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

Maria Elena del Baño-Aledo
Jacinto Javier Martinez-Payá
José Ríos-Díaz
Silvia Mejías-Suárez
Sergio Serrano-Carmona
Ana de Groot-Ferrando

1 ECOFISTEM Research Group, Health Sciences Department, Facultad de Ciencias de la Salud, UCAM, Guadalupé (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:
Maria 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.

Conclusions: 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: II.

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 tendons2,5 and plantar aponeurosis6,7. 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 examina-
tion are required³.

Although intra and inter-rater reliability of MSUS has
been reported to be very good⁷,⁹, inter-rater reliability
has been evaluated only in the context of two exam-
ners analyzing stored MSUS data acquired by a sin-
gle examiner. For example, Skou et al. (2013) found
that using a mean of two measurements of the patel-
ar 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 in-
fluence during MSUS of plantar fascia. For that, they
analyzed the reliability of two sonographers indepen-
dently imaging the plantar fascia in both longitudinal
and transverse planes. They found that in inter-rater
evaluations of patients with painful heels, the reliabil-
ity of transverse scanning was markedly worse than
that of longitudinal scanning⁷. In fact, ultrasonogra-
phy 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 cor-
rect 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 tendi-
nopathy¹¹, research into the reliability of this measure
is desirable. It was hypothesized that there would be
slight differences in thickness measurements be-
tween two sonographers with different levels of expe-
rience and between two different ultrasound ma-
chines. For this reason, the present study looks at in-
ter, intra-rater and inter-machine reliability in MSUS
image acquisition relating to several clinically impor-
tant 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 univer-
sity 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 ten-
dons 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, statisti-
cal 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. 

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 plantar calcaneal border vertically to the inferior border of the plantar fascia (Fig. 2). 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 insertion. 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 pole.

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 ECET.

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 IMB 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 (0.61-1.00).

Table I. Ultrasonography setting parameters.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Depth (cm)</th>
<th>Focal point (cm)</th>
<th>Grey Map</th>
<th>Frequency (MHz)</th>
<th>Gain (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M5</td>
<td>LE bt12</td>
<td>M5</td>
<td>LE bt12</td>
<td>M5</td>
</tr>
<tr>
<td>PF&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2.5</td>
<td>2.5</td>
<td>0.5</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td>AT&lt;sup&gt;2&lt;/sup&gt;</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>PT&lt;sup&gt;3&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td>ECET&lt;sup&gt;4&lt;/sup&gt;</td>
<td>3</td>
<td>3</td>
<td>0.5</td>
<td>0.5</td>
<td>4</td>
</tr>
</tbody>
</table>

<sup>1</sup>Plantar fascia; <sup>2</sup>Achilles tendon; <sup>3</sup>Patellar tendon; <sup>4</sup>Elbow common extensor tendon.
Intra-rater reliability was calculated using both the first and second measurements that the raters recorded from each structure. Each rater’s first measurement 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 context. Passing-Bablok (1985) regression method was applied to evaluate the possible systematic bias in measurement.
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.

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

SD: standard deviation. 95% CI: 95% confidence interval. ECET: elbow common extensor tendon.
Technical difficulties attributable to raters include probe pressure, which can influence thickness measurements, and probe inclination, which can produce image distortion\(^{17}\).

**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...

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

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>
<thead>
<tr>
<th>Agreement</th>
<th>Mean difference (%)* (95% LOA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plantar fascia</td>
</tr>
<tr>
<td><strong>Intra-rater</strong></td>
<td></td>
</tr>
<tr>
<td>Examine A</td>
<td>0.4 (0.6 %)</td>
</tr>
<tr>
<td></td>
<td>(-50 to 51)</td>
</tr>
<tr>
<td>Examine B</td>
<td>4 (0.8 %)</td>
</tr>
<tr>
<td></td>
<td>(-530 to 620)</td>
</tr>
<tr>
<td><strong>Inter-rater</strong></td>
<td>7 (0.2 %)</td>
</tr>
<tr>
<td></td>
<td>(-720.2 to 733)</td>
</tr>
<tr>
<td><strong>Inter-machine</strong></td>
<td>1 (0.1 %)</td>
</tr>
<tr>
<td></td>
<td>(-580 to 600)</td>
</tr>
</tbody>
</table>

*Mean differences in micrometers (\(\mu\)m). ECET: elbow common extensor tendon. LOA: limit of agreement.
level of experience in musculoskeletal MSUS\textsuperscript{7}. 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 sonograms\textsuperscript{18}. 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 thickness\textsuperscript{19}.

**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\textsuperscript{17,20}. O’Connor et al. (2004) found that inter-observer variability was a greater source of error than intra-observer variability\textsuperscript{17} and Rios-Díaz et al. (2010) found an excellent ICC for inter-rater reliability but wide confidence interval (ICC= 0.94; 95% CI= 0.55 to 0.98)\textsuperscript{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 AT\textsuperscript{1}. 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 tendinopathy\textsuperscript{5}. 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 length\textsuperscript{2}. 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 mm\textsuperscript{2}, with a 95\% limit of agreement of 20.14 to 0.11 mm\textsuperscript{2}, which suggests a possible bias in the measurement\textsuperscript{4}. 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 0.65\textsuperscript{21} or an ICC of 0.46 (95\% IC= 0.16 to 0.68)\textsuperscript{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 boundaries\textsuperscript{22}. 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\textsuperscript{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, 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 indepth 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