Does magnetic resonance imaging provide superior reliability for Achilles and patellar tendon cross-sectional area measurements compared to ultrasound imaging?

Lauri Stenroth[1], Sandra Sefa[1], Jari Arokoski[2], Juha Töyräs[1,3,4]

[1]Department of Applied Physics, University of Eastern Finland, Finland

[2]Department of Physical and Rehabilitation Medicine, Helsinki University Hospital, Helsinki, Finland and University of Helsinki, Helsinki, Finland.

[3]Diagnostic Imaging Center, Kuopio University Hospital, Finland

[4]School of Information Technology and Electrical Engineering, The University of Queensland

Corresponding author:

Lauri Stenroth

Department of Applied Physics, University of Eastern Finland, PO Box 1627, 70210 Kuopio, Finland

Phone: +358505649096, E-mail: lauri.stenroth@uef.fi

Abstract

This study investigated reliability of Achilles and patellar tendon cross-sectional area (CSA) measurement using ultrasound imaging (USI) and magnetic resonance imaging (MRI). Fifteen healthy adults were imaged twice on two occasions, interrupted by a tendon loading protocol. Tendon CSA segmentation were conducted by experienced and unexperienced raters blinded to the information regarding subject, session and loading status. USI provided good test-retest reliability (ICC 2,1>0.85, SEM 5-6%), while with MRI it was excellent (ICC 2,1>0.92, SEM 4%) for experienced rater. This study suggests that MRI provides superior reliability for tendon CSA measurement compared to USI. However, the difference in reliability between the methods was small and the results were inconclusive regarding objectivity and sensitivity to change as assessed by effect of loading. We conclude that both methods can be used for reliable CSA measurements of the Achilles and patellar tendons when using a highly standardized measurement protocol and when conducted by an experienced rater.

Keywords: test-retest reliability; interobserver reliability; intraobserver reliability; repeatability; measurement error; sensitivity; tendon morphology

Introduction

Reliable assessment of tendon dimensions is invaluable for studying tendon adaptations occurring in response to interventions, aging, injury or disease and has importance in both research and clinical settings. Tendon cross-sectional area (CSA) is an important two-dimensional measure that can be used to estimate the average stress that a tendon is subjected to during loading and is needed to estimate tendon material properties, such as Young's modulus. In addition, tendon CSA is an important measure that reflects physiological adaptations of a tendon, such as tissue hypertrophy in response to loading (Seynnes et al. 2009) or maturation (Kubo et al. 2014), pathophysiological adaptations in response to overloading (Arya and Kulig 2010), healing after tendon rupture (Karjalainen et al. 1997) and changes due to aging (Stenroth et al. 2012), unloading (de Boer et al. 2007) or surgical operations (Kösters et al. 2015).

Ultrasound imaging (USI) and magnetic resonance imaging (MRI) are the imaging modalities used for measuring tendon CSA *in vivo*. USI is an appealing method of choice for scientific research and clinical evaluations. It is readily available, relatively inexpensive, its temporal resolution is good allowing for dynamic imaging and fast measurements, and portable devices are available. However, it has been suggested that USI results poor reliability and objectivity for determination of Achilles (Bohm et al. 2016) and patellar tendon (Ekizos et al. 2013) CSA – the two lower limb tendons that are most often of interest in human studies. The authors speculated that the poor reliability and objectivity may be due to unclear visualization of tendon borders (Bohm et al. 2016; Ekizos et al. 2013). There is also operator dependency in USI resulting from variations in probe orientation and pressure (Kruse et al. 2017) which may in part explain the poor reliability. On the other hand, based on recent systematic reviews, previous studies have mostly reported excellent test-retest reliability (ICC>0.9) for Achilles and patellar tendon CSA measurement using USI with experienced operator. Only individual studies have reported moderate or poor reliability estimates (ICC<0.75). (Mc Auliffe et al. 2017; Thoirs and Childs 2018).

MRI presumably overcomes most of the issues contributing to the reported poor reliability of USI since contrast between tissues is good, there is no pressure applied on the tissue, image plane orientation can be set accurately, and with 3D sequences the imaging plane can be adjusted post imaging (re-slicing). In fact, in comparison to USI, MRI has been shown to produce better objectivity of tendon CSA measurement (Kruse et al. 2017). Hence, MRI has been suggested to be the gold standard for measuring tendon CSA and has been used as the measure to which USI based measurement is compared to investigate its validity (Bohm et al. 2016; Kruse et al. 2017). However, literature does not consistently support the premise that MRI provides better reliability for measurement of tendon dimensions. Brushøj et al. (2006) reported lower within rater coefficient of variations for thickness, width and CSA measurements for tibialis anterior tendon using USI compared to MRI. In addition, within and between rater limits of agreement were smaller for USI compared to MRI when determining Achilles tendon thickness. These authors also concluded that reliability was limited with both methods and that there is no clear evidence on one of the methods being preferable to detect changes in tendon dimensions.

To the best of our knowledge, only two previous studies have reported test-retest reliability (repeating both imaging and image segmentation) of Achilles tendon CSA measurement using MRI and hence including also other sources of error than image segmentation in the reliability estimates (Hansen et al. 2003; Kubo et al. 2002). These studies had a limited number of subjects for the reliability analysis (6 and 7 subjects) and reliability was estimated only for a single rater. Furthermore, we are aware of only one study reporting test-retest reliability for patellar tendon CSA using MRI (Kubo et al. 2001). In addition, to the best of our knowledge, only studies by Brushøj et al. (2006) and Kruse et al. (2017) have reported reliability estimates of USI and MRI for measurement of lower limb tendon dimensions from the same set of subjects. However, as MRI was not repeated in these studies, the reliability estimates in these studies refers only to reliability of image segmentation.

Currently, there is limited knowledge on reliability of MRI in tendon CSA measurement from a study set-up with repeated image acquisitions on separate days to include measurement variability accountable to biological variability, instrumentation and measurement procedures. Hence, conclusions on superiority of MRI compared to USI for tendon CSA measurement cannot be made based on the available evidence. Therefore, the aim of the study was to investigate reliability of Achilles and patellar tendon CSA measurement using USI and MRI allowing direct comparison of reliability of the methods. In particular, the aim was to estimate test-retest reliability, inter-rater reliability and intra-rater reliability of both imaging modalities and to investigate effect of rater experience on the reliability. Additionally, we aimed to investigate sensitivity of the methods to detect changes in tendon CSA. This was done by repeating the imaging before and after a loading protocol aimed to alter tendon CSA acutely. We hypothesized that MRI results in better reliability compared to USI for both Achilles and patellar tendons and that reliability is better for an experienced rater compared to an inexperienced rater.

Materials and Methods

Subjects

Fifteen healthy adults (8 female, 7 male, age 26±5 years) were recruited for the study. Exclusion criteria were surgical operations at the ankle or knee, any known previous Achilles or patellar tendon alterations including complete or partial rupture, diagnosed tendinopathy or tendon pain in the past two years, contraindication for MRI (cardiac pacemaker, metal objects in the body such as aneurysm clip, joint prosthesis or bone fixation devices and pregnancy), rheumatoid arthritis, gout, use of fluoroquinolone medication in the past two years and acute illness such as fever or the common cold. In addition, images of Achilles and patellar tendons obtained during the study were reviewed to exclude subjects exhibiting signs of tendinopathy (hypoechoic areas or fusiform thickening of the tendon). The ethics committee of the Hospital District of Northern Savo approved the study protocol, and all participants signed an informed consent before participation in the study. The study was conducted according to the principles set by the Declaration of Helsinki.

Protocol

Subjects' Achilles and patellar tendons were imaged using MRI and USI on two occasions separated approximately by one week (mean±SD, 7.7±3.5 days, Fig. 1). Imaging was conducted at the same time of day on both days to avoid possible diurnal effects on tendon size. The subjects avoided physical activities, other than commuting, on the day of measurements as confirmed by a questionnaire. During each session, both Achilles and patellar tendons were imaged before and directly after a loading protocol. Previous studies have reported acute changes in Achilles tendon volume after running and rope skipping (Grosse et al. 2015; Syha et al. 2013). Since volume was assessed in these studies from constant tendon length the results imply that average CSA was altered. Therefore, a loading protocol was used to induce an acute change in tendon CSA that would enable comparison of sensitivity of the methods to detect change in tendon CSA and to investigate whether the possible response is systematic across the sessions.

Tendon loading protocol

Both Achilles and patellar tendons were loaded using a loading protocol similar to those previously shown to induce acute response in tendon dimensions (Grigg et al. 2009; Wearing et al. 2013). Subjects performed 5 sets of 10 repetitions of single leg straight knee heel drops from a step and single leg incline squats with external load of approximately 20 % of body mass (19.8±1.5%) situated in a backpack. Eccentric phase of the movement lasted for 3 seconds and concentric phase 2 seconds. There was one-minute rest between the sets.

Imaging was conducted as quickly as possible after the loading. USI was conducted on average 12±2 minutes and MRI 26±2 minutes from cessation of the loading. An additional USI was conducted for the mid location of the tendon (see below for details) immediately after the MRI. This was done to verify that the possible differences in the effect of loading measured by USI and MRI would not be due to the different time intervals from loading to imaging. The additional USI imaging was conducted on average 30±4 minutes after the loading. Due to technical difficulties, the additional USI was performed for 12 subjects for Achilles tendon and for 11 subjects for patellar tendon.

Image acquisition

For the Achilles tendon imaging, subjects lay prone on an examination table with feet over the edge of the table. The ankle angle was fixed to 90° angle (tibia perpendicular to foot) using a custom made MRI compatible splint (Fig 2, Woodcast, Onbone Oy, Helsinki). For patellar tendon imaging, subjects were positioned supine on the examination table and the knee joint was positioned in 15° flexion to remove slackness of the tendon and render the patellar tendon straight. The same joint configuration was used for both USI and MRI and the same ankle splint was used for both USI and MRI of the Achilles tendon.

Tendon CSA was imaged using MRI and USI from three locations: 25, 50 and 75 % of the tendon length similarly as in several previous studies to estimate average tendon CSA (Couppé et al. 2008; Couppé et al. 2009; Couppé et al. 2016; Kongsgaard et al. 2007; Murtagh et al. 2018). These locations are later referred as distal, mid and proximal locations. Distal and proximal ends of the tendon (the proximal edge of posterior calcaneus and the most distal point of soleus muscle-tendon junction; inferior patellar pole and tibia tuberosity) were located using USI (Philips EnVisor HD, 12 MHz, 160 element, 38 mm linear transducer, image depth 30 mm, image size 539x450 pixels, 0.07 mm image resolution) and the locations were marked over the skin. Tendon length was measured using a flexible measuring tape and the locations corresponding the three measurement locations for CSA were marked with a pen after which a thin strip of tape (0.5 cm wide, Micropore, 3M, USA) was positioned transversely over the skin to these locations. These tapes were used to accurately detect the imaging location of the tendon while conducting USI (Kruse et al. 2017).

During USI, the ultrasound transducer was positioned over the tendon in transverse plane. To ensure that the imaging plane was perpendicular to the longitudinal axis of the tendon, the highest echo intensity of the tendon was found by small adjustments to the transducer angle. Three repeat images were acquired from each location with repositioning of the transducer between each image. Generous amount of ultrasound gel (Aquasonic 100, Parker Laboratories Inc, USA) and a 10 mm thick strand-off pad (ATS Laboratories, Inc., Norfolk, USA) were used to reduce to amount of pressure needed to obtain clear images (Kruse et al. 2017). Additionally, a three-second video sequence was stored from each measurement location while slightly varying the transducer angle and location. The video was later used during the segmentation as a reference to help identify borders of the tendon (Bohm et al. 2016).

MRI was performed using an open MRI device (Esaote E-Scan XQ, Esaote, Italy, 0.18 T). Imaging plane was carefully aligned perpendicular to the long axis of the tendon using scout images acquired during subject positioning. Transverse images of the tendon were acquired using the following imaging parameters: Spin Echo T1, TR/TE 750/26, 5 mm slice thickness, 0.5 mm slice gap, number of averages: 2 and 0.59 mm in-plane resolution. A fish oil capsule positioned over the lateral aspect of the tendon at the level of the mid location (long axis of the capsule in transverse plane) was used to identify the slices corresponding to the locations of USI (Kruse et al. 2017).

Image analysis

Tendon CSA was manually segmented using polygon tool on OsiriX software (OsiriX Lite v.9.0, Pixmeo SARL, Switzerland, fig. 3 and 4). Two raters independently analyzed the images. Rater 1 was considered as an experienced rater with more than five years of experience in musculoskeletal imaging and segmentation. Rater 2 was considered as inexperienced with no previous experience on musculoskeletal radiography. The experienced rater conducted all measurements and imaging described above and trained the inexperienced rater to conduct the segmentation. The training included viewing a set of example images together to unify the analysis and segmentation of a training set of images. The training segmentation was conducted twice to verify repeatable analysis and to identify possible sources for variability, which were checked together with the experienced rater. After the training, the segmentations were performed independently by the raters.

Each subject was given a random identification number separately for both sessions and for both pre and post loading conditions. This ensured that the segmentation process was conducted blinded and in a random order, excluding possibility of systematic error between sessions due to learning or other systematic change in the segmentation. Mean of the three repeated images per location obtained using USI were used as the CSA value for the location. Rater 1 repeated analysis of the ultrasound and magnetic resonance images obtained during the first session before loading for the analysis of intra-rater reliability. This analysis was done approximately three months after the first analysis to make the repeated analysis as independent as possible.

Statistical analysis

Due to technical difficulties, data from one subject for patellar tendon USI after loading in the first session was missing. Hence, the number of subjects for analysis of effect of loading for patellar tendon was 14.

Repeated measures two-way ANOVA was used to test the effect of loading on tendon CSA, separately for both raters and imaging modalities (within subject factors: session and loading). Partial eta squared (η^2) and percentage change were used as measures of effect size for this analysis. Reliability of tendon CSA measurement was assessed by analysis of test-retest reliability (Session 1 vs Session 2), inter-rater reliability (Rater 1 vs Rater 2) and intra-rater reliability (repeated image segmentation). Test-retest reliability analyses used the images obtained before the loading in both sessions. The images obtained before and after the loading in both sessions were used for inter-rater reliability analyses. Intra-rater reliability analyses used the images obtained before the loading during the first session. Repeated measures t-test was used to test for systemic errors in the test-retest reliability analysis and in the intra-rater reliability analysis. Repeated measures three-way ANOVA was used to test for systemic errors in the inter-rater reliability analysis (within subject factors: session, loading and rater). In the reliability analyses, intraclass correlation coefficient (ICC), standard error of measurement (SEM) and minimal detectable change with 95% confidence level (MDC) were calculated as measures of random error according to Weir (2005). The ICC model used was two-way random effects model for absolute agreement and single rater (ICC 2,1). ICC values were interpreted according to Koo and Li (2016) with the following cut points: <0.5 poor, 0.5-0.75 moderate, 0.75-0.9 good and >0.90 excellent reliability. SEM was calculated as the square root of the mean square error from the ANOVA. MDC was calculated as $MDC=SEM*1.96*\sqrt{2}$. Both SEM and MDC are presented in the units of the measurement and as a percentage of the mean. To evaluate agreement between the methods (USI and MRI), Bland-Altman plots were produced with limits of agreement. Two-way ANOVA was used to test the difference between the imaging modalities, separately for both raters (within subject factors: session and modality). Additionally, Pearson correlation coefficients were calculated for the tendon CSA values measured by USI and MRI. The level of statistical significance was set at $p<0.05$. All statistical analyses were conducted using IBM SPSS Statistics software (version 25, SPSS Inc., IBM Company, Armonk, NY, USA).

Results

Loading significantly reduced Achilles tendon CSA measured from ultrasound images by Rater 1 (marginal means: 56.2±2.0 vs 54.8±2.1 mm^2, p=0.021). This effect was not observed for the measurements performed by Rater 2 (p=0.277) or for CSA measured from MRI (Rater 1 p=0.381, Rater 2 p=0.560, Table 1). The additional USI conducted after MRI revealed that Achilles tendon CSA had returned to the pre loading value at this time point (marginal means: pre loading 54.1±2.2 mm^2, post loading: 52.0±2.6 mm^2 and post MRI: 55.0±2.9 mm^2, pre loading vs post loading p=0.033, pre loading vs post MRI p=0.524). No significant effect of loading was observed for patellar tendon with either imaging method or for analyses performed by either rater (Table 1).

No systematic differences were observed between the sessions (test-retest) in the analyses performed by either of the raters (Table 2). Test-retest reliability estimates were consistently better for MRI compared to USI and better for experienced (Rater 1) compared to unexperienced (Rater 2) rater. No clear difference was observed for test-retest reliability between Achilles and patellar tendons. MDC estimates were approximately 90% larger for unexperienced (Rater 2) compared to experienced (Rater 1) rater. Within rater, there were not clear differences between tendons in MDC. Inter-rater reliability analysis revealed systematic differences between the raters for the Achilles tendon CSA measured from USI and MRI and for the patellar tendon measured from MRI ($p<0.05$, Table 3). MRI performed better than USI also regarding inter-rater reliability. This was reflected in approximately 40% smaller MDC for MRI compared to USI. In addition, larger ICC values were observed consistently for Achilles tendon compared to patellar tendon. However, for USI there was no clear difference in SEM% or MDC% between tendons in the inter-rater reliability. There was a systematic difference in the Achilles tendon CSA between the repeated image segmentations (i.e. intra-rater reliability) performed by the Rater 1 from MRI ($p<0.001$, Table 4). No differences were observed between the repeated analyses in Achilles tendon CSA measured from USI or in patellar tendon CSA measured from USI or MRI. Regardless of the systematic error observed for Achilles tendon CSA from MRI in the intra-rater reliability analysis, measures of relative (ICC) and absolute (SEM) reliability were consistently better for MRI compared to USI. MDC values estimated from intra-rater reliability analysis were approximately 50% and 60% smaller for MRI compared to USI, for Achilles and patellar tendons, respectively.

Bland-Altman plots in figure 5 illustrates agreement between the imaging modalities for both raters separately. Systematic differences between the methods were observed for Achilles tendon in the analyses performed by Rater 1 ($p<0.001$) and for patellar tendon in the analysis performed by Rater 2 ($p=0.002$). Agreement between the methods was better for Rater 1 as indicated by the smaller limits of agreement and higher correlation between the methods.

Discussion

To the best of our knowledge, this is the first study allowing for direct comparison of test-retest reliability estimates of USI and MRI for measurement of Achilles and patellar tendon CSA. As hypothesized, MRI resulted in slightly better test-retest, intra-rater and inter-rater reliability for both tendons investigated. This was true for both relative (ICC estimates) and absolute (SEM values) reliability. However, the differences in reliability were not large and other factors, such as rater experience, may have larger effect on reliability than imaging modality. The loading protocol significantly reduced Achilles tendon CSA measured from USI by the experienced rater. This was not observed from MRI but the results of an additional USI performed directly after MRI suggested that the CSA was recovered to pre loading values already by the time of MRI. No significant effect of loading was observed for patellar tendon CSA. Hence, the loading protocol did not allow us to investigate differences between the imaging methods in sensitivity to detect change in tendon CSA.

Effects of loading

Loading did not have clear effect on the measured tendon CSA that would have been observed with both imaging modalities or by both raters. A small reduction of 2.5% in Achilles tendon CSA was observed with USI by Rater 1. The lack of consistent effect of loading on tendon CSA was unexpected since several previous studies have reported acute reduction in tendon dimensions (Grigg et al. 2009; Grigg et al. 2012; Wearing et al. 2011; Wearing et al. 2013; Wearing et al. 2014). The possible reason for the lack of effect in the current study could be insufficient volume of loading. Here, we used total of 50 repetitions with effective loading of 120% of body mass (20% external load) and 5 second loading duration per repetition. Series of studies by Wearing et al., in which large reduction in both Achilles and patellar tendon thickness has been observed, used comparable effective loading but the subjects performed in total 90-100 repetitions of the exercises (Grigg et al. 2009; Grigg et al. 2012; Wearing et al. 2011; Wearing et al. 2013; Wearing et al. 2014). Still, the same research group have reported significant reduction in patellar tendon thickness after only 45 repetitions of similar loading (Wearing et al. 2015). Moreover, reduction in Achilles tendon volume has been reported after cross-country running and rope skipping (Grosse et al. 2015; Syha et al. 2013).

The lack of change in tendon CSA due to loading prevented us from assessing differences in sensitivity to change between the methods. Still, we opted to report the results regarding effects of loading, since they provide valuable information for designing future studies. The results showed that the effect of strenuous loading (subjects reported considerable fatigue and some subjects were forced to limit the range of motion during the final set to complete the exercise) on tendon CSA could be recovered 30 minutes after the loading. Therefore, a 30-minute rest prior to the measurement of tendon CSA should be enough to remove loading history dependent effect on tendon CSA from a loading similar to the one performed in the current study. However, it should be noted that, time course of the recovery may differ depending on the type of loading and extremely intensive loading may require as long as 24 hours for the recovery of tendon dimensions (Wearing et al. 2014).

Test-retest reliability

No systematic differences were observed between the CSA measured on different sessions indicating that there were no systematic sources of error in the test setup. For the more experienced rater the mean difference was small in each case (<1% of the mean). The ICC estimates for Rater 1 suggests good reliability for USI and excellent reliability for MRI, for both tendons. For Rater 2 the ICC estimates suggests good reliability for Achilles tendon USI and MRI and for patellar tendon MRI. Moderate reliability was obtained for patellar tendon USI. However, confidence intervals for the estimates of ICC were large and thus the relative reliability values should be interpreted with caution (Table 2).

Previous reports on test-retest reliability of Achilles tendon CSA measurement using USI, as summarized by recent systematic review, have consistently showed SEM values around 5% for experienced rater. The review identified only one exception in which case larger SEM values were reported for measurement taken 2-4 cm from Achilles insertion to calcaneus (Thoirs and Childs 2018). This is consistent with the current study in which we report SEM of 5.3% for Achilles tendon for the experienced rater. Previous studies have reported ICC ranging from 0.59 to 0.99 for patellar tendon test-retest reliability using USI (Mc Auliffe et al. 2017; Wiesinger et al. 2016). We reported ICC estimate of 0.851 for the experienced rater 0.504 for the unexperienced rater.

Only few previous studies have reported test-retest reliability estimates for MRI based tendon CSA measurements. Those studies reported the reliability as coefficient of variation and yielded values of 1.5% (Kubo et al. 2002) and 4.5-7.5% depending on location (Hansen et al. 2003) for Achilles tendon. From our data the corresponding values were 3.4% and 6.6% for the experienced and unexperienced raters, respectively. Kubo et al. (2001) reported coefficient of variation of 1.6% for patellar tendon CSA measurement for repeated measures performed 12 weeks apart. In our study these values were 4.1% and 6.0% for the experienced and unexperienced raters, respectively.

Inter-rater reliability

The unexperienced rater (Rater 2) evaluated Achilles tendon to be larger using both USI and MRI compared to the experienced rater. ICC estimates for inter-rater reliability indicated moderate relative reliability. SEM values were similar for the inter-rater reliability compared to test-retest reliability for Rater 2 and suggests that the variability between the raters was mostly due to variability in the segmentations conducted by Rater 2. Better ICC and SEM estimates for MRI compared to USI indicate smaller random error of measurement between the raters for MRI. However, only patellar tendon measurement using USI did not show systematic difference between the raters. Hence, while segmentation of tendon CSA from MRI involves less random error, the segmentations of different raters may differ systematically. Therefore, we cannot conclude that MRI based CSA measurement would be more objective compared to USI based CSA measurement. Inter-rater reliability in the current study was worse than reported for the Achilles tendon CSA measurement using USI in many previous reports (Bleakney et al. 2002; Kruse et al. 2017; Ying et al. 2003). However, the reliability was similar than previously reported for unexperienced raters (Dudley-Javoroski et al. 2010). Hence, as stated above, the values for inter-rater reliability in the current study are probably mostly affected by the variability in the segmentations conducted by the unexperienced rater. The systematic differences between the raters along with the poorer test-retest reliability for the Rater 2 highlight the need for a single experienced rater.

The aim of the current study was to analyze reliability when imaging is done by a single experienced operator. Therefore, our analysis of inter-rater reliability only accounts for variation due to image segmentation. Although, we did not conduct investigation of intra-operator reliability, we suggest that single experienced operator should be used for USI. Kruse et al. (2017)

examined inter-operator reliability of Achilles tendon CSA measurement with USI. Intra-operator reliability was found to be better compared to inter-operator reliability, and they concluded that a single operator should be used.

Intra-rater reliability

Reliability of the image segmentation was estimated from repeat analysis of a subset of the images by Rater 1 (intra-rater reliability). The analysis showed comparable ICC estimates to the test-retest reliability analysis and the SEM estimates were only marginally better for the intra-rater reliability analysis compared to test-retest analysis. The result suggests that most of the measurement error in the test-retest setting can be attributed to the image segmentation. Significant mean difference was observed for Achilles tendon CSA measured from MRI. This may be due to the relatively long time separation between the repeated segmentations resulting in a systematic change in the way the tendon borders were identified during the segmentation. However, the time separation between segmentations was deemed necessary to be able to consider the repeated segmentations independent. From perspective of random measurement error, both relative and absolute intra-rater reliability was better for MRI compared to USI. As there is possibility for systematic changes in the segmentations performed by a single rater over time, it is suggested that in research settings the segmentations are performed within as short time period as possible.

Agreement between USI and MRI

Achilles tendon CSA values measured from USI were systematically larger compared to CSA measured from MRI for Rater 1 (mean difference 7.6%, $p<0.001$), but not for Rater 2 (mean difference 4.1%, $p=0.214$). On the other hand, patellar tendon CSA values measured by Rater 2 from USI were smaller compared to CSA measured from MRI (mean difference 13.9%, p=0.002), but this systematic difference was not observed for Rater 1 (mean difference 1.2%, p=0.405). Therefore, a clear conclusion cannot be made on which imaging method would overestimates or underestimates tendon CSA. There was a significant correlation between CSA values measured with USI and MRI. The correlations were larger and limits of agreement between the methods smaller for the more experienced rater (Rater 1) for both tendons. Magnitudes of the limits of agreement as a percentage of the mean value were similar between the tendons indicating that the agreement between USI and MRI is similar for both Achilles and patellar tendons. In comparison to study by Bohm et al. (2016), the limits of agreement obtained for the Achilles tendon CSA measurement in the current study were slightly smaller for the experienced rater (Rater 1), but larger for the unexperienced rater. Although, significant mean differences were observed between the methods the differences were not systematic between the raters. Therefore, CSA values, and by extension calculated tendon stresses and Young's modulus values, may be comparable between different studies regardless whether USI or MRI was used. However, possible systematic difference between raters, which may be substantial, should be taken into account in the interpretation and this is not limited to only between methods comparisons, but also comparisons of results from different studies using the same imaging modality.

Assessment of differences within or between individuals

MDC is an estimate of the minimal difference that can be observed from single measures and is therefore indicative of the method's ability to detect changes within an individual or to compare two individuals. We observed MDC values ranging from 12 to 27% in the inter-rater analysis, which were larger than in the test-retest reliability analysis for Rater 1 (MDC range from 10-16%), but comparable to that observed for Rater 2 (MDC range from 17-28%). Therefore, we suggest that the same experienced rater will analyze the images for a particular subject. In addition, when making inference regarding adaption within an individual or when comparing two different individuals the difference in two measurements should exceed MDC to be considered real difference. The MDC values are relatively large. However, as summarized by the review by Wiesinger et al. (2015), several cross-sectional studies revealed that tendon CSA was consistently larger (approximately 34%) in long-term athletes than that in controls. Hence, accuracy of tendon CSA measurement may be sufficient for assessing individual adaptations or for comparing two individuals. Due to smaller MDC estimates for MRI compared to USI, MRI should be used for these purposes when possible.

Limitations

The results of the current study may not generalize to different imaging devices. In the current study, a low field MRI device was used (0.18 T) due to practical reasons. MRI devices with higher magnetic field strengths have better signal to noise ratio at comparable imaging parameters. This allows use of smaller voxel size and may yield better delineation of tendon borders and hence better reliability of the CSA measurement. However, better signal to noise ratio may not directly translate to better measurement reliability as a study comparing a 0.2 T MRI device to a 1 T MRI device reported that, regardless of the better signal to noise ratio with the 1 T magnet, contrast was similar with the two systems (Trattnig et al. 1997). To overcome the limitations of the low magnetic field device used in the current study, we used imaging parameters that resulted good signal to noise ratio and contrast between the tendon and surrounding tissues. To obtain this, relatively large voxel size was used with a 5 mm slice thickness. Since tendon was imaged perpendicular to its longitudinal axis, we assumed that partial volume effect would be minimal regardless of the thick slices. Additionally, the slice thickness was comparable to that used in several previous studies to measure Achilles and patellar tendon CSA (Couppé et al. 2009; Couppé et al. 2014; Hansen et al. 2003; Magnusson et al. 2001). As high as 10 mm slice thickness has been used in patellar tendon CSA measurements (Kubo et al. 2002). In addition, we increased imaging time to reach better image quality and hence average of two acquisitions was used.

The explanation for the discrepancy between USI and MRI in detecting the effect of loading was probably the difference in the timespan from the cessation of loading to image acquisition. This is supported by the observation that USI repeated after the MRI at the 50% of tendon length did not reveal the same reduction in Achilles tendon CSA that was observed in the images acquired before MRI. Therefore, it is likely that the loading induced reduction in Achilles tendon CSA had already recovered by the time of MRI which prevented us from making solid conclusion regarding the sensitivity to change between MRI and USI. Counterbalanced study design for the order, in which MRI and USI was conducted after the loading, would have removed order effect from the analyses. However, this might have also masked transient effects since it would have lowered statistical power of our analyses. We wanted to retain sufficient statistical power to observe small transient changes in tendon CSA and hence opted not to use counterbalanced study design.

Finally, the result of the current study cannot be directly transferred to other studies. Factors such as devices used, experience of operator and different measurement sites may affect measurement reliability. Furthermore, the results of the current study are not transferrable to measurements taken from pathological tendons, e.g. in case of tendinopathy.

Conclusions

Compared to USI, MRI provides slightly better relative and absolute reliability for measuring Achilles and patellar tendon CSA. Most of the measurement variability in both USI and MRI based measurements comes from segmentation errors. Rater experience has significant effect on reliability regardless of the imaging methods. It remains inconclusive if the better reliability observed for MRI also leads to superior ability of MRI to detect changes in tendon CSA compared to USI since lower image resolution and the resulting partial volume effect may hinder ability of MRI to detect subtle changes. We conclude that both USI and MRI can be used for reliable measurement of Achilles and patellar tendon CSA when using a highly controlled measurement protocol. In future, development of automatic segmentation techniques for USI, that are already available for MRI (Kruse et al. 2017; Syha et al. 2012), could further improve reliability of USI based measurements. This study allows direct comparison of MRI and USI in measurement of tendon CSA. The results of the current study can be used for calculating required sample sizes for future studies considering the measurement errors associated with the particular method.

Acknowledgements

L.S. acknowledges financial support from Finnish Cultural Foundation and Emil Aaltonen Foundation. The funding sources did not have any influence on the study.

References

Arya S, Kulig K. Tendinopathy alters mechanical and material properties of the Achilles tendon. J Appl Physiol 2010;670–675.

Bleakney RR, Tallon C, Wong JK, Lim KP, Maffulli N. Long-term ultrasonographic features of the Achilles tendon after rupture. Clin J Sport Med 2002;273–278.

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

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

Couppé C, Hansen P, Kongsgaard M, Kovanen V, Suetta C, Aagaard P, Kjaer M, Magnusson SP. Mechanical properties and collagen cross-linking of the patellar tendon in old and young men. J Appl Physiol 2009;880–886.

Couppé C, Kongsgaard M, Aagaard P, Hansen P, Bojsen-Møller J, Kjær M, Magnusson SP. Habitual loading results in tendon hypertrophy and increased stiffness of the human patellar tendon. J Appl Physiol 2008;805–10.

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

Couppé C, Svensson RB, Sødring-Elbrønd V, Hansen P, Kjær M, Magnusson SP. Accuracy of MRI technique in measuring tendon cross-sectional area. Clin Physiol Funct Imaging

2014;237–241.

de Boer MD, Maganaris CN, Seynnes OR, Rennie MJ, Narici M V. Time course of muscular, neural and tendinous adaptations to 23 day unilateral lower-limb suspension in young men. J Physiol 2007;1079–1091.

Dudley-Javoroski S, McMullen T, Borgwardt MR, Peranich LM, Shields RK. Reliability and Responsiveness of Musculoskeletal Ultrasound in Subjects with and without Spinal Cord Injury. Ultrasound Med Biol 2010;1594–1607.

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

Grigg NL, Wearing SC, Smeathers JE. Eccentric calf muscle exercise produces a greater acute reduction in Achilles tendon thickness than concentric exercise. Br J Sports Med 2009;280–3.

Grigg NL, Wearing SC, Smeathers JE. Achilles Tendinopathy Has an Aberrant Strain Response to Eccentric Exercise. Med Sci Sport Exerc 2012;12–17.

Grosse U, Syha R, Gatidis S, Grözinger G, Martirosian P, Partovi S, Nikolaou K, Robbin MR, Schick F, Springer F. MR-based in vivo follow-up study of Achilles tendon volume and hydration state after ankle-loading activity. Scand J Med Sci Sport 2015;1200–1208.

Hansen P, Aagaard P, Kjaer M, Larsson B, Magnusson SP. Effect of habitual running on human Achilles tendon load-deformation properties and cross-sectional area. J Appl Physiol 2003;2375–2380.

Karjalainen PT, Aronen HJ, Pihlajamäki HK, Soila K, Paavonen T, Böstman OM. Magnetic resonance imaging during healing of surgically repaired Achilles tendon ruptures. Am J

Sports Med 1997;164–71.

Kongsgaard M, Reitelseder S, Pedersen TG, Holm L, Aagaard P, Kjaer M, Magnusson SP. Region specific patellar tendon hypertrophy in humans following resistance training. Acta Physiol 2007;111–21.

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

Kösters A, Rieder F, Wiesinger H-P, Dorn U, Hofstaedter T, Fink C, Müller E, Seynnes OR. Alpine Skiing With total knee ArthroPlasty (ASWAP): effect on tendon properties. Scand J Med Sci Sports 2015;67–73.

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

Kubo K, Kanehisa H, Fukunaga T. Effects of different duration isometric contractions on tendon elasticity in human quadriceps muscles. J Physiol 2001;

Kubo K, Kanehisa H, Fukunaga T. Effects of resistance and stretching training programmes on the viscoelastic properties of human tendon structures in vivo. J Physiol 2002;219–26.

Kubo K, Teshima T, Hirose N, Tsunoda N. A cross-sectional study of the plantar flexor muscle and tendon during growth. Int J Sports Med 2014;828–834.

Magnusson SP, Aagaard P, Dyhre-Poulsen P, Kjaer M. Load-displacement properties of the human triceps surae aponeurosis in vivo. J Physiol 2001;277–288.

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

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

Seynnes OR, Erskine RM, Maganaris CN, Longo S, Simoneau EM, Grosset JF, Narici M V. Training-induced changes in structural and mechanical properties of the patellar tendon are related to muscle hypertrophy but not to strength gains. J Appl Physiol 2009;523–530.

Stenroth L, Peltonen J, Cronin NJNJ, Sipilä S, Finni T, Sipila S, Finni T, Sipilä S, Finni T, Sipila S, Finni T, Sipilä S, Finni T. Age-related differences in Achilles tendon properties and triceps surae muscle architecture in vivo. J Appl Physiol 2012;1537–1544.

Syha R, Springer F, Grözinger G, Würslin C, Ipach I, Ketelsen D, Schabel C, Gebhard H, Hein T, Martirosian P, Schick F, Claussen CD, Grosse U, Würslin C, Schick F. Short-term exercise-induced changes in hydration state of healthy achilles tendons can be visualized by effects of off-resonant radiofrequency saturation in a three-dimensional ultrashort echo time MRI sequence applied at 3 tesla. J Magn Reson Imaging 2013;1400–1407.

Syha R, Würslin C, Ketelsen D, Martirosian P, Grosse U, Schick F, Claussen CD, Springer F. Automated volumetric assessment of the Achilles tendon (AVAT) using a 3D T2 weighted SPACE sequence at 3T in healthy and pathologic cases. Eur J Radiol 2012;1612–1617.

Thoirs KA, Childs J. Are Ultrasound Measurements of Achilles Tendon Size Reliable? A Systematic Review of Rater Reliability. Ultrasound Med Biol 2018;

Trattnig S, Kontaxis G, Breitenseher M, Czerny C, Rand T, Turetschek K, Barth M, Imhof H. MRI on low-field tomography systems (0.2 Tesla). A quantitative comparison with equipment of medium-field strength (1.0 Tesla). Radiologe 1997;773–777.

Wearing SC, Grigg NL, Hooper SL, Smeathers JE. Conditioning of the Achilles tendon via ankle exercise improves correlations between sonographic measures of tendon thickness and body anthropometry. J Appl Physiol 2011;1384–1389.

Wearing SC, Hooper SL, Purdam C, Cook J, Grigg N, Locke S, Smeathers JE. The acute transverse strain response of the patellar tendon to quadriceps exercise. Med Sci Sports Exerc 2013;772–777.

Wearing SC, Locke S, Smeathers JE, Hooper SL. Tendinopathy alters cumulative transverse strain in the patellar tendon after exercise. Med Sci Sports Exerc 2015;264–271.

Wearing SC, Smeathers JE, Hooper SL, Locke S, Purdam C, Cook JL. The time course of in vivo recovery of transverse strain in high-stress tendons following exercise. Br J Sports Med 2014;383–387.

Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Cond. Res. 2005. pp. 231–240.

Wiesinger HP, Rieder F, Kösters A, Müller E, Seynnes OR. Are Sport-Specific Profiles of Tendon Stiffness and Cross-Sectional Area Determined by Structural or Functional Integrity? PLoS One 2016;e0158441.

Ying M, Yeung E, Li B, Li W, Lui M, Tsoi CW. Sonographic evaluation of the size of Achilles tendon: The effect of exercise and dominance of the ankle. Ultrasound Med Biol 2003;637–642.

Figure captions

Figure 1. Experimental protocol. Achilles and patellar tendons were imaged on two sessions separated by approximately one week. During each session, the tendons were first imaged using USI and then using MRI. Then, a loading protocol was conducted first for the Achilles tendon (immediately followed by USI and MRI) and then for the patellar tendon (immediately followed by USI and MRI). Mean (±SD) time from the cessation of loading to the end of imaging is presented for both USI and MRI. * denotes the additional USI performed after the loading to confirm that the potential acute change in tendon CSA remained stable throughout imaging. Due to technical difficulties, the additional USI was performed for 12 subjects for Achilles tendon and for 11 subjects for patellar tendon.

Figure 2. Determination of the imaging locations for Achilles (A and B) and patellar tendon (C and D). Distal and proximal ends of the tendon and distal, mid and proximal measurement locations corresponding 25, 50 and 75% of tendon length were marked over the skin. Thin strips of tape were positioned over the imaging locations to help identifying the correct location while performing ultrasound imaging. Notice also the mark on the foot and the corresponding mark on the splint (A) that ensured consistent splint positioning between USI and MRI and between pre and post loading measurements.

Figure 3. Examples of native and segmented ultrasound and magnetic resonance images of Achilles tendon. Bright areas on ultrasound images represent area that produce high echo intensity. T1 sequence was used for MRI. In this sequence fat is visualized with bright and tendon with dark pixels. Fish oil capsule used to identify the corresponding imaging planes from USI and MRI is partially seen in the mid image.

Figure 4. Examples of native and segmented ultrasound and magnetic resonance images of patellar tendon. Bright areas on ultrasound images represent area that produce high echo intensity. T1 sequence was used for MRI. In this sequence fat is visualized with bright and tendon with dark pixels. Fish oil capsule used to identify the corresponding imaging planes from USI and MRI is partially seen in the mid image.

Figure 5. Bland-Altman plots visualizing the agreement between ultrasound and magnetic resonance based measurement of Achilles and patellar tendon cross sectional area. Data is presented for both sessions and for pre and post loading. Mean difference, limits of agreement (1.96 times standard deviation) and Pearson correlation coefficients are presented in the figure. P-value is for the marginal mean difference between the modalities from three-way analysis of variance. P-values for the Pearson correlation were all $p<0.001$.

Table 1. Mean values of tendon CSA

	Session 1			Session 2		Effect of loading	
	Rater 1	Rater 1 Analysis 2	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Achilles							
USI, pre loading	56.2±8.4	57.8±8.8	61.9±13.3	56.3±7.7	62.8±12.1	p=0.021	p=0.277
USI, post loading	54.2±8.0		60.9±13.4	55.5±9.0	61.0±15.0	η^2=0.327, -2.5%	η^2=0.084, -2,3%
MRI, pre loading	51.4±8.4	48.5±8.4	59.1±10.5	51.3±8.8	59.8±9.8	p=0.381	p=0.560
MRI, post loading	51.7±8.9		59.5±10.2	51.5±8.7	58.3±9.8	η^2=0.055, +0.5%	η^2=0.025, -0.9%
Patellar							
USI, pre loading	89.2±13.3	91.0±16.0	85.6±12.1	89.5±12.0	91.6±13.9	p=0.463	p=0.542
USI, post loading	89.8±14.2		87.8±14.0	90.4±13.2	86.5±12.0	η^2=0.042, +0.8%	η^2=0.029, -1.7%
MRI, pre loading	87.6±12.2	87.2±12.8	101.8±16.3	87.9±12.5	101.2±17.5	p=0.075	p=0.519
MRI, post loading	88.9±13.7		102.3±14.1	90.0±15.7	98.0±19.1	η^2=0.224, +1.9%	η^2=0.033, -1.3%

Descriptive values are expressed as mean±SD (mm^2). P-value and effect size (η^2 and percentage change) for the main effect of loading from repeated measures two-way ANOVA is presented on the right.

Table 2. Test-retest reliability (Session 1 vs Session 2, pre loading)

	Mean difference (mm^2)	Mean difference (% of mean)	P-value	ICC	SEM (mm^2)	SEM%	MDC (mm^2)	MDC%
Rater 1								
Achilles, USI	0.1	0.2	0.914	0.873 (0.661-0.956)	3.0	5.3	8.2	14.6
Achilles, MRI	0.1	0.2	0.890	0.958 (0.880-0.986)	1.8	3.5	5.0	9.8
Patellar, USI	0.5	0.5	0.807	0.851 (0.611-0.948)	5.0	5.6	13.8	15.6
Patellar, MRI	0.5	0.5	0.716	0.921 (0.784-0.973)	3.5	4.0	9.7	11.2
Rater 2								
Achilles, USI	0.9	1.4	0.704	0.786 (0.471-0.923)	6.0	9.6	16.7	26.7
Achilles, MRI	0.7	1.1	0.663	0.851 (0.613-0.947)	4.0	6.7	11.1	18.7
Patellar, USI	5.7	6.4	0.105	0.504 (0.050-0.796)	8.9	10.2	24.7	28.2
Patellar, MRI	0.2	0.2	0.939	0.868 (0.650-0.954)	6.2	6.2	17.2	17.1

P-value is for the mean difference. ICC values were calculated using two-way random effects model for absolute agreement and single rater (ICC 2,1).

Table 3. Inter-rater reliability (Rater 1 vs Rater 2)

	Mean difference (mm^2)	Mean difference (% of mean)	P-value	ICC	SEM (mm^2)	SEM%	MDC (mm^2)	MDC%
Achilles, USI	6.1	10.5	0.004	0.630 (0.227-0.813)	5.7	9.8	15.8	27.0
Achilles, MRI	7.7	13.9	<0.001	0.687 (-0.066-0.908)	2.5	4.5	6.9	12.4
Patellar, USI	1.8	2.0	0.512	0.553 (0.349-0.707)	8.5	9.6	23.6	26.7
Patellar, MRI	12.2	13.0	<0.001	0.618 (-0.074-0.859)	6.0	6.4	16.7	17.7

Mean difference is based on marginal means. P-value is for the mean difference. ICC values were calculated using two-way random effects model for absolute agreement and single rater (ICC 2,1).

Table 4. Intra-rater reliability for the Rater 1 (images from session 1 pre loading segmented twice)

	Mean difference (mm^2)	Mean difference (% of mean)	P-value	ICC	SEM (mm^2)	SEM%	MDC (mm^2)	MDC%
Achilles, USI	1.6	2.8	0.128	0.892 (0.708-0.962)	2.7	4.7	7.5	13.2
Achilles, MRI	2.9	5.8	<0.001	0.923 (0.100-0.984)	1.3	2.6	3.6	7.1
Patellar, USI	2.5	2.8	0.250	0.835 (0.589-0.941)	5.8	6.4	16.0	17.8
Patellar, MRI	0.6	0.7	0.452	0.974 (0.925-0.991)	2.1	2.4	5.7	6.6

P-value is for the mean difference. ICC values were calculated using two-way random effects model for absolute agreement and single rater (ICC 2,1).

Figure 1

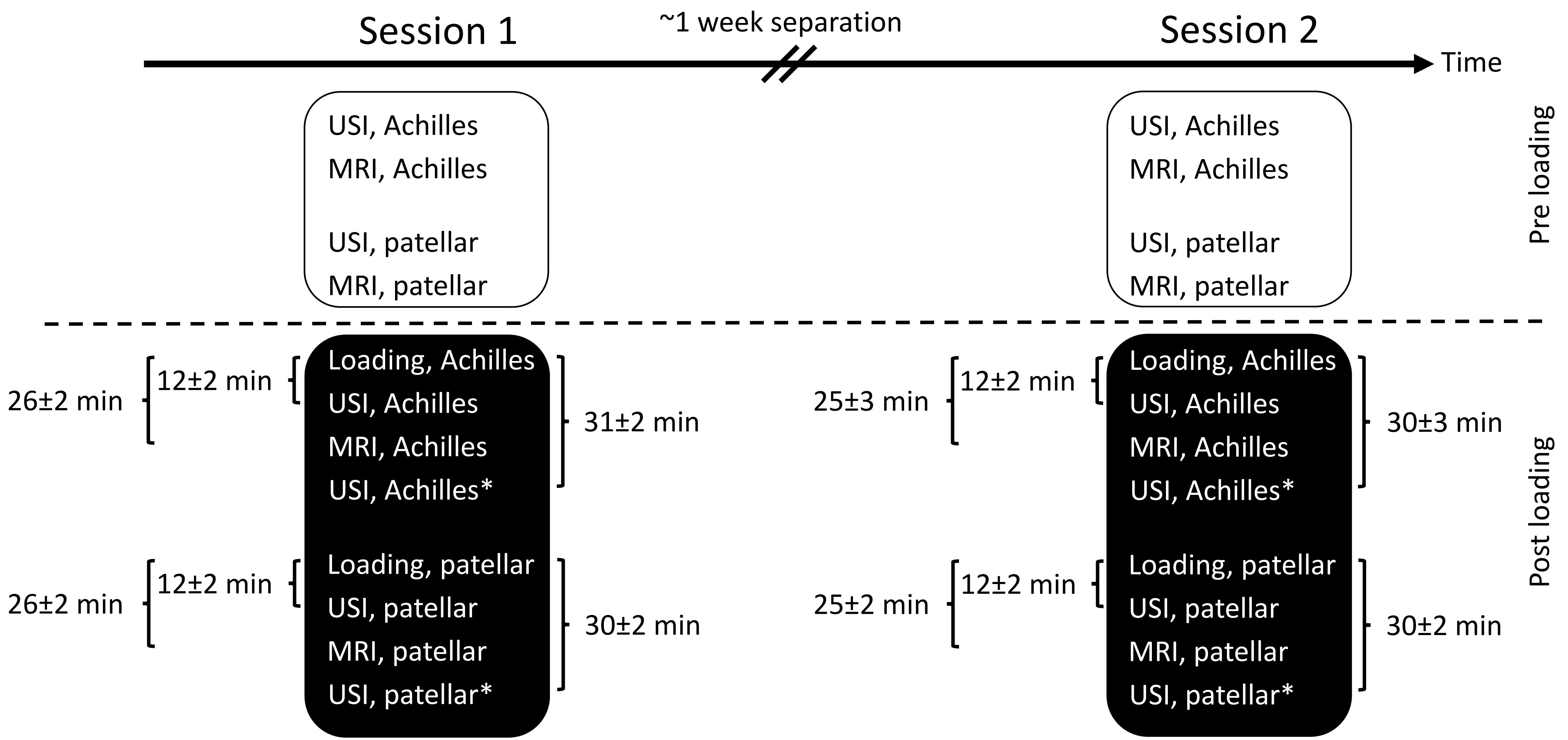

Session 1
~1 week separation
Session 2
Time
USI, Achilles
MRI, Achilles
USI, patellar
MRI, patellar
USI, Achilles
MRI, Achilles
USI, patellar
MRI, patellar
Pre loading
26±2 min
12±2 min
Loading, Achilles
USI, Achilles
MRI, Achilles
USI, Achilles*
31±2 min
25±3 min
12±2 min
Loading, Achilles
USI, Achilles
MRI, Achilles
USI, Achilles*
30±3 min
26±2 min
12±2 min
Loading, patellar
USI, patellar
MRI, patellar
USI, patellar*
30±2 min
25±2 min
12±2 min
Loading, patellar
USI, patellar
MRI, patellar
USI, patellar*
30±2 min
Post loading

Figure 2

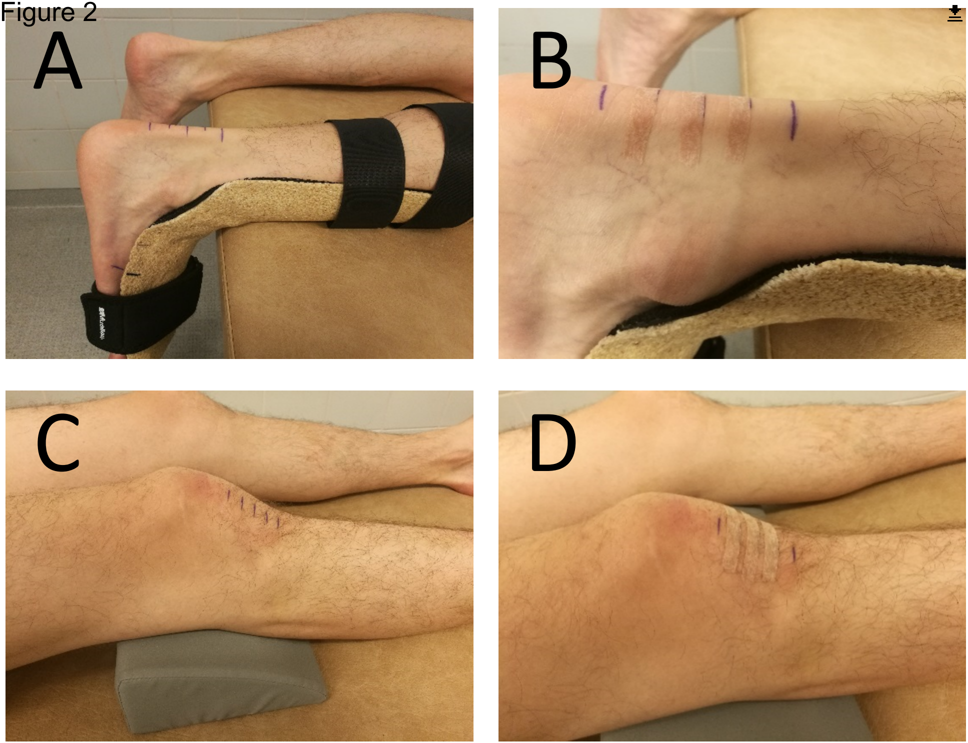
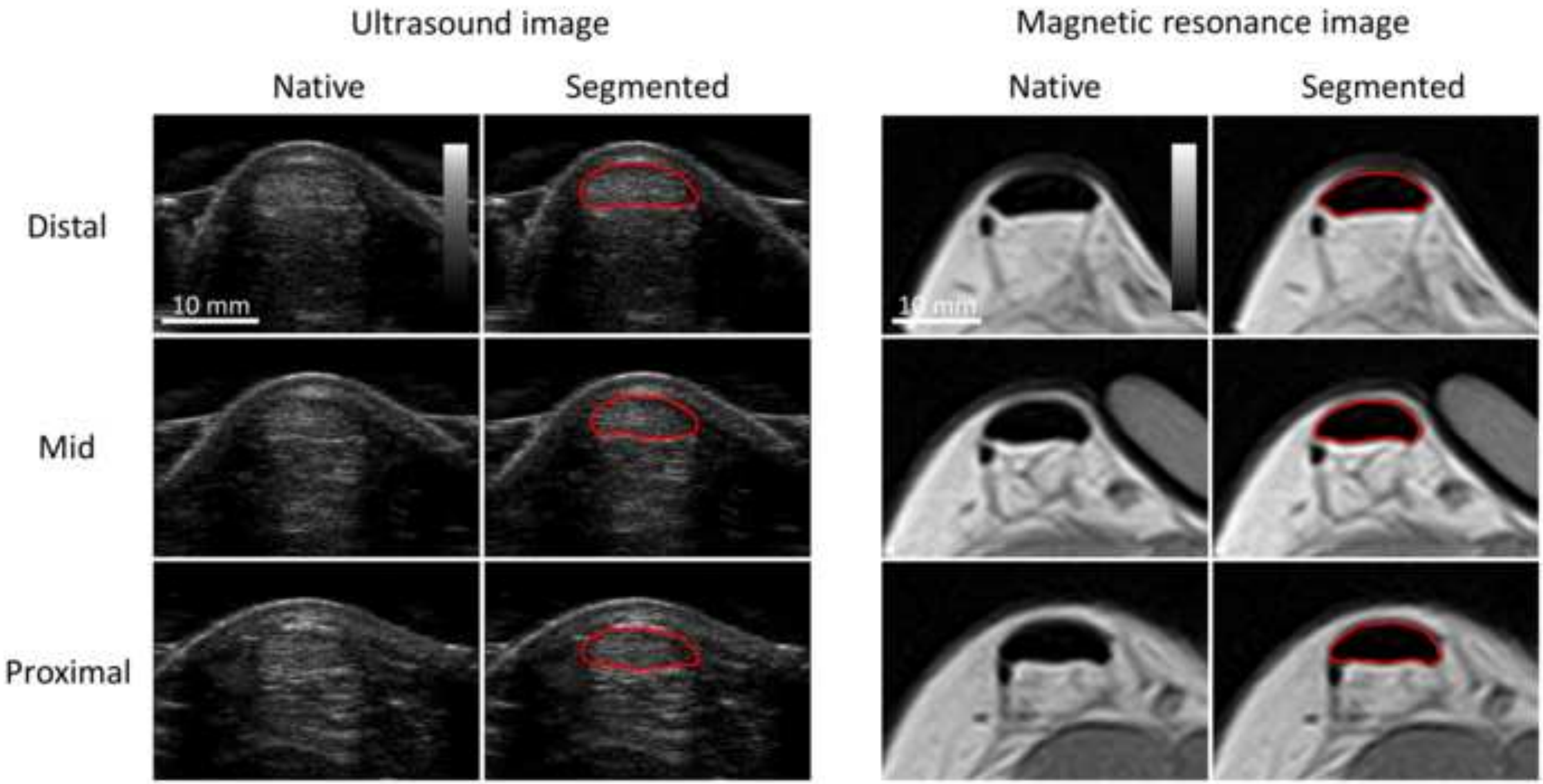


Figure 3



Ultrasound image
Magnetic resonance image
Native
Segmented
Native
Segmented
Distal
Mid
Proximal
10 mm

Figure 4

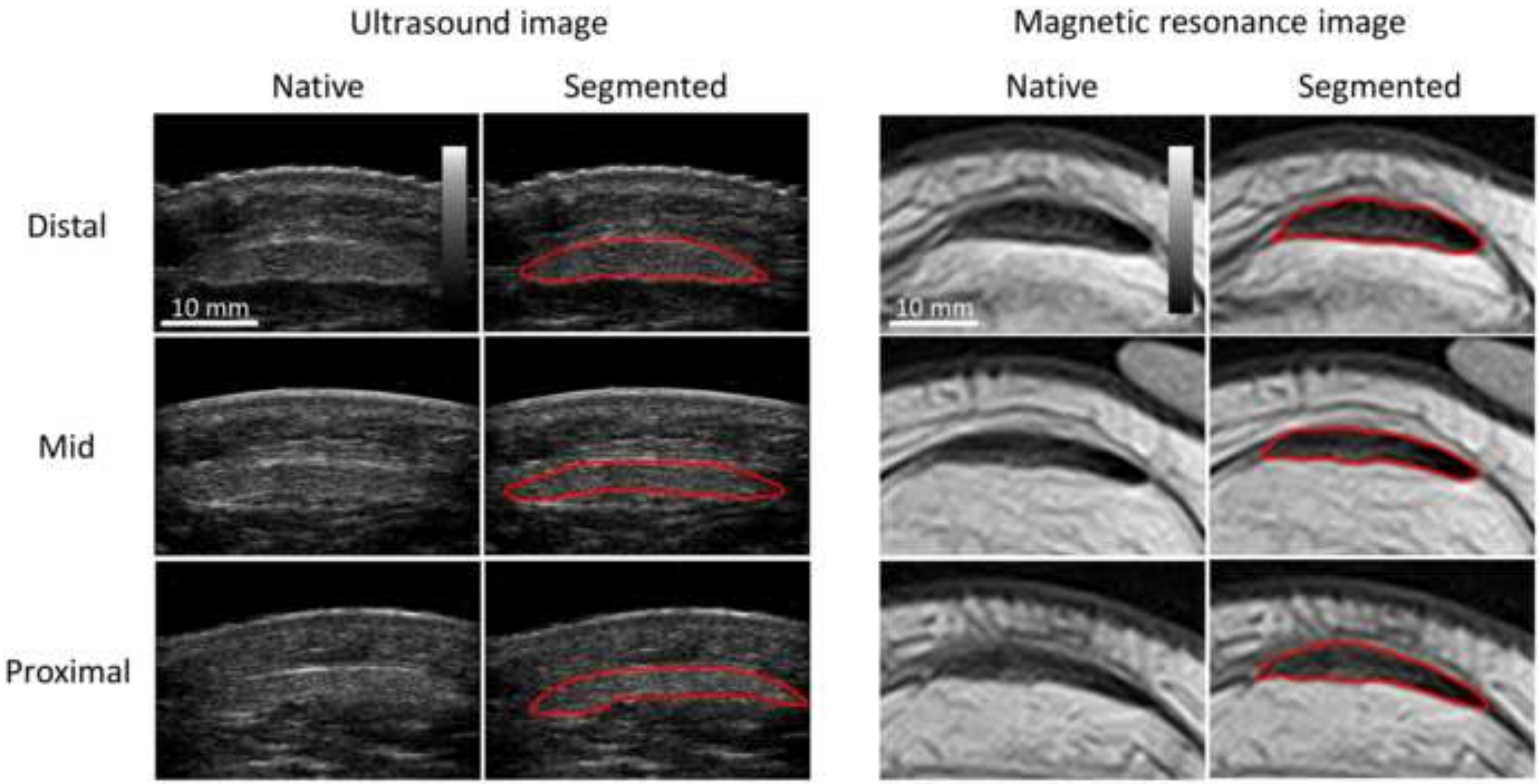
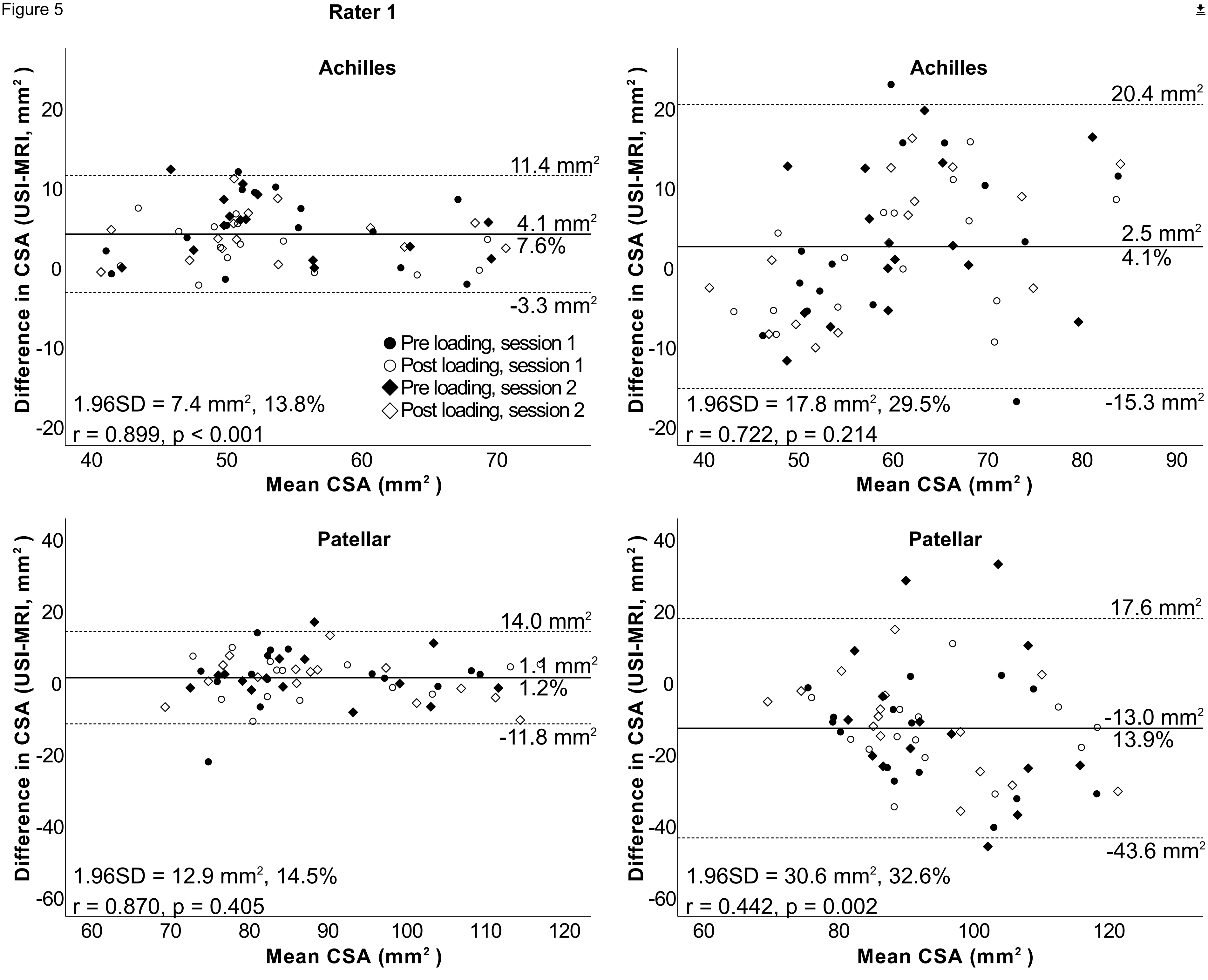

Ultrasound image
Magnetic resonance image
Native
Segmented
Native
Segmented
Distal
Mid
Proximal
10 mm
10 mm

Figure 5



Rater 1
Achilles
11.4 mm²
4.1 mm²
7.6%
-3.3 mm²
Pre loading, session 1
Post loading, session 1
Pre loading, session 2
Post loading, session 2
1.96SD = 7.4 mm², 13.8%
r = 0.899, p < 0.001
Achilles
20.4 mm²
2.5 mm²
4.1%
-15.3 mm²
1.96SD = 17.8 mm², 29.5%
r = 0.722, p = 0.214
Patellar
14.0 mm²
1.1 mm²
1.2%
-11.8 mm²
1.96SD = 12.9 mm², 14.5%
r = 0.870, p = 0.405
Patellar
17.6 mm²
-13.0 mm²
13.9%
-43.6 mm²
1.96SD = 30.6 mm², 32.6%
r = 0.442, p = 0.002
Difference in CSA (USI-MRI, mm²)
Mean CSA (mm²)