B-Mode Ultrasonography Is a Reliable and Valid Alternative to Magnetic Resonance Imaging for Measuring Patellar Tendon Cross-Sectional Area

MARSHALL, Steven J <http://orcid.org/0000-0003-0458-3824>, HICKS, Kirsty M, HOWATSON, Glyn, COX, Lisa and THOMAS, Kevin

Ultrasound in Med. & Biol., Vol. 49, No. 2, pp. 578–587, 2023

0301-5629/$ - see front matter

https://doi.org/10.1016/j.ultrasmedbio.2022.10.011

● *Original Contribution*

B-MODE ULTRASONOGRAPHY IS A RELIABLE AND VALID ALTERNATIVE TO MAGNETIC RESONANCE IMAGING FOR MEASURING PATELLAR TENDON CROSS-SECTIONAL AREA

STEVEN J. MARSHALL,*,† KIRSTY M. HICKS,† GLYN HOWATSON,†,‡ LISA COX,§ and KEVIN THOMAS†

* Department of Service Sector Management, Sheffield Hallam University, Howard Street, Sheffield, South Yorkshire, UK;
† Department of Sport, Exercise and Rehabilitation, Northumbria University, Newcastle City Campus, Newcastle upon Tyne, UK;
‡ Water Research Group, North West University, Potchefstroom, South Africa; and § Newcastle Clinic, Independent House, Team Valley, Gateshead, UK

(*Received* 21 *June* 2022; *revised* 27 *September* 2022; *in final form* 14 *October* 2022)

Abstract—This study investigated the validity and reliability of measuring patellar tendon (PT) cross-sectional area (CSA) using magnetic resonance imaging (MRI) and ultrasound (US) imaging. Nineteen healthy participants (10 women, 9 men) participated in three imaging sessions of the PT, once via MRI and twice via US, with image acquisition conducted by two raters, one experienced (rater 2) and one inexperienced (rater 1). All PT segmentations were analyzed by both raters. The validity of US-derived estimates of PT CSA against MRI estimates was analyzed using linear regression. Within-day reliability of US and MRI measurements and between-day reliability of US measurements were quantified using typical error (TE) and intra-class correlation coefficients ($ICC_{3,1}$). There was good agreement between US- and MRI-derived estimations of PT CSA (standard errors of the estimate of 3.3 mm^2 for rater 1 and 2.6 mm^2 for rater 2; Pearson's $r = 0.97$ and 0.98 for raters 1 and 2, respectively). Within-session reliability for estimations of total PT CSA from US and MRI were excellent ($ICC_{3,1} > 0.95$, coefficient of variation [CV] <4.1%, TE = 1.3–3.6 mm^2. Between-day reliability for US was excellent ($ICC_{3,1} > 0.97$, CV <2.7%, TE = 1.6–2.3 mm^2), with little difference between raters. These findings suggest that MRI and US both provide reliable estimates of PT CSA and that US can provide a valid measure of PT CSA.

Key Words: **Reproducibility, Inter-rater, Intra-rater, Tendon morphology, Repeatability, Imaging, Magnetic resonance imaging.**

INTRODUCTION

The human patellar tendon (PT) plays a crucial role in locomotion by transmitting force from the quadriceps muscle group to the tibia, via the patellar. Tendon is a viscoelastic tissue and will deform under loading (Magnusson et al. 2008), with the degree of loading corresponding to the structural properties of the tissue (Maganaris and Paul 1999). These structural properties, such as tendon stiffness and Young's modulus (YM), determine the compliance of the tendon, which in turn can affect the behavior of the muscle–tendon unit during locomotion (Fukunaga et al. 2002). To calculate tendon stiffness and YM, the cross-sectional area (CSA) of the PT needs to be accurately measured. Moreover, measuring PT CSA can determine adaptations of the PT in response to mechanical loading (Kongsgaard et al. 2007; Couppe et al. 2008) or immobilization (Maganaris et al. 2006). Therefore, accurate and reliable measurements of PT CSA must be obtained to quantify the associated properties and adaptations of the PT.

Methods of assessing tendon morphology *in vivo* include magnetic resonance imaging (MRI) and 2-D B-mode ultrasound imaging (US). Several studies have validated the accuracy of MRI in measuring tendon properties (Berthoty et al. 1989; Sonin et al. 1996; Carrino et al. 1997), and it is considered the "gold standard" tool in validating other measurement techniques (Bohm et al. 2016; Kruse et al. 2017). However, previous research suggested that US outperformed MRI with respect to the reliability of

Address correspondence to: Steven J. Marshall, Food and Nutrition Subject Group, Department of Service Sector Management, Sheffield Hallam University, Howard Street, Sheffield, South Yorkshire, S1 1WB, UK.

measuring tendon morphology. Brushøj et al. (2006) found that US, when compared with MRI, had smaller within- and between-rater limits of agreement for Achilles tendon (AT) thickness. Moreover, the same study reported that US measures of AT thickness, CSA and width resulted in lower within-rater coefficients of variation, when compared with MRI. The use of US is recommended as a first-line imaging modality according to the last clinical indications of the European Society of Musculoskeletal radiology (Klauser et al. 2012). In addition, US is an attractive alternative to assess tendon properties because of its affordability, time efficiency, portability and non-invasive nature. Despite the widespread use of US in musculoskeletal research, the reliability of US tendon measures is debated within the literature (Gellhorn and Carlson 2013; McAuliffe et al. 2017). For example, US measures of PT CSA have been reported to be reliable when measured on multiple days (Reeves and Narici 2003), by multiple operators with different experience, using multiple machines (Gellhorn and Carlson 2013). In contrast, more recent studies have found US to be unreliable when measuring PT and AT CSA (Ekizos et al. 2013; Bohm et al. 2016), which was attributable, in part, to poor definition of tendon borders. With respect to the relationship between US and MRI, conflicting results have been reported in the literature. Albano et al. (2017) reported excellent agreement between MRI and US measures of AT (ICC = 0.986). Kruse et al. (2017) reported that intra-rater US measures AT were reliable, but not interchangeable with MRI measures, as US underestimated AT CSA by ~5.5%. Additionally, recent research by Stenroth et al. (2019) revealed that systematic differences between US and MRI measures of the PT were noted for inexperienced raters, with US underestimating PT CSA by 13.9% compared with MRI, but not for more experienced raters with more than 5 y of experience in musculoskeletal imaging and segmentation. This suggests a need to investigate the experience of the rater when assessing the reliability of US and MRI estimations of tendon measurements.

A typical approach when assessing tendon CSA is to measure the tendon at multiple sites, typically 25%, 50% and 75% of tendon length, and calculate an average based on those collective measures (Onambele et al. 2007; Hicks et al. 2013). However, studies investigating the reliability of US- and MRI-derived measures of tendons have only reported the results of the combined averages of the tendon and not each specific measurement site, despite taking multiple measurements along the tendon (Kruse et al. 2017; Stenroth et al. 2019). Therefore, whether reliability differs between measurement sites for both US and MRI remains unknown, warranting further investigation.

An additional consideration when using estimates of tendon CSA to calculate structural properties such as tendon stiffness and YM is joint angle. Typically, PT stiffness and YM are calculated with the participant performing a ramped, isometric maximal voluntary contraction (iMVC) in an isokinetic dynamometer with the knee angle fixed at 90°. As PT CSA is an integral part of the equation used to calculate PT YM (PT stiffness × [PT length {mm}/PT CSA {mm^2}]) (Onambele et al. 2007), it would be prudent to calculate PT CSA at the knee angle relevant to the iMVC being performed. This would minimize any miscalculations caused by changes in PT CSA as a result of Poisson's ratio, whereby diameter would decrease at a constant to the strain (Poisson 1827), which would occur with an increase in knee joint angle. However, to date, no other study investigating the validity and reliability of PT CSA via US or MRI has employed a knee angle of 90°; therefore, the effects of knee joint angle on the accuracy of these measures remain unknown.

Collectively, these data suggest that the reliability and validity of US and MRI measures of tendon CSA are inconsistent and require further investigation. Therefore, the aims of this study were threefold: (i) to determine the agreement between US and MRI measures of PT CSA for two independent raters; (ii) to determine the within-day, inter- and intra-rater reliability for US and MRI measures of PT CSA; and (iii) to determine the between-day, inter- and intra-rater reliability of US measures of PT CSA.

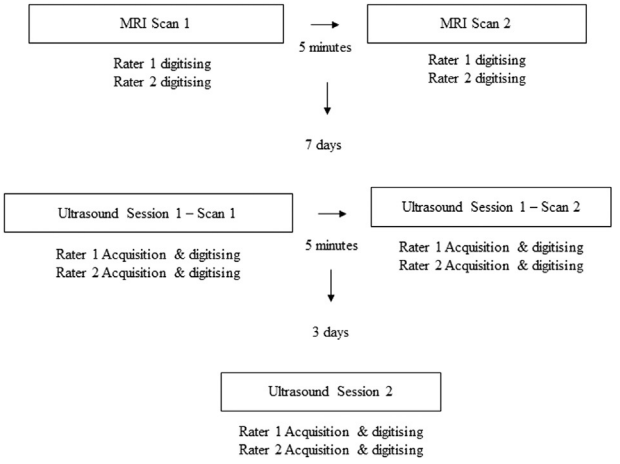
METHODS

Participants

Nineteen healthy participants, 10 women and 9 men, participated in the study (age: 25 ± 6 y, stature: 1.71 ± 0.10 m, mass: 71.3 ± 12.5 kg). Participants completed a pre-test questionnaire and were included in the study only if they had had no neuromuscular or musculoskeletal impairments in the lower limbs within the last 6 mo. Contraindications to MRI included cardiac pacemaker, metal objects in the body (such as aneurysm clips or a programmable shunt in the brain), joint prostheses, bone fixation devices and pregnancy. Institutional ethical approval was received from the Northumbria University Faculty of Health & Life Sciences Research Ethics Committee in accordance with the Declaration of Helsinki. Participants were supplied with a participant information sheet detailing the purpose of the study and provided written consent before participating.

Experimental design

Participants were asked to visit the laboratory on three occasions. Figure 1 is a schematic of the experimental protocol. The first session was the imaging of the PT using MRI. On the second (1 wk after the first visit)



Comparison type	Comparisons made
	US vs MRI
Within-day, intra-rater	Rater 1 (MRI scan 1) vs Rater 1 (US session 1, scan 1)
Within-day, intra-rater	Rater 2 (MRI scan 1) vs Rater 2 (US session 1, scan 1)
Within-day, intra-rater	Combined (MRI scan 1) vs Combined (US session 1, scan 1)
	MRI Reliability
Within-day, intra-rater	Rater 1 (MRI scan 1) vs Rater 1 (MRI scan 2)
Within-day, intra-rater	Rater 2 (MRI scan 1) vs Rater 2 (MRI scan 2)
Within-day, inter-rater	Rater 1 (MRI scan 1) vs Rater 2 (MRI scan 1)
	US Reliability
Within-day, intra-rater	Rater 1 (US session 1, scan 1) vs Rater 1 (US session 1, scan 2)
Between-day, intra-rater	Rater 1 (US session 1, scan 1) vs Rater 1 (US session 2)
Within-day, intra-rater	Rater 2 (US session 1, scan 1) vs Rater 2 (US session 1, scan 2)
Between-day, intra-rater	Rater 2 (US session 1, scan 1) vs Rater 2 (US session 2)
Within-day, inter-rater	Rater 1 (US session 1, scan 1) vs Rater 2 (US session 1, scan 1)
Between-day, inter rater	Rater 1 (US session 1, scan1) vs Rater 2 (US session 2)

Fig. 1. Study design highlighting the role of each rater. The table explains each of the comparisons made to assess validity and reliability. MRI = magnetic resonance imaging; PT = patellar tendon; US = ultrasound.

and third visits, the PT was imaged using US twice, with a 3-d interval. In session 1, two MRI scans of the PT were performed, separated by a 5-min interval. To determine the reliability of the MRI measures, the participant was then removed from the MRI scanner before being repositioned and scanned again. In sessions 2 and 3, two raters each performed two US scans of the PT on the same leg. The participant was then removed from the scanning position, before being repositioned and undergoing US scans again. This resulted in four US scans per rater, per visit (eight US scans in total). Rater 1 was considered less experienced; however, training in image acquisition using US, and image digitisation and analysis was provided in depth before the study by rater 2, who had >5 y of experience in musculoskeletal radiography. Imaging was performed at the same time of day in each session to remove the potential diurnal effects on tendon size (Stenroth et al. 2019). Prior to each visit, participants were asked to refrain from strenuous lower body exercise for 48 h to reduce possible deformations in the PT structure caused by fluid ecchymosis.

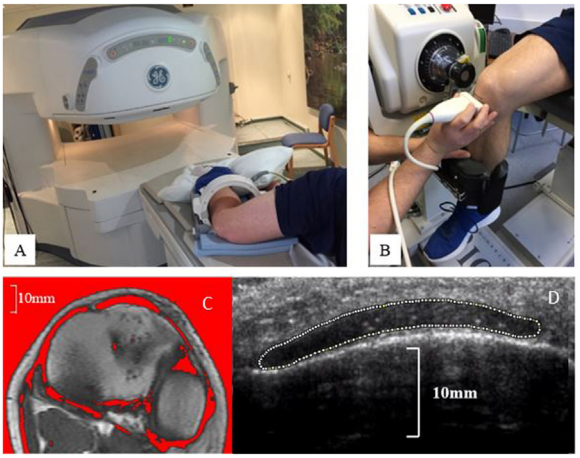


Fig. 2. Scanning position prior to the MRI scan (A) and scanning position for US imaging (B). The axial MRI image reveals the automatically outlined PT CSA (C) and the corresponding axial US image reveals the manually outlined PT CSA (D) using ImageJ software (ImageJ 1.45, National Institutes of Health, Bethesda, MD, USA). CSA = cross-sectional area; MRI = magnetic resonance imaging; PT = patellar tendon; US = ultrasound.

Procedures

MRI examinations. Participants were placed in an open MRI device (GE Ovation 0.35 T open MRI scanner, GE Healthcare, Little Chalfont, UK) in a left decubitus position, with the right hip and knee flexed to 85° and 90° (0° = full extension), respectively (Fig. 2a), confirmed using a goniometer. This positioning was chosen to mirror the hip and knee angles of participants during the US measurements. All MRI procedures were performed by a qualified radiographer after positioning of the knee had been confirmed by rater 1.

The MRI scanning procedure was divided into three sections: a localizer scan to confirm the correct field of view, a sagittal plane scan (spin echo T1, TR/TE 500/24, field of view 20 × 15 cm, slice thickness 4 mm, spacing between slices = 0 mm, 3:45-min scan time) and an axial scan (spin echo T1, TR/TE 450/24, field of view 20 × 15 cm, slice thickness 4 mm, spacing between slices 0 mm, 5:31-min scan time).

Ultrasound examinations. Participants were positioned in an isokinetic dynamometer (System 4 Pro, Biodex Medical Systems Inc., Shirley, NY, USA) in a seated position, with the hip flexed at 85° and the knee flexed at 90° (0° = full extension) (Fig. 2b). A real-time B-mode ultrasound (HDI 5000 SonoCT, Philips,

Amsterdam, Netherlands) and conductance gel (Aquasonic 100, Parker Laboratories INC, NJ, USA) were used to assess PT CSA and PT length. Sagittal images of the PT were obtained using a US probe (7.5-MHz linear array probe, 55-mm width) to locate the apex of the patellar and tibial tuberosities, with marks placed on the skin at each site. The distance between the two sites was measured via an inextensible anthropometric tape measure and taken as PT length. Patellar tendon CSA was measured in the axial plane at 25% (proximal), 50% (mid) and 75% (distal) of PT length, with the scan locations clearly marked on the skin using a permanent marker. Ultrasound images were captured live using image acquisition software (AVer Media Capture Studio, AVer Media Technologies, New Taipei City, Taiwan) and analyzed offline. Patellar tendon CSA images were obtained by two US operators. Within each US session, when the participant was removed from the dynamometer for 5 min, the scan location marks were removed from the skin before the patient was repositioned. The procedure was then repeated to allow for within-session reliability assessment.

MRI image analysis. Sagittal MRI images, which had a corresponding axial image that could be used to measure PT CSA, were used to locate the apex of the patellar and tibial tuberosities. This was to ensure the consistency of anatomical landmarks used to determine PT length during US examinations. The number of images between the axial image for the apex of the patellar and tibial tuberosities was used to determine the PT CSA image at 25%, 50% and 75% of PT length. For example, if 12 images lay between the apex of the patellar and the tibial tuberosity, images 3, 6 and 9 were analyzed for 25%, 50% and 75% PT length, respectively. When the appropriate point lay between two images, the image toward the proximal region of the PT was analyzed.

Images were exported and analyzed by digitizing software (ImageJ 1.45, National Institutes of Health, Bethesda, MD, USA). Images were first converted to 32-bit grayscale. An adjustable threshold cutoff method was used to determine PT borders (Kruse et al. 2017). The threshold was adjusted until the smallest natural appearance of the PT was achieved (Fig. 2c); PT CSA was taken as the area within this border. Both raters performed this sequence twice for each image, with the mean PT CSA recorded for further analysis. All images were independently blinded and randomised for both raters prior to analysis to reduce researcher bias.

Ultrasound image analysis. Ultrasound videos were exported to video editing software (Adobe Premier Elements version 15, Adobe, Mountain View, CA, USA) for frame-by-frame analysis. The images at the appropriate PT CSA location were manually assessed before being exported for analysis in ImageJ software. The tendon border was then manually outlined, and the CSA was calculated (Fig. 2d). Each rater manually analyzed each image twice, with the mean PT CSA used for further analysis. All US images were independently blinded before both raters analyzed all images in a randomised order to prevent the possibility of systematic bias resulting from recalling previous analysis.

Statistical analysis

Data are expressed as the mean ± standard deviation. The level of significance was set to $\alpha = 0.05$. Data were analyzed using a published spreadsheet (Hopkins 2015) in Microsoft Excel (Microsoft Excel 2016, Microsoft, Washington DC, USA) as follows:

Agreement between MRI- and US-derived measures of PT CSA was assessed for each rater individually, and the collapsed scores of both raters, via linear regression (Hopkins, 2015). Pearson's correlation coefficients and the standard error of the estimate (SEE) were calculated to quantify agreement, and paired sample *t*-tests were used to assess for systematic error. The standard error of measurement (SEM) was calculated as the square root of the mean square error from a one-way analysis of variance (Stenroth et al. 2019).

Within-day, intra-rater reliability was assessed for MRI and US images by comparing the PT CSA scores from each scan performed on the respective visits (two scans per visit). Between-day, intra-rater reliability was assessed for US images by comparing the PT CSA scores of the first scan for each rater (before the participant was repositioned) for each visit. Within-day inter-rater reliability for MRI and US images was assessed by comparing the PT CSA scores of raters 1 and 2 during visit 1. Reliability was assessed for the proximal, mid and distal PT CSA images individually, in addition to the mean of all three PT CSA scores. The relative reliability of MRI and US measurements was assessed using $ICC_{3,1}$, while absolute reliability was assessed by calculating the SEM and calculating the TE (95% confidence intervals) expressed as raw units and as a coefficient of variation (CV %). Paired sample *t*-tests were implemented to assess for systematic error. Reliability via ICC was interpreted as follows: ICC 0.5−0.75, moderately reliable; ICC 0.75−0.9, good reliability; ICC >0.9, excellent reliability (Koo and Li 2016).

RESULTS

Ultrasound versus MRI

Mean ± SD PT CSA measures for US and MRI for both raters are presented in Table 1. No systematic differences between US and MRI were present when

Table 1. Mean values of patellar tendon cross-sectional area*

	Rater 1			Rater 2			Combined†
	Session 1		Session 2	Session 1		Session 2	Session 1
	Measure 1	Measure 2	Measure 1	Measure 1	Measure 2	Measure 1	Measure 1
Ultrasound Imaging							
Proximal	81.4 ± 12.5	82.6 ± 14.0	83.3 ± 14.7	83.5 ± 14.3	84.9 ± 14.2	84.9 ± 14.2	84.2 ± 14.3
Mid	87.5 ± 13.5	87.8 ± 14.1	88.4 ± 14.2	89.3 ± 14.7	89.0 ± 14.2	89.0 ± 14.2	88.0 ± 15.5
Distal	91.3 ± 14.2	92.0 ± 13.1	90.9 ± 14.8	91.1 ± 14.1	91.2 ± 13.6	91.2 ± 13.6	91.5 ± 14.6
Mean‡	86.7 ± 12.6	87.5 ± 13.1	87.6 ± 13.9	88.0 ± 14.0	88.4 ± 13.5	88.4 ± 13.5	87.4 ± 13.2
Magnetic resonance imaging							
Proximal	84.0 ± 14.7	82.4 ± 14.0	—	84.5 ± 14.0	83.1 ± 13.0	—	82.5 ± 13.2
Mid	87.8 ± 14.9	88.7 ± 15.3	—	88.3 ± 16.2	88.3 ± 15.4	—	88.4 ± 13.8
Distal	91.1 ± 16.6	90.4 ± 13.8	—	90.8 ± 15.4	91.9 ± 13.9	—	91.2 ± 13.9
Mean‡	87.6 ± 14.3	87.2 ± 13.5	—	87.8 ± 13.9	87.8 ± 13.1	—	87.7 ± 14.0

Data are expressed as the mean ± standard deviation in meters.
† Combined mean values from raters 1 and 2.
‡ Mean of the proximal, mid and distal values.

proximal, mid and distal values were averaged for both raters individually and when values were averaged across both raters ($p = 0.055-0.785$) (Table 2). For rater 1, there was evidence of a small systematic bias, as US underestimated MRI PT CSA by 2.6 mm^2 ($p = 0.017$) and 5.3 mm^2 ($p = 0.008$) for proximal and mid measurements, respectively, compared with MRI. Pearson's r ranged from 0.88 to 0.96 for location-specific measures and from 0.97 to 0.98 for combined scores, with similar scores between raters. Combined scores between both raters exhibited good agreement between US and MRI, with SEEs that were in the range 3.8 to 4.7 mm^2 for location-specific measures and 2.4 mm^2 when scores were combined. Pearson's r ranged from 0.95 to 0.97 for location-specific measures and 0.98 for combined scores (Table 2). Figure 3 illustrates an excellent association between US and MRI for rater 1 ($r^2 = 0.95$), rater 2 ($r^2 = 0.97$) and combined rater 1 and 2 measurements ($r^2 = 0.97$).

Table 2. Agreement between US and MRI measures of patellar tendon cross-sectional area

	Bias (95% CI) (mm^2)	p	SEE (mm^2)	r
Rater 1				
Proximal	−2.6 (−4.6, 0.5)	0.017*	4.0	0.96
Mid	−5.3 (−9.0, 1.5)	0.008*	7.9	0.86
Distal	0.2 (−3.6, 4.0)	0.905	8.1	0.88
Mean†	−0.9 (−2.6, 0.8)	0.278	3.3	0.97
Rater 2				
Proximal	−0.9 (−3.1, 1.3)	0.395	4.6	0.95
Mid	1.1 (−1.8, 3.9)	0.431	6.1	0.93
Distal	0.3 (−3.1, 3.7)	0.863	7.3	0.89
Mean†	0.2 (−1.1, 1.4)	0.793	2.6	0.98
Combined				
Proximal	−1.7 (−3.5, 0.1)	0.055	3.8	0.97
Mid	0.4 (−1.5, 2.3)	0.673	3.9	0.97
Distal	−0.3 (−2.5, 1.9)	0.785	4.7	0.95
Mean†	0.4 (−0.9, 1.6)	0.541	2.4	0.98

CI = confidence interval; MRI = magnetic resonance imaging; p = probability using paired sample t-test; SEE = standard error of the estimate; r = Pearson correlation coefficient; US = ultrasound.
* Significant difference between US and MRI.
† Mean of the proximal, mid and distal values.

Ultrasound and MRI within-day intra-rater reliability

For rater 2, US analysis overestimated PT CSA by 1.4 mm^2, in measure 2 compared with measure 1 ($p = 0.028$), for the proximal PT CSA. No other systematic differences between the first and second measures were found for US or MRI analysis for raters 1 and 2 ($p = 0.117-0.997$).

The mean TE, CV and ICC were similar for both raters for both US and MRI (Table 3). Within-day measures were good (ICC ≥ 0.81) for rater 1 distal MRI, rater 2 proximal MRI and distal MRI analysis. All other within-day measures were considered excellent (ICC ≥0.91). Association between measurements 1 and 2 was excellent for US ($r^2 = 0.98$) for both raters (Fig 4a). Association between measurements 1 and 2 for MRI was similar for both raters (rater 1 = 0.89, rater 2 = 0.91; Fig. 3b).

Ultrasound between-day intra-rater reliability

There were no systematic differences between visits for either rater ($p = 0.096-0.737$). The typical error for rater 1 ranged from 3.2 to 3.5 mm^2 for location-specific measures and was 2.3 mm^2 for combined scores (Table 4). The typical error for rater 2 ranged from 2.6 to 3.7mm^2 for location-specific measures and was 1.6 mm^2 for combined scores. All between-day measures were considered excellent (ICC ≥0.94). Figure 4 depicts excellent association for between-day measurements for rater 1 ($r^2 = 0.95$) and rater 2 ($r^2 = 0.98$).

Ultrasound within-day inter-rater reliability

No systematic differences were found between raters for within-day MRI or US analysis

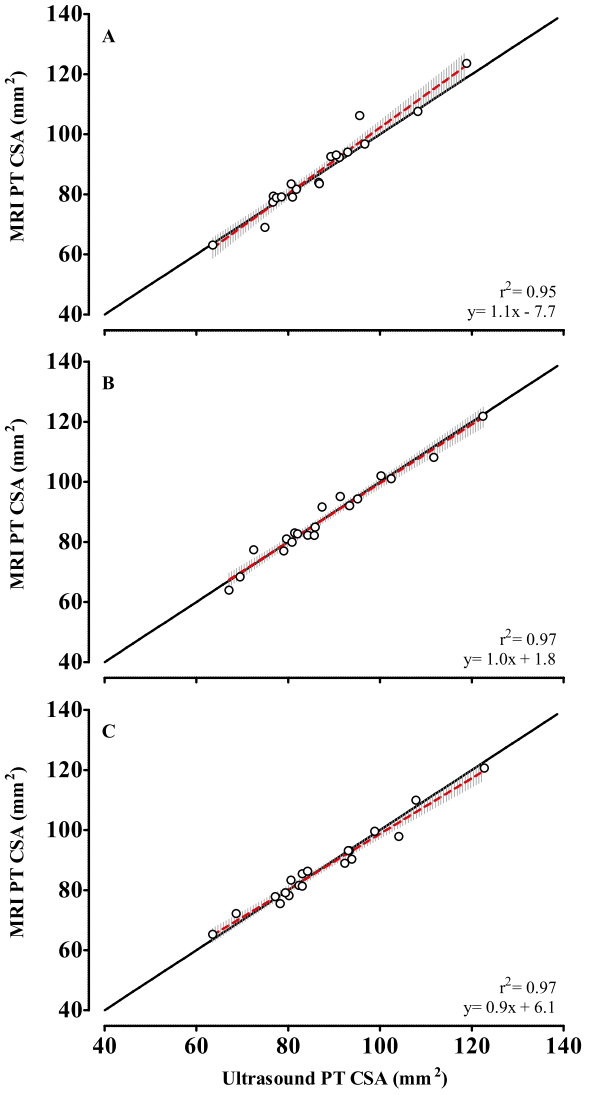


Fig. 3. Individual data points for the agreement between estimations of PT CSA using US imaging and MRI. (A) Rater 1. (B) Rater 2. (B) Raters 1 and 2 combined. The *solid line* represents the line of equality. The *dashed red line* denotes the regression line with corresponding 95% confidence intervals represented by the gray hashed area. CSA = cross sectional area; MRI = magnetic resonance imaging; PT = patellar tendon.

(p = 0.127−0.890). Typical errors for US analysis ranged from 3.3 to 4.3 mm^2 for location-specific measures and averaged 2.4 mm^2 for combined scores (Table 5). Typical errors for MRI analysis ranged from 2.2 to 2.8 mm^2 for location-specific measures and averaged 1.5 mm^2 for combined scores. All within-day, inter-rater scores were considered excellent (ICC ≥0.92). Figure 5a illustrates that within-day associations between raters was excellent for both MRI (r^2 = 0.98) and US (r^2 = 0.94).

Ultrasound between-day inter-rater reliability

Rater 1 underestimated PT CSA by 2.7 mm^2 in comparison to rater 2 (p = 0.033), at the proximal PT CSA site (Table 5). There were no other systematic differences between raters (p = 0.351−0.572). Typical errors for US analysis ranged from 3.5 to 4.0 mm^2 for location-specific measures and averaged 2.5 mm^2 for combined scores. All between-day, inter-rater scores were considered excellent (ICC ≥0.93). Figure 5b illustrates an excellent between-day association between raters 1 and 2 (r^2 = 0.94).

DISCUSSION

The aims of this study were to determine the agreement between US and MRI measures of PT CSA for two independent raters, determine the within-day inter- and intra-rater reliability for US and MRI measures of PT CSA and determine the between-day inter- and intra-rater reliability of US measures of PT CSA. This study indicates that there are high levels of agreement between US- and MRI-derived measures of PT CSA. Moreover, both US and MRI provide reliable within-day inter- and intra-rater measures of PT CSA. Finally, US provides reliable between-day, inter- and intra-rater measures of PT CSA. These findings illustrate that US provides a valid and reliable assessment of PT CSA, which increases confidence in downstream measures of tendon properties, such as tendon stiffness and Young's modulus.

Validity of ultrasound versus MRI

Previous studies investigating US *versus* MRI have reported conflicting results, with US both over-estimating (Stenroth et al. 2019) and under-estimating (Kruse et al. 2017) tendon CSA when compared with MRI. However, this study indicated that high levels of agreement existed between US and MRI and that similar tendon CSA measures were produced. Though there was systematic under-reporting of proximal and mid PT CSA measures by US, there were no systematic differences between US and MRI for either rater when all sites (proximal, mid and distal) were combined for each participant. This is an important finding, as the mean score is commonly used to estimate average tendon CSA and subsequently calculate tendon stiffness and YM (Maganaris and Paul 1999; Kongsgaard et al. 2007; Onambele et al. 2007; Couppe et al. 2008; Hicks et al. 2013; Couppe et al. 2016; Murtagh et al. 2018; Stenroth et al. 2019). This high level of agreement with MRI suggests that the more convenient and cost-effective method of US can be confidently used to measure PT CSA.

Table 3. Within-day intra-rater reliability for estimates of patellar tendon cross-sectional area using US and MRI

	Bias (95% CI) (mm^2)	*p*	TE (95% CI) (mm^2)	CV (95% CI)	ICC	SEM (mm^2)	SEM%
Rater 1 US							
Proximal	1.2 (−0.5, 2.9)	0.158	2.5 (1.9, 3.7)	2.6 (1.9, 3.8)	0.97	3.6	4.2
Mid	0.3 (−0.8, 1.2)	0.606	1.5 (1.1, 2.2)	1.7 (1.3, 2.5)	0.99	0.8	0.8
Distal	0.7 (−1.1, 2.5)	0.407	2.7 (2.0, 3.9)	3.0 (2.3, 4.5)	0.97	2.3	2.3
Mean†	0.7 (−0.2, 1.6)	0.117	1.4 (1.0, 2.0)	1.6 (1.2, 2.4)	0.99	2.2	2.5
Rater 1 MRI							
Proximal	−1.6 (−4.7, 1.6)	0.318	4.6 (3.5, 6.9)	6.0 (4.5, 8.9)	0.91	4.8	5.3
Mid	0.9 (−2.3, 4.1)	0.560	4.6 (3.5, 6.9)	5.8 (4.3, 8.7)	0.92	2.8	2.9
Distal	−0.6 (−5.2, 4.0)	0.778	6.7 (5.1, 9.9)	7.6 (5.7, 11.4)	0.82	1.9	1.9
Mean†	−0.4 (−2.7, 1.9)	0.700	3.6 (2.5, 5.0)	4.1 (3.1, 6.1)	0.95	1.3	1.4
Rater 2 US							
Proximal	1.4 (0.2, 2.5)	0.028*	1.8 (1.3, 2.6)	2.1 (1.6, 3.1)	0.99	4.2	4.6
Mid	−0.4 (−1.8, 1.0)	0.574	2.0 (1.5, 3.0)	2.3 (1.7, 3.4)	0.98	1.2	1.2
Distal	0.1 (−1.9, 2.2)	0.894	3.0 (2.3, 4.4)	3.2 (2.4, 4.8)	0.96	0.4	0.4
Mean†	0.4 (−0.5, 1.3)	0.400	1.3 (1.0, 2.0)	1.5 (1.1, 2.2)	0.99	1.1	1.2
Rater 2 MRI							
Proximal	−1.3 (−4.8, 2.1)	0.423	5.1 (3.8, 7.5)	6.4 (4.8, 9.6)	0.87	4.1	4.6
Mid	0.0 (−3.1, 3.1)	0.997	4.5 (3.4, 6.7)	5.5 (4.1, 8.2)	0.93	0.0	0.0
Distal	1.1 (−3.5, 5.7)	0.626	6.7 (5.1, 9.9)	8.0 (6.0, 12.0)	0.81	3.3	3.4
Mean†	−0.1 (−2.1, 2.0)	0.931	3.0 (2.3, 4.5)	3.7 (2.8, 5.6)	0.96	0.3	0.3

CI = confidence interval; CV = coefficient of variation expressed as a percentage; ICC = intra-class correlation coefficient; *p* = paired sample *t*-test; MRI = magnetic resonance imaging; SEM = standard error of measurement expressed as a percentage of the mean; TE = typical error; US = ultrasound imaging.

* Significant difference between measures 1 and 2.

† Mean of the proximal, mid and distal values.

Within-day intra-rater reliability

The within-day, intra-rater reliability for both raters was excellent for both US and MRI, with slightly more favorable ICC estimates, relative reliability and absolute reliability for US compared with MRI. However, rater 2 produced a smaller estimation of PT CSA on measure 2 in comparison to measure 1 (1.4 mm^2) for US, whereas no systematic differences between measures for MRI were reported. This small systematic difference, in US measures, could be attributed to a small adjustment in probe orientation while scanning, as this can result in an increased diameter when positioned slightly askew (Gellhorn and Carlson 2013). Nevertheless, the systematic difference in this study was confined to the proximal site of the PT, with no difference occurring when the three locations were combined.

To the best of our knowledge, this study is the first to investigate the within-day, intra-rater reliability of MRI estimates of PT CSA. Two comparisons that could be made are from Kubo et al. (2001), who reported a CV of 1.6%, and Stenroth et al. (2019), who reported CVs of 4.1% and 6.0% for experienced and inexperienced raters, respectively; both studies assessed PT CSA estimations by MRI over 2 separate days. In comparison, CVs in the current study were 4.1% and 3.7% for the experienced and inexperienced raters, respectively. It is possible that the higher reliability displayed by Kubo et al. (2001) is due to the small sample size of 6 participants, which can affect estimates of error (Springate 2011), in comparison to the 19 in this study and the 15 participants in the study by Stenroth et al. (2019). Nevertheless, the data from the current study suggest both MRI and US measures of PT CSA indicate excellent within-day intra-rater reliability.

Between-day intra-rater reliability

This study reported that US resulted in excellent relative and absolute between-day intra-rater reliability, by both raters, comparing more favorably than in previous work. For example, Stenroth et al. (2019) reported higher absolute reliability in comparison to the two raters in the current study, with SEMs of 5.0 and 8.9 mm^2 *versus* 1.5 and 2.6 mm^2, respectively. Reliability assessed by ICC in this study was higher for both raters (ICC = 0.94−1.00) in comparison to other studies. For example, Stenroth et al. (2019) reported ICCs of 0.87 and 0.50 for experienced and inexperienced raters, respectively. Ekizos et al. (2013) also reported lower reliability (mean ICC 0.60) than the current study, which was attributed to limited visibility of the tendon border, making structure identification difficult. In this study and previous work (Stenroth et al. 2019), anatomical landmarks were used to define the origin and insertion of the PT, whereby the proximal, mid and distal sites were calculated based on these measurements, which was repeatedly done on each visit. This highlights the importance of a rigorous testing protocol which might, in turn, improve reliability (Thoirs and Childs 2018).

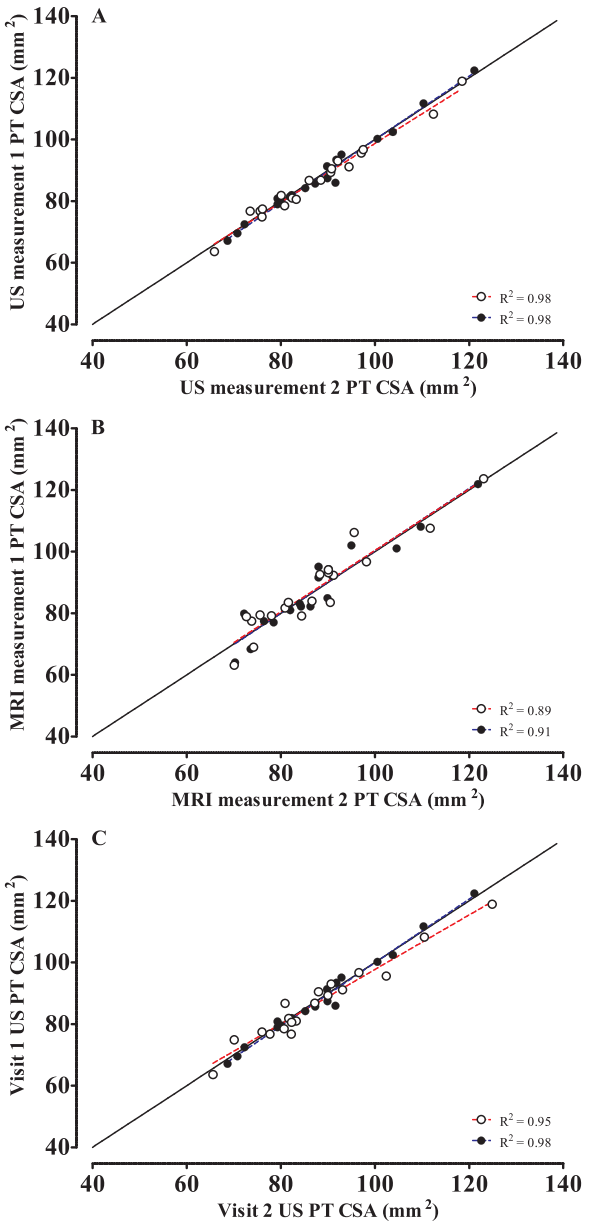


Fig. 4. (A) Within-day intra-rater reliability of estimates of PT CSA using US imaging. (B) Within-day, intra-rater reliability of estimates of PT CSA using MRI. (C) Between-day intra-rater reliability of estimates of PT CSA using US. The *solid black line* represents the line of equality. The *dashed red line* denotes the regression line for rater 1, and the *dashed blue line* denotes the regression line for rater 2. The *white dots* represent individual data points for rater 1. The *black dots* represent individual data points for rater 2. PT = patellar tendon; CSA = cross sectional area; US = ultrasound imaging; MRI = magnetic resonance imaging.

Inter-rater reliability

For both US and MRI, within- and between-day inter-rater reliability was excellent (ICC ≥0.92), with no systematic differences present for within-day measures. Despite a systematic difference between raters at the proximal site for between-day measures, this did not result in a systematic difference when the three measurement sites were combined. Inter-rater reliability was considerably higher in this study than in Stenroth et al. (2019) for both relative (US ICCs 0.97 *vs.* 0.56, MRI ICCs 0.99 *vs.* 0.62) and absolute (US SEM 0.7 mm^2 *vs.* 6.0 mm^2) reliability. The large inter-rater differences between the two studies might be attributable to differences in the experience of the raters in Stenroth et al. (2019), with the inexperienced rater having no prior experience in musculoskeletal radiography. In the current work, although rater 1 was less experienced than rater 2, there was a substantial level of practice with the digitisation process prior to the study onset. There is little doubt that experience can improve the reliability of US measures (Dudley-Javoroski et al. 2010), although it remains to be determined exactly what level of experience might be needed to produce high levels of reliability, but demonstrable high levels of reliability seem to be a good index of competence.

Limitations

The present study provides important methodological evidence which will allow the valid and reliable use of US and MRI in estimating PT CSA. However, this study is not without its limitations. Specifically, the estimation of PT CSA for both US and MRI were based on the judgements of the raters and their interpretation of the tendon borders. Although agreement between the two studies was excellent, it cannot be ruled out that the true CSA is what was measured by MRI analysis. It is difficult to ascertain if both the US and MRI images included the paratenon because of it not being clearly identifiable (Bohm et al. 2016). This gross over- or underestimation might have consequences for subsequent mechanical calculations pertaining to PT CSA (*e.g.,* Young's modulus), and while within-study comparisons would not be affected, extrapolation to other populations might be difficult.

Another limitation is the time period between the test days of the US measurements. With only 3 d between measures, the test−retest reliability of scores over longer periods is not known. While this approach ensures that the US measures are comparable, it does not consider the potential change in diameter of tendons that can occur over time with exercise (Tardioli et al. 2012). Finally, caution must be taken if future research utilizes equipment different from that used in the present study or uses raters with different musculoskeletal radiography experience, as this might affect the reliability of any subsequent results.

Table 4. Between-day intra-rater reliability for estimates of patellar tendon cross-sectional area using ultrasound imaging

	Bias (95% CI) (mm^2)	*p*	TE (95% CI) (mm^2)	CV (95% CI)	ICC	SEM (mm^2)	SEM%
Rater 1							
Proximal	2.0 (−0.4, 4.3)	0.096	3.4 (5.6, 5.0)	4.1 (3.1, 6.1)	0.94	6.0	6.9
Mid	0.9 (−1.5, 3.4)	0.426	3.5 (2.7, 5.2)	4.1 (3.1, 6.2)	0.94	2.9	3.1
Distal	−0.4 (−2.5, 1.8)	0.737	3.2 (2.4, 4.7)	3.4 (2.6, 5.1)	0.96	1.1	1.1
Mean†	0.9 (−0.7, 2.4)	0.278	2.3 (1.8, 3.5)	2.7 (2.0, 4.0)	0.97	2.6	2.9
Rater 2							
Proximal	0.5 (−1.5, 2.6)	0.599	3.0 (2.3, 4.5)	3.8 (2.9, 5.7)	0.96	1.6	1.8
Mid	−1.2 (−3.7, 1.3)	0.335	3.7 (2.8, 5.5)	4.2 (3.2, 6.3)	1.00	3.7	4.2
Distal	−0.8 (−2.6, 1.0)	0.339	2.6 (2.0, 3.9)	3.0 (2.2, 4.4)	0.97	2.6	2.7
Mean†	1.6 (1.2, 2.3)	0.344	1.6 (1.2, 2.3)	1.8 (1.4, 2.7)	1.00	1.5	1.6

CI = confidence interval; *p* = paired sample *t*-test; CV = coefficient of variation expressed as a percentage; ICC = intraclass correlation coefficient; SEM = standard error of measurement; SEM% = standard error of measurement expressed as a percentage of the mean; TE = typical error.
† Mean of the proximal, mid and distal values.

Table 5. Rater 1 *versus* rater 2: Inter-rater reliability for estimates of patellar tendon cross-sectional area using US and MRI

	Bias (95% CI) (mm^2)	*p*	TE (95% CI) (mm^2)	CV (95% CI)	ICC	SEM (mm^2)	SEM%
			Within-day				
US							
Proximal	2.2 (−0.4, 4.7)	0.890	3.7 (2.8, 5.5)	4.6 (3.4, 6.8)	0.93	6.6	7.6
Mid	1.8 (−1.0, 4.8)	0.203	4.3 (3.2, 6.3)	4.8 (3.6, 7.2)	0.92	5.7	6.0
Distal	−0.2 (−2.4, 2.0)	0.858	3.3 (2.5, 4.8)	3.9 (3.0, 5.9)	0.95	0.6	0.6
Mean†	1.3 (−0.4, 2.9)	0.127	2.4 (1.8, 3.6)	3.0 (2.3, 4.5)	0.97	3.9	4.3
MRI							
Proximal	0.5 (−1.0, 2.0)	0.503	2.2 (1.7, 3.3)	3.0 (2.3, 4.5)	0.98	1.5	1.7
Mid	0.4 (−1.5, 2.4)	0.649	2.8 (2.2, 4.2)	3.2 (2.4, 4.8)	0.97	1.3	1.4
Distal	−0.3 (−1.6, 1.1)	0.691	2.0 (1.5, 2.9)	2.1 (1.6, 3.1)	0.99	0.8	0.8
Mean†	0.2 (−0.8, 1.2)	0.653	1.5 (1.1, 2.2)	1.6 (1.2, 2.4)	0.99	0.7	0.8
			Between-day				
US							
Proximal	2.7 (0.2, 5.1)	0.033*	3.6 (2.7, 5.3)	4.6 (3.4, 6.8)	0.94	3.6	4.3
Mid	0.7 (−1.7, 3.0)	0.572	3.5 (2.6, 5.2)	4.2 (3.1, 6.2)	0.94	3.5	4.0
Distal	−1.0 (−3.8, 1.8)	0.440	4.0 (3.0, 5.9)	4.9 (3.7, 7.3)	0.93	4.0	4.4
Mean†	0.8 (−0.9, 2.5)	0.351	2.5 (1.7, 2.5)	3.2 (2.4, 4.7)	0.97	2.5	2.8

CI = confidence interval; CV = coefficient of variation expressed as a percentage; ICC = intraclass correlation coefficient; MRI = magnetic resonance imaging; *p* = paired sample t-test; SEM = standard error of measurement; SEM% = standard error of measurement expressed as a percentage of the mean; TE = typical error; US = ultrasound imaging.
* Significant difference between raters
† Mean of the proximal, mid and distal values.

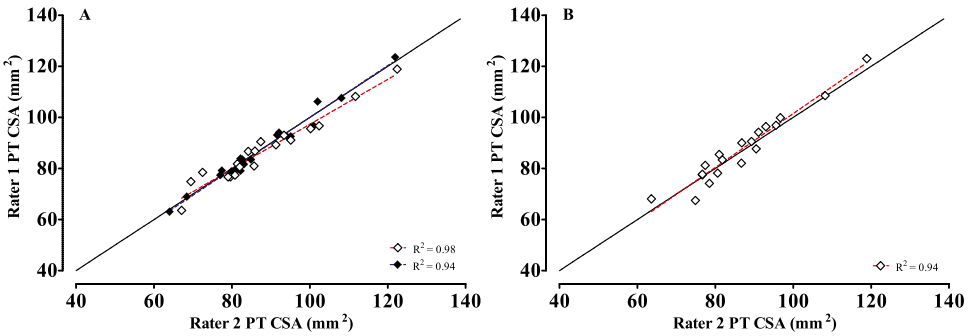


Fig. 5. (A) Within-day inter-rater reliability of estimates of PT CSA using US imaging and MRI. (B) Between-day inter-rater reliability of estimates of PT CSA using US imaging. The *solid black line* represents the line of equality. In (A), the *dashed red line* denotes the regression line for US, and the *dashed blue line* denotes the regression line for MRI. In (B), the *dashed red line* denotes the regression line for US imaging. The *white diamonds* represent data points for US imaging. The *black diamonds* represent data points for MRI. PT = patellar tendon; CSA = cross sectional area; US = ultrasound imaging; MRI = magnetic resonance imaging.

CONCLUSIONS

Measures of PT obtained using US are comparable to those obtained with MRI, which gives researchers confidence that US technology is a viable alternative to the more expensive and less available MRI when measuring tendon morphology. Moreover, US delivers reliable measures of PT CSA both within and between raters and days of measurement. However, it is recommended that the methodological protocols are rigorous and the operators are well trained in image acquisition and interpretation should manual segmentation methods be implemented.

CONFLICT OF INTEREST DISCLOSURE

The authors declare no conflicts of interest.

Acknowledgments—The work was funded by the collaborative doctoral scheme between Northumbria University, UK, and the Talented Athlete Scholarships Scheme (TASS), UK, and by the Research Development Fund (RDF) Studentship Scheme. The magnetic resonance imaging costs were met by Newcastle Clinic, Newcastle, UK.

Data availability statement—The data that support the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

Albano D, Messina C, Usuelli FG, De Girolamo L, Grassi M, Maccario C, Bignotti B, Tagliafico A, Sconfienza LM. Magnetic resonance and ultrasound in Achilles tendinopathy: Predictive role and response assessment to platelet-rich plasma and adipose-derived stromal vascular fraction injection. Eur J Radiol 2017;95:130–135.

Berthoty D, Haghighi P, Sartoris DJ, Resnick D. Osseous invasion by soft-tissue sarcoma seen better on MR than on CT. AJR Am J Roentgenol 1989;152:1131.

Bohm S, Mersmann F, Schroll A, Mäkitalo N, Arampatzis A. Insufficient accuracy of the ultrasound-based determination of Achilles tendon cross-sectional area. J Biomech 2016;49:2932–2937.

Brushøj C, Henriksen BM, Albrecht-Beste E, Hölmich P, Larsen K, Bachmann Nielsen M. Reproducibility of ultrasound and magnetic resonance imaging measurements of tendon size. Acta Radiol 2006;47:954–959.

Carrino JA, McCauley TR, Katz LD, Smith RC, Lange RC. Rotator cuff: Evaluation with fast spin-echo versus conventional spin-echo MR imaging. Radiology 1997;202:533–539.

Couppe C, Kongsgaard M, Aagaard P, Hansen P, Bojsen-Moller J, Kjaer M, Magnusson SP. Habitual loading results in tendon hypertrophy and increased stiffness of the human patellar tendon. J Appl Physiol 2008;105:805–810.

Couppe C, Svensson RB, Kongsgaard M, Kovanen V, Grosset JF, Snorgaard O, Bencke J, Larsen JO, Bandholm T, Christensen TM, Boesen A, Helmark IC, Aagaard P, Kjaer M, Magnusson SP. Human Achilles tendon glycation and function in diabetes. J Appl Physiol 2016;120:130–137.

Dudley-Javoroski S, McMullen T, Borgwardt MR, Peranich LM, Shields RK. Reliability and responsiveness of musculoskeletal ultrasound in subjects with and without spinal cord injury. Ultrasound Med Biol 2010;36:1594–1607.

Ekizos A, Papatzika F, Charcharis G, Bohm S, Mersmann F, Arampatzis A. Ultrasound does not provide reliable results for the measurement of the patellar tendon cross sectional area. J Electromyogr Kinesiol 2013;23:1278–1282.

Fukunaga T, Kawakami Y, Kubo K, Kanehisa H. Muscle and tendon interaction during human movements. Exerc Sport Sci Rev 2002;30:106–110.

Gellhorn AC, Carlson MJ. Inter-rater, intra-rater, and inter-machine reliability of quantitative ultrasound measurements of the patellar tendon. Ultrasound Med Biol 2013;39:791–796.

Hicks KM, Onambele-Pearson GL, Winwood K, Morse CI. Gender differences in fascicular lengthening during eccentric contractions: The role of the patellar tendon stiffness. Acta Physiol 2013;209:235–244.

Hopkins WG. Spreadsheets for analysis of validity and reliability. 2015. https://www.thefreelibrary.com/Spreadsheets+for+Analysis+of+Validity+and+Reliability.-a0562004416. Accessed January 11, 2022.

Klauser AS, Tagliafico A, Allen GM, Boutry N, Campbell R, Court-Payen M, Grainger A, Guerini H, McNally E, O'Connor PJ, Ostlere S, Petroons P, Reijnierse M, Sconfienza LM, Silvestri E, Wilson DJ, Martinoli C. Clinical indications for musculoskeletal ultrasound: A Delphi-based consensus paper of the European Society of Musculoskeletal Radiology. Eur Radiol 2012;22:1140–1148.

Kongsgaard M, Reitelseder S, Pedersen TG, Holm L, Aagaard P, Kjaer M, Magnusson SP. Region specific patellar tendon hypertrophy in humans following resistance training. Acta Physiol 2007;191:111–121.

Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chirop Med 2016;15:155–163.

Kruse A, Stafilidis S, Tilp M. Ultrasound and magnetic resonance imaging are not interchangeable to assess the Achilles tendon cross-sectional-area. Eur J Appl Physiol 2017;117:73–82.

Kubo K, Kanehisa H, Kawakami Y, Fukunaga T. Effects of repeated muscle contractions on the tendon structures in humans. Eur J Appl Physiol 2001;84:162–166.

Maganaris CN, Paul JP. In vivo human tendon mechanical properties. J Physiol 1999;521(Pt 1):307–313.

Maganaris CN, Reeves ND, Rittweger J, Sargeant AJ, Jones DA, Gerrits K, De Haan A. Adaptive response of human tendon to paralysis. Muscle Nerve 2006;33:85–92.

Magnusson SP, Narici MV, Maganaris CN, Kjaer M. Human tendon behaviour and adaptation, in vivo. J Physiol 2008;586:71–81.

McAuliffe S, McCreesh K, Purtill H, O'Sullivan K. A systematic review of the reliability of diagnostic ultrasound imaging in measuring tendon size: Is the error clinically acceptable?. Phys Ther Sport 2017;26:52–63.

Murtagh CF, Stubbs M, Vanrenterghem J, O'Boyle A, Morgans R, Drust B, Erskine RM. Patellar tendon properties distinguish elite from non-elite soccer players and are related to peak horizontal but not vertical power. Eur J Appl Physiol 2018;118:1737–1749.

Onambele GNL, Burgess K, Pearson SJ. Gender-specific in vivo measurement of the structural and mechanical properties of the human patellar tendon. J Orthop Res 2007;25:1635–1642.

Poisson SD. Note sur l'extension des fils et des plaques élastiques. Ann Chim Phys 1827;36:384–385.

Reeves ND, Narici MV. Behavior of human muscle fascicles during shortening and lengthening contractions in vivo. J Appl Physiol 2003;95:1090–1096.

Sonin AH, Peduto AJ, Fitzgerald SW, Callahan CM, Bresler ME. MR imaging of the rotator cuff mechanism: Comparison of spin-echo and turbo spin-echo sequences. AJR Am J Roentgenol 1996;167:333–338.

Springate SD. The effect of sample size and bias on the reliability of estimates of error: a comparative study of Dahlberg's formula. Eur J Orthod 2011;34:158–163.

Stenroth L, Sefa S, Arokoski J, Toyras J. Does magnetic resonance imaging provide superior reliability for Achilles and patellar tendon cross-sectional area measurements compared with ultrasound imaging?. Ultrasound Med Biol 2019;45:3186–3198.

Tardioli A, Malliaras P, Maffulli N. Immediate and short-term effects of exercise on tendon structure: Biochemical, biomechanical and imaging responses. Br Med Bull 2012;103:169–202.

Thoirs KA, Childs J. Are ultrasound measurements of Achilles tendon size reliable? A systematic review of rater reliability. Ultrasound Med Biol 2018;44:2476–2491.