

Reliability of a Novel Semiautomated Ultrasound Segmentation Technique for Assessing Average Regional Femoral Articular Cartilage Thickness

Caroline Lisee, Melanie L. McGrath, Christopher Kuenze, Ming Zhang,
Matt Salzler, Jeffrey B. Driban, and Matthew S. Harkey

Context: Ultrasound imaging is a clinically feasible tool to assess femoral articular cartilage and may have utility in tracking early knee osteoarthritis development. Traditional assessment techniques focus on measurements at a single location, which can be challenging to adopt for novice raters. **Objective:** To introduce a novel semiautomated ultrasound segmentation technique and determine the intrarater and interrater reliability of average regional femoral articular cartilage thickness and echo intensity of a novice and expert rater. **Design:** Descriptive observational study. **Setting:** Orthopedic clinic. **Patients or Other Participants:** Fifteen participants (mean [SD]; age 23.5 [4.6] y, height = 172.6 [9.3] cm, mass = 79.8 [15.7] kg) with a unilateral history of anterior cruciate ligament reconstruction participated. **Intervention:** None. **Main Outcome Measures:** One rater captured anterior femoral cartilage images of the participants' contralateral knees using a transverse suprapatellar ultrasound assessment. The total femoral cartilage cross-sectional area of each image was segmented by a novice and expert rater. A novel custom program automatically separated the cartilage segmentations into medial, lateral, and intercondylar regions to determine the cross-sectional area and cartilage length. The average cartilage thickness in each region was calculated by dividing the cross-sectional area by the cartilage length. Echo intensity was calculated as the average gray-scale pixel value of each region. Two-way random effect intraclass correlations coefficient (ICC) for absolute agreement were used to determine the interrater reliability between a novice and expert rater, as well as the intrarater reliability of the novice rater. **Results:** The novice rater demonstrated excellent intrarater (ICC [2,k] range = .993–.997) and interrater (ICC [2,k] range = .944–.991) reliability with the expert rater of all femoral articular cartilage average thickness and echo intensity regions. **Conclusions:** The novel semiautomated average cartilage thickness and echo-intensity assessment is efficient, systematic, and reliable between an expert and novice rater with minimal training.

Keywords: knee, body region(s), measurement properties, research designs/area(s), diagnostic ultrasound, discipline area(s)

Knee osteoarthritis (OA) is characterized by declining synovial joint health, especially in the articular cartilage.¹ Clinically feasible assessments of early articular cartilage changes may overcome the barriers of technically demanding or expensive biomarkers and encourage implementation of timely interventions to prevent or delay the progression of the chronic disease.¹ Diagnostic ultrasound is a valid assessment of resting anterior femoral articular cartilage structure that quantifies tissue thickness or cross-sectional area (CSA).² Although not yet validated, femoral articular cartilage echo intensity may also provide unique information about water content changes associated with early knee OA development.³ Individuals with anterior cruciate ligament (ACL)

injuries at elevated risk for OA demonstrate differences in ultrasound-based femoral articular cartilage thickness,⁴ which provides preliminary evidence for the assessment's utility.

Traditional manual femoral cartilage thickness segmentation demonstrates good to excellent intrarater and test-retest reliability of medial, lateral, and intercondylar thickness in an expert rater,³ but interrater reliability with a novice rater has not been established. It is important to understand the reliability between individuals with different image-processing training experience, as cartilage ultrasound assessment is adopted in research. The traditional technique requires raters to draw a vertical line with consistent perpendicular alignment between cartilage borders at a subjective location within 3 regions, which may vary between raters. Small deviations in the location and orientation selection of thickness lines can result in large thickness differences and large measurement variance between raters.⁵ In addition, a single thickness location may not represent thickness throughout the entire cartilage region.^{3,4} We have developed a novel semiautomated segmentation technique using a manual segmentation of the entire femoral cartilage CSA to automatically separate the cartilage into standardized regions normalized to cartilage length and calculate the average thickness within each region. This novel segmentation technique improves the ability of ultrasonography to assess femoral cartilage by providing an outcome that is representative of the thickness throughout standardized femoral regions while reducing the burden of manually measuring additional cartilage outcomes

Lisee and Kuenze are with the Department of Kinesiology, College of Education, Michigan State University, East Lansing, MI, USA. McGrath is with the School of Integrative Physiology and Athletic Training, College of Health Professions and Biomedical Sciences, University of Montana, Missoula, MT, USA. Kuenze is with the Department of Orthopedics, College of Osteopathic Medicine, Michigan State University, East Lansing, MI, USA. Zhang is with the Department of Computer Science & Networking, Wentworth Institute of Technology, Boston, MA, USA. Zhang, Driban, and Harkey are with the Division of Rheumatology, Allergy & Immunology, Tufts Medical Center, Boston, MA, USA. Salzler is with the Department of Orthopaedics, Tufts Medical Center, Boston, MA, USA. Harkey is also with the Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA. Lisee (liseecar@msu.edu) is corresponding author.

(ie, cartilage length or multiple compartments). Therefore, the purpose of this technical report was to (1) thoroughly describe ultrasound assessment procedures, (2) introduce a novel semiautomated technique to assess average cartilage thickness and echo intensity within standardized femoral regions, and (3) determine the intrarater and interrater reliability of a novice and expert rater using the semiautomated technique.

Methods

Participants

A convenience sample of 15 participants between 18 and 35 years old with a primary unilateral anterior ACL injury were recruited at their preoperative visit to an orthopedic surgeon (M.S.). Participants were excluded if they had a history of lower-extremity surgery, injured either knee within the last 6 months (other than ACL injury), or had previously been diagnosed with any form of arthritis. The participants' uninjured knee was used for this study.

Ultrasonographic Assessment of Femoral Articular Cartilage

Participant Positioning, Probe Positioning, and Imaging Acquisition. Three femoral articular cartilage images were collected using a LOGIQ E ultrasound machine and 12L-RS linear probe (GE Healthcare, Chicago, IL) in the participants' uninjured knees by an expert rater with 5 years of imaging experience, who has demonstrated excellent image intrasession reliability.³ After 30 minutes of sitting, the participants were positioned in maximum knee flexion, and the ultrasound probe was positioned perpendicular to the femoral cartilage surface, similar to previous methods (Supplementary Figure 1A [available online]).³ Image acquisition, using the transparency grid over the monitor image display to capture similar medial and lateral femoral condyle positions between knee images (Supplementary Figure 1B and 1C [available online]), was also similar to previous methods.³

Novel Semiautomated Technique to Assess Average Cartilage Thickness. Ultrasound images were manually segmented with freely available ImageJ software (<https://imagej.nih.gov/>). The Supplementary Video (available online) provides an explanation of important points to consider when segmenting total cartilage CSA. The images were rotated to align the cartilage parallel to the horizontal plane to ensure similar cartilage orientation between the participants. The entire cartilage CSA was segmented between the synovial–cartilage border and the cartilage–bone border.⁴ Segmented images were exported to Microsoft Paint (Microsoft, Redmond, WA) to mark the central point of the cartilage intercondylar notch at the deepest point of the synovial–cartilage border (Figure 1A).

The following steps to determine average cartilage thickness (in mm) and echo intensity were automatically processed with a custom MATLAB code (version 9.2; MathWorks, Natick, MA) on the marked segmentation images. First, the overall cartilage CSA was automatically separated into standardized medial, intercondylar, and lateral regions. The intercondylar region was centered around the manually identified central point and defined as the middle 25% of the cartilage based on the overall image width (Figure 2A). The medial and lateral regions of the image were defined as the area medial or lateral to the intercondylar region, respectively (Figure 2B). Next, the custom program determined the length of the cartilage–bone interface for each region. To calculate the average cartilage thickness, the regional CSA was divided by the length of its cartilage–bone interface (Figure 2C). Regional echo intensity was defined as the average gray-scale pixel value ranging from black (ie, 0) to white (ie, 255). The cartilage outcomes from the 3 images were averaged together for statistical analysis.

The expert rater has previously demonstrated excellent intrasession and test–retest reliability of total cartilage CSA,⁴ whereas the novice rater had no prior processing experience. Upon receiving in-depth ultrasound assessment training from the expert rater, the novice rater completed manual segmentation on 3 practice sets of at least 30 images from individuals not included in this study. Afterward, the expert rater reviewed the segmentations, provided constructive feedback on how to improve, and provided his own segmentations for the novice rater to visually compare. Finally,

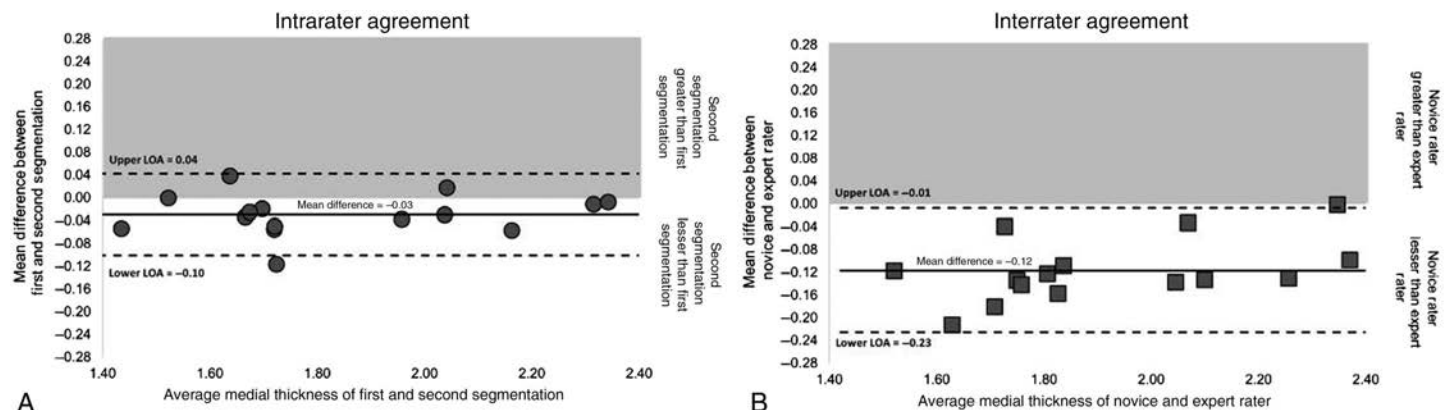


Figure 1 — (A) Average medial thickness between the first and second segmentations of the novice rater are represented by the circles. The solid black line indicates mean difference between segmentations, and the dotted lines indicate upper and lower LOA. The unshaded white area and shaded gray area represent systematic underestimation and overestimation of the second segmentation compared with the first segmentation, respectively. The novice rater's first and second segmentations demonstrate acceptable agreement. (B) Average medial thickness between segmentations of the expert and novice rater are represented by the squares. The solid black line indicates mean difference between segmentations, and the dotted lines indicate upper and lower LOA. The unshaded white area and shaded gray area represent systematic underestimation and overestimation of the novice rater's segmentation compared with the expert rater's segmentation, respectively. The novice and expert raters' segmentations demonstrate acceptable agreement, but the novice rater underestimates average medial thickness.

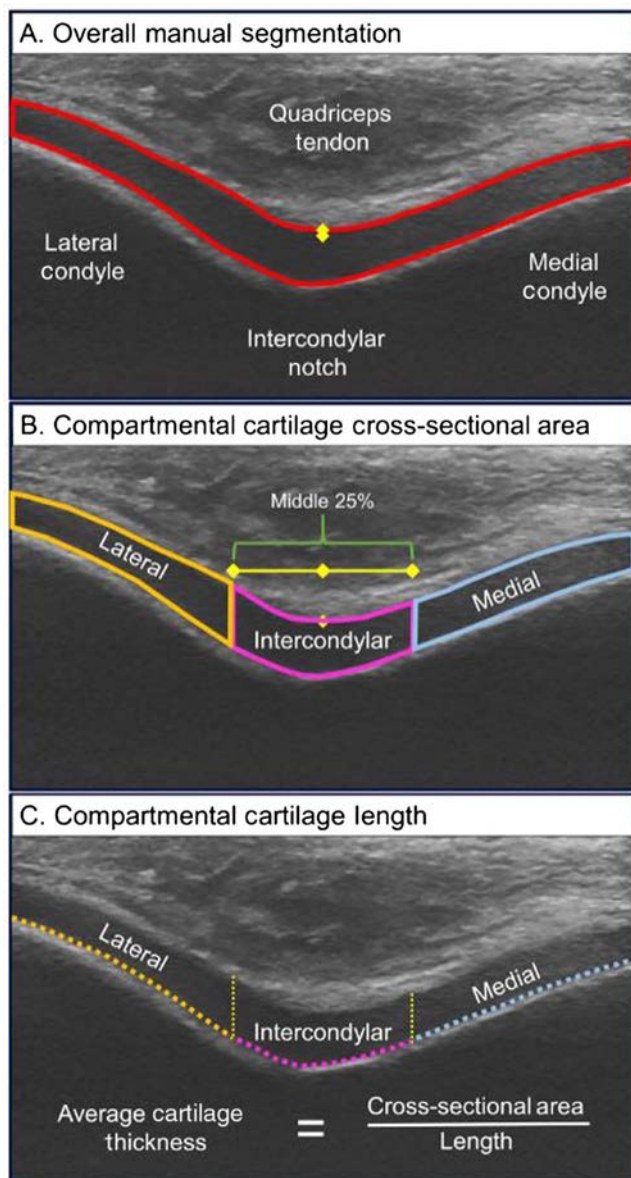


Figure 2 — (A) The overall cartilage cross-sectional area of the anterior femoral articular cartilage was outlined manually. The middle of the intercondylar notch was denoted with the central diamond. (B) The custom program uses the location of the central diamond to automatically separate the segmentation into the lateral, intercondylar, and medial cartilage regions and calculate the cross-sectional area for each. The size of the intercondylar region represented 25% of the image width centered around the central diamond. (C) The custom program also calculated the length of the cartilage–bone interface (ie, cartilage length). Regional average cartilage thickness was calculated as the cross-sectional area divided by the cartilage length.

both raters segmented the study images to determine interrater reliability. The novice rater processed the same images 2 weeks later to determine intrarater reliability.

Statistical Analysis

Two-way random effect intraclass correlations coefficient ($ICC_{2,k}$) based on absolute agreement,⁶ Bland–Altman plots with 95%

limits of agreement⁷ were used to assess intrarater and interrater reliability and agreement. The SEM and minimal detectable change were calculated, as previously reported.⁸

Results

This study included 9 men and 6 women (age 23.5 [4.6] y, height 172.6 [9.3] cm, mass 79.8 [15.7] kg). The maximal knee flexion angle ranged from 120° to 140° (135° [7°]). The novice rater demonstrated excellent intrarater reliability over a 2-week period and excellent interrater reliability with the expert rater for femoral articular cartilage average thickness and echo-intensity outcomes (Table 1). Only one or none of the data points fell outside the limits of agreement, indicating acceptable intrarater (Figure 1A) and interrater (Figure 1B) agreement for average medial thickness, respectively. The novice rater systematically underestimated the average medial thickness compared with the expert rater. Bland–Altman plots for lateral and intercondylar thickness, as well as echo intensity for all regions, are represented in [Supplementary Figures 2 and 3](#) (available online).

Discussion

This novel semiautomated technique demonstrates excellent intrarater and interrater reliability and agreement for femoral articular cartilage thickness and echo intensity in the medial, lateral, and intercondylar regions (Table 1). Our study results are similar to previous studies reporting good to excellent intrarater reliability of cartilage thickness and echo intensity using traditional thickness techniques.^{3,4} These results suggest that this technique can be used by individuals with limited image-processing training.

Traditional cartilage-processing techniques that assess thickness at a single location may result in inconsistencies in the exact location and angle of the thickness segmentation, and do not represent thickness of the entire cartilage region. The novel semiautomated technique described in this report overcomes these barriers by reducing the methodological interpretation and measuring average cartilage thickness throughout each standardized region. The novel technique requires segmentation of the total cartilage CSA and then uses an automated program to standardize regional separation and calculation of average cartilage thickness, which may reduce the variability compared to the traditional technique. In addition, our novel technique calculates an average cartilage thickness throughout each region, which replicates the approach used in magnetic resonance imaging studies that calculate average cartilage thickness as the cartilage volume divided by the subchondral bone area.⁹ Another benefit to the novel technique is that cartilage echo intensity can be quantified without additional processing. Lower echo intensity (ie, greater darkness) indicates greater water content in muscles¹⁰ and may help identify early cartilage swelling (ie, greater water content). Our results indicate that assessing echo intensity has excellent reliability, but future research is needed to validate echo intensity as a measure of cartilage composition.

Although our results highlight the excellent interrater reliability for our cartilage segmentation technique, the Bland–Altman plots highlight a potential limitation for bias between a novice and expert rater. Figure 1B indicates that the novice rater tended to underestimate the average medial cartilage thickness when compared with the expert rater. Therefore, we recommend that within-subjects segmentations (eg, longitudinal assessments, preloading/postloading) should be completed by the same rater to reduce error.

Table 1 Intraclass Correlations of Intrarater and Interrater Reliability of a Novice and Expert Rater of Novel Ultrasound-Assessed Femoral Articular Cartilage

Resting femoral articular cartilage outcomes	First segmentation of novice rater		Second segmentation of novice rater		Expert rater segmentation		Intrarater reliability of novice rater		Interrater reliability of novice and expert raters	
			Mean (SD)				ICC _{2,k} (95% CI [LB to UB])		ICC _{2,k} (95% CI [LB to UB])	
							SEM	MDC	SEM	MDC
Average thickness, mm	Medial	1.86 (0.28)	1.83 (0.28)	1.97 (0.25)		.993 (.959 to .998)	0.02	0.05	.944 (-.039 to .989)	0.06
	Lateral	1.76 (0.33)	1.72 (0.32)	1.86 (0.35)		.994 (.795 to .999)	0.03	0.06	.969 (.846 to .998)	0.06
	Intercondylar	2.23 (0.46)	2.17 (0.45)	2.30 (0.47)		.994 (.947 to .999)	0.03	0.08	.990 (.332 to .993)	0.05
Echo intensity	Medial	77.10 (6.78)	76.90 (6.82)	78.72 (7.24)		.999 (.996 to 1.000)	0.22	0.50	.983 (.352 to .997)	0.91
	Lateral	75.69 (3.98)	75.38 (3.96)	76.51 (4.40)		.997 (.812 to .999)	0.22	0.51	.978 (.961 to .997)	0.62
	Intercondylar	66.15 (4.08)	65.77 (4.05)	66.60 (4.36)		.997 (.955 to .999)	0.22	0.52	.991 (.860 to .994)	0.40

Abbreviations: CI, confidence interval; ICC, intraclass correlations coefficient; LB, lower bound; MDC, minimal detectable change; UB, upper bound.

The novice rater may underestimate the cartilage CSA by compensating to prevent capturing the white portions of the cartilage borders. Future training should aim to address this compensation in novice raters.

The novel semiautomated average cartilage thickness and echo intensity assessment is systematic and demonstrates excellent intrarater and interrater reliability between an expert and novice rater. This approach to ultrasound outcome processing offers a reliable, interpretable, and clinically feasible assessment in high-risk OA populations moving forward.

Acknowledgments

IRB approval no.: Tufts Medical Center IRB #12679. The authors report no conflicts of interest to disclose.

References

1. Kraus VB, Burnett B, Coindreau J, et al. Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis. *Osteoarthritis Cartilage*. 2011;19(5):515–542. PubMed ID: [21396468](#) doi:[10.1016/j.joca.2010.08.019](#)
2. Naredo E, Acebes C, Moller I, et al. Ultrasound validity in the measurement of knee cartilage thickness. *Ann Rheum Dis*. 2009; 68(8):1322–1327. PubMed ID: [18684742](#) doi:[10.1136/ard.2008.090738](#)
3. Harkey MS, Blackburn JT, Hackney AC, et al. Comprehensively assessing the acute femoral cartilage response and recovery after walking and drop-landing: an ultrasonographic study. *Ultrasound Med Bio*. 2018;44(2):311–320. doi:[10.1016/j.ultrasmedbio.2017.10.009](#)
4. Harkey MS, Blackburn JT, Nissman D, et al. Ultrasonographic assessment of femoral cartilage in individuals with anterior cruciate ligament reconstruction: a case-control study. *J Athl Train*. 2018; 53(11):1082–1088. PubMed ID: [30615493](#) doi:[10.4085/1062-6050-376-17](#)
5. Roberts HM, Moore JP, Thom JM. The reliability of suprapatellar transverse sonographic assessment of femoral trochlear cartilage thickness in healthy adults. *J Ultrasound Med*. 2019;38(4):935–946. PubMed ID: [30208236](#) doi:[10.1002/jum.14775](#)
6. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428. PubMed ID: [18839484](#) doi:[10.1037/0033-2909.86.2.420](#)
7. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17(4):571–582. PubMed ID: [17613642](#) doi:[10.1080/10543400701329422](#)
8. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1): 231–240. PubMed ID: [15705040](#)
9. Eckstein F, Ateshian G, Burgkart R, et al. Proposal for a nomenclature for magnetic resonance imaging based measures of articular cartilage in osteoarthritis. *Osteoarthritis Cartilage*. 2006;14(10):974–983. PubMed ID: [16730462](#) doi:[10.1016/j.joca.2006.03.005](#)
10. Cartwright MS, Kwayisi G, Griffin LP, et al. Quantitative neuromuscular ultrasound in the intensive care unit. *Muscle Nerve*. 2013; 47(2):255–259. PubMed ID: [23041986](#) doi:[10.1002/mus.23525](#)