Journal of Bodywork & Movement Therapies 23 (2019) 344–351

Contents lists available at ScienceDirect

Journal of Bodywork & Movement Therapies

journal homepage: www.elsevier.com/jbmt

DIAGNOSTIC METHODS: Original Research

Ultrasonographic assessment of patellar tendon thickness at 16 clinically relevant measurement sites – A study of intra- and interrater reliability

Pætur M. Holm [a, b, *], Søren T. Skou [a, b], Jens L. Olesen [c], Sissel Mølbak [d], Heidi Jørgensen [d], Morten Skjoldager [e], Steffan W. Christensen [d, f]

[a] Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Denmark
[b] Research Unit for Musculoskeletal Function and Physiotherapy, Department of Sports Science and Clinical Biomechanics, University of Southern, Denmark
[c] Research Unit for General Practice in Aalborg, Department of Clinical Medicine, Aalborg University, Denmark
[d] Department of Physiotherapy, University College of Northern Denmark, Aalborg, Denmark
[e] Arkadens Fysioterapi, Aalborg, Denmark
[f] SMI, Department of Health Science and Technology, Aalborg University, Denmark

A R T I C L E I N F O

Article history:
Received 29 January 2019
Accepted 2 February 2019

Keywords:
Knee
Patellar tendon
Reliability
Ultrasonography
Musculoskeletal disorders

A B S T R A C T

Objectives: To determine intra- and interrater reliability of ultrasonographic imaging (USI) measurements of patellar tendon (PT) thickness using 16 measurement sites covering the entire tendon.
Design: Reliability study.
Setting: Physiotherapy outpatient clinic.
Participants: Twenty healthy and physically active volunteers (9 women). Mean age: 24 years (SD ± 2.73). Mean body mass: 75.8 kg (SD ± 11.8).
Main outcome measures: Intraclass correlation coefficient (ICC) and 95% limits of agreement (LOA) in cm and in percentage relative to the mean PT thickness.
Results: Intrarater reliability ranged from 0.59 to 0.87 and 0.59 to 0.93 for examiners I and II, respectively. Interrater reliability ranged from 0.37 to 0.89. Measurement precision for examiner I ranged from 0.05 to 0.09 cm (17.5%–26.7%) while ranging from 0.04 to 0.13 cm (13.3%–38.7%) for examiner II. Interrater measurement precision ranged from 0.07 to 0.15 cm (19.1%–42.5%).
Conclusion: In an attempt to replicate daily clinical USI practice, this was the first study extensively assessing reliability throughout the full range of the patellar tendon - revealing a considerable variation in intra- and interrater reliability as well as measurement precision throughout the 16 individual PT sites. In a clinical context, the low interrater reliability and precision found at the proximal tendon insertion site may have implications for USI of the symptomatic PT, as this is the site mainly associated with underlying pathologic changes. Further reliability studies are needed to clarify the region-specific reliability of the full length PT.

1. Introduction

The use of musculoskeletal ultrasound imaging (USI) in the assessment of soft tissue structures is a widely used, inexpensive imaging modality in both research and clinical settings (Finnoff, 2016). With unique dynamic properties and high-resolution imaging of soft tissue structures, musculoskeletal USI has become a valuable tool in the clinical examination of human tendons (Grassi et al., 2000). The patellar tendon (PT) is a body part that has been of particular interest in the literature due to two main reasons. Firstly, due to the prevalence of overuse injuries in the PT which regularly leads to pain and disabling symptoms (Miller, 2013) in both athlete and non-athlete populations (Lian et al., 2005), (Zwerver et al., 2011). Secondly, due to its size, linearity and superficial location above the anterior aspect of the patella, the patellar tendon (PT) is an ideal fit for diagnostic USI (Miller, 2013). Sonographically, the healthy tendon appears hyperechoic due to a strong fibrillar bundle

* Corresponding author. Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Fælledvej 7, DK-4200, Slagelse, Denmark.
E-mail address: pamh@regionsjaelland.dk (P.M. Holm).

https://doi.org/10.1016/j.jbmt.2019.02.004
1360-8592/

with parallel superficial and deep surfaces, meaning that the healthy parts of the tendon appear bright on a screen as the dense fibrillar bundle reflects a high degree of sound waves from the USI transducer (Miller, 2013).

In contrast, the pathologic PT often appears hypoechoic and thickened, meaning that the pathologic part of the tendon appears dark on the screen as it reflects nearly no sound waves from the USI transducer due to loss of the fibrillar pattern and swelling (Miller, 2013), (Kainberger et al., 1997).

Since reduction in PT thickness might predict successful treatment outcomes following PT tendinopathy (Fredberg et al., 2004) and since correlations have been found between reduced tendon thickness and a decrease in pain (Mahowald et al., 2011), evaluation of tendon thickness is a particularly important clinical procedure.

Due to the operator-dependent nature of diagnostic USI (Wakefield et al., 2005), reliability of USI measurements of the PT has been the source of attention in several studies in recent years (Black et al., 2004; Ekizos et al., 2013; Gellhorn and Carlson, 2013; Skou and Aalkjaer, 2013; Sunding et al., 2014). However, previous studies on reliability of PT measurements have focused exclusively on one (Black et al., 2004; Gellhorn and Carlson, 2013; Skou and Aalkjaer, 2013; Sunding et al., 2014) to three (Ekizos et al., 2013) sites, even though full-length PT USI is the common procedure in clinical practice. This leaves not only a big gap in the existing knowledge on reliability of USI measurements when applied to different parts of the PT with different regional characteristics, but also goes against the clinical guidelines for diagnostic USI which recommend scanning the entire PT for pathology (Martinoli, 2010). Thus, keeping in line with clinical recommendations, it is of specific importance to provide USI reliability data covering the various regional aspects of the PT to aid the clinician with region-specific reliability values when scanning the PT for pathology.

The aim of this study was to determine intra- and interrater reliability of USI assessment of PT thickness using clinically relevant measurement sites (16 in total) covering the entire PT.

2. Materials & methods

2.1. Participants

Twenty healthy and physically active volunteers (9 women) were recruited using flyers posted at University College of Northern Denmark and at a physiotherapy outpatient clinic where the study took place. Mean age was 24 years (SD ± 2.73) with a mean body mass of 74.8 kg (SD ± 11.8).

The study was carried out in a single session lasting approximately 90 min per subject (see study protocol for more details).

Participants were recruited with exclusion criteria being current or prior lower extremity pain within the past 6 weeks leading up to the test session, previous knee surgery as well as sports activities at elite level. Furthermore, participants were told to refrain from lower limb strength training on the day of testing.

This study was based on data collected for a clinical assessment study in a physiotherapy outpatient clinic; hence, approval from the Danish Data Protection Agency was not needed. Written informed consent was obtained from all participants on forms provided by the local ethics committee and the study was conducted in accordance with the Helsinki Declaration.

2.2. Examiners

USI measurements were performed with a SonoSite S-MSK (SonoSite, Inc., Bothell, WA, USA) with an HFL38x 13-6 MHz linear transducer. Two experienced examiners familiar with using the specific USI device in daily clinical practice performed the USI measurements.

Both examiners had followed and were teaching on a formal musculoskeletal USI educational program for doctors, physiotherapists and other healthcare professionals, where one part of the course has a specific focus on the knee. Examiner I had 9 years of experience performing USI measurements while examiner II had 4 years of experience. The examiners were not otherwise involved in the collection and synthesis of data.

Prior to the start of the study both examiners were carefully instructed in the test setup and the specific protocol which was based on the afore mentioned training, where scanning the entire patella tendon is an inherent part of the education. The examiners had two training sessions each, with one examiner scanning, while the other observed the scan.

The aim of the test session was to ensure that both examiners adhered to the protocol and were in agreement on where to place cursors in order to perform the measurements. The total time spent on protocol training was approximately 8 h. Throughout each test session during the study the examiners were under the supervision of a research assistant to ensure adherence to the study protocol.

2.3. Study protocol

The USI measurement protocol was set up in accordance with the recommendations from the European Society of Musculoskeletal Radiology (Martinoli, 2010).

Subjects were positioned in a supine position on an examination table with both knees flexed to approximately 30° and supported by a firm cushion underneath the popliteal area. By keeping the target knee in slight flexion (30°) as opposed to fully extended, potential concave anisotropies from the PT were avoided (Martinoli, 2010).

Two successive bilateral measurements were performed in random order on each subject by both examiners, resulting in one of the following orders: (right-left-right-left) or (left-right-left-right). This order of measurements would force the examiner to reposition the transducer between the first and second unilateral measurement. By placing tape over the measurement values on the screen, the examiners were blinded to the results while these were still visible for the research assistant who recorded each value. Four longitudinal, clinically relevant PT sites were identified for analysis: apex patella; 1 cm under apex patella; tibial tuberosity; and 1 cm over tibial tuberosity.

At each site, one longitudinal (central placement) and three transversal measurements were performed, giving a total of 16 individual sites (Table 1 & Fig. 1). The scan depth was kept constant at 1.8 cm and the measurement of PT thickness was performed by

Table 1
Measurement sites (n = 16) of the patellar tendon.

Site	Longitudinal	Transversal
AP	AP	1a: Thickest middle third
		1b: Thickest medial third
		1c: Thickest lateral third
1 u. AP	1 u. AP	2a: Thickest middle third
		2b: Thickest medial third
		2c: Thickest lateral third
TT	TT	3a: Thickest middle third
		3b: Thickest medial third
		3c: Thickest lateral third
1°. TT	1°. TT	4a: Thickest middle third
		4b: Thickest medial third
		4c: Thickest lateral third

Thickness (cm) was measured anteroposterior.
AP: Apex Patella. **1 u. AP:** 1 cm under Apex Patella.
TT: Tibial Tuberosity. **1 o. TT:** 1 cm over Tibial Tuberosity.

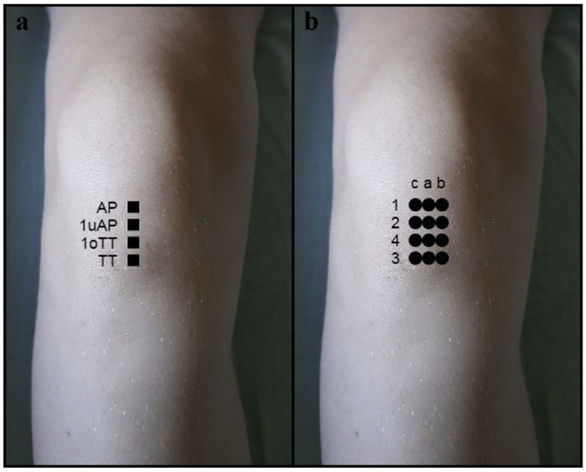

Fig. 1. Images of the knee with longitudinal (a) and transversal (b) measurement sites depicted.
Image 1a: Longitudinal plane thickness measurements (■) were conducted at apex patella (AP); 1 cm distal from apex patella (1uAP); 1 cm proximal from the tibial tuberosity (1oTT); and at the tibial tuberosity (TT).
Image 1b: Transversal plane thickness measurements (•) were conducted at the thickest part of the middle- (a), medial- (b) and lateral (c) third of the tendon at; AP (1), 1uAP (2), TT (3) and 1oTT (4).

the built-in software (Figs. 2–4). Each test session lasted approximately 45 min and was performed simultaneously on two subjects by the two examiners in different rooms.

After 45 min, the two examiners switched rooms and thereby subjects. The straight succession of each examiner's measurements on the same subject ensured that no activity-induced effects on tendon thickness was present between measurements.

2.4. Statistical analysis

All analyses were performed using IBM SPSS Statistics Version 24.0. Only the participant's dominant leg was chosen for statistical analysis. Data on the non-dominant leg are available as a supplementary appendix (appendix A). Visual inspection of QQ-plots was conducted to ensure normal distribution of data. Changes in PT thickness between the first and second measurement was investigated using a paired samples *t*-test. To investigate potential learning effects over the range of measurements of the 20 subjects, the randomly generated blocks of participants received consecutive identification numbers.

Evaluation of intrarater reliability for examiner I and II was conducted via two-way random effects, single measure model (2,1), absolute agreement type intraclass correlation coefficient (ICC). Interrater reliability using the mean of first and second measurements was analyzed via two-way random effects, average measure

Fig. 2. Sonographic image of measurement site (AP & 1uAP).
Image above: Longitudinal anteroposterior thickness measurement of AP (C) and 1uAP (A), respectively. 1uAP was found by tracing 1 cm distally from the posterior border of AP parallel with the tendon fibres (B).

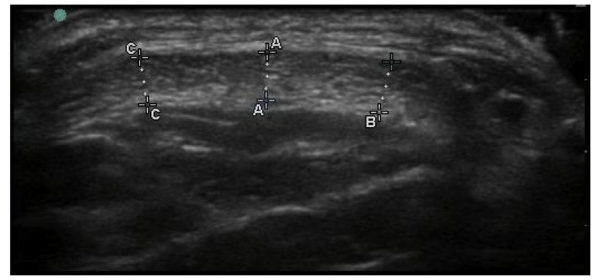

Fig. 3. Sonographic image of measurement site (Transversal).
Image above: Transversal anteroposterior thickness measurements of the patellar tendon. C represents the thickest lateral third, A represents the thickest middle third and B represents the thickest medial third. Transversal thickness measurements were performed at the following sites.
-Apex patella, represented by the above image (1a, 1b & 1c)-1 cm distal to apex patella (2a, 2b & 2c)-Tibial tuberosity (3a, 3b & 3c)-1 cm proximal to the tibial tuberosity (4a, 4b & 4c).

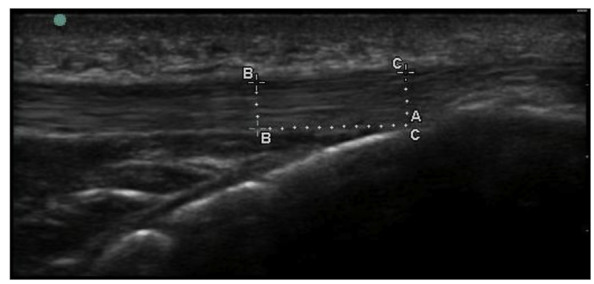

Fig. 4. Sonographic image of measurement site (TT & 1oTT).
Image above: Longitudinal anteroposterior thickness measurement of the TT (C) and 1uTT (B). 1uTT was found by tracing 1 cm proximally from the posterior border of TT parallel with the tendon fibres (B).

model (2, *k*), absolute agreement type ICC.

Since mean ratings have shown to improve measurement precision compared to single ratings (Skou and Aalkjaer, 2013), (Rathleff et al., 2011), only means from examiner I and II respectively were derived for interrater reliability analysis. A Bland-Altman plot was constructed to graphically assess agreement between the two examiners. Measurement precision was evaluated by plotting 1.96 standard deviations (SD) above and below the mean of the difference scores, respectively. The confidence interval ranging from 1.96 SD above the mean of the difference scores to 1.96 SD below represents 95% limits of agreement (LOA). LOA was also presented in percentage relative to the mean PT thickness (LOA-%).

3. Results

Results for PT thickness measurements for examiner I and II are found in Tables 2 and 3, respectively.

For examiner II, the measurement sites 3b ($p = 0.025$) and 4a ($p = 0.033$) were significantly different between the two measurements. Hence, these measurement sites were excluded from further reliability testing for examiner II. There was no consistent evidence of learning effects over the range of the 16 measurement sites when plotting the difference in PT thickness between each examiner's first and second measurements against the consecutive identification number of subjects.

3.1. Reliability

Results for intra- and interrater reliability for examiner I and II

Table 2
Patellar tendon thickness (cm) on dominant leg from. all 16 measurement sites for examiner I. Results presented in mean ± SD.

Sites (n = 16)	Men (n = 11) mean (SD)	Women (n = 9) mean (SD)	Total (n = 20) mean (SD)
AP	0.47 (0.08)	0.40 (0.06)	0.44 (0.08)
1a	0.33 (0.05)	0.30 (0.06)	0.32 (0.06)
1b	0.35 (0.06)	0.32 (0.04)	0.34 (0.05)
1c	0.32 (0.04)	0.29 (0.03)	0.31 (0.04)
1 u. AP	0.39 (0.07)	0.34 (0.05)	0.37 (0.06)
2a	0.33 (0.06)	0.29 (0.05)	0.31 (0.06)
2b	0.33 (0.05)	0.30 (0.03)	0.32 (0.05)
2c	0.33 (0.04)	0.29 (0.04)	0.31 (0.04)
TT	0.45 (0.07)	0.42 (0.07)	0.44 (0.07)
3a	0.34 (0.06)	0.30 (0.05)	0.32 (0.06)
3b	0.34 (0.05)	0.32 (0.05)	0.33 (0.05)
3c	0.34 (0.03)	0.31 (0.06)	0.33 (0.05)
1°. TT	0.37 (0.06)	0.31 (0.05)	0.34 (0.06)
4a	0.33 (0.05)	0.30 (0.05)	0.31 (0.05)
4b	0.33 (0.04)	0.31 (0.04)	0.32 (0.04)
4c	0.32 (0.03)	0.29 (0.04)	0.31 (0.04)

Table 3
Patellar tendon thickness (cm) on dominant leg from. all 16 measurement sites for examiner II. Results presented in mean ± SD.

Sites (n = 16)	Men (n = 11) mean (SD)	Women (n = 9) mean (SD)	Total (n = 20) mean (SD)
AP	0.32 (0.06)	0.35 (0.08)	0.33 (0.07)
1a	0.37 (0.09)	0.29 (0.04)	0.33 (0.08)
1b	0.38 (0.10)	0.30 (0.04)	0.35 (0.09)
1c	0.35 (0.09)	0.28 (0.04)	0.32 (0.08)
1 u. AP	0.30 (0.05)	0.31 (0.04)	0.30 (0.05)
2a	0.33 (0.06)	0.28 (0.03)	0.31 (0.06)
2b	0.33 (0.06)	0.28 (0.03)	0.31 (0.06)
2c	0.35 (0.05)	0.27 (0.03)	0.31 (0.06)
TT	0.42 (0.06)	0.38 (0.07)	0.40 (0.06)
3a	0.38 (0.06)	0.33 (0.05)	0.36 (0.06)
3b	0.36 (0.05)	0.31 (0.05)	0.34 (0.06)*
3c	0.39 (0.06)	0.32 (0.04)	0.36 (0.06)
1°. TT	0.36 (0.05)	0.32 (0.05)	0.34 (0.05)
4a	0.34 (0.05)	0.29 (0.05)	0.31 (0.05)*
4b	0.33 (0.04)	0.29 (0.04)	0.31 (0.05)
4c	0.35 (0.05)	0.28 (0.05)	0.32 (0.06)

*Significant different patellar tendon thickness between first and second measurements ($p < 0.05$).

are presented in Table 4.

For illustrative purposes, the ICC results are categorized as low (0.0–0.50), moderate (0.50–0.75) and good (>0.75) (Portney and Watkins, 2014). However, we are well aware of the difficulties in determining appropriate quality cut-offs for reliability (Weir, 2005). As such, the current categorization should be viewed as crude estimates only.

The ICC with 95% confidence interval (95% CI) for examiner I ranged from moderate to good (0.59–0.87) for all 16 measurements. The ICC with 95% CI for examiner II ranged from moderate to good (0.59–0.93) for 14 out of 16 measurements with two measurement sites (3b & 4a) excluded due to a significant difference between the first and second measurement. The interrater ICC for examiner I and II ranged from low to good (0.37–0.89) for all 14 interrater reliability tested measurements.

Table 4
Intra- & Interrater reliability results for both examiners. ICC, (95% CI).

	Examiner I	Examiner II	Examiner I + II
AP	▲ 0.86 (0.68–0.94)	▲ 0.81 (0.59–0.92)	▼ 0.37 (−0.11 – 0.58)
1a	▲ 0.80 (0.57–0.92)	▲ 0.79 (0.55–0.91)	► 0.66 (0.15–0.86)
1b	▲ 0.85 (0.57–0.92)	► 0.72 (0.42–0.88)	► 0.62 (0.02–0.85)
1c	► 0.75 (0.47–0.89)	► 0.71 (0.40–0.88)	► 0.65 (0.09–0.86)
1 u. AP	▲ 0.84 (0.65–0.93)	► 0.59 (0.21–0.81)	► 0.50 (−0.25 - 0.82)
2a	▲ 0.83 (0.62–0.93)	▲ 0.82 (0.61–0.93)	▲ 0.89 (0.73–0.96)
2b	▲ 0.77 (0.51–0.90)	▲ 0.93 (0.84–0.97)	▲ 0.87 (0.67–0.95)
2c	► 0.66 (0.32–0.85)	▲ 0.83 (0.62–0.93)	▲ 0.86 (0.65–0.94)
TT	► 0.75 (0.47–0.89)	▲ 0.88 (0.72–0.95)	► 0.71 (0.24–0.89)
3a	▲ 0.87 (0.71–0.95)	▲ 0.86 (0.68–0.94)	▲ 0.79 (0.09–0.94)
3b	► 0.68 (0.36–0.86)	*	*
3c	► 0.59 (0.21–0.82)	▲ 0.84 (0.64–0.93)	▼ 0.48 (−0.15 - 0.78)
1°. TT	▲ 0.84 (0.65–0.94)	▲ 0.78 (0.53–0.91)	▲ 0.86 (0.65–0.95)
4a	► 0.72 (0.43–0.88)	*	*
4b	► 0.72 (0.42–0.88)	▲ 0.89 (0.74–0.95)	▲ 0.80 (0.50–0.92)
4c	► 0.71 (0.41–0.87)	▲ 0.81 (0.58–0.92)	▲ 0.84 (0.61–0.94)

▲ (ICC > 0.75) = good reliability.
► (ICC 0.50–0.75) = moderate reliability.
▼ (ICC 0.0–0.50) = low reliability.
*Excluded from the reliability analysis due to significant difference in PT thickness between first and second measurement.

3.2. Measurement precision

Results for intra- and interrater LOA and LOA-% are presented in Table 5.

Visual inspection of the intrarater Bland Altman plots for both examiners revealed no proportional differences over the measurement range. For examiner I, the LOA (LOA-%) ranged from 0.05 to 0.09 cm (17.5%–26.7%) for intrarater reliability over the range of the 16 PT measurement sites. For examiner II, the LOA (LOA %) for intrarater reliability ranged from 0.04 to 0.13 cm (13.3%–38.7%) over the range of the 14 PT measurement sites. LOA (LOA-%) for interrater reliability ranged from 0.07 to 0.15 cm (21%–42.5%) over the range of 14 PT measurement sites.

Table 5
Intra- & interrater reliability presented as 95%. LOA (cm) and as percentage of the mean PT thickness (%).

	Examiner I		Examiner II		Examiner I + II	
	LOA (cm)	LOA (%)	LOA (cm)	LOA (%)	LOA (cm)	LOA (%)
AP	0.09	19.5	0.09	26.1	0.15	37.7
1a	0.07	22.1	0.11	32.6	0.14	41.9
1b	0.05	16.1	0.13	38.7	0.15	42.5
1c	0.05	17.6	0.12	37.4	0.12	36.9
1 u. AP	0.07	19.2	0.09	28.8	0.10	28.6
2a	0.07	21.0	0.07	22.3	0.07	21.7
2b	0.07	20.7	0.04	13.3	0.07	21.9
2c	0.07	22.5	0.07	21.2	0.07	21.3
TT	0.09	21.4	0.06	15.5	0.11	25.9
3a	0.06	17.5	0.06	17.8	0.07	21.0
3b	0.08	23.4	*	*	*	*
3c	0.09	26.7	0.07	20.2	0.12	33.9
1°. TT	0.07	20.2	0.07	21.1	0.08	22.5
4a	0.08	23.9	*	*	*	*
4b	0.07	20.5	0.04	14.0	0.07	21.5
4c	0.06	20.2	0.07	22.5	0.07	22.1

*Excluded from measurement precision analysis due to significant difference in PT thickness between first and second measurements.

Visual inspection of the interrater Bland Altman plots suggested a proportional difference of measurements for “AP” and “1uAP”, however when fitted into a linear regression analysis, normal distribution of the difference scores was confirmed ($p > 0.05$) (Fig. 5). The remaining Bland Altman plots revealed no systematic graphical differences, represented by Fig. 6.

A summary of results for measurement reliability and precision is presented in Table 6.

4. Discussion

This is the first study evaluating intra- and interrater reliability and measurement precision when measuring PT thickness of the entire tendon using USI. Intrarater reliability ranged from 0.59 to 0.93 for both examiners over the range of 14 and 16 sites, respectively. Results for interrater reliability were within a wider range (0.37–0.89) for the 14 assessed sites. Precision for intrarater measurements varied from 0.04 cm to 0.13 cm (13.3%–38.7%) while ranging from 0.06 cm to 0.15 cm (19.1%–42.5%) for interrater measurements.

Previous intrarater- and interrater USI studies on muscle- and tendon thickness reveal a cumulative ICC range from 0.64 to 0.97 (Bentman et al., 2010; Cheng et al., 2012; Costa et al., 2009; Craig et al., 2008; Gellhorn and Carlson, 2013; Koppenhaver et al., 2009; Liang et al., 2007; O'Sullivan et al., 2007; Rathleff et al.,

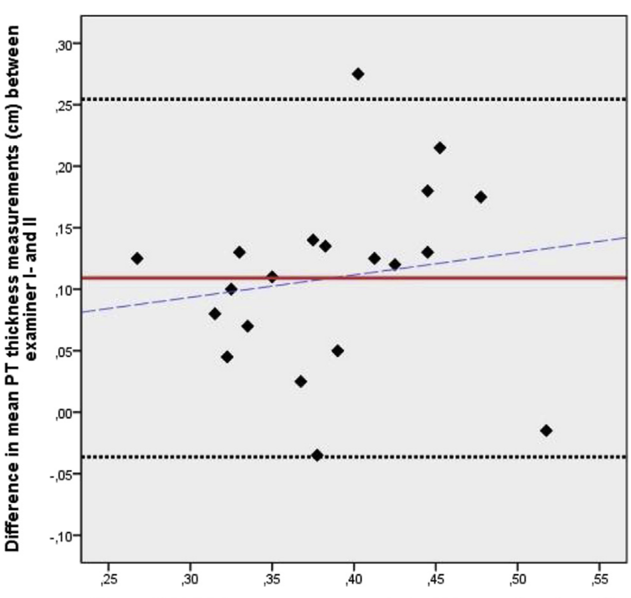

Fig. 5. Interrater Bland Altman plot with 95% limits of agreement (dotted black lines) for the measurement site AP.
The regression analysis (dotted blue line) revealed no systematic different distribution of scores ($p = 0.519$). The other measurement site (1 cm. u. AP) with a visually suggested proportional biased trend of difference scores was similarly non-significant in the regression analysis ($p = 0.073$) (plot not shown).

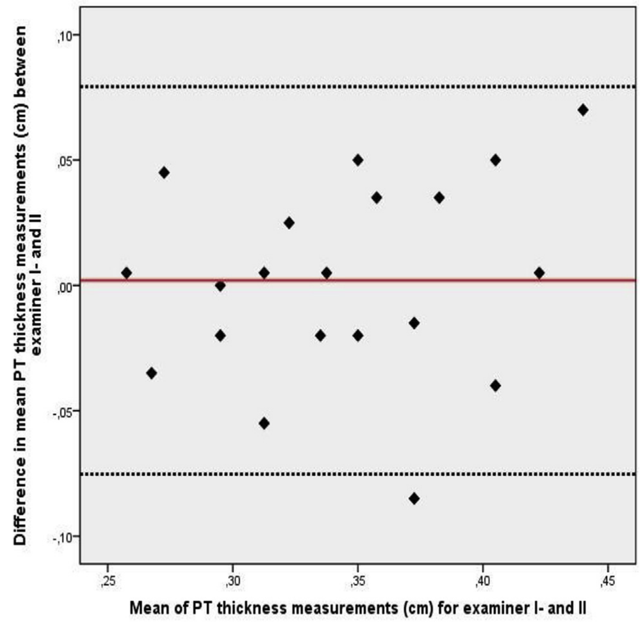

Fig. 6. Interrater Bland Altman plot with 95% limits of agreement (dotted black lines) of the measurement site 1oTT.
This representative Bland Altman plot showed no systematic different distribution of measurements. Bland Altman plots of the other measurement sites had a similar appearance, revealing no systematic difference over the measurement range (plots not shown).

Table 6
Intra- and interrater reliability and precision of the four longitudinal plane measurement sites.

	Longitudinal			Transversal		
	EX I	EX II	EX I + II	EX I	EX II	EX I + II
			AP			
ICC	0.86	0.81	0.37	0.80	0.74	0.64
LOA	0.09	0.09	0.15	0.06	0.12	0.14
LOA-%	19.5	26.1	37.7	18.6	36.2	40.4
			1uAP			
ICC	0.84	0.59	0.50	0.75	0.86	0.87
LOA	0.07	0.09	0.10	0.07	0.06	0.07
LOA-%	19.2	28.8	28.6	21.4	18.9	21.6
			TT			
ICC	0.75	0.88	0.71	0.71	0.79*	0.64*
LOA	0.09	0.06	0.11	0.08	0.07*	0.10*
LOA-%	21.4	15.5	25.9	22.5	19.0*	27.5*
			1oTT			
ICC	0.84	0.78	0.86	0.72	0.85*	0.82*
LOA	0.07	0.07	0.08	0.07	0.06*	0.07*
LOA-%	20.2	21.1	22.5	21.5	18.3*	21.8*

Results are divided between longitudinal axis scans and transverse axis scans. Transverse axis results are presented as the mean of the transverse'measurements: a (middle third), b (medial third) and c (lateral third) at each of the four sites. EX: Examiner.
*Since there was a significant difference in PT thickness between first and second measurement on 3b and 4a for examiner II, only two transversal sites were used for the derivation of means at these sites.

2011; Skou and Aalkjaer, 2013; Wallwork et al., 2007) and from 0.40 to 0.97, respectively (Bentman et al., 2010; Cheng et al., 2012; Gellhorn and Carlson, 2013; O'Sullivan et al., 2007; Rathleff et al., 2011; Skou and Aalkjaer, 2013; Wallwork et al., 2007). Previous results on measurement precision (LOA-%) reveal a cumulative range from 1.8% to 53% for intrarater (Bentman et al., 2010; Bjordal et al., 2003; Costa et al., 2009; Koppenhaver et al., 2009; O'Connor et al., 2004; O'Sullivan et al., 2007; Rathleff et al., 2011; Skou and Aalkjaer, 2013; Springer et al., 2006; Wallwork et al., 2007; Ying et al., 2003) and 15.8%–49% for interrater (Bentman et al., 2010; O'Sullivan et al., 2007; Rathleff et al., 2011; Skou and Aalkjaer, 2013; Wallwork et al., 2007; Ying et al., 2003) for USI-derived measures of muscle- and tendon thickness. The considerable variation in reliability and measurement precision in USI studies is in part reflective of the different structures being measured, with proximity to bone, depth and adjacent soft tissue, to a varying degree, influencing the quality of the sonographic image. When compared to deeper and irregular soft-tissue structures, the PT is considered relatively feasible for USI examinations (Henderson et al., 2015). Yet, the wide reliability and precision range found in our study indicates some degree of uncertainty.

In a study on PT measurement precision in healthy adults, O'Connor et al. measured LOA-% 1 cm above the tendons insertion onto the tibia tubercle corresponding to "1°. TT" in our study and found an intrarater LOA-% of 19% and 22% for measures of long transverse axis width and 24% and 32% for short axis transverse diameter. Interrater LOA-% was 22% and 27% for long axis and short axis, respectively (O'Connor et al., 2004). The present findings on transverse axis PT thickness at this measurement point revealed a similar trend with intrarater LOA-% values from 20.2% to 23.9% (examiner I) and from 14% to 22.5% (examiner II) and interrater LOA-% from 21.5% to 22.1%. When studying intra- and interrater reliability of PT measurements, Gellhorn et al. found an intrarater reliability from 0.87 to 0.96 and interrater reliability of 0.90 and 0.92, when measuring the cross-sectional area of the tendon 1 cm distal to apex patella (Gellhorn and Carlson, 2013). In our study, the longitudinal measurements 1 cm distal to apex patella (1 u. AP) revealed considerably lower intra- and interrater reliability of 0.59–0.84 and 0.50, respectively. The highly standardized measurement protocol including strapwires as external markers in the Gellhorn study might partially explain the higher reliability values found in this study. However, such extensive measures of standardization might not adequately reflect the use of USI in daily clinical practice. Still, using a similar examination protocol as in the current study, Skou and Aalkjær found an intrarater reliability of 0.89–0.94 and interrater reliability of 0.78 with intrarater LOA of 0.07 cm and interrater LOA of 0.10 cm for two examiners measuring PT thickness in a longitudinal plane 1 cm distal to apex patella (Skou and Aalkjaer, 2013). The findings in the Skou and Aalkjær study reveal somewhat higher ICC reliability scores for intrarater (0.89–0.94 vs. 0.59–0.84) and interrater (0.78 vs. 0.50) when compared to our findings. This might indicate that the task of identifying several measurement sites in one scanning session may compromise specificity at each individual site and thereby reduce the reliability of measurements.

When comparing longitudinal and transverse plane findings at the same site in our study, ICC and LOA results are largely consistent (Table 6). However, there is a notable exception at the measurement site 1 cm distal to Apex Patella (1uAP), with this site showing a considerably higher interrater ICC for the transverse plane measurements (0.87 vs. 0.50).

As such, transverse axis scans at 1uAP seem to have provided more consistent PT thickness measurements in our study. Interestingly, since the proximal PT region is aligned in a cone shaped structure, transversal thickness measurements without a metric tendon reference point at the transversal axis, naturally becomes more difficult for the examiner to reproduce than longitudinal measurements, where thickness can be measured with the aid of metric reference points.

Coincidentally, higher measurement variability has been found for transverse axis PT thickness scans (Fredberg et al., 2008) - making the present findings somewhat surprising.

As mentioned earlier, this is the first USI reliability study using measurement sites covering the entire PT, making the results more clinically relevant. Previously, when studying the reliability of more than one PT measurement site, Ekizos et al. found an average interrater reliability of 0.59, combining findings on PT cross-sectional area from three examiners on the proximal, and distal PT borders as well as the metric midpoint (Ekizos et al., 2013).

In particular, measurements at the proximal portion of the tendon showed the highest variability in the study by Ekizos and colleagues (root mean square range from $7.9 \pm 3.9\,\text{mm}^2$ to $16.1 \pm 11.3\,\text{mm}^2$). In our study, reliability values at the proximal border (AP, 1a, 1b & 1c) revealed a mean interrater reliability for both measurement planes ranging from 0.37 to 0.64 with an LOA range of 0.14–0.15 cm (37.7%–40.4%) (Table 6). In comparison, the corresponding measurements at the distal border (TT, 3a & 3c) revealed an interrater reliability ranging from 0.64 to 0.71 with LOA ranging from 0.10 to 0.11 (25.9%–27.5%). However, the significant difference between first and second measurement for 3b indicates a certain degree of measurement variability at the distal border as well, and may point to difficulties in reproducing measurements near the tendon insertion sites.

Consequently, the frequently reported pathological findings at the posterior, proximal aspect of the PT (Helland et al., 2013; Johnson et al., 1996; Khan et al., 1996) might be limited by the reliability of measurements. One possible contributing factor to the limited interrater reliability at various PT points in our study might be that the mean was derived from two measurements as opposed to three. Previously, a mean consisting of three measurements as opposed to two has resulted in improved intra- and interrater reliability as well as measurement precision when applied in the evaluation of plantar fascia (Rathleff et al., 2011) and transversus abdominis and lumbar multifidi muscles (Koppenhaver et al., 2009). However, Skou and Aalkjær found no further improvement on intra- and interrater reliability or measurement precision when using a mean of three measurements compared to two in the evaluation of PT thickness (Skou and Aalkjaer, 2013). Yet, regional PT characteristics might influence the reliability of the mean of measurements to a varying degree. This might especially hold true for measurements at the proximal PT region. As such, inclusion of means of two and three measurements in this study would have clarified this relationship.

Both examiners were experienced USI examiners and produced largely comparable intrarater reliability scores over the range of measurements. However, examiner II had two sites (3b & 4a) with statistically significant different PT thickness at the first compared to the second measurement. This might be suggestive of a need for a more standardized examination protocol and should be considered in future studies.

Put into a clinical context, this study provides region-specific reliability data covering the entire PT. This is important information for the clinician when assessing the different parts of the PT for pathology.

The limited agreement between examiners when assessing the proximal portion of the PT suggests some degree of uncertainty and should be taken into consideration when assessing the frequently found pathologic changes in this region. Future studies should explore the reliability of measurements retrieved at the proximal portion of PT in patients with specific overuse injuries as this might be of clinical relevance when considering treatment effects.

It is important to note that the current study is limited by the fact that only healthy tendons were studied, hence the findings cannot be extrapolated fully to pathological or degenerated tendons. However, since pathological and degenerated tendons often display a heterogeneous tendon structure with blurred tendon margins (Grassi et al., 2000), reliability may be even more compromised. Further studies testing the protocol used in this study on pathological and degenerated PT are warranted and will aid with important clinical insights into potential reliability issues.

In conclusion, USI reliability assessment of PT thickness, using 16 measurement points covering the entire PT revealed contrasting degrees of reliability and measurement precision. Especially inter-rater reliability and agreement fluctuated throughout the range of measurements.

Further reliability studies on the different aspects of PT are needed to clarify the uniformity of region-specific PT examinations using USI.

Ethical approval

Obtainment of informed and written consent from all participants on forms provided by the local ethical committee was sufficient for the scope of this study.

Conflicts of interest

The authors have no conflict of interests.

Declarations of interests

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A

Reliability and precision for non-dominant leg

Table 1
Patellar tendon thickness (cm) on non-dominant leg from all 16 measurement points for examiner I. Results presented in mean ± SD.

PT measurement points	Men (n = 11) Mean (SD)	Women (n = 9) Mean (SD)	Total (n = 20) Mean (SD)
AP	0.45 (0.08)	0,42 (0,05)	0.43 (0.07)*
1a	0.35 (0.06)	0,32 (0,04)	0.33 (0.06)
1b	0.36 (0.06)	0.31 (0.04)	0.34 (0.06)
1c	0.35 (0.05)	0.31 (0.03)	0.33 (0.05)
1 u. AP	0.38 (0.07)	0.34 (0.06)*	0.36 (0.07)*
2a	0.34 (0.07)	0.31 (0.03)	0.32 (0.06)
2b	0.34 (0.05)	0.30 (0.03)	0.32 (0.05)
2c	0.33 (0.06)	0.32 (0.04)	0.32 (0.05)
TT	0.45 (0.07)	0.40 (0.06)	0.43 (0.07)
3a	0.35 (0.05)	0.30 (0.05)	0.33 (0.06)
3b	0.34 (0.06)	0.29 (0.04)	0.32 (0.06)
3c	0.35 (0.07)	0.33 (0.04)	0.34 (0.06)
1°. TT	0.35 (0.05)	0.31 (0.05)	0.34 (0.06)
4a	0,34 (0,07)	0,30 (0,04)	0,32 (0,06)
4b	0.34 (0.05)*	0.29 (0.05)	0.32 (0.06)*
4c	0.33 (0.05)	0.30 (0.03)	0.31 (0.05)

*Significantly different patellar tendon thickness between first and second measurement ($p < 0.05$).

Table 2
Patellar tendon thickness (cm) on non-dominant leg from all 16 measurement points for examiner II. Results presented in mean ± SD.

PT measurement points	Men (n = 11) Mean (SD)	Women (n = 9) Mean (SD)	Total (n = 20) Mean (SD)
AP	0.34 (0.06)	0,35 (0,06)	0.35 (0.06)
1a	0.39 (0.08)	0,31 (0,05)	0.36 (0.08)
1b	0.42 (0.10)	0.33 (0.06)*	0.38 (0.09)*
1c	0.38 (0.09)	0.29 (0.04)	0.34 (0.08)
1 u. AP	0.33 (0.07)	0.32 (0.07)	0.32 (0.07)
2a	0.34 (0.06)	0.29 (0.03)	0.32 (0.06)*
2b	0.34 (0.05)	0.28 (0.03)	0.32 (0.05)
2c	0.34 (0.04)	0.28 (0.03)	0.31 (0.05)
TT	0.42 (0.06)	0.35 (0.05)	0.39 (0.07)
3a	0.40 (0.06)	0.33 (0.05)	0.37 (0.07)
3b	0.37 (0.05)	0.32 (0.04)	0.35 (0.05)
3c	0.40 (0.05)	0.34 (0.03)	0.37 (0.05)
1°. TT	0.38 (0.05)	0.32 (0.04)*	0.35 (0.05)
4a	0,36 (0,05)	0,29 (0,04)	0,33 (0,06)
4b	0.35 (0.06)	0.28 (0.03)	0.32 (0.06)
4c	0.36 (0.06)	0.30 (0.03)	0.33 (0.06)

*Significantly different patellar tendon thickness between first and second measurement ($p < 0.05$).

Table 3
Intra- & Interrater reliability on non-dominant leg for both examiners. ICC, (95% CI).

	Examiner I	Examiner II	Examiner I + II
AP	*	► 0.68 (0.37–0.86)	*
1a	► 0.61 (0.25–0.83)	▲ 0.80 (0.56–0.91)	▲ 0.85 (0.62–0.94)
1b	► 0.71 (0.39–0.87)	*	*
1c	► 0.61 (0.25–0.82)	▲ 0.83 (0.61–0.93)	► 0.74 (0.35–0.90)
1 u. AP	*	▲ 0.77 (0.51–0.90)	*
2a	▲ 0.87 (0.70–0.94)	*	*
2b	▲ 0.76 (0.50–0.90)	► 0.70 (0.40–0.87)	▲ 0.89 (0.72–0.96)
2c	► 0.75 (0.47–0.89)	▲ 0.82 (0.61–0.93)	▲ 0.84 (0.61–0.94)
TT	► 0.72 (0.42–0.88)	▲ 0.80 (0.56–0.92)	► 0.58 (0.07–0.83)
3a	▲ 0.90 (0.77–0.96)	▲ 0.90 (0.78–0.96)	► 0.72 (0.20–0.89)
3b	► 0.74 (0.47–0.89)	► 0.75 (0.48–0.89)	▲ 0.85 (0.36–0.95)
3c	▲ 0.80 (0.57–0.92)	► 0.66 (0.31–0.85)	► 0.67 (0.15–0.87)
1°. TT	▲ 0.83 (0.62–0.93)	▲ 0.78 (0.53–0.91)	▲ 0.78 (0.45–0.91)
4a	▲ 0.89 (0.74–0.96)	▲ 0.82 (0.61–0.93)	▲ 0.91 (0.76–0.96)
4b	*	► 0.73 (0.44–0.88)	*
4c	▲ 0.78 (0.52–0.90)	▲ 0.80 (0.56–0.92)	▲ 0.83 (0.53–0.94)

▲ (ICC > 0.75) = good reliability.
► (ICC 0.50–0.75) = moderate reliability.
▼ (ICC 0.0–0.50) = low reliability.
*Excluded from reliability due to a significant difference.
In PT thickness between first and second measurement.

Table 4
Intra- & interrater reliability for non-dominant leg presented as 95% LOA and as percentage of the mean PT thickness (LOA %)

	Examiner I		Examiner II		Examiner I + II	
	LOA (cm)	LOA (%)	LOA (cm)	LOA (%)	LOA (cm)	LOA (%)
AP	*	*	0,10	30,3	*	*
1a	0,11	32,4	0,10	28,5	0,09	25,9
1b	0,09	25,5	*	*	*	*
1c	0,09	27,6	0,10	29,6	0,12	35,5
1 u. AP	*	*	0,10	29,9	*	*
2a	0,06	18,9	*	*	*	*
2b	0,07	21,3	0,09	27,5	0,06	20,2
2c	0,08	23,9	0,06	18,8	0,07	22,2
TT	0,12	28,0	0,09	22,3	0,15	35,6
3a	0,05	15,4	0,06	16,2	0,10	29,5
3b	0,09	28,6	0,07	21,1	0,06	19,1
3c	0,08	22,1	0,10	26,6	0,10	28,7
1°. TT	0,07	20,3	0,08	21,6	0,09	26,5
4a	0,06	18,7	0,07	22,5	0,07	21,9
4b	*	*	0,09	28,4	*	*
4c	0,06	20,7	0,08	22,6	0,07	22,1

*Excluded from measurement precision analysis due to a significant Difference in PT thickness between first and second measurement.

References

Bentman, S., O'Sullivan, C., Stokes, M., 2010. Thickness of the middle trapezius muscle measured by rehabilitative ultrasound imaging: description of the technique and reliability study. Clin. Physiol. Funct. Imaging 30, 426–431.
Bjordal, J.M., Demmink, J.H., Ljunggren, A.E., 2003. Tendon thickness and depth from skin for supraspinatus, common wrist and finger extensors, patellar and achilles tendons. Physiotherapy 89, 375–383.
Black, J., Cook, J., Kiss, Z.S., Smith, M., 2004. Intertester reliability of sonography in patellar tendinopathy. J. Ultrasound Med. Off. J. Am. Inst. Ultrasound Med. 23, 671–675.
Cheng, J.-W., Tsai, W.-C., Yu, T.-Y., Huang, K.-Y., 2012. Reproducibility of sonographic measurement of thickness and echogenicity of the plantar fascia. J. Clin. Ultrasound JCU 40, 14–19.
Costa, L.O.P., Maher, C.G., Latimer, J., Hodges, P.W., Shirley, D., 2009. An investigation of the reproducibility of ultrasound measures of abdominal muscle activation in patients with chronic non-specific low back pain. Eur. Spine J. Off. Publ. Eur. Spine Soc. Eur. Spinal Deform. Soc. Eur. Sect. Cerv. Spine Res. Soc. 18, 1059–1065.
Craig, M.E., Duffin, A.C., Gallego, P.H., Lam, A., Cusumano, J., Hing, S., Donaghue, K.C., 2008. Plantar fascia thickness, a measure of tissue glycation, predicts the development of complications in adolescents with type 1 diabetes. Diabetes Care 31, 1201–1206.
Ekizos, A., Papatzika, F., Charcharis, G., Bohm, S., Mersmann, F., Arampatzis, A., 2013. Ultrasound does not provide reliable results for the measurement of the patellar tendon cross sectional area. J. Electromyogr. Kinesiol. Off. J. Int. Soc. Electrophysiol. Kinesiol. 23, 1278–1282.
Finnoff, J.T., 2016. The evolution of diagnostic and interventional ultrasound in sports medicine. PM&R 8, S133–S138.
Fredberg, U., Bolvig, L., Pfeiffer-Jensen, M., Clemmensen, D., Jakobsen, B.W., Stengaard-Pedersen, K., 2004. Ultrasonography as a tool for diagnosis, guidance of local steroid injection and, together with pressure algometry, monitoring of the treatment of athletes with chronic jumper's knee and Achilles tendinitis: a randomized, double-blind, placebo-controlled study. Scand. J. Rheumatol. 33, 94–101.
Fredberg, U., Bolvig, L., Andersen, N.T., Stengaard-Pedersen, K., 2008. Ultrasonography in evaluation of Achilles and patella tendon thickness. Ultraschall Med. Stuttg. Ger. 29, 60–65, 1980.
Gellhorn, A.C., Carlson, M.J., 2013. Inter-rater, intra-rater, and inter-machine reliability of quantitative ultrasound measurements of the patellar tendon. Ultrasound Med. Biol. 39, 791–796.
Grassi, W., Filippucci, E., Farina, A., Cervini, C., 2000. Sonographic imaging of tendons. Arthritis Rheum. 43, 969–976.
Helland, C., Bojsen-Møller, J., Raastad, T., Seynnes, O.R., Moltubakk, M.M., Jakobsen, V., Visnes, H., Bahr, R., 2013. Mechanical properties of the patellar tendon in elite volleyball players with and without patellar tendinopathy. Br. J. Sports Med. 47, 862–868.
Henderson, R.E.A., Walker, B.F., Young, K.J., 2015. The accuracy of diagnostic ultrasound imaging for musculoskeletal soft tissue pathology of the extremities: a comprehensive review of the literature. Chiropr. Man. Ther. 23, 31.
Johnson, D.P., Wakeley, C.J., Watt, I., 1996. Magnetic resonance imaging of patellar tendonitis. J. Bone Joint Surg. Br. 78, 452–457.
Kainberger, F., Mittermaier, F., Seidl, G., Parth, E., Weinstabl, R., 1997. Imaging of tendons–adaptation, degeneration, rupture. Eur. J. Radiol. 25, 209–222.
Khan, K.M., Bonar, F., Desmond, P.M., Cook, J.L., Young, D.A., Visentini, P.J., Fehrmann, M.W., Kiss, Z.S., O'Brien, P.A., Harcourt, P.R., et al., 1996. Patellar tendinosis (jumper's knee): findings at histopathologic examination, US, and MR imaging. Victorian Institute of Sport Tendon Study Group. Radiology 200, 821–827.
Koppenhaver, S.L., Parent, E.C., Teyhen, D.S., Hebert, J.J., Fritz, J.M., 2009. The effect of averaging multiple trials on measurement error during ultrasound imaging of transversus abdominis and lumbar multifidus muscles in individuals with low back pain. J. Orthop. Sports Phys. Ther. 39, 604–611.
Lian, O.B., Engebretsen, L., Bahr, R., 2005. Prevalence of jumper's knee among elite athletes from different sports: a cross-sectional study. Am. J. Sports Med. 33, 561–567.
Liang, H.-W., Wang, T.-G., Chen, W.-S., Hou, S.-M., 2007. Thinner plantar fascia predicts decreased pain after extracorporeal shock wave therapy. Clin. Orthop. 460, 219–225.
Mahowald, S., Legge, B.S., Grady, J.F., 2011. The correlation between plantar fascia thickness and symptoms of plantar fasciitis. J. Am. Podiatr. Med. Assoc. 101, 385–389.
Martinoli, C., 2010. Musculoskeletal ultrasound: technical guidelines. Insights Imag. 1, 99–141.
Miller, T.T., 2013. The patellar tendon. Semin. Muscoskel. Radiol. 17, 56–59.
O'Connor, P.J., Grainger, A.J., Morgan, S.R., Smith, K.L., Waterton, J.C., Nash, A.F.P., 2004. Ultrasound assessment of tendons in asymptomatic volunteers: a study of reproducibility. Eur. Radiol. 14, 1968–1973.
O'Sullivan, C., Bentman, S., Bennett, K., Stokes, M., 2007. Rehabilitative ultrasound imaging of the lower trapezius muscle: technical description and reliability. J. Orthop. Sports Phys. Ther. 37, 620–626.
Portney, L.G., Watkins, M.P., 2014. Foundations of Clinical Research: Applications to Practice. Pearson, Harlow.
Rathleff, M.S., Moelgaard, C., Lykkegaard Olesen, J., 2011. Intra- and interobserver reliability of quantitative ultrasound measurement of the plantar fascia. J. Clin. Ultrasound JCU 39, 128–134.
Skou, S.T., Aalkjaer, J.M., 2013. Ultrasonographic measurement of patellar tendon thickness–a study of intra- and interobserver reliability. Clin. Imaging 37, 934–937.
Springer, B.A., Mielcarek, B.J., Nesfield, T.K., Teyhen, D.S., 2006. Relationships among lateral abdominal muscles, gender, body mass index, and hand dominance. J. Orthop. Sports Phys. Ther. 36, 289–297.
Sunding, K., Fahlström, M., Werner, S., Forssblad, M., Willberg, L., 2014. Evaluation of Achilles and patellar tendinopathy with greyscale ultrasound and colour Doppler: using a four-grade scale. Knee Surg. Sports Traumatol. Arthrosc. Off. J. ESSKA.
Wakefield, R.J., Balint, P.V., Szkudlarek, M., Filippucci, E., Backhaus, M., D'Agostino, M.-A., Sanchez, E.N., Iagnocco, A., Schmidt, W.A., Bruyn, G.A.W., et al., 2005. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. J. Rheumatol. 32, 2485–2487.
Wallwork, T.L., Hides, J.A., Stanton, W.R., 2007. Intrarater and interrater reliability of assessment of lumbar multifidus muscle thickness using rehabilitative ultrasound imaging. J. Orthop. Sports Phys. Ther. 37, 608–612.
Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Cond. Res. Natl. Strength Cond. Assoc. 19, 231–240.
Ying, M., Yeung, E., Li, B., Li, W., Lui, M., Tsoi, C.-W., 2003. Sonographic evaluation of the size of Achilles tendon: the effect of exercise and dominance of the ankle. Ultrasound Med. Biol. 29, 637–642.
Zwerver, J., Bredeweg, S.W., van den Akker-Scheek, I., 2011. Prevalence of Jumper's knee among nonelite athletes from different sports: a cross-sectional survey. Am. J. Sports Med. 39, 1984–1988.