

Ultrasound measures of tendon thickness: Intra-rater, Inter-rater and Inter-machine reliability

María Elena del Baño-Aledo¹
 Jacinto Javier Martínez-Payá¹
 José Ríos-Díaz^{1,2}
 Silvia Mejías-Suárez³
 Sergio Serrano-Carmona⁴
 Ana de Groot-Ferrando⁵

¹ ECOFISTEM Research Group, Health Sciences Department, Facultad de Ciencias de la Salud, UCAM, Guadalupe (Murcia), Spain

² Centro Universitario de Ciencias de la Salud San Rafael-Nebrija, Madrid, Spain

³ Physio Consulting, Guanarteme, Las Palmas de Gran Canaria, Spain

⁴ FISIOCONTROL, Sevilla, Spain

⁵ Sport Clinic Arena, Alicante, Comunidad Valenciana, Spain

Corresponding author:

María Elena del Baño-Aledo
 ECOFISTEM Research Group, Health Sciences Department, Facultad de Ciencias de la Salud, UCAM UCAM Campus de Cartagena
 Calle Porto Alegre
 30310 Cartagena, Spain
 E-mail: mbano@ucam.edu

Summary

Background: Ultrasound imaging is often used by physiotherapists and other healthcare professionals but the reliability of image acquisition with different ultrasound machines is unknown. The objective was to compare the intra-rater, inter-rater and intermachine reliability of thickness measurements of the plantar fascia (PF), Achilles tendon (AT), patellar tendon (PT) and elbow common extensor tendon (ECET) with musculoskeletal ultrasound imaging (MSUS).

Methods: Tendon thickness was measured in four anatomical structures (14 participants, 28 images per tendon) by two sonographers and with two different ultrasound machines. Intraclass Correlation Coefficients (ICCs) and Bland-Altman plots were calculated. The standard error of measurement (SEM) and minimum detectable difference (MDD) were calculated.

Results: Inter-rater reliability was excellent for AT

(ICC=0.98; 95% CI= 0.96-0.99) and very good for PT (ICC=0.85; 95% CI = 0.67-0.93) and ECET (ICC=0.81; 95% CI= 0.72-0.94). Reliability for PF was moderate, with an ICC of 0.63 (CI 95%= 0.20-0.83). Bland-Altman plot for inter-machine reliability showed a mean difference of 1 mm for PF measurements and a mean difference of 4 mm and 20 mm for AT and PT. The relative SEMs were below 7% and the MDCs were below 0.7 mm.

Conclusion: The MSUS reliability in measuring thickness of the four tendons is confirmed by the homogeneous readings intra sonographers, between operators and between different machines. **Level of evidence:** Tendon thickness can be measured reliably on different ultrasound devices, which is an important step forward in the use of this technique in daily clinical practice and research. **Level of evidence:** III.

KEY WORDS: quantitative ultrasound, reliability, tendinopathy, ultrasound imaging.

Introduction

Normal tendons are not painful and their ultrasonographic appearance shows parallel margins, whereas tendinopathy has been associated with tendon enlargement¹. Moreover, a thickening of the tendon is highly correlated with pain, so quantifying changes in tendon thickness may provide an indirect means of measuring function and monitoring treatments designed to reduce tendon hypertrophy².

Musculoskeletal ultrasound imaging (MSUS) is increasingly used in rehabilitation services to evaluate soft tissue structures both in research and clinical settings to guide rehabilitation decisions³. MSUS has the advantages that it can be carried out rapidly, it is non-invasive, the apparatus is portable and relatively inexpensive and it is easily applicable for the routine assessment of the tendinopathy in clinical practice⁴.

In fact, ultrasound has been used widely for the assessment of tendon pathology, particularly the Achilles and patellar tendons^{2,5} and plantar aponeurosis^{6,7}. In the upper limb, several Authors have argued in favour of the diagnostic value of elbow common extensor tendon (ECET) thickness in lateral elbow tendinopathy⁸. Several studies have investigated the reliability of morphological measurements of clinically relevant tendons, but none included a physiotherapist as the operator. MSUS has been traditional-

ly used by rehabilitation professionals to assist in the application of therapeutic interventions providing feedback to the patient. Over the last decade there has been a rapid development of this technique to evaluate muscle and related soft tissue morphology, for which purpose, new skills in ultrasound examination are required³.

Although intra and inter-rater reliability of MSUS has been reported to be very good^{7,9}, inter-rater reliability has been evaluated only in the context of two examiners analyzing stored MSUS data acquired by a single examiner. For example, Skou et al. (2013) found that using a mean of two measurements of the patellar tendon thickness increased intratester (from ICC 0.84 and 0.70 to 0.94 and 0.89) and inter-tester (from ICC 0.70 to 0.78) reliability⁹.

Cheng et al. (2012) explored the effect of operator influence during MSUS of plantar fascia. For that, they analyzed the reliability of two sonographers independently imaging the plantar fascia in both longitudinal and transverse planes. They found that in inter-rater evaluations of patients with painful heels, the reliability of transverse scanning was markedly worse than that of longitudinal scanning⁷. In fact, ultrasonography is frequently described as an operator-dependent imaging modality, so operators with different levels of experience following the same protocols can lead to significant variations in the image obtained.

Given the large number of ultrasound machines that are available, and that sonographic standards have been established for the measurement of anatomical structures, it is critical that these measurements are comparable from one machine to another. Studies on inter-machine reliability are scarce⁴, while in those that exist, the statistical methods used are not always correct or appropriate, or the Authors did not adequately report the methodology. Not providing information

about blinding, the distribution of data, the calculation of the sample size, the use of t-test or Pearson's linear correlation coefficient to compare measurements are the more common failing in this respect¹⁰.

As tendon thickness can be a good indicator of tendinopathy¹¹, research into the reliability of this measure is desirable. It was hypothesized that there would be slight differences in thickness measurements between two sonographers with different levels of experience and between two different ultrasound machines. For this reason, the present study looks at inter-, intra-rater and inter-machine reliability in MSUS image acquisition relating to several clinically important tendons (plantar fascia, Achilles tendon, patellar tendon and elbow common extensor tendon).

Materials and methods

Participants and sample size

This study was approved by the institutional ethical board and all procedures were performed according to what described in Padulo et al¹². All participants provided written informed consent.

Fourteen participants were recruited from the university staff and students [5 males, 9 females; mean age 32.1 years (SD 9.47); mean height 1.68 (SD 0.07); mean weight 63.6 kg. (SD 8.41) and mean BMI 22.4 (SD 2.06)]. Subjects were excluded if they had history of tendinopathy or fascia disorders or painful episodes affecting the tendons being imaged, All tendons were imaged bilaterally so data were obtained for four sets of 28 tendons each one (Fig. 1).

The estimation of sample size was 27 images for each tendon based on an alpha-error of 0.05, statistical power of 0.80 and a minimum ICC of 0.6 and size effect of 0.2.

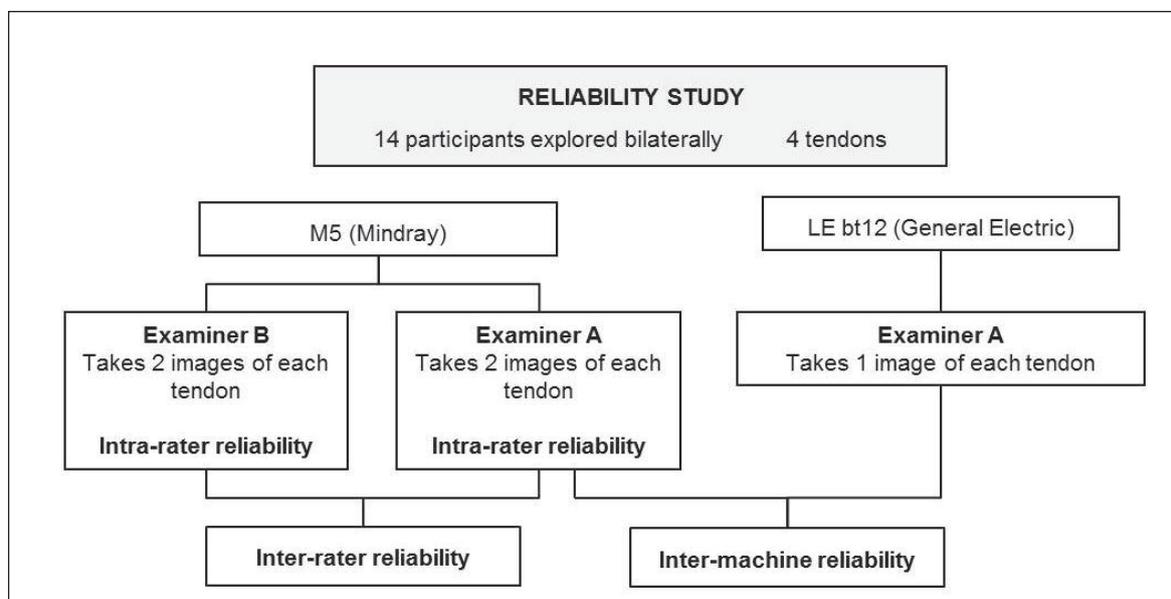


Figure 1. Sequence of intra-, inter-rater and inter-machines reliability procedure.

Investigators

One of the two sonographers was a physiotherapist with more than 12 years' experience in performing MSUS (Examiner A), while the other examiner had no prior experience in the use of ultrasound for examination purposes (Examiner B). The examiners were instructed and trained in the test setup before the start of the study to ensure that they followed the study protocol. The training, undertaken by both operators involved a one hour session. The first part of the session was used to remember the theoretical setting, including the physical basis of examinations and palpation and measurement methods. In a second part, the sonographers had to analyze the four structures by ultrasound examination following the protocols defined in previous studies, which include the positioning of the patient and probe, and recognition of tendon borders in the resulting images^{1,6,8}.

Ultrasound protocol

Two portable ultrasound machines were used: a M5 (Mindray, Shenzhen, China) with a 5-10 MHz linear transducer (7L4S) and a Logiq E (General Electric, Germany, 2013) with a 5-13 MHz linear transducer (12L RS). All system-setting parameters were kept in B-mode and constant for each structure analyzed (Table I). Compared with the AT or PT, the PF is a deeper structure, meaning that the sonographic examination requires lower frequency and changes in grey map to avoid artifacts. Time gain compensation was performed in neutral position (slides in the middle).

For the plantar fascia (PF) examination each participant lay prone with the ankle in a neutral position and toes extended against the examination couch. In a longitudinal view, the thickness of the PF was measured from the anterior edge of the inferior calcaneal border vertically to the inferior border of the plantar fascia (Fig. 2)⁶.

During Achilles tendon (AT) examination, participants were in a prone position with their ankles in neutral position with feet hanging freely over the examination couch. The AT thickness was measured in a longitudinal scan 3 cm proximal to the calcaneus tendon insertion¹. This point was identified with a mark on the skin.

Longitudinal patellar tendon (PT) images were obtained with the participant in a supine position with both knees flexed at 30°. The PT thickness was measured 1 cm distal to the inferior patellar pole¹.

For the elbow common extensor tendon (ECET) examination, the participant was seated in front of the sonographer with the elbow flexed at 90° and the forearm resting on the examination couch in the prone position. The longitudinal scan was performed with the transducer on the lateral epicondyle with the linear array parallel to the tendon fibers. Thickness was measured by drawing a line from the humeral cortical bone to the thickest point of the tendon introducing the elbow's collateral ligamentous complex, which is intimately associated with the deepest fibers of the ECET⁸.

The Figure 2 (a-h) shows a MSUS image for each structure.

At each tendon, two independent images were taken and measured by each examiner using the M5 ultrasound machine, and one image was performed by examiner A with the LogiqE to analyze inter-machine reliability. As physical activity and time-of-day might influence the thickness of the PF, we chose a period of about 30 minutes between test and retest and all participants remained seated between sessions.

All sonograms were stored in the ultrasound machine and measured one week after by both examiners using the built-in software in the scanner with an electronic calliper. Each examiner was blinded to the other's measurements and to their previous results.

Analysis of data

Statistical analyses were performed using IBM SPSS Statistics 19.0 (IBM SPSS Inc. USA, 2010). The mean and standard deviation were used for descriptive statistics with 95% confidence intervals (CI). Data were normally distributed, as confirmed by visual inspection of the Q-Q plots, kurtosis and skewness coefficients and the Kolmogorov-Smirnov test.

To determine the reliability of measurements between examiners and between machines, ICCs were calculated using a two-way random effects model and absolute agreement type. Following criteria were used to judge the reliability coefficients: very low (<0.20), low (0.21-0.40), moderate (0.41-0.60), good (0.61-

Table I. Ultrasonography setting parameters.

Structure	Depth (cm)		Focal point (cm)		Grey Map		Frequency (MHz)		Gain (dB)	
	M5	LE bt12	M5	LE bt12	M5	LE bt12	M5	LE bt12	M5	LE bt12
PF ¹	3	3	1	1	6	H	5	12	100	98
AT ²	2	2	0.5	0.5	4	H	10	12	100	98
PT ³	3	3	0.5	0.5	4	H	10	12	100	98
ECET ⁴	2.5	2.5	0.5	0.5	4	H	10	12	100	98

¹Plantar fascia; ²Achilles tendon; ³Patellar tendon; ⁴Elbow common extensor tendon.

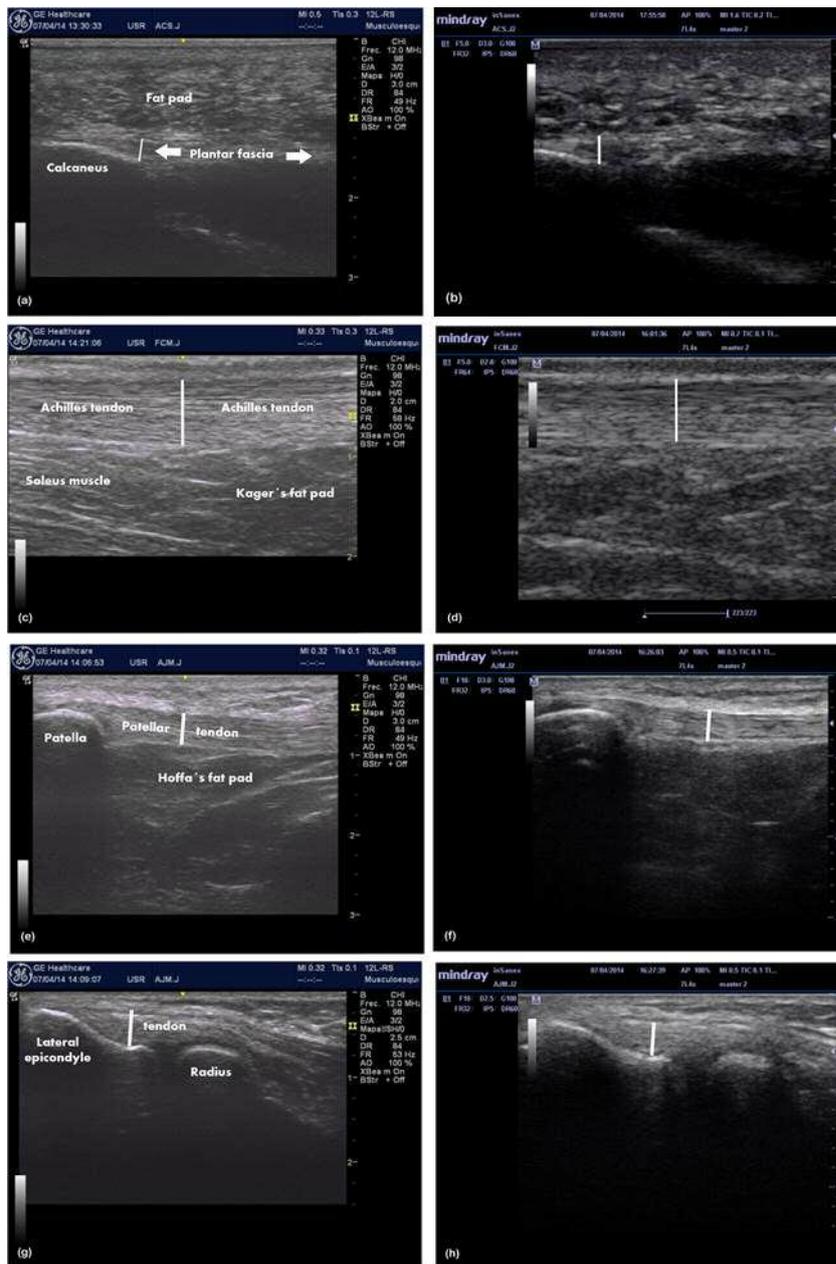


Figure 2. Ultrasound images for Plantar fascia, Achilles, Patellar and Elbow common extensor tendons obtained by examiner A with General Electric LEbt12 and Mindray M5 ultrasound device: (a) Plantar fascia with LEbt12; (b) Plantar fascia with M5; (c) Achilles tendon with LEbt12; (d) Achilles tendon with M5; (e) Patellar tendon with LEbt12; (f) Patellar tendon with M5; (g) Elbow common extensor tendon with LEbt12; (h) Elbow common extensor tendon with M5.

0.80) and excellent (0.81-1.00)¹³. Intra-rater reliability was calculated using both the first and second measurements that the raters recorded from each structure. Each rater's first measure was used for inter-rater reliability assessment to control any learning effect. Examiner A's first measurement was also used to evaluate inter-machine reliability. Measurement precision was evaluated using standard error of measurement (SEM) and its relative value with respect the average of all measurements and the minimum detectable difference (MDD). MDD is useful for determining whether a change in the thickness of a structure is due to a real change or falls

within the limits of error of the measuring method. Additionally, in a Bland-Altman plot, 95% limits of agreement (LOA) and LOA as a percentage of the mean (LOA %) were calculated. Bland-Altman plots were constructed to evaluate the presence of systematic errors. The difference between each pair of measures (Y-axis) is plotted against mean of both of them (X-axis), also global mean of differences and 95% limits of agreement complete the graphical information. LOA is presented as the difference between the mean difference and the upper and lower LOA to understand the result in a clinical context¹⁴. Passing-Bablok (1985) regression method was applied to evaluate the possible systematic bias in measurement¹⁵.

Results

The results of the MSUS measurements are summarized in Table II. No systematic biases were found by Passing-Bablok regression for intra-rater or inter-rater reliability.

Intra-rater Reliability

The ICCs for examiner A were very good (over 0.8) in all the anatomical locations. The intra-rater ICCs for examiner B were slightly lower than for examiner A in the case of PF and PT, for which the ICC of examiner B were 0.65 (95% CI= 0.25 to 0.84) and 0.82 (95% CI= 0.61 to 0.92), respectively.

Inter-rater Reliability

Inter-rater reliability was excellent when measuring AT thickness (ICC=0.98; 95% CI= 0.96 to 0.99) and very good for PT (ICC=0.85; 95% CI = 0.67 to 0.93) and ECET (ICC=0.81; 95% CI= 0.72 to 0.94). The reliability for PF thickness was moderate, with an ICC of 0.63 (CI 95%= 0.20 to 0.83).

Inter-machine Reliability

Inter-machine reliability was excellent for AT, PT and ECET, with ICC values of between 0.94 and 0.99 (Tab. III) and good for PF measurements with an ICC of 0.80 (95% CI =0.57 to 0.91).

Measurement Precision

The relative SEMs were below 4% for examiner A and 6% for examiner B. Examiner B was less precise in PF and PT measurements. In these regions the inter-rater SEM was slightly higher in the fascia (7%). For inter-machine SEM, the values were below 5% (Tab. III).

In all regions and comparisons, the MDD was below of

0.7 mm. The maximum value recorded was for inter-rater comparison of ECET. In the AT, the error in all comparisons was practically non-existent (Tab. III).

The Bland-Altman method describes the difference between measurements, as well as any bias that may be present. These measurements can be interpreted as the accuracy of assessments (Tab. IV). The limits of agreement (LOA) for intra- and inter-rater reliability in the measurement of tendon and fascia thicknesses showed mean difference that represented less than 2% of the corresponding thickness. The results were similar when the measurements taken with both ultrasound machines were compared.

For instance, the Bland-Altman plot for inter-machine reliability showed a mean difference of 1 µm (95% LOA= 58 to 60 µm) for PF measurements and a mean difference of 4 µm (95% LOA= 120 to 130 µm) and 20 µm (95% LOA= 420 to 460 µm) for AT and PT, respectively (Fig. 2 a-h).

Discussion

This study analyzed the reproducibility of thickness measurements for four structures of clinical interest: PF, AT, PT and ECET. The results suggest that ultrasound measurements of tendon thickness taken in healthy participants, with the exception of PF measurement, are reliable, even when they are established with different ultrasound machines. These are important findings, as they will improve the clinical applicability of quantitative MSUS.

The difficulties in imaging these tendons can be divided into anatomical and technical. Anatomical difficulties include a thick subcutaneous adipose layer that makes the tendon boundary less clear, and bones morphology which can generate an ultrasonographic

Table II. Results of the US measurement of plantar fascia, Achilles, patellar and elbow common extensor tendons thickness.

	Thickness (mm) Mean (SD) (95 % CI)			
	Plantar fascia	Achilles tendon	Patellar tendon	ECET
Machine 1. Mindray				
Examiner A 1 st evaluation	3.23 (0.320) (2.80-4.10)	4.73 (0.414) (4.57-4.89)	4.39 (0.584) (4.17-4.62)	5.61 (0.517) (5.41-5.81)
Examiner A 2 nd evaluation	3.24 (0.250) (2.70-3.80)	4.76 (0.454) (4.59-4.94)	4.44 (0.584) (4.21-4.66)	5.57 (0.620) (5.33-5.81)
Examiner B 1 st evaluation	3.27 (0.376) (2.60-4.40)	4.75 (0.502) (4.56-4.95)	4.45 (0.515) (4.25-4.65)	5.78 (0.618) (5.54-6.02)
Examiner B 2 nd evaluation	3.27 (0.289) (2.90-4.20)	4.75 (0.491) (4.56-4.94)	4.38 (0.462) (4.20-4.56)	5.72 (0.604) (5.49-5.96)
Machine 2. Logiq-E				
Examiner A 1 st evaluation	3.26 (0.361) (2.50-4.40)	4.71 (0.482) (4.52-4.90)	4.33 (0.564) (4.11-4.55)	5.63 (0.574) (5.40-5.85)

SD: standard deviation. 95% CI: 95% confidence interval. ECET: elbow common extensor tendon.

Table III. Intra-class correlation coefficients and SEM for inter- and intra-rater and inter-machine reliability measuring four thickness.

Comparison	Statistic	Plantar fascia	Achilles tendon	Patellar tendon	ECET	
Intra-rater						
Examiner A	ICC	0.83	0.98	0.91	0.91	
	95% CI	0.64 to 0.92	0.95 to 0.99	0.80 to 0.96	0.81 to 0.96	
	SEM (mm)	0.12 (4%)	0.06 (1%)	0.18 (4%)	0.17 (3%)	
	MDD (mm)	0.33	0.17	0.49	0.47	
Examiner B	ICC	0.65	0.99	0.82	0.92	
	95% CI	0.25 to 0.84	0.99 to 1.0	0.61 to 0.92	0.83 to 0.96	
	SEM (mm)	0.20 (6%)	<0.001 (0%)	0.21 (5%)	0.17 (3%)	
	MDD (mm)	0.55	<0.001	0.57	0.48	
Inter-rater						
	ICC	0.63	0.98	0.85	0.81	
	95% CI	0.20 to 0.83	0.96 to 0.99	0.67 to 0.93	0.72 to 0.94	
	SEM (mm)	0.21 (7%)	0.065 (1%)	0.21 (5%)	0.25 (4%)	
	MDD (mm)	0.59	0.18	0.59	0.69	
	Inter-machine					
		ICC	0.80	0.99	0.96	0.94
95% CI		0.57 to 0.91	0.99 to 1.0	0.92 to 0.97	0.91 to 0.97	
SEM (mm)		0.15 (5%)	<0.001 (0%)	0.11 (3%)	0.13 (2%)	
	MDD (mm)	0.42	<0.001	0.32	0.37	

ICC: Intra-class correlation coefficient for two-way mixed effects model and absolute agreement. SEM: Standard Error of Measure and relative error respect of the average of measures. MDD: Minimum Detectable Difference. 95% CI: 95% confidence interval. ECET: elbow common extensor tendon.

Table IV. Limits of agreement and percentage of the mean thickness for intra- and inter-rater reliability.

Agreement	Mean difference (%)* (95% LOA)			
	Plantar fascia	Achilles tendon	Patellar tendon	ECET
Intra-rater				
Examiner A	0.4 (0.6 %)	1 (0.1 %)	70 (2.0 %)	20 (0.2 %)
	(-50 to 51)	(-100 to 110)	(-590 to 723)	(-700 to 710)
Examiner B	4 (0.8 %)	30 (0.6 %)	70 (1.0 %)	30 (0.5 %)
	(-530 to 620)	(-220 to 280)	(-690 to 820)	(630 to 570)
Inter-rater				
	7 (0.2 %)	20 (0.4 %)	70 (1.7%)	120 (1.8 %)
	(-720.2 to 733)	(-230 to 270)	(-680 to 820)	(-680 to 920)
Inter-machine				
	1 (0.1%)	4 (0.1 %)	20 (0.6 %)	40 (0.7%)
	(-580 to 600)	(-120 to 130)	(-420 to 460)	(-550 to 480)

*Mean differences in micrometers (µm). ECET: elbow common extensor tendon. LOA: limit of agreement.

shadow¹⁶. Technical difficulties attributable to raters include probe pressure, which can influence thickness measurements, and probe inclination, which can produce image distortion¹⁷.

Plantar fascia Reliability

We analyzed PF thickness in healthy participants and, although reliability was good, correct identification of the structure was more complicated than for

the other three regions. PF is a deep structure and consequently its visualization is probably influenced by sonographic attenuation of the overlying soft tissue, including the hyperkeratotic skin and heel fat. Indeed, there is considerable heterogeneity between studies in this respect and not all provide information on reproducibility^{7,18,19}.

Cheng et al. (2012) obtained ICC values slightly higher than ours (ICC of 0.86 for inter-rater reliability and ICC of 0.77 and 0.78 for intra-rater reliability for each rater), probably because both examiners had similar

level of experience in musculoskeletal MSUS⁷. Rathleff et al. (2011) explored 40 plantar fasciae thicknesses to investigate improvements in measurement precision when using the mean value of three measurements compared to one. They obtained an ICC of 0.62 for interobserver reliability when using one sonogram, which increased to 0.82 when using the average of three sonograms¹⁸. To reduce discrepancies in the interpretation of the thickness measurements of the PF, the mean value of different measurement rather than one measurement should be considered in both clinical practice and research. Despite these differences in reliability values, Mohseni-Bandpei et al. (2014), in a recent systematic review, concluded that MSUS can be considered an useful imaging modality for assessing PF thickness¹⁹.

Achilles tendon reliability

Several Authors have related the increase in AT thickness with pathological processes and, although the structure is easily observable, intra- and interrater error due to the examiner's experience can create a high degree of variability in the results of the examination of thickness using ultrasound images^{17,20}. O'Connor et al. (2004) found that inter-observer variability was a greater source of error than intra-observer variability¹⁷ and Ríos-Díaz et al. (2010) found an excellent ICC for inter-rater reliability but wide confidence interval (ICC= 0.94; 95% IC= 0.58 to 0.98)²⁰. An explanation for why we found higher reliability in all analyses (with an ICC ranged from 0.98 to 0.99) could be that these Authors obtained transverse view of the AT, whereas the present study scanned the tendons longitudinally. Some Authors recommended measuring normal AT thickness longitudinally 20 mm from the attachment at the calcaneus since this distance can be measured without moving the transducer and because this point is proximal from the blurred and poorly defined distal part of the AT¹. Effectively, when measuring tendons, it seems necessary to apply a longitudinal scan as this is the only method allowing the examiner to record the distance from the point where the thickness is measured to the bony attachment.

Patellar tendon reliability

Patellar tendon thickness is a clinically relevant measurement when evaluating the effect of a given treatment protocol since a reduction in PT thickness has been observed following successful treatment of patellar tendinopathy⁹. Gellhorn et al. (2012) obtained excellent inter-rater reliability for PT length (ICC= 0.96; 95% CI= 0.87 to 0.99) with a mean difference that represented less than 0.9% of the corresponding length². In our study, the mean difference of inter-rater measurements of PT thickness represented 1.7% of the corresponding thickness. PT length is a more accurate measurement than thickness because

the measuring points are bony prominences that are easily identifiable in the ultrasound image.

We have found only one study about inter-machine reliability for the measurement of PT length and cross-sectional area in which the results were excellent, with an ICC between 0.98 and 0.91 respectively. Similarly, the Bland-Altman plot for inter-machine reliability in the measurement of tendon cross-sectional area showed a possible bias of 20.01 mm², with a 95% limit of agreement of 20.14 to 0.11 mm², which suggests a possible bias in the measurement⁴. To our knowledge, no studies have evaluated inter-machine reliability of ultrasound systems for PT thickness measurement, so direct comparison is difficult.

Elbow common extensors tendon reliability

Finally, we evaluated reproducibility for the ECET thickness measurement. Other studies have shown a moderate level of inter-rater reliability when measuring ECET thickness with an ICC ranged from 0.45 to 0.65²¹ or an ICC of 0.46 (95% IC= 0.16 to 0.68)²². That our results were better was perhaps due to the fact that we used healthy participants. Measuring affected tendons has been reported to be more difficult with poor agreement as regard tendon thickening, probably as a consequence of ambiguous tendon boundaries²².

With training, rehabilitation professionals may be able to perform reliable MSUS in a clinical context. In our study, the two examiners had different degree of experience in MSUS exploration, as can be seen in the results of intra-rater reliability analysis. Previous studies also found differences in reliability comparing examiners with different degrees of experience in MSUS assessment^{4,22}. This highlights the importance of standardizing MSUS assessment and for continuous training beyond basic training in image acquisition and measurements. Currently, there are no specific training guidelines relating to rehabilitation professionals so future research needs to determine optimal training strategies to ensure that health care professionals using MSUS are properly trained to use and interpret MSUS as an effective adjunct to traditional rehabilitation. These guidelines should include MSUS training on basic principles, the recognition of artifacts, parameter selection, participant screening, prudent use, and in-depth anatomy education.

Study limitations

This study has a number of limitations. First, the inter-machine reliability study is limited to only two machines. An exhaustive study of inter-machine reliability is difficult due to the wide number of manufacturers that currently design ultrasound devices. However, the high level of agreement we found between our two machines provides evidence that the measurements performed in clinical practice with different machines are reliable enough to be compared.

Second, we did not address the reproducibility of pathological findings in this study. Healthy tendons often have better-defined borders and the collagen pattern is more easily visualized. Therefore, a reliability study of individuals with healthy tendons may result in inflated reliability estimates that would not translate to tendons with tendinopathy which may be more difficult to image.

Third, since it is now also possible to perform quantitative MSUS analyses, it might be interesting to check reproducibility in parameters such as echogenicity, echovariation or echotexture^{16,20}, especially because these variables are much more sensitive to changes in the image than thickness.

In conclusion, our study shows that MSUS allows the reliable measurement of thickness of the PF and AT, PT and ECET, as confirmed by the homogeneous readings between sonographers. Tendon thickness can be measured reliably on different ultrasound devices, which is an important step forward in the use of this technique in daily clinical practice and research. For that purpose, we present an exhaustive and in-depth descriptive method of examining reliability of tendon thickness measurement using MSUS, which can be used in future studies.

Although reliability is context-dependent, these results are of clinical relevance because they will improve the clinical applicability of quantitative MSUS when evaluating the effects of a given treatment protocol. Intra-rater reliability slightly differed depending on the level of experience of the sonographer, which emphasizes the importance of standardization of MSUS assessment and the continuous training of operators.

Conflict of interest

The Authors have no financial or personal relationships with other people or organizations that could inappropriately influence their work.

References

1. Fredberg U, Bolvig L, Andersen NT, et al. Ultrasonography in evaluation of Achilles and patella tendon thickness. *Ultraschall Med.* 2008;29:60-65.
2. Gellhorn AC, Morgenroth DC, Goldstein BA. Novel Sonographic Method of Measuring Patellar Tendon Length. *Ultrasound Med Biol.* 2012;38:719-726.
3. Potter CL, Cairns MC, Stokes M. Use of ultrasound imaging by physiotherapists: A pilot study to survey use, skills and training. *Man Ther.* 2012;17:39-46.
4. 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.
5. Bakkegaard M, Johannsen FE, Højgaard B, et al. Ultrasonography as a prognostic and objective parameter in Achilles tendinopathy: a prospective observational study. *Eur J Radiol.* 2015;84:458-462.
6. Wu CH, Chang KV, Mio S, et al. Sonoelastography of the plantar fascia. *Radiology.* 2011;259:502-507.
7. Cheng JW, Tsai WC, Yu TY, et al. Reproducibility of sonographic measurement of thickness and echogenicity of the plantar fascia. *J Clin Ultrasound.* 2012;40:14-19.
8. Toprak U, Başkan B, Üstüner E, et al. Common extensor tendon thickness measurements at the radiocapitellar region in diagnosis of lateral elbow tendinopathy. *Diagn Interv Radiol Ank Turk.* 2012;18: 566-570.
9. Skou ST, Aalkjaer JM. Ultrasonographic measurement of patellar tendon thickness—a study of intra- and interobserver reliability. *Clin Imaging.* 2013;37:934-937.
10. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res Natl Strength Cond Assoc.* 2005;19:231-240.
11. Ríos-Díaz J, de Groot-Ferrando A, Martínez-Payá JJ, et al. Reliability and reproducibility of a morpho-textural image analysis method over a patellar ligament ultrasonography. *Reumatol Clin.* 2010;6:278-284.
12. Padulo J, Oliva F, Frizziero A, Maffulli N. *Muscles, Ligaments and Tendons Journal - Basic principles and recommendations in clinical and field science research: 2016 update.* MLTJ. 2016;6(1):1-5.
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1997;33:159-174.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.
15. Bablok W, Passing H. Application of statistical procedures in analytical instrument testing. *J Autom Chem.* 1985;7:74-79.
16. Ríos-Díaz J, Martínez-Payá JJ, del Baño-Aledo ME, et al. Sonoelastography of Plantar Fascia: Reproducibility and Pattern Description in Healthy Subjects and Symptomatic Subjects. *Ultrasound Med Biol.* 2015;41:2605-2613.
17. O'Connor PJ, Grainger AJ, Morgan SR, et al. Ultrasound assessment of tendons in asymptomatic volunteers: a study of reproducibility. *Eur Radiol.* 2004;14:1968-1973.
18. Rathleff MS, Moelgaard C, Lykkegaard OJ. Intra- and interobserver reliability of quantitative ultrasound measurement of the plantar fascia. *J Clin Ultrasound.* 2011;39:128-34.
19. Mohseni-Bandpei MA, Nakhaee M, Mousavi ME, et al. Application of ultrasound in the assessment of plantar fascia in patients with plantar fasciitis: a systematic review. *Ultrasound Med Biol.* 2014;40:1737-1754.
20. Ríos-Díaz J, Martínez-Payá JJ, del-Baño-Aledo ME, et al. Intra and inter-observer reliability and reproducibility of a semi-automatic image analysis method over Achilles tendon ultrasonography. *Cuest Fisioter.* 2010;39:190-198.
21. Krogh TP, Fredberg U, Christensen R, et al. Ultrasonographic assessment of tendon thickness, Doppler activity and bony spurs of the elbow in patients with lateral epicondylitis and healthy subjects: a reliability and agreement study. *Ultraschall Med.* 2013;34:468-474.
22. Poltawski L, Ali S, Jayaram V, et al. Reliability of sonographic assessment of tendinopathy in tennis elbow. *Skeletal Radiol.* 2012;41:83-89.