agreement, we instead relied on generalised Procrustes analysis, a standard morphometric approach to align the measured coordinates to a mean reference (Cavaignac et al., 2016; Harmon, 2007; Nicholson and Harvati, 2006). A potential drawback of the aforementioned approach was that large local variations could spread to other measurements through distance based least squares alignment (Fruciano, 2016). Resistant fit Procrustes analysis could mitigate this effect if most of the variation occurs in less than half of the landmarks (Rohlf and Slice, 1990). However, based on the available literature (Esfandiarpour et al., 2009; Victor et al., 2009) and our own results, it is apparent that only a few landmarks on the knee vary very little, while the remainder does not exhibit excessive variation. We therefore conclude that generalised Procrustes analysis remains the more suitable approach.

Victor et al. (2009) studied inter- and intra-observer agreement and reliability for a large set of landmarks defined on six CT knee reconstructions. They report mean intra-observer variability ranging from 0.4 to 1.4 mm, with their condyle centers also showing relatively low deviation at between 0.2 and 0.5 mm. Their inter-observer agreement ranges from 0.3 mm to 3.5 mm, with the joint centers' deviation again lower at between 0.3 and 0.6 mm. Intra-class ICC values were derived from the total variance of measurements by different observers and the variance between subjects only, ranging between 0.986 and 1. Similarly, Esfandiarpour et al. (2009) observed inter- and intra- observer ICCs above 0.97, with standard errors of measurement ranging between 1.03 and 4.74 mm for the femoral epicondyles identified on 16 MRI reconstructions. In comparison, while we obtain similar orders of magnitude and trends for agreement, our reliability measures contain fewer excellent correlations between measurements. We speculate that differences in results may be attributed to different methods, with ours perhaps more conservative in nature.

Many joint-related studies require an embedded coordinate system with clear definitions of the origin and various axes. In such cases we recommend the use of landmarks that deviate the least between specimens in order to avoid affecting the variance of downstream measurements. However, many landmarks exhibit the greatest variation in one or perhaps two principal directions, and very few have good all round inter-specimen agreement. A compromise would therefore be to avoid having the selected landmarks' deviation occur in at least the most *important* directions. To illustrate, consider trying to fit a medio-lateral axis to the femur. The medial and lateral epicondyles are often used for this purpose, and from Fig. 3 and Table 3 we see that they vary least along the

directions orthogonal to the desired axis. These are our example's *important* directions as any variation in the x-direction would not affect rotational alignment of the axis, but deviation in the y and z directions would.

In fact, we suggest that the best candidates for medio-lateral axis definition are the posterior condyle centers of the femur as well as the cortical rim centers of the tibia. Not only are these landmarks' inter-specimen variation concentrated along the x direction, but they also have good inter- and intra-observer agreement and reliability. We speculate that this may be due to the global 'smoothing' effect during least squares approximation of spheres and circles to otherwise locally complex geometries. As for coordinate systems' origins, the femoral intercondylar notch apex and the tibial spine center are good candidates. Even though they exhibit poor agreement and reliability, their inter-specimen variation is small compared to the alternatives. Therefore, their effect on downstream measurements' variation may not be so severe.

Once a suitable coordinate system which is robust against inter-specimen variation has been defined, subsequent measurements should prioritise observer agreement and reliability. By way of example, the femoral terminal sulci are good candidates for defining the anterior boundaries of the condyles' articular surfaces. We *do* expect them to vary between targets, and could therefore use them to study the variation in overall shape of the condyles as part of a downstream process. However, the question we should concern ourselves with is whether or not they are repeatable and reliable enough should we or someone else wish to duplicate the work, which they appear to be.

Selection of the imaging modality from which 3D models may be reconstructed is usually driven by considerations such as whether or not soft tissue must be visible, and not necessarily the influence that it may have on landmark variation. However, it has been shown that, though segmentations from CT and MRI scans both exhibit global reconstruction errors of less than 0.6 mm, the prior overestimates bony geometry while the latter underestimates it (Van den Broeck et al., 2014). While this motivates the investigation of modality-specific landmark variability as we did here for MRI, the present study did not attempt to quantify this effect. Furthermore, the MRI scans used in this study had varying in-plane pixel resolutions, which may have influenced the results from the segmentation procedure and therefore subsequent investigations of variability. Nevertheless, each individual scan was of diagnostic quality and, we therefore argue, representative of what might be encountered in practice. However, as we had attempted to remove the influence of image segmentation on observer agreement and reliability by having a single expert segment all scans prior to identifying landmarks, future work might benefit from including the effect of observer-specific segmentations as part of an investigation of landmark variability. Each new study would typically require segmentation of new data, and quantifying the influence of the segmentation processes or even different segmentation protocols could better inform landmark selection in a particular case.

5. Conclusion

MRI derived surface mesh reconstructions of bones are often required to inform landmark-based in silico processes when studying the knee. We used generalised Procrustes analysis to evaluate the inter-subject variation as well as inter- and intra-observer agreement and reliability of landmark coordinates defined on MRI models of the distal femur and proximal tibia. We suggest that landmarks that deviate little between subjects may be suitable for coordinate system definition, minimising contamination of subsequent measurements' variation. Landmarks with good observer agreement and reliability should be preferred when investigating inter-specimen variability or for the definition of reliable features such as regions of interest. Our results showed that landmarks computed from spheres or circles fitted to the joint centers were robust against these criteria, while other landmarks could vary by more than 5 mm depending on the axis of measurement, suggesting careful consideration of their application in future studies.

Declaration of Competing Interest

The authors have no conflict of interest to declare.

Acknowledgements

This work was supported financially by the South African Medical Research Council. The sponsor had no role in the study design, collection, analysis and interpretation of the data, nor in the preparation of the manuscript or the decision to submit it for publication.

References

Andrews, J., Séquin, C.H., 2014. Type-constrained direct fitting of quadric surfaces. Comput.-Aided Des. Appl. 11 (1), 107–119.
Bartlett, J.W., Frost, C., 2008. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultras. Obstetr. Gynecol. 31 (4), 466–475.
Bowes, M.A., Vincent, G.R., Wolstenholme, C.B., Conaghan, P.G., 2015. A novel method for bone area measurement provides new insights into osteoarthritis and its progression. Ann. Rheum. Dis. 74 (3), 519–525.
Braun, H.J., Gold, G.E., 2012. Diagnosis of osteoarthritis: imaging. Bone 51 (2), 278–288.
Bredbenner, T.L., Eliason, T.D., Potter, R.S., Mason, R.L., Havill, L.M., Nicolella, D.P., 2010. Statistical shape modeling describes variation in tibia and femur surface geometry between control and incidence groups from the osteoarthritis initiative database. J. Biomech. 43 (9), 1780–1786.
Cavaignac, E., Savall, F., Faruch, M., Reina, N., Chiron, P., Telmon, N., 2016. Geometric morphometric analysis reveals sexual dimorphism in the distal femur. Foren. Sci. Int. 259, 246.e1–246.e5.
Choi, J.-A., Gold, G.E., 2011. MR Imaging of articular cartilage physiology. Magn. Reson. Imag. Clin. North Am. 19 (2), 249–282.
Cobb, J.P., Dixon, H., Dandachli, W., Iranpour, F., 2008. The anatomical tibial axis. J. Bone Joint Surg. 90 (8), 1032–1038. British volume.
DeFrate, L.E., Sun, H., Gill, T.J., Rubash, H.E., Li, G., 2004. In vivo tibiofemoral contact analysis using 3D MRI-based knee models. J. Biomech. 37 (10), 1499–1504.
Della Croce, U., Cappozzo, A., Kerrigan, D.C., 1999. Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles. Med. Biol. Eng. Comput. 37 (2), 155–161.
Eckstein, F., Kwoh, C.K., Link, T.M.OAI investigators, 2014. Imaging research results from the Osteoarthritis Initiative OAI: a review and lessons learned 10 years after start of enrolment. Ann. Rheum. Dis. 73 (7), 1289–1300.
Esfandiarpour, F., Olyaei, G., Shakouri Rad, A., Farahmand, F., Talebian, S., Makhsous, M., Parnianpour, M., 2009. Reliability of determination of bony landmarks of the distal femur on MR Images and MRI based 3D models. Iran. J. Radiol. 6 (4), 225–230.
Fruciano, C., 2016. Measurement error in geometric morphometrics. Devel. Gen. Evol. 226 (3), 139–158.
Frye, B.M., Najim, A.A., Adams, J.B., Berend, K.R., Lombardi, A.V., 2015. MRI is more accurate than CT for patient-specific total knee arthroplasty. Knee 22 (6), 609–612.
Grood, E.S., Suntay, W.J., 1983. A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. J. Biomech. Eng. 105 (2), 136–144.
Harmon, E.H., 2007. The shape of the hominoid proximal femur: a geometric morphometric analysis. J. Anat. 210 (2), 170–185.
Howell, S.M., Kuznik, K., Hull, M.L., Siston, R.A., 2008. Results of an initial experience with custom-fit positioning total knee arthroplasty in a series of 48 patients. Orthopedics 31 (9), 857–863.
Iranpour, F., Merican, A.M., Dandachli, W., Amis, A.A., Cobb, J.P., 2010. The geometry of the trochlear groove. Clin. Orthop. Relat. Res. 468 (3), 782–788.
Kedgley, A.E., McWalter, E.J., Wilson, D.R., 2015. The effect of coordinate system variation on in vivo patellofemoral kinematic measures. Knee 22 (2), 88–94.
Kurosawa, H., Walker, P.S., Abe, S., Garg, A., Hunter, T., 1985. Geometry and motion of the knee for implant and orthotic design. J. Biomech. 18 (7), 487–499.
Nicholson, E., Harvati, K., 2006. Quantitative analysis of human mandibular shape using three-dimensional geometric morphometrics. Am. J. Phys. Anthropol. 131 (3), 368–383.

Osis, S.T., Hettinga, B.A., Macdonald, S.L., Ferber, R., 2015. A novel method to evaluate error in anatomical marker placement using a modified generalized procrustes analysis. Comput. Meth. Biomech. Biomed. Eng. 18 (10), 1108–1116.
Pena, E., Calvo, B., Martinez, M., Doblare, M., 2006. A three-dimensional finite element analysis of the combined behavior of ligaments and menisci in the healthy human knee joint. J. Biomech. 39 (9), 1686–1701.
Rohlf, F.J., Slice, D., 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. Syst. Biol. 39 (1), 40–59.
Searle, S.R., Casella, G., McCulloch, C.E., 2009. Varian. Comp. John Wiley and Sons.
Van den Broeck, J., Vereecke, E., Wirix-Speetjens, R., Vander Sloten, J., 2014. Segmentation accuracy of long bones. Med. Eng. Phys. 36 (7), 949–953.
Van den Heever, D., Scheffer, C., Erasmus, P., Dillon, E., 2012. Development and testing of patient-specific knee replacements. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 4875–4878.
Van der Merwe, J., Scheffer, C., Van den Heever, D., Erasmus, P., 2013. Reverse engineering the human knee. In: International Conference on Competitive Manufacturing (COMA)'13.
Victor, J., 2009. Rotational alignment of the distal femur: a literature review. Orthop. Traumatol. Surg. Res. 95 (5), 365–372.
Victor, J., Van Doninck, D., Labey, L., Innocenti, B., Parizel, P., Bellemans, J., 2009. How precise can bony landmarks be determined on a CT scan of the knee? Knee 16 (5), 358–365.
Williams, T.G., Holmes, A.P., Waterton, J.C., Maciewicz, R.A., Nash, A.F., Taylor, C.J., 2006. Regional quantitative analysis of knee cartilage in a population study using MRI and model based correspondences. In: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006. IEEE, pp. 311–314.
Yue, B., Varadarajan, K.M., Moynihan, A.L., Liu, F., Rubash, H.E., Li, G., 2011. Kinematics of medial osteoarthritic knees before and after posterior cruciate ligament retaining total knee arthroplasty. J. Orthop. Res. 29 (1), 40–46.