Reliability of sonoelastography measurements of lower limb tendon properties

Schneebeli, Alessandro; Folli, Anna; Falla, Deborah; Barbero, Marco

DOI:
10.1016/j.ultrasmedbio.2020.12.018

Reliability of sonoelastography measurements of lower limb tendon properties: a systematic review.

Schneebeli Alessandro[a,b]; Folli Anna[b]; Falla Deborah[a]; Barbero Marco[b].

[a] Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), School of Sport, Exercise Rehabilitation Sciences, College of Life Environmental Sciences, University of Birmingham, Birmingham, UK.
[b] Rehabilitation Research Laboratory 2rLab, Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, Manno/Landquart, Switzerland.

Corresponding author: Alessandro Schneebeli; Tel.: +41586666443; email: alessandro.schneebeli@supsi.ch; Rehabilitation Research Laboratory, Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, SUPSI, Stabile Piazzetta, Via Violino, 6928 Manno, Switzerland.

ABSTRACT:

Reliability of sonoelastography techniques in quantifying lower limb elasticity was investigated. A literature search was conducted using PubMed, Web of Science and CINHAL. Quality of the selected papers were evaluated using the “Guidelines for the Reporting of Reliability and Agreement Studies” and the “Quality Appraisal tool for studies of Diagnostic Reliability” checklist. Reliability values were extracted and synthesized. Twenty-four studies were included and divided into the two main technologies; strain and shear wave elastography. The overall methodological quality was questionable; all studies were at risk of bias. Highly variable results ranging from poor to excellent reliability were found for both technologies and for all tendons considered.

Intra-rater reliability of strain elastography on the Achilles tendon and shear wave elastography on the patellar and quadriceps tendon was adequate. Inter-rater, inter-session and inter-machine reliability was insufficient. Caution should be used when interpreting results from sonoelastography studies measuring lower limb tendon elasticity

Keywords: Achilles tendon, patellar tendon, quadriceps tendon, strain sonoelastography, shear wave sonoelastography, elastography.

INTRODUCTION

Understanding how tendon mechanical properties, such as elasticity, change as a response to load is crucial to better comprehend the behavior of tendons after exercise or other rehabilitation interventions. Tendons are highly responsive to mechanical stimuli (Bohm, et al. 2015, Heinemeier and Kjaer 2011, Magnusson, et al. 2008, Wiesinger, et al. 2015) therefore, the application or elimination of either tensile strain, compression or shear stress could lead to a change of the tendon's mechanical properties (Docking and Cook 2019). Although the relationship between mechanical changes and pathology or performance is still debated, the importance of measuring these parameters is not (Docking and Cook 2019). Clinicians would benefit from technologies that measure mechanical changes during or after their treatments, this would lead to (case-by-case optimization) tailoring of the interventions according to objective measurements.

Nowadays, many different measurement techniques are used to measure tendon mechanical properties with sonoelastography being the most common technique since it can be used to directly assess free tendon (Coombes, et al. 2018, Prado-Costa, et al. 2018). Based on the calculation of tissue internal displacement as a response to the application of a stress force, sonoelastography can describe the elastic properties of different materials (Ophir, et al. 1991). Two main technologies have exploited this physical phenomenon to estimate mechanical parameters using different excitation application (stress). Shear-wave elastography (SWE) uses a focused radiation force, which induces a shear wave within the tissues (Bercoff, et al. 2004, Sandrin, et al. 2002). Conversely, strain elastography (SE) or compression elastography uses a manual generated force to induce an internal displacement of the underneath tissues (Drakonaki, et al. 2012). Both SWE and SE return values describing the elastic properties of the material, in the first case a Young's elastic modulus (E) or shear elastic modulus (G), while in the latter case a color elastogram, can be analyzed either qualitatively or semi-quantitatively using strain ratio between two regions of interest (ROI).

The elastogram automatically constructed by the ultrasound device displays different colors, from red (soft) to blue (hard), indicating different elasticity values. Strain ratios are calculated as the mean between the range of color in the target structure divided by a reference structure.

A recent review (Prado-Costa, et al. 2018), analyzing the feasibility and applicability of SWE and SE to assess tendon pathology and injury, showed potential application in the evaluation of tendinopathies. However, the authors suggest that a more structured and systematic approach is needed to overcome the limitations of both technologies. Even if these technologies have been largely used to describe elasticity in different tendons and in different conditions, methodological studies aiming to evaluate the reliability of these techniques are scarcely present in literature and results are often conflicting.

Reliability, as well as other measurement properties, must be determined before any measurement device or assessment tool can be applied in a clinical or research setting (Koo and Li 2016). When no accepted reference standard is available, as for in-vivo tendon elasticity measures, and validity or accuracy measures are not possible, reliability studies of high methodological quality become crucial to guarantee the applicability of these measures. Criteria for conducting and reporting robust reliability studies are present within the literature (Kottner, et al. 2011, Lucas, et al. 2010, Mokkink, et al. 2010) and guidelines for the interpretation of results are also available (Koo and Li 2016, Shrout and Fleiss 1979). Nevertheless, to date no attempt to synthesize current knowledge on reliability of sonoelastography measurement has been done. Therefore, the aim of this systematic review was to determine the reliability of sonoelastography techniques for measuring lower limb tendon elasticity and to investigate the quality of these studies.

METHODS

Registration

The present systematic review was performed according to the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement (Moher, et al. 2009) and was registered in the International Prospective Register of Systematic Reviews (PROSPERO nr. CRD42020148572) on the 3th of February 2020.

Eligibility criteria

The PICOS framework (Population, intervention, comparator, outcome and study design) was used and adapted for this type of study (Akers 2009). Articles were included if: the full text article was published in peer-reviewed journals, one of the aims of the study was to investigate the reliability of sonoelastography on lower limb tendon and the population were healthy individuals or patients with tendon disease (i.e. patients with lower limb tendinopathies or tendon rupture). Studies using animal models were excluded. We also excluded studies using ex-vivo or human models, as this review is focused on the clinical application of the technology. We included observational studies, such as cross-sectional studies and repeated measurement studies. Studies were included if an ethics committee or institutional review board approved the protocol. Observational studies where reliability was not the primary aim but investigated in a preliminary or pilot study and not described in the methods were excluded. Randomized controlled trial were excluded. A PRISMA flow chart illustrating the study screening and selection process is presented in Figure 1.

Search strategy

A comprehensive search of the databases MEDLINE (PubMed), Web of Science and CINHAL was conducted by the first reviewer (AS) the 26th of February. The full electronic search strategy used for PubMed is presented as an example: (“Achilles tendon” OR (“Achilles” AND “tendon”) OR (“calcaneal” AND “tendon”) OR (“triceps surae” AND “tendon”) OR (“gastrocnemius” AND “tendon”) OR (“soleus” AND “tendon”) OR (“plantarflexor” AND “tendon”) OR (“calf” AND “tendon”) OR (“patellar” AND “tendon”) OR (“patellar” AND “ligament”) OR (“hamstring” AND “tendon”) OR (“lower extremities” AND “tendon”) OR (“lower limb” AND “tendon”) OR “tendons”) AND (“Elastic modulus” OR “Elasticity” OR “Elastic” OR “viscoelastic” OR (“visco” AND “elastic”) OR (“young” AND “modulus” OR “moduli”) OR “compliance” OR “lengthening” OR “stiffness” OR “properties” OR “strain” OR “mechanical”) AND (“reliability” OR “reliabilities” OR “reproducibility” OR “test-retest”)

Additionally, the search strategy was adapted for Web of Science and CINHAL databases. To avoid the exclusion of relevant papers, search filters (i.e. animal studies, type of study, in vitro) were not included (Doust, et al. 2005). No publication date or language restrictions were applied. Moreover, reference lists of included studies were checked. Studies obtained using the search strategies were exported to EndNote X9 (Clarivate Analytics, Philadelphia, PA, USA) by the first reviewer (AS) and duplicates removed.

Study selection

Two reviewers (AS and AF) independently screened titles and abstracts to determine which studies met the inclusion and exclusion criteria. Studies whose eligibility could not be established from the screening of title or abstract were searched for full-text review. Any disagreement between reviewers were resolved by consulting a third reviewer (MB).

Quality assessment

The included studies were assessed using two different quality assessment tools for reliability studies. To assess the quality of the reporting of the included studies, the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (Kottner, et al. 2011) checklist was used. To guarantee the correct interpretation and synthesis of the results of other studies, the importance of an adequate and clear reporting is mandatory.
The items covered by GRRAS evaluate if the information about the sample selection, study design, rater characteristics and statistical analysis are properly reported. This 15-items checklist was used and a quality score was given depending on the number of items reported in the correct section of the article. GRRAS items are described in detail in Figure 2-3.

The quality in the design and conduct (internal quality) of the studies was assessed using the quality appraisal tool for studies of diagnostic reliability (QAREL). This checklist was developed by Lucas et al. (Lucas, et al. 2010) with the aim to critically appraise reliability studies. Previous checklists investigating the diagnostic accuracy i.e. the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) (Whiting, et al. 2006) have been developed, nevertheless, when there is no accepted reference standard to relate to, as in the case of elasticity measures, reliability studies become crucial. The QAREL is an 11-items checklist covering the spectrum of subjects, examiners, appropriate test application and statistical analysis specific for reliability studies. QAREL items can be answered as “no”, “unclear” and “yes”; additionally, the option “not applicable” can be used in some items. The more “yes” responses, the higher the quality of the study, conversely a single “no” answer may invalidate the study findings. Five out of 11 items can be scored as “not applicable” therefore the final score of the article will take into account the real total value (11 items – N/A). A detailed explanation of all the items is presented in Figure 4.

Disagreements in the rating of the QAREL and the GRAAS checklists among the two reviewers (AS and AF) were discussed and resolved consulting the third reviewer (MB).

Data extraction

Information were independently extracted by a reviewer (AS) and controlled by a second reviewer (AF). Study characteristics, including number of participants, tendons analyzed, ultrasound device and technique used, reliability type, operators experience and measurement protocol were extracted and collected in a descriptive table. Study results are reported in separate tables divided by the type of sonoelastography technique. These tables summarize the condition evaluated, and the relative and absolute reliability results. Relative reliability calculated with intraclass correlation coefficient (ICC) were interpreted as follows:

- ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 show good reliability and values greater than 0.9 indicate excellent reliability (Portney and Watkins 2015).
- As suggested by Koo et al.,(Koo and Li 2016) where 95% confidence interval (CI) of ICC were reported, the rating considers not only the point estimation but the lower and upper limit of the 95% CI.
- a cut-off point of ICC >0.6 was used to consider the clinical relevance of the measurements (Chinn 1990, Munro 2005).

Where present, agreement or correlation where interpreted as follows:

- Cohen's Kappa value for categorical data were interpreted following the Landis and Koch scale (Landis and Koch 1977) where <0 indicate poor agreement, 0.00-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement and 0.81-1.00 almost perfect agreement.
- Correlation between measures using Spearman's rank was considered weak when $r_s<0.3$, moderate when r_s 0.31-0.70 and strong when $r_s>0.71$ (Dancey and Reidy 2002).

Absolute reliability by mean of standard error of measurement (SEM), coefficient of variation (CV) and Bland and Altman plots were reported in the Tables (2-5) when presented in the studies.

RESULTS

Literature search and selection of studies

The initial electronic search through MEDLINE, Web of Science and CINHAL identified 828 potentially eligible articles. After the removal of duplicates, 626 articles were screened by title and abstracts with an agreement between the two reviewers of K=0.73. Eighty-nine full text articles were assessed for eligibility. Twenty-four studies met all the inclusion criteria and were included in the final qualitative synthesis (Figure 1). The agreement on the full text screening between the two reviewers was K=0.97. The majority of the studies excluded after full text screening were for the following reasons: not using sonoelastography, not assessing reliability and not examining a tendon of the lower limb.

Characteristics of included studies

Characteristics of included studies are presented in Table 1. The number of participants included in the included studies ranged from a minimum of 5 to a maximum of 50. In the majority of the studies participants were healthy (n=23). Only one study evaluated reliability in patients with Achilles tendinopathy (Ooi, et al. 2015). In the included studies, the following tendons were evaluated: Achilles tendon (AT) (n=15); patellar tendon (PT) (n=5); patellar and quadriceps tendon (QT) (n=3); Achilles, patellar and quadriceps tendon (n=1).

Two major sonoelastography techniques were evaluated; SE (n=11) and SWE (n=13). Nine out of 11 studies using SE analyzed the values by mean of the strain ratio. Strain ratio is defined as the elasticity ratio between the tendon and another material, i.e. fat tissue (n=5), subcutaneous tissue (n=2) and an external reference material (n=2). Of those, 3 studies added the evaluation of the color scale by mean of the color grading.

One study used only color grading (Capalbo, et al. 2016) and another study used the percentage of displacement between the compressions (Payne, et al. 2017).
Nine out of 13 studies using SWE calculated the elasticity values by the mean of shear modulus measured in kPa or mPa whereas 2 studies used velocity of the shear wave propagation (m.s-1). Two studies (Aubry, et al. 2013, Hsiao, et al. 2015) used both parameters; shear modulus and shear wave speed.

Type of reliability and raters

In 19 out of 24 studies, the primary study aim was to test the reliability, whereas in 5 out of 24, reliability was a secondary aim. In these latter studies, the primary aim was quantification of tendon elasticity (n=4) and the effect of exercise on tendon elasticity (n=1).

Intra-rater reliability was assessed in 21 studies, inter-rater reliability in 14, intra-rater inter-session reliability in 5 and inter-machine reliability in 2 studies. In the majority of the studies (n=22), ICC was used to analyze the relative reliability, while in 2 studies (Capalbo, et al. 2016, Petitpierre, et al. 2018), reliability was assessed using Cohen k or p-values. Of the 22 studies using ICC, only 12 studies reported the 95% confidence interval (CI), despite the recommendation to do so in reporting guidelines (Koo and Li 2016). Other studies used p-values (n=4), only the point value of ICC (n=6) or the ICC range (n=1). After formal request of additional data, authors of 5 studies (Dickson, et al. 2020, Dickson, et al. 2019, Helfenstein-Didier, et al. 2016, Payne, et al. 2018, Payne, et al. 2017) provided 95% CI for ICC values.

Absolute reliability was assessed in 13 out of 24 studies, with SEM (n=4), CV (n=9), mean difference (MD) with upper and lower limit of agreement (LOA) (n=2) and smallest detectable changes (SDC) (n=1).

In the majority of studies, the rater was described as a general operator and their profession was not reported (n=11), whereas 7 studies declared the raters to be a radiologist or sonographer, 4 studies involved medical doctors as raters, 2 studies a physical therapist and 1 study a sport scientist. The years of ultrasound experience of the raters ranged from 0 years (novice) to more than 20 years (senior). Ten out of 24 studies did not report the years of experience of the raters.

Methodological and reporting quality

Reporting quality (GRRAS)

The mean score for all the selected studies was 7.9 out of 15 items with a lowest score of 5/15 (Aubry, et al. 2013, Mannarino, et al. 2018, Petitpierre, et al. 2018) and a highest score of 12/15 (Alsiri, et al. 2020). None of the included studies provided auxiliary material to detail the results of the study therefore item 15 was not reported in any article. Item 12 regarding the description in the results section of the rater and subject characteristics, was not reported in any of the included studies. However, in 15 of the selected studies, this information was reported in the methods section.

There were a high number of studies missing on one or more of the following items: Item 4: only 1 study (Dickson, et al. 2020) specified the rater population of interest in the introduction, Item 6: only 3 studies (Alsiri, et al. 2020, Ünal, et al. 2019, Yamamoto, et al. 2016) explained in the methods how sample size was chosen, Item 7: 8 out of 24 studies described the sampling methods, Item 11: 5 out of 24 studies stated, in the results section, the actual number of raters and subjects included.

Detailed scores for each item are reported in Figure 2. Conversely, a large amount of the studies included (>21) reported the items regarding the description of the study design (item 1) and of the technology included (item 2), specified the subject population of interest (item 3), described the statistical analysis (item 10) and discussed the practical relevance of the results (item 14). All the items of the GRRAS scale divided for each study are reported in the Figure 3. The mean agreement between the two reviewers scoring on the GRRAS scale was 95.8%. Disagreement on one item of three different studies were resolved by the third reviewer.

Methodological quality (QAREL)

QAREL checklist was used to appraise the methodological quality of the included studies. Item 5 regarding the blinding of the disease status was rated N/A in all studies whereas item 7 regarding the blinding of additional cues was rated as unclear in all studies. Once the N/A score was removed, the maximum score possible ranged from 7 to 10. The maximum of “yes” answers were obtained by Siu et al. (Siu, et al. 2016) with a score of 7 out of 9. A minimum of 1 and a maximum of 6 “unclear” answers were reported in all the 24 studies selected, indicating uncertainty in the methodological quality exposing all the studies to a possible risk of bias. Detailed scores for each item are reported in Figure 4.

The study that presented with the highest “yes” answers net of the total score is the study of Alsiri et al. (Alsiri, et al. 2020) with only one “unclear” answer. One study (Petitpierre, et al. 2018) obtained a “no” answer for the item 11 regarding the appropriate statistical analysis. As defined by Lucas et al. (Lucas, et al. 2010) the items rated as “no” can invalidate the study and compromise the findings. Therefore, the aforementioned study was not included in the qualitative syntheses of the present review.

QAREL items divided for each study are reported in Figure 5. The mean agreement between the two reviewers scoring on the QAREL checklist was 96.2%. Disagreement on one item of two different studies were resolved by the third reviewer.

Reliability

The heterogeneity of the analyzed studies precluded a quantitative analysis, therefore a qualitative analysis was performed dividing the studies by the type of technology used (SE or SWE) and the tendon analyzed (Achilles, patellar and quadriceps). Relative and absolute reliability for the different tendons and technologies are reported in Tables 2-5. A visual representation of relative reliability with ICC values is presented in Figures 6-9.

The qualitative synthesis of the studies was performed considering the following criteria:

- The ICC punctual value has to be ≥ 0.6
- The lower limit of the 95% CI has to exceed 0.5

If one of the following criteria was not fulfilled, this study was considered to have inconclusive reliability. Moreover, where 95% CI were not reported, the study was cautiously evaluated.

SE - Achilles tendon

Eight studies evaluated the reliability of AT elasticity measures using SE. Moderate to excellent intra-rater reliability was found for the longitudinal measurement of strain ratio; results ranged from an ICC of 0.66 to 0.99. Inter-rater reliability was moderate to good (ICC= 0.51 to 0.84) with some studies showing values beneath the 0.6 threshold.

A moderate inter-rater correlation was also found in one study (r_s 0.61) (Yamamoto, et al. 2016). One study (Ünal, et al. 2019) analyzing the inter-machine reliability showed an ICC of 0.85 (0.26-0.95), indicating good reliability but with very large 95%CI. Agreement on the color grading of SE images were considered substantial (k=0.70) for intra-rater evaluation and from substantial to almost perfect (k= 0.70-0.91) for inter-rater evaluation.

Poor reliability was found for intra-rater (ICC=0.06-0.11) and for inter-session (ICC=0.01-0.10) when using the % of displacement values (Payne, et al. 2017). Axial scan using SE was evaluated in two studies (Drakonaki, et al. 2009, Payne, et al. 2017) showing poor intra and inter-rater reliability (ICC<0.45).

Overall, intra and inter-rater reliability for AT elasticity measured with SE strain ratios in a longitudinal scan was adequate. Inter-machine reliability tested in one study reported inconclusive results. Color grading intra and inter-rater agreement was substantial.

SE- patellar and quadriceps tendons

Two studies (Dickson, et al. 2020, Dickson, et al. 2019) shared part of the dataset, therefore intra-rater reliability values obtained with the GE ultrasound device were reported only for the study by Dickson et al. (2019) eliminating the other study (Dickson, et al. 2020).

PT strain ratio was evaluated in three studies showing intra-rater reliability values ranging from poor to good (ICC= 0.40-0.85) for the proximal, (ICC= 0.13-0.67), middle (ICC= 0.15-0.91) and distal part of the tendon. Inter-rater reliability ranged from poor to moderate (ICC= 0.11-0.64) for the proximal (ICC= 0.08-0.60), middle (ICC= 0.22-0.96) and distal part of the tendon.

Inter-rater agreement of strain ratio using a t-test (p values) (Dickson, et al. 2020) revealed mixed results with good values only documented for the proximal and distal part of the PT (Table 3). Inter-machine association using strain ratios was evaluated in one study (Dickson, et al. 2020) showing significant results indicating no association between the two ultrasound devices for the PT.

Mean % of agreement on the color grading for PT was >80% for intra-rater, between 68%-95% for inter-rater and between 64%-86% for inter-machine (Dickson, et al. 2020, Dickson, et al. 2019).

QT strain ratio was evaluated in two studies revealing poor intra-rater (ICC<0.5) and inter-rater (ICC<0.5) reliability and no inter-machine association (p>0.05). Mean % of agreement on the color grading for QT was between 63%-85% for intra-rater, between

48%-68% for inter-rater and between 58%-64% for inter-machine (Dickson, et al. 2020, Dickson, et al. 2019).

Overall, reliability of SE measures using strain ratios on the PT was adequate only for intra-rater reliability when measured in the proximal part of the PT, and if analyzed by expert operators. QT measures were not reliable. Intra-rater agreement in the color grading was good for the PT and QT. Inter-rater agreement was adequate in the PT and lower on the QT

SWE - Achilles tendon

Eight studies evaluated the reliability of using SWE to measure AT elasticity in a longitudinal scan. Results indicated poor to excellent intra-rater reliability (ICC= 0.17-1.00) for the elasticity or shear modulus (kPa), moderate intra-rater reliability (ICC=0.54-0.71) for velocity (m.s-1) and poor to good intra-rater reliability (ICC=0.42-0.85) for compression modulus (MPa). In two studies (Aubry, et al. 2013, Helfenstein-Didier, et al. 2016), reliability results were also divided according to angle of the ankle when the measurements were performed (Table 4). Inter-rater reliability was poor to moderate (ICC=-0.02-0.59) for shear modulus (kPa) and poor (ICC<0.5) for velocity (m.s-1). Inter-session reliability was poor to moderate (ICC= 0.42-0.60) for shear modulus (kPa) and moderate (ICC=0.55-0.67) for velocity (m.s-1).

One study (Payne, et al. 2018) showed moderate to good intra-rater and inter-session reliability for velocity axial values (ICC=0.62-0.85) whereas another study (Aubry, et al. 2013) showed poor inter-rater axial values (ICC<0.5) for shear modulus and for velocity.

Overall, inter-rater and inter-session reliability of SWE measures for AT was inconclusive.

Intra-rater reliability was conflicting with some studies indicating good and other poor reliability.

SWE - patellar and quadriceps tendon

Six studies evaluated intra-rater reliability of PT measurements with poor to excellent results (ICC= 0.45-1.00) for shear modulus (kPa) and excellent intra-rater reliability (ICC=0.91-0.92) for velocity (m.s-1). Inter-session reliability was moderate (ICC=0.63-0.65) for shear modulus (kPa) and good (ICC=0.81-0.83) for velocity (m.s-1). Inter-rater reliability ranged from moderate to excellent (ICC=0.71-0.97) for both shear modulus and velocity.

Two studies evaluated the QT intra-rater reliability, one with moderate results (ICC=0.72-0.77) (Peltz, et al. 2013) and the other with excellent results (ICC= 0.99-1.00) (Zardi, et al. 2019).

Overall, intra-rater reliability of SWE was satisfactory for the PT and QT. Inter-rater and inter-session reliability results were conflicting.

DISCUSSION

The present systematic review aimed to synthesize the evidence regarding the reliability of sonoelastography for the assessment of lower limb tendon mechanical properties.

Reliability as well as validity is a prerequisite to the clinical applicability of a measure. These measurement properties should be confirmed before this technology is

used in clinical settings. The results of this systematic review suggest that reliability is highly variable across the studies and therefore the use of this technology in a clinical setting should be further evaluated in high-quality reliability studies.
Furthermore, other clinimetric properties (i.e. validity and responsiveness) should be further investigated.

Twenty-four articles were included and reviewed in subgroups of studies divided by technology used and tendon analyzed. High heterogeneity was found between studies even when using the same technology on the same tendon. Different joint angles and position, different probe orientation (longitudinal or axial scan), different parameters used to calculate elasticity and different ultrasound devices used, challenged the possibility to compare results between studies. Nevertheless, a qualitative synthesis was conducted and overall, there was a large variability of the results regardless of the technology used or the tendon analyzed with variable reliability findings; these conflicting results could be attributed to the low methodological quality of the included studies.

As expected, intra-rater reliability results were superior to inter-rater or inter-session values, confirming that this type of ultrasound-based technology, whether with qualitative or quantitative elastogram analysis, are particularly dependent on the operator's manual skills and experience (Prado-Costa, et al. 2018). Although, we cannot exclude that the lower reliability found for inter-rater and inter-session is related to the paucity of studies analyzing this type of reliability. Different results were found between the lower limb tendons, suggesting that the reliability of sonoelastography could be influenced by the structural and morphological characteristics of the tissues evaluated.

Very wide 95% confidence interval of the ICC values were found across the majority of the studies questioning the sample size used. The use of inadequate sample size is confirmed by the low score on the GRRAS scale (item 6).

The quality of the reporting in the selected papers was poor with only two studies with more than ten out of fifteen points on the GRRAS scale (Alsiri, et al. 2020, Dickson, et al. 2020). The majority of the studies lacked a description of the raters and subjects of interest, in the definition of the sampling methods and the calculation of the sample size. Although the reviewed papers were designed specifically to evaluate reliability, many were not reported adequately. Moreover, the comparison between studies and the replication of an individual study is compromised by the lack of detailed reporting of the procedures. Furthermore, even the methodological quality of the studies included, screened with the QAREL checklist, showed a possible risk of bias due to shortcomings in the study design. Considering the real total score without “N/A” answers, the majority of the included studies (17 out of 24) scored at maximum 2 or 3 point less than the possible highest value. Nevertheless, in almost all of the included studies, one or more items were scored as “unclear” indicating possible methodological bias (Figure 5). The lack of clarity in the experimental design mainly concerned the blinding of the raters to the finding of other raters (item 3), the blinding to their own prior findings (item 4), the blinding to additional cues that were not part of the test (items 7) and the varied order of examination (item 8). The “unclear” score did not invalidate the study but made it vulnerable to possible bias. These limitations may be attributed to insufficient reporting rather than real methodological bias. Given the low quality in reporting and the possible methodological bias present in all studies, caution should be used in the interpretation of the results.

Future studies evaluating reliability should adhere to validated quality assessment tools to improve methodological quality and reporting of the results.

SE - Achilles tendon

The majority of the included studies showed that SE has acceptable (ICC >0.6) intra-rater reliability except for one study (Payne, et al. 2017) showing very low reliability values. High intra-rater reliability values were found when the AT was analyzed using semi-quantitative analysis using strain ratio with an external reference material (Schneebeli, et al. 2016, Yamamoto, et al. 2016). Variable results (moderate to excellent) were obtained when using strain ratio for the Kager fat pad (Drakonaki, et al. 2009, Ooi, et al. 2015, Ünal, et al. 2019) or subcutaneous fat (Alsiri, et al. 2020) and low results when using the % of displacement calculation (Payne, et al. 2017).

Inter-rater results were variable (moderate to high) between studies with one study presenting results lower than the threshold of 0.6 (Drakonaki, et al. 2009) and two studies with moderate to good reliability (Ooi, et al. 2015, Ünal, et al. 2019).

Overall, both intra and inter-rater SE measurements of the AT appear reliable when using semi-quantitative strain ratio analysis. Conversely, calculating the % of displacement in SE seems unreliable (Payne, et al. 2017). When analyzing the tendon qualitatively with color grading, the agreement is substantial to almost perfect.

In addition to strain ratio measurements, Alsiri et al. (Alsiri, et al. 2020) presented another method to quantify SE images using the color pixel analysis. This method, is based on the quantification of the percentages of each color in the SE image. Intra-rater reliability of this method on the AT appeared higher than the strain ratio measure. However, caution should be used in the analysis of elastograms without an external reference material not affected by changes in elasticity due to pathology, recovery or different measurement conditions. A simple count of the color pixel within the same structure disregards the algorithm implemented in SE ultrasound devices that uses an “auto-scale” color range (Schneebeli, et al. 2019).

Ünal et al.(Ünal, et al. 2019) evaluated the inter-machine reliability showing variable results ranging from ICC of 0.26 to 0.95. Given this result, the author suggests to evaluate the tendon separately for each device. The latter study emphasizes one of the major limitations of SE. The semi-quantitative analysis obtained by a color scale may differ depending on the algorithm implemented by different companies and from the amount of compression with the probe permitted by different devices.

The differences between devices in the generation of the color scale may hinder the clinical applicability of SE. Caution should be taken when comparing results obtained from different devices on the same structure.

SE - patellar and quadriceps tendon

In contrast to the AT, SE using strain ratios on the PT and QT have questionable intra and inter-rater reliability results. 95%CI of the reliability measurement often fell below the 0.6 threshold indicating poor to moderate reliability and weak clinical relevance. Porta et al. (Porta, et al. 2014) showed moderate to good intra-rater reliability with almost all the CI values below 0.6 and poor inter-rater reliability. Two studies (Dickson, et al. 2020, Dickson, et al. 2019) reported mixed results with measurements on some part of the PT appearing to be reliable when measured by an expert rater. Overall the results of those studies indicate poor reliability except for the proximal PT when analyzed by expert operators.

When using color grading, agreement between operators appears to be better compared to using semi-quantitative analysis (Dickson, et al. 2020, Dickson, et al. 2019). Operator experience seems to play an important role to be able to measure the PT and QT in a reliable way. The different reliability values obtained with SE on the AT, QT and PT could be attributed to differences in anatomy between tendons, where the first two are tendons with one bone insertion and muscle while the latter is attached to a bony structure on both sides. Compression with the ultrasound probe on a “free” tendon rather than on a more solid structure with the proximity of prominent bone projection as for PT, could affect reliability results. Furthermore the different structures used for the strain ratio, subcutaneous tissue (Porta, et al. 2014) and Hoffa fat pad (Dickson, et al. 2020, Dickson, et al. 2019) could have led to a difference in reliability when compared to AT measurements using an external reference material (Schneebeli, et al. 2016, Yamamoto, et al. 2016) or Kager fat pad (Alsiri, et al. 2020, Drakonaki, et al. 2009, Ooi, et al. 2015, Ünal, et al. 2019).

To date, only three studies with a small sample size are present in the literature. Given the high variability of these results, further studies with high methodological quality are needed to establish the reliability of SE when measuring these tendons.

SWE - Achilles tendon

SWE measurements of the AT showed conflicting results with two studies (Aubry, et al. 2013, Peltz, et al. 2013) showing poor to moderate intra and inter-rater reliability for longitudinal and axial scans. One of these studies (Peltz, et al. 2013) indicated a reliability slightly over the 0.6 threshold but only for the left tendon.

Two other studies (Corrigan, et al. 2019, Helfenstein-Didier, et al. 2016) revealed moderate to excellent intra-rater reliability for compression and shear modulus (Helfenstein-Didier, et al. 2016) except for the compression modulus at 45° and 35°plantarflexion and moderate to excellent intra-rater reliability for dynamic modulus and viscosity (Corrigan, et al. 2019) with inconclusive results for the static modulus.

Furthermore, two studies (Lima, et al. 2017, Siu, et al. 2016) showed good to excellent intra-rater reliability but poor or moderate inter-rater (Siu, et al. 2016) and inter-session reliability (Lima, et al. 2017). Finally, Payne et al. (Payne, et al. 2018) showed acceptable intra-rater reliability in the longitudinal scan when the ankle was fixed at 0° plantarflexion, but doubtful reliability for the relaxed position. Surprisingly, this study showed higher values (moderate to high) when the tendon was analyzed in the axial scan. Indeed, waves propagate along the fibers, as for longitudinal scan, more easily than across the fiber, as for axial scan. Moreover, during an axial scan, inhomogeneous compression with the edges of the probe could occur which may affect the reliability. Considering these results, further studies are needed to better understand the value of axial scans in the analysis of tendon elasticity.

Although the majority of the SWE studies used the Aixplorer ultrasound equipment, two studies (Corrigan, et al. 2019, Payne, et al. 2018) analyzed AT with a different SWE device but obtained similar results to the other studies. Overall the results indicated relatively good intra-rater reliability but poor inter-rater and inter-session reliability.

The main reason for the lack of inter-rater reliability are, according to the authors, due to the difficulties in the stabilization of the ankle joint position (Aubry, et al. 2013, Peltz, et al. 2013), measurement bias related to positioning, the direction of the transducer or to the positioning of the ROI (Aubry, et al. 2013). Moreover, anisotropic changes of the tendon due to a slightly different position of the ankle can occur. Furthermore, some authors (Aubry, et al. 2013, Chino and Takahashi 2015, DeWall, et al. 2014) suggested that the higher shear modulus achievable with this technology (600-800kPa) was not sufficient to calculate tendon elasticity in a position near to 0° plantarflexion or in dorsiflexion because of saturation of the signal.

SWE - patellar and quadriceps

The six studies evaluating the PT and QT with SWE showed high intra and inter-rater reliability with all the point values of ICC over 0.6, except for the PT in the study of

Peltz et al. (Peltz, et al. 2013). Inter-session reliability was also considered moderate to high in two studies (Mannarino, et al. 2018, Tas, et al. 2017).

Two studies (Zardi, et al. 2019, Zhang and Fu 2013) reported excellent intra and inter-rater reliability values for PT and QT. Nevertheless, considering the 95%CI of all the ICC results (see Figure 9), the majority of the studies included, especially for inter-rater and inter-session reliability, presented CI values substantially lower than 0.5, and therefore should be considered as poor reliability. These values are often neglected by authors and readers, but are critical when analyzing and interpreting ICC values. Indeed, the CI gives the range within which the true ICC value is supposed to be, and this could be at any point between the lower and the higher limits. Therefore, the true ICC value could largely be lower than the estimate value (Koo and Li 2016, Shrout and Fleiss 1979). Considering this large confidence interval, Hsiao et al. (Hsiao, et al. 2015) concluded that the only suitable part of the PT to make SWE evaluation is the mid tendon and that inter-rater reliability should be cautiously interpreted. Overall, intra-rater reliability was adequate for PT and QT but dubious for inter-rater and inter-session. According to the authors, the reason for this variability between studies could be attributed to the difference in the ROI placement and size. Furthermore, some authors (Chen, et al. 2013, Drakonaki, et al. 2012, Hsiao, et al. 2015) suggested that another possible cause of heterogeneity in the results are the saturation of the signal with high shear modulus (kPa) values and the excessive noise signal when measuring soft tissue near bone. Finally, the small sample size present in some of these studies may have led to an increased 95% CI.

Limitations:

To improve the quality of the included articles, the search was restricted to full peer-reviewed published papers; relevant grey literature may have been excluded. The high heterogeneity of the included studies, which used varying joint angles, tendon portion, conditions, ultrasound scan, extracted parameters and statistical methods, made generalization of the results difficult. Graphical synthesis of the results was done in a sub-portion of the data of the included studies. Given the high heterogeneity of the data reported especially regarding the parameter used to calculate elasticity a quantitative synthesis with meta-analysis was not possible. This systematic review included only reliability measures on lower limb tendons and therefore the results may not be generalized to other tendons. Other measurement properties apart from reliability (i.e. validity, diagnostic accuracy, sensitivity, specificity) were not evaluated in this systematic review; different sonoelastography measurement properties need to be reviewed in the future. Moreover, the majority of the studies assessed tendon in healthy participants, and therefore it is not possible to generalize these results to a clinical context in the assessment of pathological tendons.

CONCLUSIONS

Although sonoelastography measures of lower limb tendons are commonly used in research and clinical settings, overall the reliability of these measures remains highly variable across studies. SE on AT is reliable when using strain ratio and good agreement was found in the color grading analysis, this was not the case for the PT.

The reliability of SWE measurements were highly variable between studies, nevertheless it seems to be more reliable on PT than on AT, especially if different operators are compared. However, weakness in the reporting and possible methodological bias were present in all studies, making it difficult to recommend one technology over another. The interpretation of reliability results was often overestimated given to a lack of consideration of the confidence interval. The lack of inter-rater, inter-session and inter-machine reliability found in this systematic review raises important doubts about the clinical applicability of these technologies. Caution should be used when interpreting results from sonoelastography measures.

Further reliability studies of high methodological quality are needed to better understand the value of these technologies in the assessment of lower limb tendon mechanical properties to be able to transfer the use of these measures in a clinical setting providing meaningful information about tendon status.

AKNOWLEDGEMENTS

We thank Matteo Beretta-Piccoli for the methodological support.

COMPETING INTERESTS

None declared.

FUNDING

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

References:

Akers J. Systematic reviews : CRD's guidance for undertaking reviews in health care. CRD Centre for Reviews and Dissemination, University of York, York, 2009.

Alsiri N, Al-Obaidi S, Asbeutah A, Palmer S. Intra-rater reliability and smallest detectable change of compression sonoelastography in quantifying the material properties of the musculoskeletal system. J Anat 2020.

Aubry S, Risson JR, Kastler A, Barbier-Brion B, Siliman G, Runge M, Kastler B. Biomechanical properties of the calcaneal tendon in vivo assessed by transient shear wave elastography. Skeletal Radiol 2013; 42:1143-50.

Bercoff J, Tanter M, Fink M. Supersonic shear imaging: a new technique for soft tissue elasticity mapping. IEEE Trans Ultrason Ferroelectr Freq Control 2004; 51:396-409.

Bohm S, Mersmann F, Arampatzis A. Human tendon adaptation in response to mechanical loading: a systematic review and meta-analysis of exercise intervention studies on healthy adults. Sports Med Open 2015; 1:7.

Capalbo E, Peli M, Stradiotti P. Sonoelastography of the distal third of the Achilles tendon in asymptomatic volunteers: correlation with anthropometric data, ultrasound findings and reproducibility of the method. Radiol Med 2016; 121:667-74.

Chen XM, Cui LG, He P, Shen WW, Qian YJ, Wang JR. Shear wave elastographic characterization of normal and torn achilles tendons: a pilot study. J Ultrasound Med 2013; 32:449-55.

Chinn S. The assessment of methods of measurement. Stat Med 1990; 9:351-62.

Chino K, Takahashi H. The association of muscle and tendon elasticity with passive joint stiffness: In vivo measurements using ultrasound shear wave elastography. Clin Biomech (Bristol, Avon) 2015; 30:1230-5.

Coombes BK, Tucker K, Vicenzino B, Vuvan V, Mellor R, Heales L, Nordez A, Hug F. Achilles and patellar tendinopathy display opposite changes in elastic properties: A shear wave elastography study. Scand J Med Sci Sports 2018; 28:1201-08.

Corrigan P, Zellers JA, Balascio P, Silbernagel KG, Cortes DH. Quantification of Mechanical Properties in Healthy Achilles Tendon Using Continuous Shear Wave Elastography: A Reliability and Validation Study. Ultrasound Med Biol 2019; 45:1574-85.

Dancey CP, Reidy J. Statistics without maths for psychology : using SPSS for Windows. Harlow ; New York: Prentice Hall, 2002.

DeWall RJ, Slane LC, Lee KS, Thelen DG. Spatial variations in Achilles tendon shear wave speed. J Biomech 2014; 47:2685-92.

Dickson DM, Fawole HO, Hendry GJ, Smith SL. Intermachine Variation of Ultrasound Strain Elastographic Measures of the Quadriceps and Patellar Tendons in Healthy Participants: Implications for Clinical Practice. J Ultrasound Med 2020.

Dickson DM, Fawole HO, Newcombe L, Smith SL, Hendry GJ. Reliability of ultrasound strain elastography in the assessment of the quadriceps and patellar tendon in healthy adults. Ultrasound 2019; 27:252-61.

Docking SI, Cook J. How do tendons adapt? Going beyond tissue responses to understand positive adaptation and pathology development: A narrative review. J Musculoskelet Neuronal Interact 2019; 19:300-10.

Doust JA, Pietrzak E, Sanders S, Glasziou PP. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. J Clin Epidemiol 2005; 58:444-9.

Drakonaki EE, Allen GM, Wilson DJ. Real-time ultrasound elastography of the normal Achilles tendon: reproducibility and pattern description. Clin Radiol 2009; 64:1196-202.

Drakonaki EE, Allen GM, Wilson DJ. Ultrasound elastography for musculoskeletal applications. Br J Radiol 2012; 85:1435-45.

Heinemeier KM, Kjaer M. In vivo investigation of tendon responses to mechanical loading. J Musculoskelet Neuronal Interact 2011; 11:115-23.

Helfenstein-Didier C, Andrade RJ, Brum J, Hug F, Tanter M, Nordez A, Gennisson JL. In vivo quantification of the shear modulus of the human Achilles tendon during passive loading using shear wave dispersion analysis. Phys Med Biol 2016; 61:2485-96.

Hsiao MY, Chen YC, Lin CY, Chen WS, Wang TG. Reduced Patellar Tendon Elasticity with Aging: In Vivo Assessment by Shear Wave Elastography. Ultrasound Med Biol 2015; 41:2899-905.

Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016; 15:155-63.

Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64:96-106.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-74.

Lima K, Martins N, Pereira W, Oliveira L. Triceps surae elasticity modulus measured by shear wave elastography is not correlated to the plantar flexion torque. Muscles Ligaments Tendons J 2017; 7:347-52.

Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol 2010; 63:854-61.

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

Mannarino P, Lima KMM, Fontenelle CRC, Matta TT, de Salles BF, Simao R, Oliveira LF. Analysis of the correlation between knee extension torque and patellar tendon elastic property. Clin Physiol Funct Imaging 2018; 38:378-83.

Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009; 6:e1000097.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010; 19:539-49.

Munro BH. Statistical methods for health care research. Philadelphia, Penn. ; London: Lippincott Williams & Wilkins, 2005.

Ooi CC, Schneider ME, Malliaras P, Chadwick M, Connell DA. Diagnostic performance of axial-strain sonoelastography in confirming clinically diagnosed Achilles tendinopathy: comparison with B-mode ultrasound and color Doppler imaging. Ultrasound Med Biol 2015; 41:15-25.

Ophir J, Cespedes I, Ponnekanti H, Yazdi Y, Li X. Elastography: a quantitative method for imaging the elasticity of biological tissues. Ultrason Imaging 1991; 13:111-34.

Payne C, Watt P, Cercignani M, Webborn N. Reproducibility of shear wave elastography measuresof the Achilles tendon. Skeletal Radiol 2018; 47:779-84.

Payne C, Webborn N, Watt P, Cercignani M. Poor reproducibility of compression elastography in the Achilles tendon: same day and consecutive day measurements. Skeletal Radiol 2017; 46:889-95.

Peltz CD, Haladik JA, Divine G, Siegal D, van Holsbeeck M, Bey MJ. ShearWave elastography: repeatability for measurement of tendon stiffness. Skeletal Radiol 2013; 42:1151-6.

Petitpierre F, Perez J-T, Bise S, Fournier C, Hauger O, Dallaudiere B. Quantitative elastography of Achilles tendon using Shear Wave Elastography (SWE): correlation with zonal anatomy. Muscles, Ligaments & Tendons Journal (MLTJ) 2018; 8:229-37.

Porta F, Damjanov N, Galluccio F, Iagnocco A, Matucci-Cerinic M. Ultrasound elastography is a reproducible and feasible tool for the evaluation of the patellar tendon in healthy subjects. Int J Rheum Dis 2014; 17:762-6.

Portney LG, Watkins MP. Foundations of clinical research : applications to practice. Upper Saddle River, N.J.: Pearson/Prentice Hall, 2015.

Prado-Costa R, Rebelo J, Monteiro-Barroso J, Preto AS. Ultrasound elastography: compression elastography and shear-wave elastography in the assessment of tendon injury. Insights Imaging 2018; 9:791-814.

Sandrin L, Tanter M, Catheline S, Fink M. Shear modulus imaging with 2-D transient elastography. IEEE Trans Ultrason Ferroelectr Freq Control 2002; 49:426-35.

Schneebeli A, Del Grande F, Falla D, Cescon C, Clijsen R, Barbero M. A novel application of strain sonoelastography can detect changes in Achilles tendon elasticity during isometric contractions of increasing intensity. J Foot Ankle Res 2019; 12:30.

Schneebeli A, Del Grande F, Vincenzo G, Cescon C, Clijsen R, Biordi F, Barbero M. Real-time sonoelastography using an external reference material: test-retest reliability of healthy Achilles tendons. Skeletal Radiol 2016; 45:1045-52.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86:420-8.

Siu WL, Chan CH, Lam CH, Lee CM, Ying M. Sonographic evaluation of the effect of long-term exercise on Achilles tendon stiffness using shear wave elastography. J Sci Med Sport 2016; 19:883-87.

Tas S, Onur MR, Yilmaz S, Soylu AR, Korkusuz F. Shear Wave Elastography Is a Reliable and Repeatable Method for Measuring the Elastic Modulus of the Rectus Femoris Muscle and Patellar Tendon. J Ultrasound Med 2017; 36:565-70.

Ünal Ö, GÜZel Ş, ŞİĞİT İKİZ S, ŞİRİN ÖZcan A. Evaluation of Sonoelastography Compatibility on the Achilles Tendon Between Different Devices and Observers. Duzce Medical Journal 2019; 21:156-61.

Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006; 6:9.

Wiesinger HP, Kosters A, Muller E, Seynnes OR. Effects of Increased Loading on In Vivo Tendon Properties: A Systematic Review. Med Sci Sports Exerc 2015; 47:1885-95.

Yamamoto Y, Yamaguchi S, Sasho T, Fukawa T, Akatsu Y, Nagashima K, Takahashi K. Quantitative Ultrasound Elastography With an Acoustic Coupler for Achilles Tendon Elasticity: Measurement Repeatability and Normative Values. J Ultrasound Med 2016; 35:159-66.

Zardi EM, Franceschetti E, Giorgi C, Palumbo A, Franceschi F. Reliability of quantitative point shear-wave ultrasound elastography on vastus medialis muscle and quadriceps and patellar tendons. Med Ultrason 2019; 21:50-55.

Zhang ZJ, Fu SN. Shear Elastic Modulus on Patellar Tendon Captured from Supersonic Shear Imaging: Correlation with Tangent Traction Modulus Computed from Material Testing System and Test-Retest Reliability. PLoS One 2013; 8:e68216.

Figure legends

Figure 1. PRISMA Flowchart of study identification, screening and eligibility.

Abbreviations: LL, lower limbs.

Figure 2. Bar chart showing the presence of the different GRRAS items in the selected studies (n=24)

Title/abstract; Item 1: Identify in title or abstract that interrater/intrarater reliability or agreement was investigated. Introduction; Item 2: Name and describe the diagnostic or measurement device of interest explicitly. Item 3: Specify the subject population of interest. Item 4: Specify the rater population of interest (if applicable). Item 5: Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable). Methods; Item 6: Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations. Item 7: Describe the sampling method. Item 8: Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding). Item 9: State whether measurements/ratings were conducted independently. Item 10: Describe the statistical analysis. Results; Item 11: State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted. Item 12: Describe the sample characteristics of raters and subjects (e.g. training, experience). Item 13: Report estimates of reliability and agreement including measures of statistical uncertainty. Discussion; Item 14: Discuss the practical relevance of results. Auxiliary material; Item 15; Provide detailed results if possible (e.g. online).

Figure 3. Bar chart showing the total GRRAS score of the different selected studies

Figure 4. Bar chart showing the presence of the different QAREL items in the selected studies (n=24). *Abr.: N/A, not applicable*

Item 1: Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? Item 2: Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? Item 3: Were raters blinded to the findings of other raters during the study? Item 4: Were raters blinded to their own prior findings of the test under evaluation? Item 5: Were raters blinded to the results of the reference standard for the target disorder (or variable) being evaluated? Item 6: Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? Item 7: Were raters blinded to additional cues that were not part of the test? Item 8: Was the order of the examination varied? Item 9: Was the time interval between repeated measurements compatible with the stability (or theoretical stability) of the variable being measured? Item 10: Was the test applied correctly and interpreted appropriately? Item 11: Were appropriate statistical measures of agreement used?

Figure 5. Bar chart showing the total QAREL score of the different selected studies

Abr.: N/A, not applicable

Figure 6. Forest plot representing the ICC values of the study analyzing SE reliability in the Achilles tendon in the longitudinal scan. Dotted line; ICC =0.6 clinical relevance threshold. Red, orange, light green and dark green areas represent poor, moderate, good and excellent reliability values respectively

■ ; intra-rater reliability

▲ ; inter-rater reliability

◆ ; inter-session reliability

● ; inter-machine reliability

Abr.: OP, operator ; SI, strain index.

Figure 7. Forest plot representing the ICC values of the study analyzing SWE reliability in the Achilles tendon in the longitudinal scan. Dotted line; ICC =0.6 clinical relevance threshold. Red, orange, light green and dark green areas represent poor, moderate, good and excellent values respectively

■ ; intra-rater reliability

▲ ; inter-rater reliability

◆ ; inter-session reliability

Abr.: PF, plantar flexion; DF, dorsi flexion; A, anisotropy coefficient; OP, operator ; Mod., modulus ; D, day; T, trial.

Figure 8. Forest plot representing the ICC values of the study analyzing SE reliability in the Patellar and quadriceps tendon in the longitudinal scan. Dotted line; ICC =0.6 clinical relevance threshold. Red, orange, light green and dark green areas represent poor, moderate, good and excellent reliability values respectively

■ ; intra-rater reliability

▲ ; inter-rater reliability

Abr.: PT, patellar tendon; QT, quadriceps tendon; Prox., proximal; Dist., distal; OP, operator; Dom., dominant; N-Dom, non-dominant; D, day; E, Esaote ultrasound.

** for GE ultrasound values of this study see Dickson et al. (l) OP1 and OP3*

Figure 9. Forest plot representing the ICC values of the study analyzing SWE reliability in the patellar and quadriceps tendon in the longitudinal scan. Dotted line; ICC =0.6 clinical relevance threshold.

Red, orange, light green and dark green areas represent poor, moderate, good and excellent reliability values respectively

■ ; intra-rater reliability

▲ ; inter-rater reliability

◆ ; inter-session reliability

Abr.: PT, patellar tendon; QT, quadriceps tendon; Prox., proximal; Dist., distal; OP, operator; Dom., dominant; N-Dom, non-dominant; D, day; T, trial.

Table 1. Study characteristics of included articles

Autors (y)	Subj.	Tendon	SE/SWE	Device	Reliability type	Protocol (measure distance)	Operators (experience)
Drakonaki et al. (2009)	50	AT	SE	Hitachi	Intra-Inter rater	1 session 2 measurements (2 h)	2 musuloskeletal radiologists (NR)
Aubry et al. (2013)	30	AT	SWE	Aixplorer	Inter-rater	1 session 3 measurements (NR)	3 medical doctors (>6 yr)
Peltz et al. (2013)	12	AT PT QT	SWE	Aixplorer	Intra-rater	2 sessions D1: 2 measurements (4 h) D2: 2 measurements (4 h) 2 days between D1 and D2	1 trained operator (NR)
Zhang and Fu (2013)	11	PT	SWE	Aixplorer	Intra-Inter rater	1 session 2 measurements (1 h)	1 operator (5 yr) and 1 physical therapist (2 yr)
Porta et al. (2014)	11	PT	SE	Esaote	Intra-Inter rater	1 session 3 measurements (NR)	2 medical doctors (>8 yr)
Hsiao et al. (2015)	5	PT	SWE	Aixplorer	Intra-Inter rater	2 sessions D1: 1 measurement D2: 1 measurement 1 week between D1 and D2	2 operators (>1 yr)
Ooi et al. (2015)	10	AT	SE	Philips	Intra-Inter rater	2 sessions D1: 1 measurement D2: 1 measurement 1 week between D1 and D2	2 sonographer (>10 and >20 yr)
Capalbo et al. (2016)	40	AT	SE	Hitachi	Inter-rater	NR	2 radiologists (10 and 4 yr)
Helfenstein-Didier et al. (2016)	7	AT	SWE	Aixplorer	Intra-rater	1 session 5 measurements (immediate)	1 operator (NR)
Schneebeli et al. (2015)	24	AT	SE	Esaote	Intra-rater	1 session 2 measurements (1 min)	1 Physical therapist (NR)
Siu et al.(2016)	36	AT	SWE	Aixplorer	Intra-Inter rater	1 session 3 measurements (NR)	3 sonographer (>1 yr)
Yamamoto et al. (2017)	25	AT	SE	Hitachi	Intra-Inter rater	1 session 3 measurements (NR)	2 medical doctors (7 and 8 yr)
Lima et al. (2017)	24	AT	SWE	Aixplorer	Intra-rater Inter-session	2 sessions D1: 1 measurement D2: 1 measurement 1 week between D1 and D2	1 operator (NR)

Table 1. *(continued)*

Autors (y)	Subj.	Tendon	SE/SWE	Device	Reliability type	Protocol (measure distance)	Operators (experience)
Payne et al. (2017)	8	AT	SE	Siemens	Intra-rater Inter-session	6 sessions D1: 5 measurements (1 h) D2-D6: 1 measurement	1 operator (NR)
Tas et al. (2017)	12	PT	SWE	Siemens	Intra-rater Inter-rater Inter-session	2 sessions D1: 2 measurements (20 min) D2: 1 measurement 1 week between D1 and D2	1 radiologist and 1 physical therapist with different experience(5-1 yr)
Mannarino et al. (2018)	23	PT	SWE	Aixplorer	Intra-rater Inter-session	2 sessions D1: 2 measurements D2: 2 measurements 1 week between D1 and D2	1 operator (NR)
Payne et al. (2018)	14	AT	SWE	Siemens	Intra-rater Inter-session	6 sessions D1: 5 measurements (1 h) D2-D6: 1 measurement	1 sport scientist (3 yr)
Petitpierre et al.(2018)	15	AT	SWE	Aixplorer	Inter-rater	1 session 2 measurements (NR)	2 radiologists (6 and 3 yr)
Corrigan et al. (2019)	20	AT	SWE	Ultrasonix	Intra-rater	2 sessions D1: 4 measurements (30 min; 2 / 4 h) D2: 1 measurement 2 weeks between D1 and D2	1 operator (NR)
Dickson et al. (2019)	20	PT QT	SE	GE	Intra-Inter rater	1 session 6 measurements (NR)	3 operators (12, 7 and 0 yr)
Ünal et al. (2019)	42	AT	SE	Toshiba and Hitachi	Intra-Inter rater Inter-machine	2 sessions D1: 2 measurements with 2 devices D2: 2 measurements with 2 devices 10 days between D1 and D2	2 operators (NR)
Zardi et al. (2019)	18	PT QT	SWE	Esaote	Intra-rater	1 session 2 measurements (15 min)	1 sonographer (>10 yr)
Alsiri et al. (2020)	18	AT	SE	GE	Intra-rater	1 session 2 measurements (1 h)	1 Physical therapist (4 yr)
Dickson et al. (2020)	20	PT QT	SE	GE and Esaote	Intra-Inter rater Inter-machine	1 session 8 measurements (15 min)	2 operators (12 and 0 yr)

Abbreviations: AT, Achilles tendon; PT, patellar tendon; QT, quadriceps tendon; SE, strain elastography; SWE, shear wave elastography; NR, not reported; D, day

Table 2. Strain elastography reliability results on Achilles tendon

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-session reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Drakonaki et al. (2009)	Strain ratio (Tendon/fat)		ICC	CV %			ICC	CV %
		Axial	0.43 ● OP1	39%			0.41 ●	30%
			0.45 ● OP2	37%				
		Longitudinal	0.78 ● OP1	30.5%			0.51 ●	29.6%
			0.66 ● OP2	30.1%				
Ooi et al. (2015)	Strain ratio (Tendon/fat)	Longitudinal	ICC 0.87 ●				ICC 0.79 ●	
	Color grading		Cohen K 0.70				Cohen K 0.70	
Capalbo et al. (2016)	Color grading	Longitudinal					Cohen K	
		Left					0.89	
		Right					0.91	
Schneebeli et al. (2016)	Strain ratio (Tendon/ reference material)	Longitudinal	ICC (95%CI)	MD (LoA)				
		Relaxed	0.87 (0.75-0.93) ● ●	0.13 (0.73/-0.47)				
		Contracted	0.94 (0.90-0.97) ● ●	0.00 (0.16/-0.17)				
Yamamoto et al. (2016)	Strain ratio (Tendon/ reference material)	Longitudinal	ICC (95%CI)	SEM			Spearman rank	
			0.93 (0.89-0.96) ● ● OP1	0.06			0.61	
			0.87 (0.80-0.92) ● ● OP2	0.07				
Payne et al. (2017)	% of displacement	Longitudinal	ICC	CV%	ICC	CV%		
		Fixed	0.11 (-0.07-0.54) ● ●	111.5%	0.01 (-0.14-0.40) ● ●	92.8%		
		Relaxed	0.06 (-0.12-0.49) ● ●	105.7%	0.10 (-0.08-0.53) ● ●	80.8%		
		Axial						
		Fixed	0.00 (-0.16-0.38) ● ●	112.4%	0.02 (-0.14-0.43) ● ●	80.4%		
		Relaxed	0.00 (-0.19-0.31) ● ●	53.6%	0.11 (-0.10-0.55) ● ●	60.9%		

Table 2. *(continued)*

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-machine reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Ünal et al. (2019)	Strain ratio (tendon/fat)	Longitudinal	ICC (95%CI)				ICC (95%CI)	
		Hitachi	0.98 (0.95-0.99) ● ● OP1				0.84 (0.74-0.89) ● ● M1	
			0.98 (0.96-0.99) ● ● OP2				0.81 (0.70-0.88) ● ● M2	
		Toshiba	0.97 (0.92-0.99) ● ● OP1				0.73 (0.57-0.83) ● ● M1	
			0.96 (0.93-0.98) ● ● OP2				0.75 (0.61-0.84) ● ● M2	
		Hitachi-Toshiba			ICC(95%CI)			
					0.85 (0.26-0.95) ● ●			
Alsiri et al. (2020)	Strain ratio (tendon/subc. fat)	Longitudinal	ICC (95%CI)	MD (LoA)				
			0.82 (0.53-0.93) ● ●	-0.33(2.70/-3.24)				
		Lower SI	0.91 (0.76-0.97) ● ●	0.00(0.03/-0.02)				
	Strain index	Higher SI	0.91 (0.78-0.97) ● ●	0.02(0.15/-1.12)				
	Color Pixel %	Red pixel %	0.99 (0.97-1.00) ● ●	-0.33(3.51/-4.17)				
		Green pixel %	0.91 (0.77-0.97) ● ●	-2.07(20.07/-24.21)				
		Blue pixel %	0.94 (0.85-0.98) ● ●	2.41(25.36/-20.54)				

Table 2. Strain elastography reliability results on Achilles tendon

Color dots in the table describes the ICC value. ● poor reliability, ● moderate reliability, ● good reliability, ● excellent reliability. When 2 dots are present, they indicate the lower and upper limit of the 95%CI.

Abbreviations: ICC, Intraclass correlation coefficient; CI, confidence interval; CV, coefficient of variation; OP, operator; MD, mean difference; LoA, limit of agreement; SEM, standard error of measurement; SI, strain index; M, measurement.

Table 3. Strain elastography reliability results on patellar and quadriceps tendon.

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-session reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Porta et al. (2014)	Strain ratio (tendon/subcutaneous tissue)	Longitudinal	ICC (95%CI)				ICC (95%CI)	
		Proximal	0.79 (0.63-0.90) ● ● OP1				0.64 (0.31-0.84) ● ●	
			0.75 (0.56-0.88) ● ● OP2					
		Middle	0.65 (0.43-0.83) ● ● OP1				0.59 (0.24-0.81) ● ●	
			0.74 (0.55-0.87) ● ● OP2					
		Distal	0.71 (0.50-0.85) ● ● OP1				0.53 (0.15-0.77) ● ●	
			0.67 (0.44-0.83) ● ● OP2					
Dickson et al. (2019)	Strain ratio (tendon/fat)	Longitudinal	ICC				ICC	
		PQT	0.49 (0.07-0.76) ● ● OP1				0.10 (-0.05-0.34) ● ● M1	
		DQT	0.43 (0.00-0.73) ● ● OP1				0.25 (-0.60-0.39) ● ● M1	
		PPT	0.85 (0.66-0.94) ● ● OP1				0.15 (-0.07-0.45) ● ● M1	
		MPT	0.13 (-0.33-0.53) ● ● OP1				0.08 (-0.10-0.35) ● ● M1	
		DPT	0.91 (0.78-0.96) ● ● OP1				0.10 (-0.05-0.33) ● ● M1	
		PQT	0.30 (-0.16-0.65) ● ● OP2				0.14 (-0.14-0.50) ● ● M2	
		DQT	-0.20 (-0.58-0.26) ● ● OP2				0.38 (-0.08-0.45) ● ● M2	
		PPT	0.49 (0.07-0.76) ● ● OP2				0.11 (-0.10-0.40) ● ● M2	
		MPT	0.44 (0.01-0.74) ● ●OP2				0.18 (-0.02-0.45) ● ● M2	
		DPT	0.41 (0.08-0.76) ● ● OP2				0.22 (0.00-0.50) ● ● M2	
		PQT	0.17 (-0.29-0.56) ● ● OP3					
		DQT	0.01 (-0.43-0.44) ● ● OP3					
		PPT	0.40 (-0.01-0.72) ● ● OP3					
		MPT	0.21 (-0.24-0.59) ● ● OP3					
		DPT	0.22 (-0.24-0.60) ● ● OP3					
	Color grading		Mean % agreement				Mean % agreement	
		PQT	72%				60% M1 / 57% M2	
		DQT	77%				50% M1 / 48% M2	
		PPT	83%				83% M1 / 83% M2	
		MPT	82%				68% M1 / 80% M2	
		DPT	93%				87% M1 / 93% M2	

Table 3. *(continued)*

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-machine reliability		Inter-rater agreement	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Dickson et al. (2020)	Strain ratio (tendon/fat)	Longitudinal	ICC		Pearson and Spearman (p values)&		t-Test # (p values)	
		PQT	0.49 (0.07-0.76) ● ● OP1 G *		0.575 OP1 M1		**<0.001** G M1	
		DQT	0.43 (0.00-0.73) ● ● OP1 G *		0.794 OP1 M1		**<0.001** G M1	
		PPT	0.85 (0.66-0.94) ● ● OP1 G *		0.267 OP1 M1		0.896 G M1	
		MPT	0.13 (-0.33-0.53) ● ● OP1 G *		0.711 OP1 M1		**0.03** G M1	
		DPT	0.91 (0.78-0.96) ● ● OP1 G *		0.138 OP1 M1		0.36 G M1	
		PQT	0.17 (-0.30-0.56) ● ●OP2 G *		0.574 OP2 M1		0.086 G M2	
		DQT	0.01 (-0.43-0.44) ● ● OP2 G *		0.846 OP2 M1		**<0.001** G M2	
		PPT	0.40 (-0.01-0.72) ● ● OP2 G *		0.122 OP2 M1		0.765 G M2	
		MPT	0.21 (-0.24-0.59) ● ● OP2 G *		0.901 OP2 M1		**0.048** G M2	
		DPT	0.22 (-0.24-0.60) ● ● OP2 G *		0.493 OP2 M1		0.09 G M2	
		PQT	0.27 (-0.19-0.63) ● ● OP1 E		0.963 OP1 M2		0.575 E M1	
		DQT	0.51 (0.10-0.77) ● ● OP1 E		0.324 OP1 M2		0.213 E M1	
		PPT	0.48 (0.05-0.75) ● ● OP1 E		0.224 OP1 M2		0.455 E M1	
		MPT	0.67 (0.32-0.85) ● ● OP1 E		0.369 OP1 M2		**0.035** E M1	
		DPT	0.15 (-0.31-0.60) ● ● OP1 E		0.619 OP1 M2		0.25 E M1	
		PQT	0.35 (-0.10-0.68) ● ● OP2 E		0.983 OP2 M2		0.533 E M2	
		DQT	0.20 (-0.26-0.58) ● ● OP2 E		0.964 OP2 M2		0.092 E M2	
		PPT	0.46 (0.04-0.75) ● ● OP2 E		0.092 OP2 M2		0.615 E M2	
		MPT	0.41 (-0.03-0.71) ● ● OP2 E		0.756 OP2 M2		**0.044** E M2	
		DPT	0.57 (0.18-0.80) ● ● OP2 E		0.441 OP2 M2		0.145 E M2	
	Color grading		Mean % agreement		Mean % agreement		Mean % agreement	
		PQT	81% G / 63% E		58%		68% G / 63% E	
		DQT	85% G / 75% E		64%		53% G / 53% E	
		PPT	80% G / 80% E		86%		80% G / 93% E	
		MPT	83% G / 88 % E		81%		73% G / 95 % E	
		DPT	95% G / 80 % E		64%		95% G / 70 % E	

Table 3. Strain elastography reliability results on patellar and quadriceps tendon.
Color dots in the table describes the ICC value. ● poor reliability, ● moderate reliability, ● good reliability, ● excellent reliability. When 2 dots are present they indicate the lower and upper limit of the 95%CI.
& significant values indicated inter-machine association
# significant values **(bolt)** indicated no inter-rater reliability.
* same dataset as Dickson et al. 2019 OP1 and OP3

Abbreviations: ICC, Intraclass correlation coefficient; CI, confidence interval; OP, operator; M, measurement; PQT, proximal quadriceps tendon; DQT, distal quadriceps tendon; PPT, proximal patellar tendon; MPT, mid patellar tendon; DPT, distal patellar tendon; G, GE ultrasound; E, Esaote ultrasound.

Table 4. Shear wave elastography reliability results on Achilles tendon.

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-session reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Aubry et al. (2013)	Velocity (m.s-1)	Axial					ICC	
		P1 max PF					0.15 ●	
		P2 45° PF					0.06 ●	
		P3 0° PF					0.29 ●	
		P4 45° DF					0.19 ●	
		Longitudinal					ICC	
		P1 max PF					0.46 ●	
		P2 45° PF					0.22 ●	
		P3 0° PF					-0.01 ●	
		P4 45° DF					0.03 ●	
	Shear modulus (kPa)	Axial					ICC	
		P1 max PF					0.16 ●	
		P2 45° PF					0.02 ●	
		P3 0° PF					0.31 ●	
		P4 45° DF					0.17 ●	
		Longitudinal					ICC	
		P1 max PF					0.43 ●	
		P2 45° PF					0.01 ●	
		P3 0° PF					-0.02 ●	
		P4 45° DF					0.04 ●	
	Anisotropy Coefficent (A)	Longitudinal					ICC	
		P1 max PF					0.21 ●	
		P2 45° PF					0.10 ●	
		P3 0° PF					0.23 ●	
		P4 45° DF					0.20 ●	
Peltz et al. (2013)	Shear modulus (kPa)	Longitudinal	ICC					
		Achilles Left	0.54 ● D1 T1					
			0.31 ● D1 T2					
			0.66 ● D2 T1					
			0.82 ● D2 T2					
		Achilles Right						
			0.19 ● D1 T2					
			0.00 ● D1 T2					
			0.11 ● D2 T1					
			0.37 ● D2 T2					

Table 4. *(continued)*

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-session reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Helfenstein-Didier et al. (2016)	Compression modulus (MPa)	Longitudinal						
			ICC	CV% and SEM				
		45°	0.42 (0.05-0.83)	0.81% and 2.14				
		35°	0.73 (0.40-0.94)	0.60% and 2.04				
		25°	0.85 (0.60-0.97)	0.48% and 1.37				
	Shear modulus (kPa)	45°	1.00 (1.00-1.00)	0.07% and 10.69				
		35°	1.00 (1.00-1.00)	0.05% and 6.45				
		25°	1.00 (1.00-1.00)	0.05% and 3.5				
Siu et al.(2016)	Shear modulus (kPa)	Longitudinal	ICC	CV %			ICC	CV %
			0.85 OP1	0.09%			0.59	0.13%
			0.84 OP2	0.1%				
			0.80 OP3	0.11%				
Lima et al. (2017)	Shear modulus (kPa)	Longitudinal	ICC	CV %	ICC			
		Left	0.93	23%	0.42			
		Right	0.82	25%	0.60			
Payne et al. (2018)	Velocity (m.s-1)	Longitudinal	ICC	CV%; SEM and MDC	ICC	CV%; SEM and MDC		
		Fixed	0.71 (0.50-0.83)	4.9%; 0.23 and 0.65	0.67 (0.44-0.81)	5.2%; 0.27 and 0.76		
		Relaxed	0.54 (0.21-0.73)	2.9%; 0.19 and 0.51	0.55 (0.24-0.73)	3.1%; 0.20 and 0.55		
		Axial						
		Fixed	0.62 (0.29-0.78)	3.9%; 0.19 and 0.53	0.78 (0.62-0.87)	4.4%; 0.17 and 0.46		
		Relaxed	0.71 (0.50-0.82)	6.0%; 0.26 and 0.71	0.85 (0.74-0.91)	6.3%; 0.19 and 0.53		
Petitpierre et al.(2018)	Shear modulus (kPa)	Longitudinal PF					p values	
		MTJ					0.163	
		Pre insertional					0.153	
		Enthesis					0.075	
		Longitudinal DF						
		Body					0.236	
		Pre insertional					0.070	
		Enthesis					0.096	
		Axial PF						
		MTJ					0.305	
		Body					0.304	
		Enthesis					0.273	
		Axial DF						
		MTJ					0.134	
		Body					0.096	
		Enthesis					0.169	

Table 4. *(continued)*

Autors (y)	Parameters	Conditions	Intra-rater reliability		Inter-session reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Corrigan et al. (2019)	Static and dynamic shear modulus. (kPa) Viscosity (Pa*s)	Longitudinal Static modulus Dynamic modulus Viscosity	ICC (95%CI) 0.70 (0.39-0.87) ● ● 0.86 (0.72-0.94) ● ● 0.86 (0.72-0.94) ● ●					

Table 4. Shear wave elastography reliability results on Achilles tendon.

Color dots in the table describes the ICC value. ● poor reliability, ● moderate reliability, ● good reliability, ● excellent reliability. When 2 dots are present they indicate the lower and upper limit of the 95%CI.

Abbreviations: ICC, Intraclass correlation coefficient; CI, confidence interval; CV, coefficient of variation; SEM, standard error of measurement; MDC, minimal detectable change; OP, operator; PF, plantar flexion; DF, dorsi flexion; MTJ, musculotendinious junction.

Table 5. Shear wave elastography reliability results on patellar and quadriceps tendon.

Autors (y)	Parameters	Contitions	Intra-rater reliability		Inter-session reliability		Inter-rater reliability	
			Relative	Absolute	Relative	Absolute	Relative	Absolute
Peltz et al. (2013)	Shear modulus (kPa)	Longitudinal	ICC					
		Patellar Left	0.90 ● D1 T1					
			0.49 ● D1 T2					
			0.40 ● D2 T1					
			0.67 ● D2 T2					
		Patellar Right	0.45 ● D1 T1					
			0.74 ● D1 T2					
			0.24 ● D2 T1					
			0.23 ● D2 T2					
		Quadriceps Left	0.67 ● D1 T1					
			0.80 ● D1 T2					
			0.79 ● D2 T1					
			0.81 ● D2 T2					
		Quadriceps Right	0.67 ● D1 T1					
			0.61 ● D1 T2					
			0.87 ● D2 T1					
			0.67 ● D2 T2					
Zhang and Fu (2013)	Shear modulus (kPa)	Longitudinal Proximal	ICC (95%CI) 0.98 (0.93-0.99) ● ●	SEM / MDC (kPa) 1.54 / 4.27			ICC (95%CI) 0.97 (0.93-0.98) ● ●	SEM / MDC (kPa) 1.64 / 4.54
Hsiao et al. (2015)	Shear modulus (kPa)	Longitudinal	ICC (95%CI)				ICC (95%CI)	
		Proximal	0.87 (0.58–0.97) ● ● OP1				0.74 (0.04–0.94) ● ●	
			0.83 (0.32–0.96) ● ● OP2					
		Middle	0.97 (0.87–0.99) ● ● OP1				0.82 (0.28–0.96) ● ●	
			0.99 (0.96–1.00) ● ● OP2					
		Distal	0.84 (0.47–0.96) ● ● OP1				0.78 (0.13–0.95) ● ●	
			0.88 (0.51–0.97) ● ● OP2					
Tas et al. (2017)	Velocity (m.s-1)	Longitudinal	ICC (95%CI)	CV% / SEM	ICC (95%CI)	CV% / SEM	ICC (95%CI)	CV% / SEM
			0.92 (0.82-0.97) ● ● OP1	6.5% / 0.48	0.83 (0.57-0.95) ● ● OP1	7.5% / 0.56	0.71 (0.32-0.88) ● ●	9.3% / 0.70
			0.91 (0.80-0.96) ● ● OP2	5.9% / 0.44	0.81 (0.57-0.93) ● ● OP2	7.1% / 0.51		
Mannarino et al. (2018)	Shear modulus (kPa)	Longitudinal	ICC (95%CI)	CV%	ICC (95%CI)	CV%		
		Dominant	0.89 (0.55-0.92) ● ● D1	40.19%	0.65 (0.09-0.87) ● ●	41.26%		
			0.93 (0.80-0.97) ● ● D2	62.79%				
		Non-dominant	0.89 (0.61-0.97) ● ● D1	62.98%	0.63 (0.07-0.85) ● ●	48.19%		
			0.90 (0.65-0.97) ● ● D2	48.37%				
Zardi et al. (2019)	Shear modulus (kPa)	Longitudinal	ICC (95%CI)	CV (M1/M2)				
		Quadriceps	1.00 (0.99-1.00) ● ●	0.527 / 0.528				
		Patellar	0.99 (0.98-1.00) ● ●	0.623 / 0.629				