



Reliability and concurrent validity of three-dimensional ultrasound for quantifying knee cartilage volume



S. Papernick^{a,b,*}, R. Dima^e, D.J. Gillies^{a,b}, C.T. Appleton^{c,d}, A. Fenster^{a,b}

^a Department of Medical Biophysics, Western University Canada, London, Ontario, Canada

^b Robarts Research Institute, Western University Canada, London, Ontario, Canada

^c Department of Medicine, Western University Canada, London, Ontario, Canada

^d Department of Physiology and Pharmacology, Western University Canada, London, Ontario, Canada

^e Department of Health and Rehabilitation Sciences; Western University Canada, London, Ontario, Canada

ARTICLE INFO

Keywords:

3D ultrasound
Knee cartilage volume
Osteoarthritis
Segmentation
MRI
Validation

ABSTRACT

Objective: The goal of this study was to test the reliability and validity of a handheld mechanical three-dimensional (3D) ultrasound (US) device for quantifying femoral articular cartilage (FAC) against the current clinical standard of magnetic resonance imaging (MRI).

Design: Bilateral knee images of 25 healthy volunteers were acquired with 3D US and 3.0 T MRI. The trochlear FAC was segmented by two raters who repeated segmentations on five cases during separate sessions. MRI and 3D US segmentations were registered using a semi-automated surface-based registration algorithm, and MRI segmentations were trimmed to match the FAC region from 3D US. Intra- (n = 5) and inter-rater (n = 25) reliabilities were assessed using intraclass correlation coefficients (ICCs) calculated from FAC volumes. Relationships between MRI and 3D US were assessed using Spearman correlation and linear regression (n = 25).

Results: MRI intra-rater ICCs were 0.97 (0.79, 1.00) and 0.90 (0.25, 0.99) for each rater with an inter-rater ICC of 0.83 (0.48, 0.94). 3D US intra-rater ICCs were 1.00 (0.98, 1.00) and 0.98 (0.84, 1.00) for each rater with an inter-rater ICC of 0.96 (0.90, 0.98). Spearman correlation and linear regression revealed a strong correlation $\rho = 0.884$ (0.746, 0.949) and regression $R^2 = 0.848$ (0.750, 0.950).

Conclusion: These results suggest 3D US demonstrates excellent intra- and inter-rater reliabilities and strong concurrent validity with MRI when quantifying healthy trochlear FAC volume. 3D US may reduce imaging costs and greatly improve feasibility of quantifying knee cartilage volume during knee arthritis clinical trials and patient care.

1. Introduction

Knee osteoarthritis (KOA) is a whole-joint disease with a prevalence of 7–17% among adults 45+ years old, and is increasing with rising obesity rates and population aging [1,2]. KOA affects all knee joint tissues, leading to cartilage degradation, subchondral bone remodeling, and muscle atrophy [3]. Cartilage degradation, a hallmark of KOA, has motivated efforts to characterize disease severity through measures of femoral articular cartilage (FAC) loss, where decreases in FAC quality and quantity are interpreted as increased KOA severity. Semi-quantitative scoring systems, such as the Kellgren-Lawrence (KL) grading scale, define the presence of KOA using tibiofemoral (TF) joint space narrowing (JSN) as a surrogate for FAC loss. Most imaging-based

KOA scales target TF cartilage because of easy visualization with weight-bearing radiography. Although radiographic JSN may represent FAC loss, radiographic grading has poor sensitivity to detect FAC changes in early stage KOA [4]. Furthermore, radiographic JSN suffers from limited reproducibility for visualizing three-dimensional (3D) features due to variations in knee joint angulation [5]. Additionally, JSN is a composite of meniscal positioning and degeneration, which are not necessarily associated with KOA severity [6,7].

Limitations of radiographic JSN have motivated magnetic resonance imaging (MRI) investigations of FAC as a discriminative and evaluative KOA tool. The MRI Osteoarthritis Knee Score (MOAKS), Boston-Leeds Osteoarthritis Knee Score (BLOKS), Knee Osteoarthritis Scoring System (KOSS), and Whole-Organ MRI Score (WORMS) are all MRI-based semi-

* Corresponding author. Robarts Research Institute, Western University Canada, 1151 Richmond St. N. London, Ontario, N6A 5B7, Canada.

E-mail addresses: spaperni@uwo.ca (S. Papernick), rdima@uwo.ca (R. Dima), dgillie@uwo.ca (D.J. Gillies), tom.appleton@sjhc.london.on.ca (C.T. Appleton), afenster@robarts.ca (A. Fenster).

<https://doi.org/10.1016/j.ocarto.2020.100127>

Received 15 July 2020; Accepted 18 November 2020

2665-9131/© 2020 The Authors. Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International (OARSI). This is an open access article under the

CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

quantitative scales that have shown excellent reliabilities in OA populations [8–11]. Furthermore, compositional MRI techniques produce quantitative measurements of cartilage biochemistry and have primarily been developed to investigate early stage KOA. Due to the ability of MRI to assess the status of whole joint cartilage with reasonable spatial resolution, it has been largely accepted as the gold standard for KOA FAC assessments. While MRI has accelerated the scientific and medical communities' understanding of KOA, it has limitations. MRI is not feasible for point-of-care disease classification due to high manufacturing and operating costs, long acquisition times, and inaccessibility to all patients at all times [12]. However, while other modalities may be less expensive and more accessible than MRI, finding individuals that possess the expertise needed to interpret images in under-served areas of the world is challenging.

Conventional two-dimensional (2D) ultrasound (US) is widely accessible, relatively inexpensive, and overcomes the limitations associated with MRI. 2D US is a high-resolution imaging modality that has been increasingly used for point-of-care assessments of rheumatological diseases [13–16]. 2D US has been implemented in KOA research via Outcome Measures in Rheumatology (OMERACT) US working group's semi-quantitative grading scale [17]. However, this scale has not been formally validated, and conventional 2D US is associated with limitations. Clinicians must cognitively integrate multiple 2D images to mentally reconstruct 3D anatomy, which is inefficient and leads to operator variability [18]. Additionally, 2D US tissue volume calculations require measurements of height, width, and length in two orthogonal views and are associated with low accuracy, high variability, and large operator dependency. Furthermore, sensitivity to change is limited when using ordinal scales with small dynamic range such as 0–3 in the OMERACT scale. Alternatively, 3D US techniques involve translating a 2D US transducer while continually acquiring images that are reconstructed into a 3D image. 3D US imaging overcomes the limitations of 2D US and may fill the clinical need for an objective imaging-based point-of-care tool for assessing KOA status, progression, and response to treatment.

3D US techniques have been applied to neonatal, gynaecological, and vascular applications, among others [19–21]. We have developed a handheld mechanical 3D US device to provide point-of-care assessments of trochlear FAC (tFAC). The objectives of this cross-sectional study were to investigate the intra- and inter-rater reliabilities of our 3D US scanner for measuring tFAC volumes in healthy volunteers, and to assess its concurrent validity compared to the current clinical standard of MRI. We hypothesized that tFAC volumes measured from 3D US would demonstrate excellent reliability ($ICC > 0.90$) and be strongly correlated ($\rho > 0.80$) to MRI measurements in the same region-of-interest (ROI).

2. Methods

Twenty-five volunteers over the age of 18 without a recent history of chronic knee joint pathology (healthy knees) in the year prior to the study were recruited for MR and 3D US knee imaging. The imaging protocol was approved by the Research Ethics Board at Western University Canada and all volunteers provided written informed consent prior to imaging. Knees were deemed healthy if volunteers denied experiencing knee pain on most days of the weeks prior to this study and had not been diagnosed with any type of knee arthritis. Volunteers with prior knee injuries and/or surgeries that occurred before the year leading up to the study were not excluded from the cohort if they denied experiencing frequent knee symptoms including pain, aching, or stiffness on most days of the weeks prior to this study.

2.1. Image acquisition

MRI scans were acquired on a 3.0 T MR system (General Electric Healthcare, Milwaukee, WI, USA) using a 3D Multiple Echo Recombined

Gradient Echo (MERGE) sequence in accordance with the Osteoarthritis Research Society International (OARSI) recommendations for KOA imaging clinical trials [22]. An HD T/R Knee Array Coil (8 Channels) was used while volunteers were positioned supine with minimal knee flexion. Images were acquired in the sagittal plane with voxel sizes of $0.63 \times 0.63 \times 0.40 \text{ mm}^3$, an average of 250 slices, a reconstructed matrix size of 256 by 256 voxels, and a field-of-view of 16 cm. The excitation flip angle was 5° with a repetition time (TR) of 30 ms and an echo time (TE) of 11.71 ms. The MERGE sequence scan time for one knee was 4 min and 27 s. Total scan time was 45 min including both knees.

3D US images were acquired using an Aplio i800 US machine (Canon Medical Systems Corporation, Otawara, Tochigi, Japan) equipped with a 14L5 linear transducer with a 58 mm footprint length and an operating frequency of 10 MHz (3.8 MHz–10.0 MHz). The 2D US transducer was mounted to our 3D US scanner using a custom 3D-printed mold (Fig. 1). Our 3D US device consisted of a motorized drive mechanism that linearly translated the transducer over 4.0 cm along the patient's skin. 2D US images were continually acquired at regular spatial intervals which were reconstructed into a 3D image immediately after scanning via computer software [18]. Our 3D US scanner has previously been validated on tungsten filament phantoms and volumetric agar phantoms, demonstrating the ability to acquire Euclidean distance and volumetric measurements with errors $< 2\%$ [23]. For 3D US acquisition, volunteers were positioned supine and instructed to flex their knee to the maximum range of motion without eliciting pain. 3D US images of the tFAC were acquired at the distal end of the femur, proximal to the patella during maximum knee flexion. 120 2D US images were acquired in the transverse plane with transducer translation along the perpendicular axis. Reconstructed 3D US image voxel sizes were $0.058 \times 0.058 \times 0.33 \text{ mm}^3$ with 2D US in-plane image dimensions of 968×694 voxels. 3D US acquisition time was 15 s for one knee.

2.2. Manual segmentation

MRI voxel resampling was performed to ensure that the segmentation pixel spacings were substantially smaller than the smallest FAC image feature. Voxel resampling was conducted in MATLAB R2019b (MathWorks, Natick, Massachusetts, USA) using the *interp2* function with the spline interpolation method. The resampled voxel size was $0.15 \times 0.15 \times 0.40 \text{ mm}^3$ to provide a balance between segmentation sensitivity and computation time.

Manual tFAC segmentations were completed by two raters (SP, RD) on MRI and 3D US after receiving training during three formal calibration sessions with a rheumatologist possessing advanced diagnostic and interventional musculoskeletal ultrasonography training (CTA). One rater had no prior experience with medical image segmentation but possesses a medical physics academic background with courses in medical imaging modalities including US and MRI. The other rater is a registered diagnostic medical sonographer with formal training and clinical experience in medical imaging. Segmentations were performed in the open-source software 3D Slicer (3D Slicer 4.11.0 Preview Release) using the segment editor module and were conducted in the sagittal MRI and transverse 3D US planes [24]. Segmentations were completed using every second 2D image to decrease segmentation time for both MRI and 3D US without a reduction in sensitivity to tFAC volume changes [25]. Segmented 2D images were interpolated using a morphological contour interpolation algorithm in 3D Slicer, resulting in an average of 146 and 92 segmented 2D images per MRI and 3D US image, respectively [26].

During MRI segmentations, the posterior condylar cartilage was excluded by defining the anterior border of the posterior aspect of the lateral and medial menisci as a segmentation border to further reduce segmentation times (Fig. 2a and b). The hyperintense synovial membrane lining Hoffa's fat pad was excluded from MRI segmentations. For 3D US segmentations, the anterior hyperechoic tFAC surface and the hyperechoic border of the cortex were defined as boundaries for the anechoic

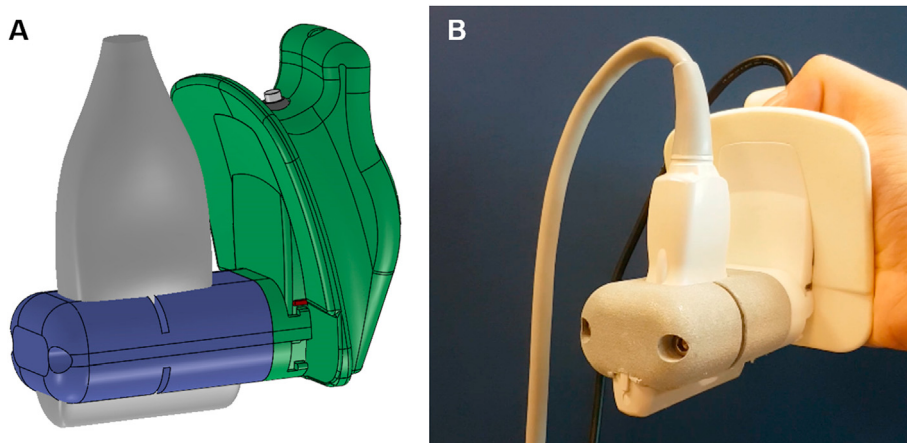


Fig. 1. (A) Schematic diagram of our handheld mechanical 3D US acquisition device. The conventional US transducer (gray) is mounted to a motorized drive mechanism (green) via a custom 3D-printed transducer mold (purple). Pressing the button located on the top of the device initiates a 3D US acquisition. (B) Image of the 3D US acquisition device in the hand of a user.

cartilage (Fig. 2c and d). With these boundaries and definitions, total segmentation times were approximately 45–60 min per knee for MRI and 20–30 min per knee for 3D US. Five knees from separate volunteers were randomly selected by each rater and re-segmented on MRI and 3D US. Repeated segmentations were conducted during sessions separated by a two-week washout period.

2.3. Reliability and validation analysis

MRI and 3D US segmentations were registered via manual initialization followed by automated surface-based registration in 3D Slicer (Fig. 3). Initialization involved manipulating 3D US tFAC models using linear transformations and rotations along the three Cartesian axes to

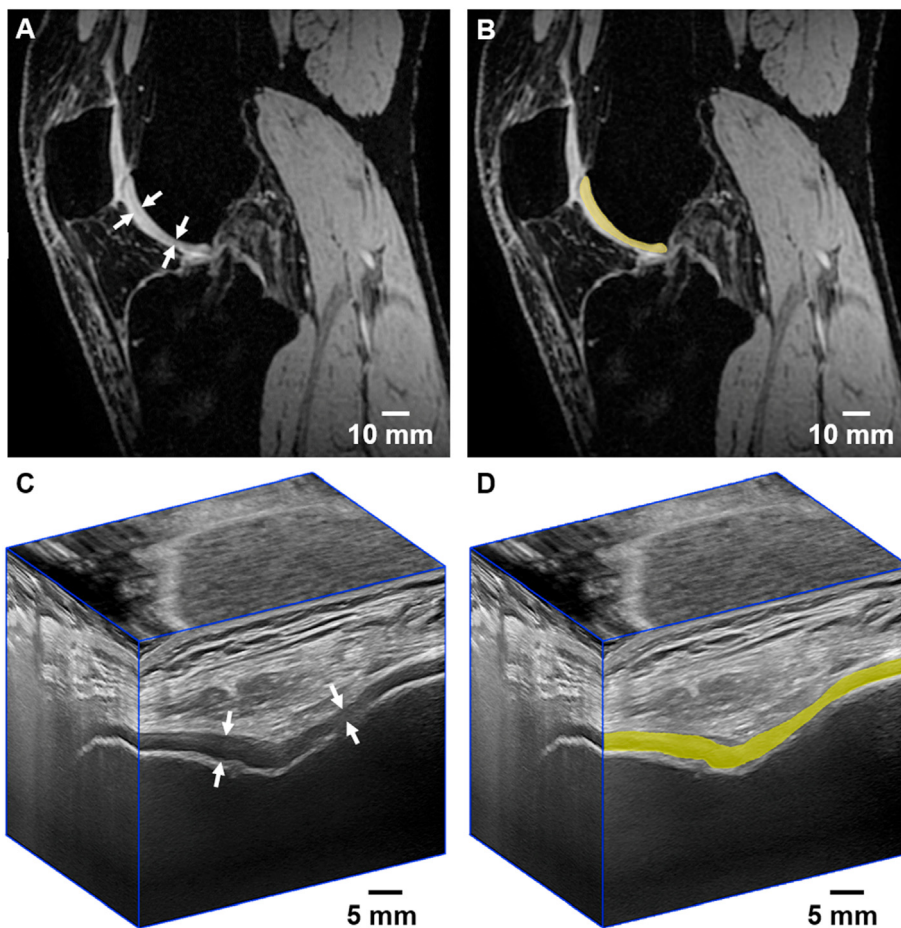


Fig. 2. MERGE MRI (A) and 3D US (C) images of the trochlear articular knee cartilage outlined by the white arrows in the sagittal MRI and transverse US planes of a healthy volunteer, accompanied by the same images with an overlaid MRI (B) and 3D US (D) slice that has been manually segmented.

align the segmentations with MRI using the intercondylar notch as an anatomical landmark. An automatic surface-based registration method (Jean-Baptiste & Vinicius Boen, University of Michigan) was applied to the segmentations to complete the registration. Intra- and inter-rater reliabilities were assessed using the same registration procedures. Reliability analysis was conducted using the entire segmented area of MRI and 3D US tFAC models, while validation between modalities involved additional trimming of MRI segmentations. MR images captured a larger FAC field-of-view than 3D US, resulting in segmentations that did not represent identical anatomical ROI when comparing modalities. Therefore, MRI segmentations were manually trimmed using the overlaid 3D US segmentations as guides, ensuring that tFAC models represented the same ROI on both modalities. Registration and trimming were repeated on five knees selected at random during sessions separated by a two-week washout period.

Segmentation volumes were computed by 3D Slicer and the percent differences between MRI and 3D US volumes were calculated. The mean surface distance (MSD), Hausdorff distance (HD), and Dice similarity coefficient (DSC) were computed as these metrics are widely used to compare and evaluate segmentations [27]. MSD represents the

mean distance from a point on one surface to the nearest corresponding point on the other surface, while HD is the largest distance from a point on one surface to the closest point on the other surface (Fig. 3e). DSC provides a measure of similarity in terms of overlap between segmentations and ranges from 0% (no overlap) to 100% (identical objects). MSD and HD values were computed using the open-source software CloudCompare (CloudCompare v2.11 beta), and DSC values were computed using the segment comparison module in 3D Slicer (Csaba Pinter, PerLab, Queen's University, Canada).

2.4. Statistical analysis

Statistical analyses were performed using SPSS (SPSS Statistics v26; IBM, Armonk, NJ). All data were initially tested for normality using the Shapiro-Wilk test. Intra- and inter-rater segmentation reliabilities from MRI and 3D US for both raters were assessed using intraclass correlation coefficients (ICCs). Intra-rater ICCs were based on a single-rating, absolute-agreement, 2-way mixed-effects model, while inter-rater ICCs were based on a single-rating, absolute-agreement, 2-way random-effects model. ICCs were interpreted as less than 0.50 indicating poor

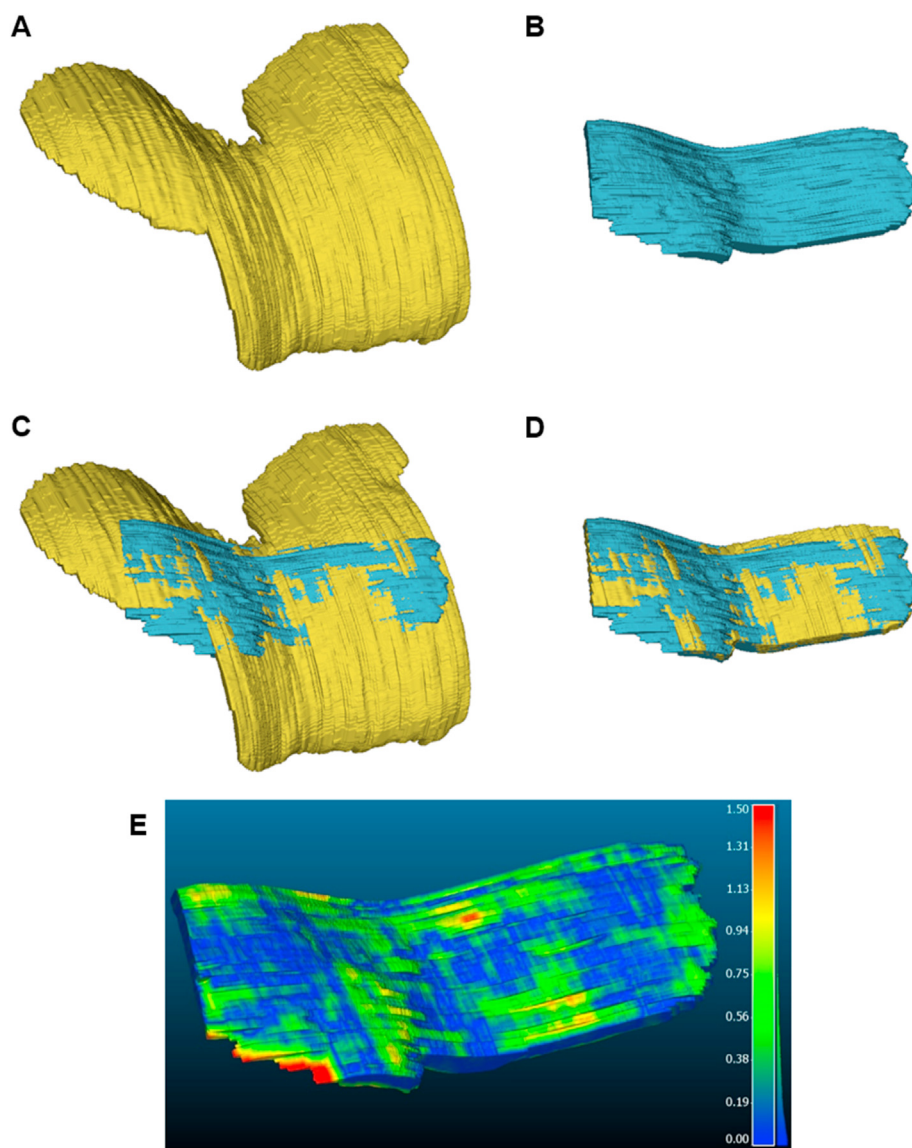


Fig. 3. Manual segmentations of the FAC from (A) MRI and (B) 3D US images. 3D US segmentations were registered to MRI using a semi-automated surface-based registration algorithm (C). MRI segmentations were then trimmed (D) to ensure both MRI and 3D US segmentations covered the same cartilage region for comparison purposes. (E) Colour map representing the absolute distance (mm) between a given MRI and 3D US segmentation pair from the same knee of a volunteer. The distance map has been overlaid on the 3D US segmentation and represents the distance from each point to the nearest points on the MRI segmentation.

reliability, between 0.50 and 0.75 indicating moderate reliability, between 0.75 and 0.90 indicating good reliability, and greater than 0.90 indicating excellent reliability [28]. Bland-Altman plots were used to assess differences between intra- and inter-rater tFAC volumes along with differences between MRI and 3D US segmentations. A cumulative percentile plot was used to observe the relationship of the differences between MRI and 3D US tFAC volumes. Correlations between tFAC volumes calculated as the mean of the two raters from MRI and 3D US segmentations were determined using Spearman Rank-Order Correlation due to the non-normal distribution of data. Linear regression analysis was conducted using MRI segmentation volumes as predictors for 3D US tFAC volumes and the enter method for equation construction.

3. Results

The demographic data of the volunteers is shown in Table 1 and was available from 24 of the 25 participants.

3.1. Reliability

Similar mean segmentation volumes and mean absolute volume differences between intra- and inter-rater comparisons were observed using the same modality for each rater (Table 2, Fig. 4). The smallest ICC was 0.83 (0.48, 0.94) and was observed for the inter-rater comparison of MRI, while the largest ICC was 1.00 (0.98, 1.00) and was observed for the intra-rater 3D US comparison for rater 1 (Table 3). Global mean MSD and HD were smaller for 3D US than MRI for intra- and inter-rater comparisons, while DSC was larger for 3D US than MRI during all comparisons (Table 3).

3.2. 3D US to MRI registration and trimming

The mean percent difference between MRI and 3D US volumes averaged across all comparisons including both raters individually, following registration and trimming, was $16.7 \pm 12.9\%$ ($n = 50$). 3D US tFAC volume measurements were larger than MRI volume measurements in 88% of the comparisons between the two modalities (Fig. 5a and b). Spearman Rank-Order Correlation revealed a strong correlation between MRI and 3D US volumes ($\rho = 0.884$ (0.746, 0.949), $p < 0.0001$), and linear regression resulted in $R^2 = 0.848$ (0.750, 0.950), $p < 0.0001$, and $Y = 1.29 \cdot X - 230$ (Fig. 5c). Global mean MSD, HD, and DSC between registered segmentations averaged between both raters were 0.375 ± 0.071 mm, 2.85 ± 1.18 mm, and $71.2 \pm 6.5\%$, respectively ($n = 25$).

4. Discussion

This is the first study investigating the reliability and validation of FAC volume measurements using 3D US in healthy volunteers. This study focused on validation of tFAC volumes, which is important when studying the status and progression of KOA affecting patellofemoral articulation. Since KOA affects the entire joint, these results are pertinent to the study of nearly all KOA phenotypes. Healthy FAC possesses a relatively smooth and continuous surface without distinct anatomical landmarks that can be used for registering segmentations, besides the

Table 1

Demographic data of twenty-four out of the twenty-five volunteers with healthy knees.

	Volunteers with healthy knees
% Women	58.3
Age [year] (mean \pm SD)	29.9 \pm 14.5
Height [m] (mean \pm SD)	1.68 \pm 0.11
Weight [kg] (mean \pm SD)	67.0 \pm 14.8
BMI [kg/m^2] (mean \pm SD)	23.4 \pm 3.3

Table 2

Mean volumes \pm standard deviations (SDs) for all intra-rater and inter-rater comparisons, along with the absolute volume difference \pm SD between MRI and 3D US. The mean volumes and absolute differences for repeated registrations and trimmings of MRI segmentations are also provided.

	Mean Volume [cm^3]	Mean Volume (repeated) [cm^3]	Absolute Difference [cm^3]
<i>Intra-rater</i>			
<i>(n = 5)</i>			
MRI (rater 1)	4.71 \pm 1.18	4.76 \pm 1.20	0.232 \pm 0.152
MRI (rater 2)	4.56 \pm 1.10	4.20 \pm 1.04	0.366 \pm 0.351
3D US (rater 1)	2.52 \pm 1.01	2.53 \pm 0.96	0.0516 \pm 0.0531
3D US (rater 2)	2.15 \pm 0.92	2.17 \pm 1.08	0.167 \pm 0.111
<i>Inter-rater</i>			
<i>(n = 25)</i>			
MRI	4.79 \pm 1.23	4.38 \pm 1.03	0.494 \pm 0.465
3D US	2.29 \pm 0.72	2.30 \pm 0.64	0.155 \pm 0.134
<i>Registration & trimming</i>			
<i>(n = 5)</i>			
Single rater	2.14 \pm 0.56	2.13 \pm 0.54	0.0173 \pm 0.0166

intercondylar notch. Therefore, tFAC images that included the intercondylar notch enabled registration of MRI and 3D US segmentations. Additionally, the intercondylar notch can be used as an anatomical landmark during longitudinal studies to ensure repeated measures are taken from the same ROI.

3D US imaging is possible in any application involving 2D US since the only modification required is mounting the 2D US transducer to a 3D US scanning device. Several studies have previously investigated the application of 2D US for evaluating femoral condylar cartilage for KOA assessments [29–32]. However, quantitative image analysis of non-invasive knee US has only been reported for cartilage thickness measurements but not entire cartilage volumes [33–35]. Quantitative image analysis may provide more sensitive information regarding early KOA than semi-quantitative grading scales, which are subjective and potentially susceptible to US operator/rater differences. However, semi-quantitative grading scales are potentially faster than manual quantitative image analysis. Therefore, this study builds on previous work and is easily implemented in similar clinical settings.

Many studies have investigated cartilage thickness measurements for assessing KOA severity [36,37]. However, thickness measurements are highly variable and dependent on the FAC ROI being measured, which can vary within subjects due to US transducer placement and angulation at different time points [38–40]. Detecting changes in cartilage loss using thickness measurements requires the ability to sample the same ROI with good test-retest reliability. Volume measurements may overcome these limitations by enabling quantification of cartilage loss in all dimensions and provide a similar metric to average cartilage thickness. Furthermore, 3D US may provide meaningful advantages over MRI for quantifying FAC volume. Our 3D US device is compatible with any commercially available US machine and is associated with low manufacturing and operating costs. Additionally, the portability of our 3D US device enables FAC volume measurements to be acquired at the patient's bedside.

4.1. Reliability

Intra-rater ICCs for MRI and 3D US demonstrated excellent reliabilities (>0.90). Inter-rater ICC for MRI demonstrated good reliability (0.75–0.90) while 3D US ICC demonstrated excellent reliability (>0.90). Intra- and inter-rater Bland-Altman plots displayed smaller volume difference variations for 3D US compared to MRI in all comparisons (Fig. 4). Additionally, global mean MSD and HD were smaller for 3D US than MRI, and mean DSC for 3D US was higher than MRI for intra- and inter-rater

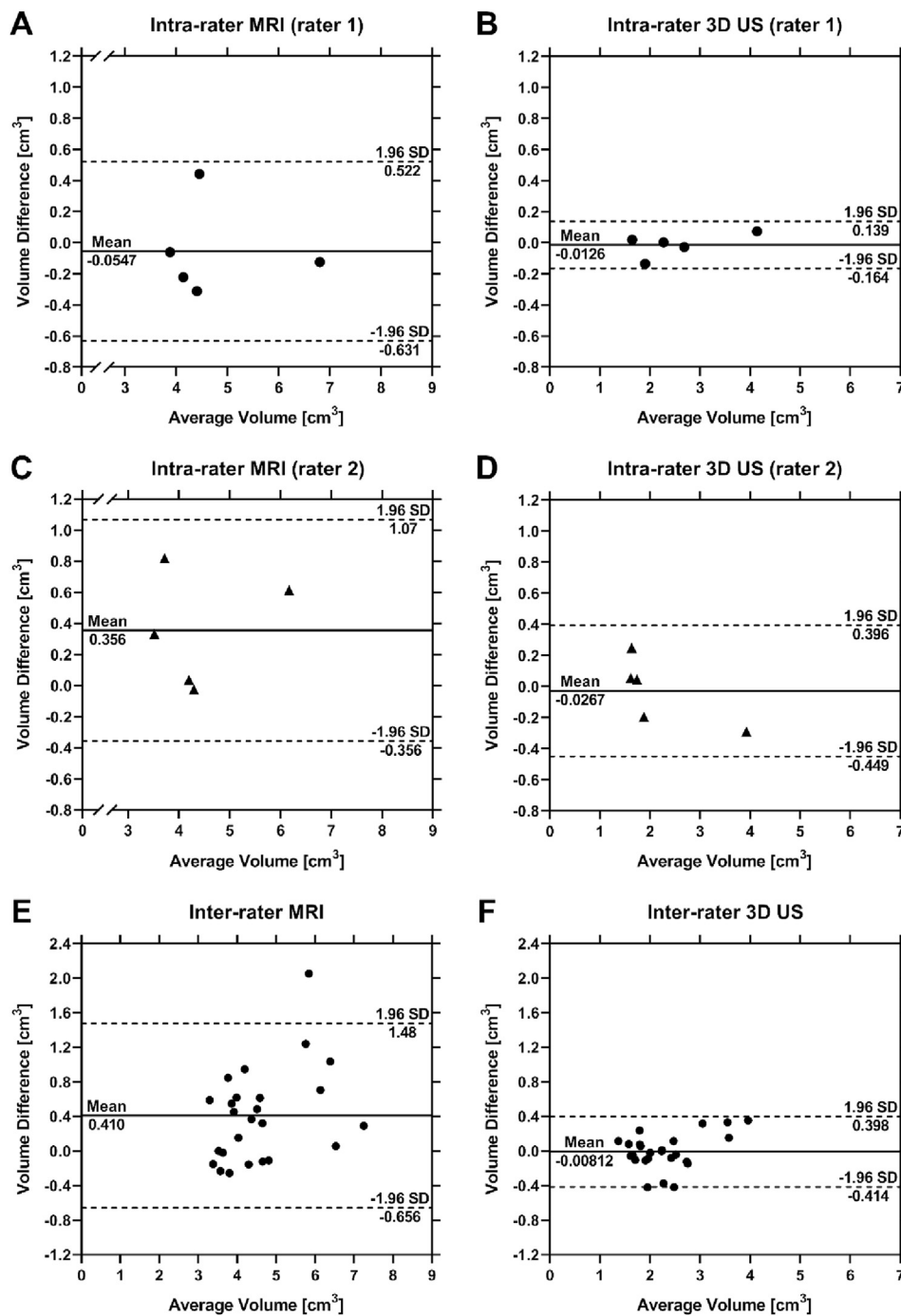


Fig. 4. Bland-Altman plots assessing intra-rater test/re-test reliability of rater 1 with MRI (A) and 3D US (B), and rater 2 with MRI (C) and 3D US (D). Bland-Altman assessing inter-rater reliability between the two raters using MRI (E) and 3D US (F) to complete segmentations. Mean differences in segmentation volumes are indicated by a solid line and mean \pm 1.96 SD are indicated by dashed lines.

comparisons (Table 3). Collectively, these results suggest that our 3D US system can quantify tFAC volume with similar or perhaps superior reliability and precision than MRI.

The higher spatial resolution of 3D US images acquired with the Canon 14L5 linear transducer compared to 3.0 T MRI may partially account for reliability and precision differences. Resolution differences between modalities were most apparent during MRI segmentations when raters attempted to define the interface between tFAC and the synovial lining of Hoffa's fat pad. Differentiating tFAC from slightly hyperintense synovial lining proved extremely difficult or impossible during MRI

segmentations despite manipulating image contrast. Additionally, the TF cartilage interface was difficult to identify on MRI as both cartilage structures were equally hyperintense. The synovial lining of Hoffa's fat pad along with the TF contact point were not within the ROI of 3D US acquisitions since images were acquired during maximum knee flexion. Healthy FAC produced ideal US images with excellent differentiation from surrounding tissues. The difficulties in identifying borders on MRI likely also contributed to higher segmentation times compared to 3D US. The MRI and 3D US resolutions were chosen to match what is routinely used in both patient care and clinical trials for OA to enable comparisons

Table 3

Intra- and inter-rater reliability ICCs with 95% confidence intervals (CIs) for manual MRI and 3D US segmentations, along with repeated MRI and 3D US registrations and trimmings. The MSD, HD, and DSC values \pm SD for all comparisons are also presented.

	ICC (95% CI)	P value	MSD [mm]	HD [mm]	DSC [%]
Intra-rater (n = 5)					
MRI (rater 1)	0.97 (0.79, 1.00)	0.001	0.218 \pm 0.109	2.88 \pm 1.37	87.3 \pm 2.8
MRI (rater 2)	0.90 (0.25, 0.99)	0.002	0.499 \pm 0.275	6.66 \pm 2.76	83.5 \pm 4.6
3D US (rater 1)	1.00 (0.98, 1.00)	<0.0001	0.126 \pm 0.024	1.76 \pm 0.35	92.9 \pm 0.2
3D US (rater 2)	0.98 (0.84, 1.00)	0.0003	0.256 \pm 0.143	3.70 \pm 2.23	88.1 \pm 2.6
Inter-rater (n = 25)					
MRI	0.83 (0.48, 0.94)	<0.0001	0.274 \pm 0.122	3.51 \pm 1.77	83.1 \pm 3.6
3D US	0.96 (0.90, 0.98)	<0.0001	0.243 \pm 0.133	2.89 \pm 1.72	86.4 \pm 3.1
Registration & trimming (n = 5)					
Single rater	1.00 (0.99, 1.00)	<0.0001	0.101 \pm 0.090	1.72 \pm 0.94	94.3 \pm 4.4

in a clinically relevant context.

4.2. Validity

3D US tFAC segmentations possessed larger volumes than MRI segmentations. Considering the higher spatial resolution of US compared to MRI, it is possible that MRI segmentations were not able to capture the true cartilage volume as effectively as 3D US. Medial and lateral portions of the tFAC and condylar cartilage become thin and difficult to delineate from thin adipose tissue and may often not be visible in MRI. Due to the high spatial resolution of 3D US, the thin medial and lateral portions of tFAC were easily identified and therefore included in segmentations. This will be of great importance in clinical studies of joint disease since thinner areas of cartilage are particularly susceptible to damage and loss in KOA. Our 3D US device was able to visualize tFAC and condylar cartilage regions that were difficult or impossible to visualize using MRI, providing a more comprehensive model of the cartilage and improved volume quantifications. Notwithstanding these differences in absolute cartilage volumes, Spearman Rank-Order Correlation and linear regression analyses revealed a strong correlation between MRI and 3D US tFAC measurements, and that MRI tFAC volumes can predict 3D US volumes.

4.3. Limitations and impact

This study was conducted on volunteers with healthy knees rather than KOA patients. Validating our 3D US system on healthy knees prior to testing with KOA patients was a necessary first step for developing image acquisition, segmentation, and analysis protocols. In KOA patients, FAC characteristically develops fissures, abrasions, and other surface irregularities whereas healthy cartilage is smooth and continuous. Therefore,

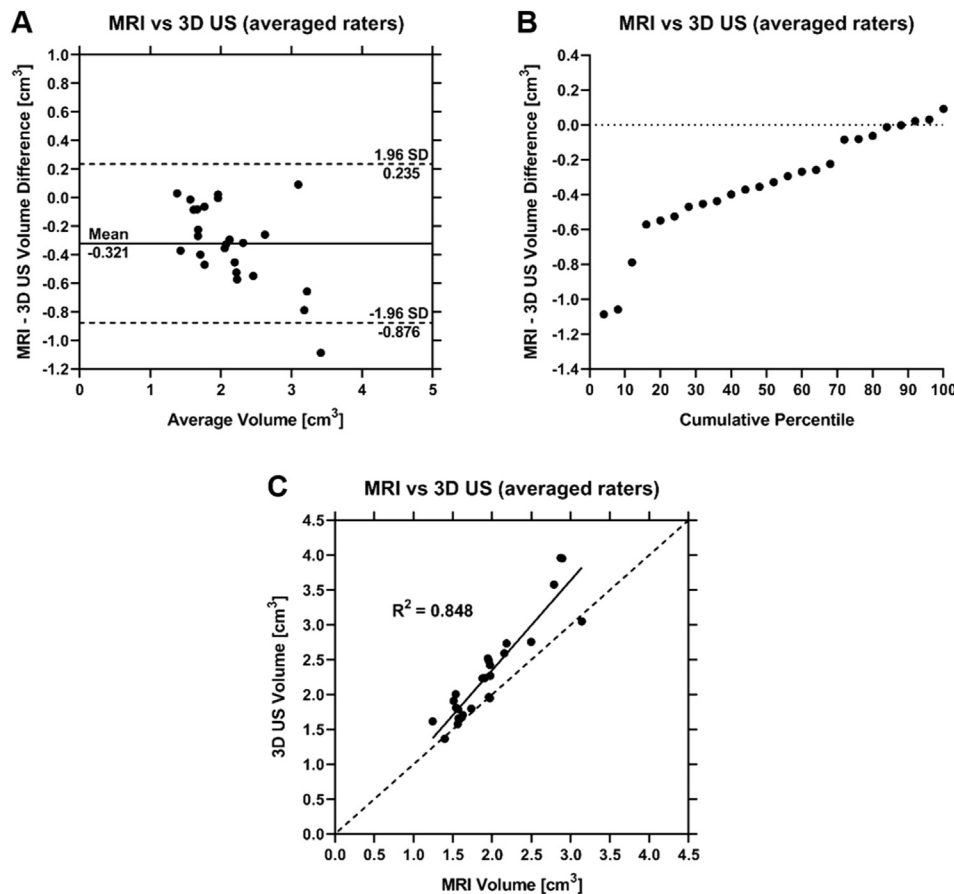


Fig. 5. (A) Bland-Altman plot assessing the relationship between MRI and 3D US segmentation volumes as the mean for both raters. Mean differences in segmentation volumes are indicated by a solid line and the mean \pm 1.96 SD are indicated by dashed lines. (B) Cumulative percentile plot depicting the volume difference between MRI and 3D US segmentations averaged between both raters. (C) Linear regression plot of MRI segmentation volumes used as a predictor for 3D US. A line of equality is represented by the dashed line.

before implementing our 3D US device clinically, the measurement properties of this system should also be evaluated in KOA patients. Results from a KOA patient study will enable us to determine if KOA cartilage pathology impacts measurement properties of our system relative to healthy cartilage. However, given the high resolution and excellent soft tissue contrast of clinical US systems, we anticipate our 3D US system will perform similarly in KOA patients.

Only a portion of FAC was captured in a single pass 3D US acquisition as the FAC cannot be visualized through the patella and tibia using US. Therefore, knees were scanned in maximum flexion proximal to the patella to capture the greatest portion of FAC possible. However, during maximum knee flexion the posterior medial and lateral condylar cartilage are in contact with the tibial cartilage and are not visible with 3D US. This limitation can be overcome if the tFAC is used as a non-invasive imaging “biopsy” of knee cartilage, providing clinicians with an indication of FAC status representative of overall joint health. Additionally, manual trimming of MRI segmentations to match the 3D US ROI was only necessary for validating our system against MRI and would not be required when using 3D US independently in future studies. While this procedure may have introduced variability or bias, repeated registrations, along with repeated trimming, revealed nearly perfect reproducibility (Table 3), indicating that our protocol results in very little variability or bias.

The weight-bearing condylar cartilage was able to be visualized using our 3D US device. However, this required additional acquisitions on the medial and lateral sides of the patella during maximum knee flexion. Since MR images of FAC were acquired during minimal knee flexion, variations in patella positioning relative to the FAC surface in MRI compared to 3D US resulted in difficulties registering 3D US condylar cartilage segmentations to MRI. Therefore, this study focused on the tFAC region for validation with MRI, but 3D US could be used for monitoring condylar cartilage volume changes over time without requiring MRI comparisons. Finally, a small subset of patients with severe KOA may experience limited range of motion, which might interfere with visualization of the most inferior aspects of the tFAC.

The greatest advantage of our 3D US system is the ability to acquire images quickly, easily, and comfortably at the patient's bedside, providing cost-effective and non-invasive assessments of FAC status for reliable longitudinal monitoring. Our 3D US device could alter the workflow of orthopedic, sports medicine, primary care, and arthritis clinics by enabling clinicians and researchers to obtain more information without added complexity or additional stress and discomfort to patients. This technology may be well-suited to longitudinal and interventional clinical studies where detecting changes in cartilage volume is required. In the future, this system will also be useful in a routine clinical care context. Currently, the use of KL grading for assessing KOA progression is insensitive to change and relies on indirect features of FAC thinning. MRI-based measures of cartilage volume are superior to radiographic measures, but are limited by cost, time, accessibility, and patient-related factors, preventing generalized use of quantitative MRI for KOA. Our study demonstrates that cartilage volume measurements acquired using 3D US represent a more feasible method to quantitatively assess tFAC volume with very high reliability and accuracy.

In conclusion, we have demonstrated the reliability and validity of a handheld mechanical 3D US device we developed to quantify tFAC volumes in healthy volunteers. We demonstrated that 3D US segmentations are associated with excellent intra- and inter-rater reliabilities and possess strong agreement with MRI tFAC volume measurements. The tFAC is a vital region of the knee joint for investigating the progression of patellofemoral OA and could also be used as a non-invasive imaging “biopsy” of the FAC to monitor KOA progression and response to treatment. Future work will assess the reliability of our 3D US device in KOA patients and the ability to monitor FAC volume changes over time. Further assessment of measurement properties including sensitivity to change is necessary before its use can be recommended in clinical trials. Future work will also assess the test-retest reliability of 3D US during

image acquisitions separated by time. In addition to longitudinal construct validity, future work will also assess the intra- and inter-rater reliability of 3D US cartilage measurements in a longitudinal study to monitor the progression of tFAC change and degradation for early detection of KOA. 3D US is a promising, inexpensive, and widely accessible imaging modality for point-of-care assessments of KOA and will enable clinicians and researchers to obtain additional information without added complexity or discomfort to patients.

Author contributions

SP, CTA, and AF contributed to the conception and design of the study. All authors contributed to the analysis and interpretation of the data. SP and RD performed the acquisition of images, segmentations, and analyses on both imaging modalities. DJG contributed to developing the image processing, segmentation, and analysis workflow. All authors contributed to drafting the article with critical revisions and final approvals before submission. SP (spaperni@uwo.ca) takes full responsibility for the integrity of the study from inception to the finished article.

Role of the funding source

This work is funded by the Canadian Institutes of Health Research (CIHR #154314) and a Collaborative Research Seed Grant from the Schulich School of Medicine and Dentistry (R5727A08). The funding agencies did not have any role in the design, analysis, interpretation of data, writing of the manuscript, or the decision to submit for publication.

Sources of support

Work is funded by the Canadian Institutes of Health Research (#154314) and a Collaborative Research Seed Grant from the Schulich School of Medicine and Dentistry (R5727A08).

Declaration of competing interest

The authors have no competing interests with respect to the work completed in this study.

Acknowledgements

The authors acknowledge Dr. David Tessier for volunteer recruitment and image coordination, and Holly Philpott for 3D US image acquisition. We especially thank all of the volunteers who made this study possible. The authors acknowledge Dr. Trevor Birmingham for providing guidance with experimental design.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ocarto.2020.100127>.

References

- [1] J.M. Jordan, C.G. Helmick, J.B. Renner, G. Luta, A.D. Dragomir, J. Woodard, et al., Prevalence of knee symptoms and radiographic and symptomatic knee osteoarthritis in african Americans and caucasians: the johnston county osteoarthritis project, *J. Rheumatol.* 34 (1) (2007) 172–180.
- [2] D.T. Felson, A. Naimark, J. Anderson, L. Kazis, W. Castelli, R.F. Meenan, The prevalence of knee osteoarthritis in the elderly: the framingham osteoarthritis study, *Arthritis Rheum.* 30 (8) (1987) 914–918, <https://doi.org/10.1002/art.1780300811>.
- [3] D.T. Felson, Osteoarthritis of the knee, *N. Engl. J. Med.* 354 (8) (2006) 841–848, <https://doi.org/10.1056/NEJMc051726>.
- [4] S. Amin, M.P. LaValley, A. Guermazi, M. Grigoryan, D.J. Hunter, M. Clancy, et al., The relationship between cartilage loss on magnetic resonance imaging and radiographic progression in men and women with knee osteoarthritis, *Arthritis Rheum.* 52 (10) (2005) 3152–3159, <https://doi.org/10.1002/art.21296>.

- [5] A. Guermazi, F.W. Roemer, D. Burstein, D. Hayashi, Why radiography should no longer be considered a surrogate outcome measure for longitudinal assessment of cartilage in knee osteoarthritis, *Arthritis Res. Ther.* 13 (6) (2011) 247, <https://doi.org/10.1186/ar3488>.
- [6] D.J. Hunter, Y.Q. Zhang, X. Tu, M. Lavalley, J.B. Niu, S. Amin, et al., Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? *Arthritis Rheum.* 54 (8) (2006) 2488–2495, <https://doi.org/10.1002/art.22016>.
- [7] E. Yusuf, M.C. Kortekaas, I. Watt, T.W.J. Huizinga, M. Kloppenburg, Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review, *Ann. Rheum. Dis.* 70 (1) (2011) 60–67, <https://doi.org/10.1136/ard.2010.131904>.
- [8] D.J. Hunter, A. Guermazi, G.H. Lo, A.J. Grainger, P.G. Conaghan, R.M. Boudreau, et al., Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score), *Osteoarthritis Cartilage* 19 (8) (2011) 990–1002, <https://doi.org/10.1016/j.joca.2011.05.004>.
- [9] D.J. Hunter, G.H. Lo, D. Gale, A.J. Grainger, A. Guermazi, P.G. Conaghan, The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score), *Ann. Rheum. Dis.* 67 (2) (2008) 206–211, <https://doi.org/10.1136/ard.2006.066183>.
- [10] P.R. Kornaat, R.Y.T. Ceulemans, H.M. Kroon, N. Riyazi, M. Kloppenburg, W.O. Carter, et al., MRI assessment of knee osteoarthritis: knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system, *Skeletal Radiol.* 34 (2) (2005) 95–102, <https://doi.org/10.1007/s00256-004-0828-0>.
- [11] C.G. Peterfy, A. Guermazi, S. Zaim, P.F.J. Tirmar, Y. Miaux, D. White, et al., Whole-organ magnetic resonance imaging Score (WORMS) of the knee in osteoarthritis, *Osteoarthritis Cartilage* 12 (3) (2004) 177–190, <https://doi.org/10.1016/j.joca.2003.11.003>.
- [12] R.J. Schmitz, H.-M. Wang, D.R. Polprasert, R.A. Kraft, B.G. Pietrosimone, Evaluation of knee cartilage thickness: a comparison between ultrasound and magnetic resonance imaging methods, *Knee* 24 (2) (2017) 217–223, <https://doi.org/10.1016/j.knee.2016.10.004>.
- [13] A. Iagnocco, C. Perella, M.A. D'Agostino, E. Sabatini, G. Valesini, P.G. Conaghan, Magnetic resonance and ultrasonography real-time fusion imaging of the hand and wrist in osteoarthritis and rheumatoid arthritis, *Rheumatol Oxf Engl* 50 (8) (2011) 1409–1413, <https://doi.org/10.1093/rheumatology/ker111>.
- [14] S. Ohndorf, M. Backhaus, Pro musculoskeletal ultrasonography in rheumatoid arthritis, *Clin. Exp. Rheumatol.* 33 (4 Suppl 92) (2015) S50–S53. PubMed PMID: 26457337.
- [15] T. Okano, K. Mamoto, M. Di Carlo, F. Salaffi, Clinical utility and potential of ultrasound in osteoarthritis, *Radiol Med (Torino)* (2019), <https://doi.org/10.1007/s11547-019-01013-z>. Published online March 4.
- [16] M. Ponikowska, J. Świerkot, B. Nowak, The importance of ultrasound examination in early arthritis, *Reumatol Wars* 56 (6) (2018) 354–361, <https://doi.org/10.5114/reum.2018.80712>.
- [17] G.A. Bruyn, E. Naredo, A. Iagnocco, P.V. Balint, M. Backhaus, F. Gandjbakhch, et al., The OMERACT ultrasound working group 10 years on: update at OMERACT 12, *J. Rheumatol.* 42 (11) (2015) 2172–2176, <https://doi.org/10.3899/jrheum.141462>.
- [18] A. Fenster, D.B. Downey, H.N. Cardinal, Three-dimensional ultrasound imaging, *Phys. Med. Biol.* 46 (5) (2001) R67–R99, <https://doi.org/10.1088/0031-9155/46/5/201>.
- [19] J. Kishimoto, S. de Ribaupierre, D.S.C. Lee, R. Mehta, K. St Lawrence, A. Fenster, 3D ultrasound system to investigate intraventricular hemorrhage in preterm neonates, *Phys. Med. Biol.* 58 (21) (2013) 7513–7526, <https://doi.org/10.1088/0031-9155/58/21/7513>.
- [20] S.H. Saravelos, G.W.S. Kong, J.P.W. Chung, J.S.M. Mak, C.H.S. Chung, L.P. Cheung, et al., A prospective randomized controlled trial of 3D versus 2D ultrasound-guided embryo transfer in women undergoing ART treatment, *Hum Reprod Oxf Engl* 31 (10) (2016) 2255–2260, <https://doi.org/10.1093/humrep/dew206>.
- [21] A. Zahalka, A. Fenster, An automated segmentation method for three-dimensional carotid ultrasound images, *Phys. Med. Biol.* 46 (4) (2001) 1321–1342, <https://doi.org/10.1088/0031-9155/46/4/327>.
- [22] D.J. Hunter, R.D. Altman, F. Cicuttini, M.D. Crema, J. Duryea, F. Eckstein, et al., OARSI Clinical Trials Recommendations: knee imaging in clinical trials in osteoarthritis, *Osteoarthritis Cartilage* 23 (5) (2015) 698–715, <https://doi.org/10.1016/j.joca.2015.03.012>.
- [23] S. Papernick, D.J. Gillies, T. Appleton, A. Fenster, Three-dimensional ultrasound for monitoring knee inflammation and cartilage damage in osteoarthritis and rheumatoid arthritis, *Proc. SPIE* 11315 (2020), <https://doi.org/10.1117/12.2549624>.
- [24] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, et al., 3D slicer as an image computing platform for the quantitative imaging network, *Magn. Reson. Imaging* 30 (9) (2012) 1323–1341, <https://doi.org/10.1016/j.mri.2012.05.001>.
- [25] W. Wirth, M. Nevitt, M.-P. Heliö Le Graverand, O. Benichou, D. Dreher, R.Y. Davies, et al., Sensitivity to change of cartilage morphometry using coronal FLASH, sagittal DESS, and coronal MPR DESS protocols – comparative data from the osteoarthritis initiative (OAI), *Osteoarthritis Cartilage* 18 (4) (2010) 547–554, <https://doi.org/10.1016/j.joca.2009.12.003>.
- [26] A.B. Albu, T. Beugeling, D. Laurendeau, A morphology-based approach for interslice interpolation of anatomical slices from volumetric images, *IEEE Trans. Biomed. Eng.* 55 (8) (2008) 2022–2038, <https://doi.org/10.1109/TBME.2008.921158>.
- [27] A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, *BMC Med. Imag.* 15 (2015), <https://doi.org/10.1186/s12880-015-0068-x>.
- [28] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, *J Chiropr Med* 15 (2) (2016) 155–163, <https://doi.org/10.1016/j.jcm.2016.02.012>.
- [29] A. Iagnocco, G. Coari, A. Zoppini, Sonographic evaluation of femoral condylar cartilage in osteoarthritis and rheumatoid arthritis, *Scand. J. Rheumatol.* 21 (4) (1992) 201–203, <https://doi.org/10.3109/03009749209099222>.
- [30] W. McCune, D. Dedrick, A. Aisen, A. MacGuire, Sonographic evaluation of osteoarthritic femoral condylar cartilage. Correlation with operative findings, *Clin. Orthop.* 254 (1990) 230–235. PubMed PMID: 2182254.
- [31] C.-L. Lee, M.-H. Huang, C.-Y. Chai, C.-H. Chen, J.-Y. Su, Y.-C. Tien, The validity of in vivo ultrasonographic grading of osteoarthritic femoral condylar cartilage: a comparison with in vitro ultrasonographic and histologic gradings, *Osteoarthritis Cartilage* 16 (3) (2008) 352–358, <https://doi.org/10.1016/j.joca.2007.07.013>.
- [32] S. Saarakkala, P. Waris, V. Waris, I. Tarkkiainen, E. Karvanen, J. Aarnio, et al., Diagnostic performance of knee ultrasonography for detecting degenerative changes of articular cartilage, *Osteoarthritis Cartilage* 20 (5) (2012) 376–381, <https://doi.org/10.1016/j.joca.2012.01.016>.
- [33] A.H. Spannow, M. Pfeiffer-Jensen, N.T. Andersen, E. Stenbø, T. Herlin, Inter- and intraobserver variation of ultrasonographic cartilage thickness assessments in small and large joints in healthy children, *Pediatr. Rheumatol.* 7 (1) (2009) 12, <https://doi.org/10.1186/1546-0096-7-12>.
- [34] A. Faisal, S.-C. Ng, S.-L. Goh, K.W. Lai, Knee cartilage segmentation and thickness computation from ultrasound images, *Med. Biol. Eng. Comput.* 56 (4) (2018) 657–669, <https://doi.org/10.1007/s11517-017-1710-2>.
- [35] M.A. Bedewi, A.A. Elsifey, M.F. Naguib, A.K. Saleh, N.B. Nwihadh, A.A. Abd-Elghany, et al., Sonographic assessment of femoral cartilage thickness in healthy adults, *J. Int. Med. Res.* 48 (8) (2020), 300060520948754, <https://doi.org/10.1177/0300060520948754>, 300060520948754.
- [36] C.-H. Yoon, H.-S. Kim, J.H. Ju, W.-H. Jee, S.-H. Park, H.-Y. Kim, Validity of the sonographic longitudinal sagittal image for assessment of the cartilage thickness in the knee osteoarthritis, *Clin. Rheumatol.* 27 (12) (2008) 1507–1516, <https://doi.org/10.1007/s10067-008-0956-3>.
- [37] E. Naredo, C. Acebes, I. Möller, F. Canillas, J.J. de Agustín, E. de Miguel, et al., Ultrasonod validity in the measurement of knee cartilage thickness, *Ann. Rheum. Dis.* 68 (8) (2009) 1322–1327, <https://doi.org/10.1136/ard.2008.090738>.
- [38] C.E. Draper, T.F. Besier, G.E. Gold, M. Fredericson, A. Fiene, G.S. Beaupre, et al., Is cartilage thickness different in young subjects with and without patellofemoral pain? *Osteoarthritis Cartilage* 14 (9) (2006) 931–937, <https://doi.org/10.1016/j.joca.2006.03.006>.
- [39] S. Koo, G.E. Gold, T.P. Andriacchi, Considerations in measuring cartilage thickness using MRI: factors influencing reproducibility and accuracy, *Osteoarthritis Cartilage* 13 (9) (2005) 782–789, <https://doi.org/10.1016/j.joca.2005.04.013>.
- [40] T.G. Williams, A.P. Holmes, M. Bowes, G. Vincent, C.E. Hutchinson, J.C. Waterton, et al., Measurement and visualisation of focal cartilage thickness change by MRI in a study of knee osteoarthritis using a novel image analysis tool, *Br. J. Radiol.* 83 (995) (2010) 940–948, <https://doi.org/10.1259/bjr/68875123>.