Journal of
Foot and Ankle Research

Quantitative ultrasound imaging of Achilles tendon integrity in symptomatic and asymptomatic individuals: reliability and minimal detectable change

Marie-Josée Nadeau[1,2], Amélie Desrochers[1,3], Martin Lamontagne[3], Christian Larivière[1,4] and Dany H. Gagnon[1,2*]

Abstract

Background: Quantifying the integrity of the Achilles tendon (AT) is a rehabilitation challenge. Adopting quantitative ultrasound measurements (QUS measurements) of the AT could fill this gap by 1) evaluating the test-retest reliability and accuracy of QUS measurements of the AT; 2) determining the best protocol for collecting QUS measurements in clinical practice.

Methods: A total of 23 ATs with symptoms of Achilles tendinopathy and 63 asymptomatic ATs were evaluated. Eight images were recorded for each AT (2 visits × 2 evaluators × 2 images). Multiple sets of QUS measurements were taken: geometric (thickness, width, area), first-order statistics (computed from a grayscale histogram distribution: echogenicity, variance, skewness, kurtosis, entropy) and texture features (computed from co-occurrence matrices: contrast, energy, homogeneity). A generalizability study quantified the reliability and standard error of measurement (accuracy) of each QUS measurement, and a decision study identified the best measurement taking protocols.

Results: Geometric QUS measurements demonstrated excellent accuracy and reliability. QUS measurements computed from the grayscale histogram distribution revealed poor accuracy and reliability. QUS measurements derived from co-occurrence matrices showed variable accuracy and moderate to excellent reliability. In clinical practice, using an average of the results of three images collected by a single evaluator during a single visit is recommended.

Conclusions: The use of geometric QUS measurements enables quantification of AT integrity in clinical practice and research settings. More studies on QUS measurements derived from co-occurrence matrices are warranted.

Keywords: Achilles tendon, Computer-assisted image analysis, Measures, Musculoskeletal system, Rehabilitation, Reproducibility, Tendinopathy, Ultrasonography, Quantitative evaluation

Abbreviations: ASYMPT, Asymptomatic; AT, Achilles tendon; CRIR, Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal; CSA, Cross sectional area; E, Evaluator; I, Image; MDC, Minimal detectable change; MDC_{ABS}, Absolute minimal detectable change; MDC_{NORM}, Normalized minimal detectable change; QUS, Quantitative ultrasound; ROI, Region of interest; S, Subject; SEM, Standard error of measurement; SYMPT, Symptomatic; UI, Ultrasound image; V, Visit; VISA-A, Victorian Institute of Sport Assessment- Achilles

* Correspondence: dany.gagnon.2@umontreal.ca
[1]Pathokinesiology Laboratory, Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR) – Institut de réadaptation Lindsay-Gingras de Montréal, Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Centre-Sud-de-l'Île-de-Montréal, Montreal, Canada
[2]School of Rehabilitation, Université de Montréal, Pavillon 7077 Avenue du Parc, Station Centre-Ville, P.O. Box 6128, Montreal, Quebec H3C 3J7, Canada
Full list of author information is available at the end of the article

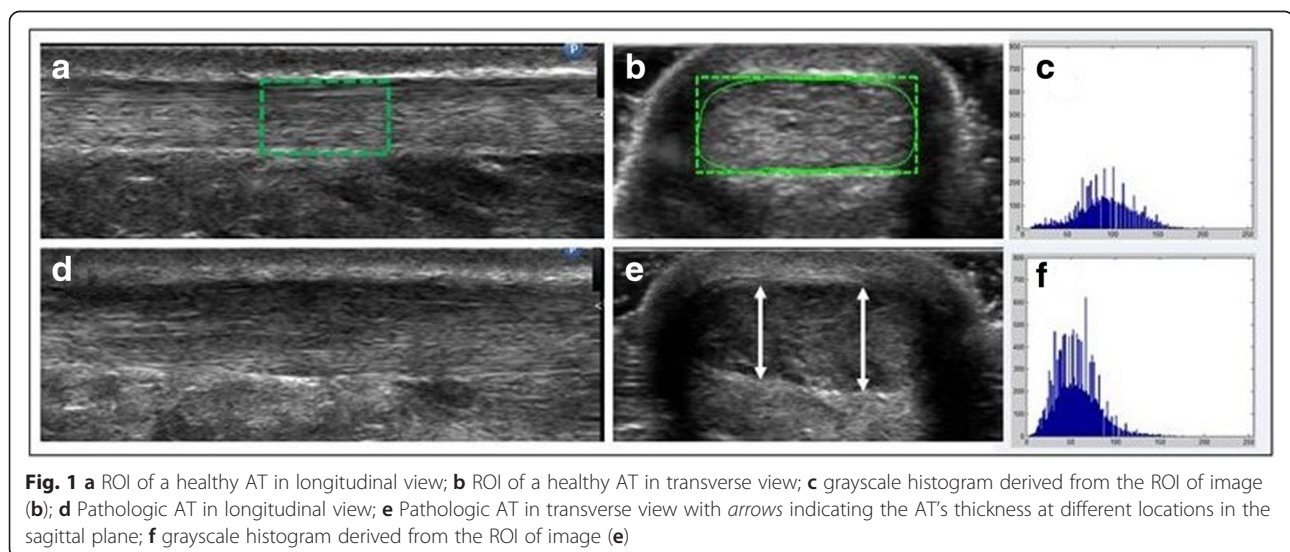Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 2 of 17

Background

The Achilles tendon (AT) is the largest and strongest tendon in the human body. The great tensile loads, occurring predominantly during its elongation or contraction, make it vulnerable to overuse injuries. Although the prevalence and incidence of midsubstance Achilles tendinopathy (i.e., in the middle third of the tendon) are high in athletes, cases are also frequently reported in sedentary individuals [1–6]. The aetiology and pathogenesis of AT tendinopathy have been the subject of much research, but with inconsistent findings [2, 3, 7]. Hence, treating people suffering from this pathology remains challenging for rehabilitation professionals and the success rate of conservative treatments is variable [8–10].

Ultrasound imaging allows in vivo visualization of the biological integrity of the tendon. It is a safe, rapid, non-invasive, relatively inexpensive and popular method used in the assessment of AT tendinopathy [11, 12]. When looking at ultrasound images (UIs) of healthy ATs, well-organized and parallel alignment of the collagen fibres (i.e., fibrillary striation) are highlighted by alternating parallel bright bands (hyperechoic) of collagen and dark bands (hypoechoic) of extracellular matrix [12, 13]. The paratenon of a healthy AT appears as an uninterrupted, well-defined bright line surrounding the tendon [12, 13] (Fig. 1). Conversely, in people with midsubstance AT tendinopathy, the fibrillar striation pattern is often altered as a result of a disorganization of the collagen fibres and a thickened and hypoechoic portion of the AT reflects an increase in the quantity of extracellular matrix and tenocytes [8, 14, 15]. This will typically translate to focal thickening along the AT, presence of dark (hypoechoic) intratendinous regions and sometimes irregular contours of the tendon on UIs [13] (Fig. 1).

Interpretation of an UI of the AT is generally semi-objective. The general appearance of the image is annotated based on the different contrasts observed (e.g., heterogeneous, homogenous, focal or diffuse abnormalities) and the maximum thickness of the AT is often measured using a two-point digital caliper function on the US machine. This interpretation is largely influenced by the evaluator's experience with the recording technique and ability to interpret an UI [16, 17]. Recent technological advances have helped to promote the development of new quantitative ultrasound (QUS) outcome measures extracted from an UI, specifically from a particular region of interest (ROI). Digital UIs can now be broken down into a multitude of micro pixels, and numerical values (e.g., average thickness, tendon width and area) can be measured. The echogenicity of a ROI within an image can also be quantified by allocating a numerical grayscale value to each of those micro pixels [18, 19].

The usefulness of new UI analysis techniques has been demonstrated in various studies on animals and humans [20]. For example, these techniques have helped to quantify changes in the composition of an exercised muscle compared to an unexercised muscle in an elderly population [21–23]. These techniques have also revealed differences in the histological composition of the supra-spinatus muscle and the quadriceps muscle in adults [24] and have been successfully used to detect structural changes in four key muscles in youths with neuromuscular disorders [25]. Moreover, new UI analysis techniques have enabled the differentiation of persons with Achilles tendinopathy from healthy individuals [26, 27] and have been effective in detecting focal and diffuse abnormalities in the AT [28].

Fig. 1 a ROI of a healthy AT in longitudinal view; **b** ROI of a healthy AT in transverse view; **c** grayscale histogram derived from the ROI of image (**b**); **d** Pathologic AT in longitudinal view; **e** Pathologic AT in transverse view with *arrows* indicating the AT's thickness at different locations in the sagittal plane; **f** grayscale histogram derived from the ROI of image (**e**)

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 3 of 17

Very few studies have been conducted to evaluate the reliability of QUS measurements of the AT. This is worrisome considering that the reliability of the QUS measurement of AT thickness, a key diagnostic criterion for Achilles tendinopathy, is rarely reported. To our knowledge, studies that have investigated test-retest reliability of QUS measurements of the AT have shown a moderate to good level of reliability [29–33]. In addition, it was shown that ultrasound image recording is greatly influenced by the evaluator, even among highly experienced ultrasonographers (weak inter-evaluator reliability [34]). Various factors such as the pressure applied on the probe and its alignment can influence recorded image properties and thus alter the quantitative values extracted [35, 36]. Information about the reliability and minimal detectable change is essential in order to develop evidence-based measurement taking protocols, empowering clinicians and researchers to quantify the tendinous changes observed in Achilles tendinopathy and incorporate these findings into clinical practice.

The primary objective of this study was to evaluate the reliability and minimal detectable change (MDC) of AT QUS measurements in people with symptoms consistent with midsubstance Achilles tendinopathy affecting at least one lower limb, as well as in completely asymptomatic individuals. The secondary objective was to recommend the best QUS measurement collection protocol possible, which could be subsequently used to characterize AT integrity in clinical practice or in research projects. It is anticipated that all QUS measurements, when collected by the same evaluator, will be reliable ($\Phi \geq 0.75$) and accurate ($MDC_{NORMALIZED} \leq 15\ \%$) and that a QUS measurement taking protocol in which a single evaluator averages the results of at least three images obtained during a single visit will be recommended in clinical practice.

Methods
Participants
A group of 20 individuals with clinical signs or symptoms of unilateral or bilateral midsubstance Achilles tendinopathy and a group of 23 asymptomatic individuals agreed to take part in this study. Individuals with symptoms consistent with Achilles tendinopathy had to have experienced pain over four weeks, evoked pain on palpation in the middle third of the AT and a VISA-A score below 100. The VISA-A questionnaire [37], completed by all the participants, is a reliable and validated measurement tool with an interest in AT pain, ability to function in daily life and during athletic activities. Eight questions are summed to produce an overall score, which is used as an indicator of the pathology's severity. Scores range from 0 to 100, with a low score indicating greater severity. Asymptomatic participants were to have no pain or previous history of pain in the AT, no observable sign of Achilles tendinopathy

or pain in the ankle, and a VISA-A score equal to 100 [38]. The criteria for the inclusion and exclusion of all participants, as well as each group's specific characteristics, are summarized in Fig. 2. Finally, before conducting any formal testing, ultrasound visualization of the two ATs was performed for each participant to verify its integrity (i.e., normal tendon structure) at and around its insertion and also to rule out complete rupture of the AT. This experiment was approved by the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR) Research Ethics Committee (Certificate: CRIR-557–1110). Participants were fully informed of the nature of the study and asked to sign a consent form before participating.

Clinical examination
Initially, all the participants underwent a clinical examination conducted by an experienced physiotherapist specialized in musculoskeletal disorders with over 10 years of experience. This examination aimed at detecting signs and symptoms typically present with midsubstance Achilles tendinopathy. Special attention was given to the visualization of the AT (without ultrasonography), with emphasis on finding the characteristic thickening sometimes present in its middle third, as well as evoked pain on palpation of the AT's middle third. A series of manoeuvres was carried out to apply passive and active tensions to selected structures with the intention of reproducing the participant's symptoms at the AT: manually resisted contraction of the sural triceps, passive stretch of the sural triceps muscle, repeated unilateral heel rise test and repeated unilateral jump.
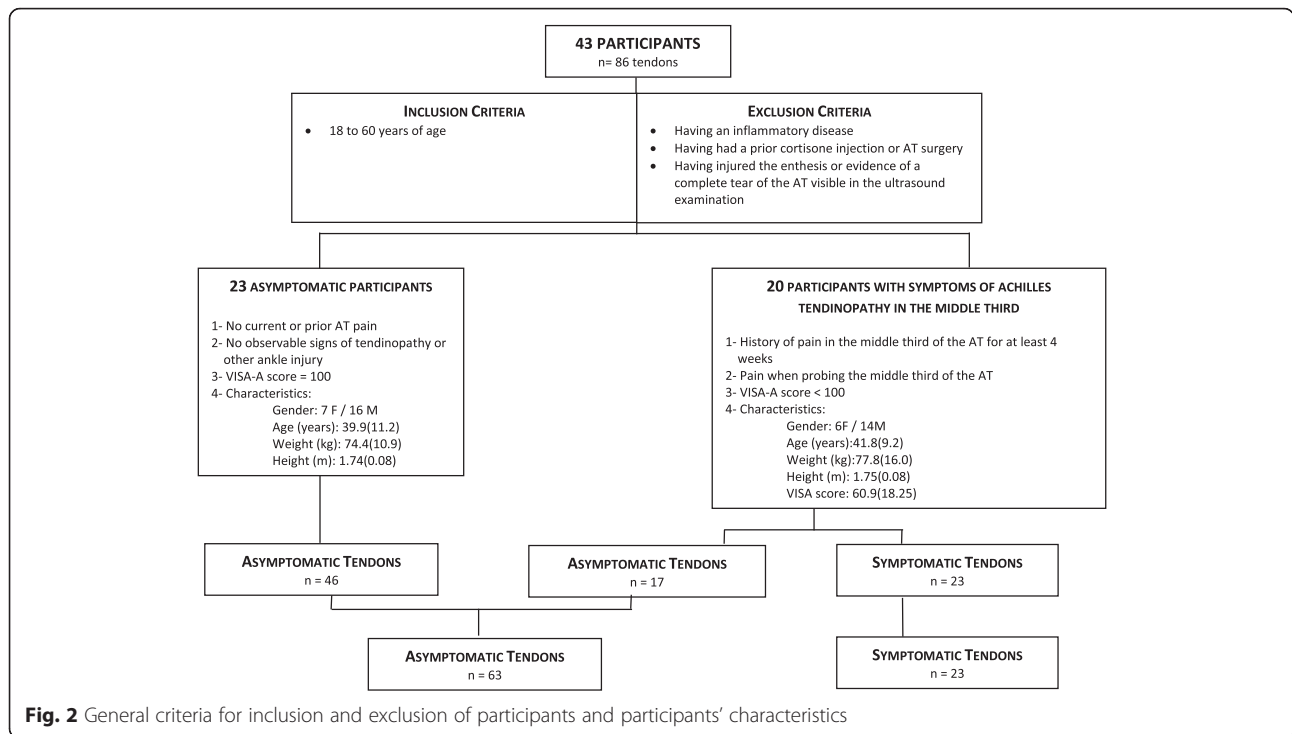
Ultrasound image recording
Device and settings
All of the ultrasound examinations were conducted using a Philips HD11 1.0.6 ultrasound machine with a 5–12 MHz 50 mm linear array transducer (Philips Medical Systems, Bothell, WA). Image field depth was set to 2 cm, gain was set to 85 dB, probe frequency to 12 MHz and a unique focal zone (set at a depth of 0,5 cm) was positioned at the level of the AT. These main machine settings, as well as all the other options (e.g., compress, map, smooth, X-resolution) were maintained across all examinations performed for all participants in order to standardize the recorded images across all participants.

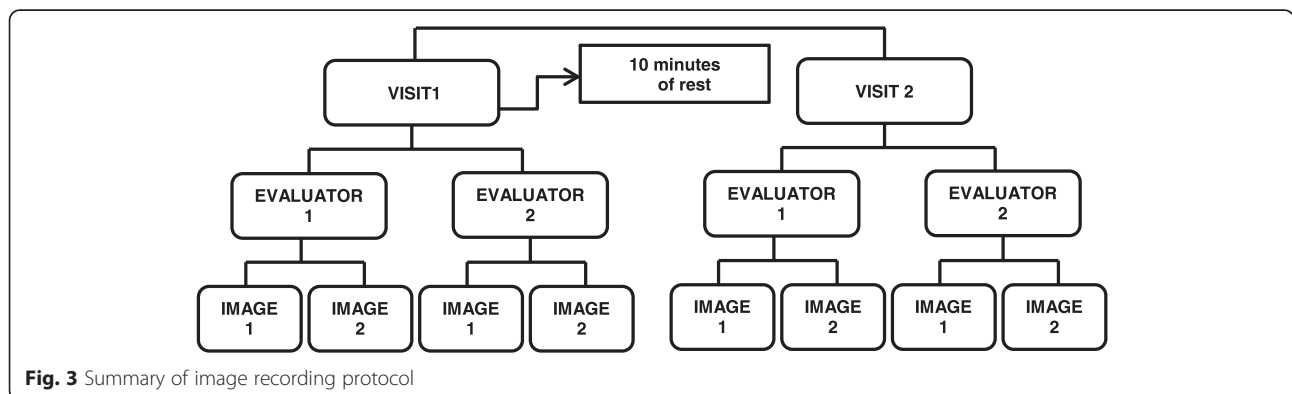#### Ultrasonographers
A physiotherapist (M-J Nadeau) and a resident in physiatry (A. Desrochers) conducted all ultrasound examinations and recorded all of the AT images using a precise protocol (see next section). Both had previously received 10 h of practical training in AT ultrasound examination from an experienced physiatrist recognized by his peers in musculoskeletal ultrasonography (M. Lamontagne).

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 4 of 17

Fig. 2 General criteria for inclusion and exclusion of participants and participants' characteristics

Image recording protocol

A summary of the image recording protocol is illustrated in Fig. 3. *Initial visit (test)*: Each participant was placed in a prone position, with both feet dangling over the end of the table, and ankles positioned at about 5° of plantar flexion using a splint to immobilize the foot. Once placed in this position, the AT's enthesis on the calcaneus was located by ultrasonography, and the skin marked at this location. The enthesis was defined as the most distal point of the insertion of the AT on the calcaneus. Another mark, made at a distance of 6 cm proximal to the enthesis, served as a standardized location for the center of the probe when performing the recording of all the ultrasound images. The images were recorded in this precise location since studies indicate that

the incidence of Achilles tendinopathy is higher at this level (i.e., middle third) [2, 7]. The first evaluator recorded two images of the AT in the longitudinal view, as well as two additional images in the transverse view. During the recording of each of these images, the probe was removed and then repositioned on the skin with the center of the probe continually aligned with the mark on the skin. Once these four images were recorded, the first evaluator erased all the marks drawn on the skin before the second evaluator repeated the same image collection protocol. Particular attention was paid to the probe's positioning on the tendon, taking care to apply minimal pressure on the probe and to align the transducer according to fiber orientation, with respect to the local referential (i.e., x, y, and z axis) defined by the tendon itself.

Fig. 3 Summary of image recording protocol

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 5 of 17

Hence, the transducer was not necessarily perfectly aligned with the traditional anatomical planes and may have deviated slightly from them (i.e., yaw, roll and pitch movements of the transducer).

Second visit (retest) After a minimal 10-min rest period, the evaluators repeated the image collection sequence described above. No significant change in QUS measurements was anticipated as each participant had to remain at rest between the two sessions.

Image analysis

To calculate the different QUS measurements and facilitate characterization of the integrity of the AT, ultrasound images initially recorded in DICOM format were converted to JPEG format. An interactive 2D viewing and image analysis software, developed by the research team using MATLAB Image Processing Toolbox (The Mathworks, Natick, MA), was used to extract the QUS measurements. The development of this program was inspired by work previously realized by a research team based at the University of Pittsburgh that used QUS measurements to characterize shoulder tendons (i.e., supraspinatus, biceps) [39–41]. Each image selected for analysis appeared on screen and the evaluators (i.e., the physiotherapist and a trained research associate) traced a standardized ROI directly on the image, using markers. For blinding purposes, all images recorded during visits 1 and 2 by a unique evaluator were uploaded as a block of images prior to starting the image analysis. Thereafter, each image was presented in a random order to the evaluator to conduct the image analysis. While conducting the image analysis, only the image appeared on the computer screen and all other information was blinded with a black frame generated by the program. As described in detail below, the anatomical landmarks defining the ROIs in the longitudinal and transverse images of the AT differed between both views (Fig. 1a and b).

In the longitudinal view images, the ROI included a 1-cm length area, centered in the middle of the image and captured 6 cm proximal to the tendon enthesis (Fig. 1a). For transverse view images, the ROI was defined by the tendon's contour (Fig. 1b). The ROI outline in both images was established to include the AT's fibres and exclude the paratenon.

These two ROIs were used to extract the following QUS measurements selected for this study: thickness, width (only for transverse images), area (only for transverse images), echogenicity, variance, skewness, kurtosis, entropy, contrast, energy and homogeneity.

Thickness The average thickness of the tendon's ROI is calculated in the longitudinal view. One hundred equidistant points are plotted respectively on the upper and lower edges of the AT and the distance between each pair of points is calculated. The 100 distance measurements are then averaged and represent the thickness. In the transverse view, the thickness of the AT is determined by encompassing the tendon with a rectangle (Fig. 1b). The height of the rectangle reflects the tendon's maximum thickness (Fig. 1e).

Width The width is determined from the rectangle that encompasses the tendon, as defined above. The rectangle's width reflects the tendon's width (Fig. 1b).

Area The tendon's area corresponds to the area of the region delimited by the tendon's outline.

To calculate the other QUS measurements, the ROI is fragmented into multiple micro pixels (micro pixel = 0,0057 mm^2) by the software and a numerical grayscale colour value is allocated to each micro pixel. The grayscale is a scale of colors used in imagery that ranges from 0 = black to white = 255 for a total of 256 possible shades.

The micro pixels' grayscale values included in the ROI are initially represented by a grey level frequency distribution curve found in the ROI (grayscale histogram) (Fig. 1c and f). The following first-order statistics can be calculated from this distribution curve: *echogenicity, variance, skewness, kurtosis* and *entropy*. Additional information on QUS measurements are provided in Table 1.

Next, a co-occurrence matrix is calculated. Texture analysis using a co-occurrence matrix is based on the repeated occurrence of a typical pixel configuration in the image's ROI. It considers how many pairs of pixels with specific grayscale values and a specific predefined spatial relationship (distance and relative orientation angle) are present in an ROI. In this study, pairs of pixels were calculated in four directions (angles = 0°, 45°, 90° 135°) and a distance of 10 pixels was determined. The following texture indicator measurements are derived from this matrix: *contrast, energy* and *homogeneity* (Table 1).

All of the QUS measurements can be classified into three categories: *geometric measurements* (thickness, width, area), *measurements computed from a grayscale histogram* (echogenicity, variance, skewness, kurtosis, entropy) and *measurements computed from a co-occurrence matrix* (contrast, energy, homogeneity).

It is expected that a healthy tendon would have a more heterogeneous appearance because of the alternation of its black and white stripes, with a larger range of values on the grayscale. A pathological area in a tendon would, in contrast, have a darker, more homogeneous appearance, with grayscale values closer to zero (black). Hence, the following QUS measurement values are expected to be found in a pathological tendon: increased thickness, width, area, skewness, kurtosis, homogeneity and energy,

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 6 of 17

Table 1 Definitions of QUS measurements and their mathematical formulas

QUS measurements computed from a grayscale histogram		Mathematical formulas
First order statistics		
Echogenicity Mean ($\bar{x}$)	Mean of grayscale values of micro pixels encompassed within the ROI (from 0 (black) to 255 (white) inclusively).	$\bar{x} = \sum_{x=1}^{M} \sum_{y=1}^{N} \frac{I(x,y)}{MN}$
Variance (σ^2)	Dispersion around the mean of the grayscale values of micro pixels encompassed within the ROI.	$\sigma^2 = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \{I(x,y)-\bar{x}\}^2}{MN}$
Skewness (S_k)	Reflects the asymmetry of the grey level frequency distribution curve around its mean. A high coefficient (in absolute value) translates in a shifted distribution relative to the mean, while a zero coefficient indicates a symmetric distribution. In a positively skewed distribution, pixels intensities are biased toward lower values (shifted distribution to the left). In a negatively skewed distribution, pixels intensities are biased toward higher values (shifted distribution to the right).	$S_k = \frac{1}{MN} \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \{I(x,y)-\bar{x}\}^3}{\sigma^3}$
Kurtosis (K_t)	Reflects the flatness of the grey level frequency distribution curve around its mean. A diffuse distribution will translate in a lower kurtosis value. Distribution concentrated around its mean will translate in a higher kurtosis value.	$K_t = \frac{1}{MN} \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \{I(x,y)-\bar{x}\}^4}{\sigma^4}$
Entropy (E)	Reflects disorder in a ROI. It considers the number of grey levels in a ROI, and the proportions of each grey level. There is an increased entropy when multiple grey level values are present in the ROI. Vice-versa, entropy equals zero if an image has a single grey level value for all its micro pixels.	$E = -\sum_{i=0}^{255} \sum_{j=0}^{255} p(i,j) \log_2(p(i,j))$
QUS Measurements computed from a co-occurence matrix		
Texture parameters		
Contrast (I_{con})	Contrast measures the difference of intensity between the grey level values of neighboring micro pixels. There is a reduced contrast in a constant image with lesser local variations of the grey level intensities. On the contrary, contrast is higher in an image containing a large amount of local sudden variations in the values of grey level intensities.	$I_{con} = \sum_{i=0}^{255} \sum_{j=1}^{255} |i-j|^2 \, p(i,j)$
Energy (I_{eng})	Energy is linked to the regularity and consistency of the patterns in an image. High energy is measured in a constant and steady picture. Vice-versa, low energy is found in an image in which the contacts of grey level values are diverse, uncoordinated and random.	$I_{eng} = \sum_{i=0}^{255} \sum_{j=0}^{255} p(i,j)^2$
Homogeneity (I_{hmg})	Homogeneity is increased in an image with a large number of pixels having the same grey level values, with little grayscale transition (i.e., increased when there is a large area of the same color).	$I_{hmg} = \sum_{i=0}^{255} \sum_{j=0}^{255} \frac{1}{1+(i-j)^2} p(i,j)$

I(x,y) denotes the grayscale intensity at the x,y coordonates in a ROI comprising M rows and N columns. p(i,j) represent the element of a grey-level co-occurrence matrix and denotes the probability that grayscale intensities i and j are adjacent

as well as reduced echogenicity, variance, entropy and contrast [40].

Statistical analysis
Outcome measures
The overall averages, standard deviations and confidence intervals of the results of the 8 images obtained for each QUS measurement in longitudinal and transverse views were calculated for all of the images for all tendons ($n = 86$) and separately for all of the images of symptomatic tendons ($n = 23$) and for all of the images of asymptomatic tendons ($n = 63$). The percentage difference between the averaged results of these two sub-groups was also calculated.

Reliability
The generalizability theory was used to determine the reliability of the different QUS measurements taken for the symptomatic and asymptomatic tendons. On the basis of the analysis of variance, this theory is mainly divided by 2 studies: the generalizability study (G-Study)

and the dependability study (D-Study) [42]. Unlike the traditional theory of reliability that provides a unique random error term, the G-Study divides the error into different facets (sources of variance) relevant to our study and allows for the magnitude of the variance attributed to each facet to be determined. Therefore, in this study, the G-study determined the magnitude of the variance attributed to the subject (S), evaluator (E), visit (V), image (I), and random errors resulting from the interactions between these different sources of variance (SE, SV, SI, EV, EI, VI, SEV, SEI, SVI, EVI), thus leaving much less unexplained variance. In the G-Study, the variance component assigned to the subject (S) represents the difference between the subjects. This proportion of variance is error-free. The unexplained residual error is solely from the interaction between all sources of error and corresponds to the combination of the variances of subjects, evaluators, visits, and images (SEVI). Unlike the traditional theory of reliability that assumes that reliability exists independently of the measurement protocol design, the D-study relies upon information

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 7 of 17

generated from the G-study to determine the reliability of specific simulated protocol designs and provides information to optimize reliability according to, for example, the context in which the measurement is being used (e.g., clinical practice versus research). In this study, the impact of different experimental protocols on the reliability coefficients (Φ), standard error of measurement (SEM) and normalized minimal detectable change (MDC_{NORM}) for each QUS measurement was determined. Since it is documented in studies that inter-rater reliability of QUS measurements is clearly inferior to intra-rater reliability, a single evaluator was used in the D-study. Improvements which may be obtained by averaging 1–3 images during a single visit or by averaging the images obtained during two visits by a single evaluator were compared. The G- and D-studies allow for the calculation of dependability coefficients (Φ), ranging from 0 (no reliability) to 1 (perfect reliability). In general, the dependability coefficient (Φ) can be interpreted as follows: poor reliability ($\Phi < 0.50$), moderate ($0.50 \leq \Phi < 0.75$), good ($0.75 \leq \Phi < 0.90$) and excellent ($\Phi \geq 0.90$). However, there is some consensus that reliability indicators must exceed 0.90 for clinical measurements on an individual basis in order to minimize error and ensure reasonable validity [43]. More liberal reliability scores are allowed on a group basis, particularly for research purposes. The generalizability analysis was conducted using PC GENOVA statistical software, Version 2.2.

Standard error of measurement
Because the dependability coefficient (Φ) can be high despite substantial variability in the measurements, the standard error of measurement (SEM) has also been reported. The absolute SEM is estimated using the same units as the primary outcome measure. The SEM, which is the square root of the error variance, reflects the accuracy of a measurement.

Minimal detectable change
The absolute minimal detectable change (MDC_{ABS}) was calculated to determine the extent of the absolute change required to detect a difference that could be interpreted as a real difference exceeding the measurement error. For a 90 % confidence level ($z = 1.65$), which is considered sufficient for clinical decision-making, the MDC_{ABS} was calculated using the following equation:

$$MDC_{ABS} = 1.65 \times \sqrt{2} \times SEM$$

In order for the MDC to be independent from the unit of measurement and to facilitate its interpretation, the MDC_{ABS} has been subsequently normalized relative to the average obtained (MDC_{NORM}) and calculated using the following equation:

$$MDC_{NORM} = (MDC_{ABS} / \text{overall group average}) \times 100.$$

$MDC_{NORM} \leq 15$ % reflects excellent measurement accuracy.

Results
The overall averages, standard deviations and confidence intervals of the results of the 8 images obtained for each QUS measurement in longitudinal and transverse views for the complete set of tendons ($n = 86$) as well as for the symptomatic tendons ($n = 23$) and asymptomatic tendons ($n = 63$) are summarized in Table 2.

Sources of variance
The magnitude of each variance component (source of error), expressed as a percentage of the total variance for each QUS measurement for symptomatic and asymptomatic tendons, is presented in Tables 3 and 4 for images recorded in longitudinal and transverse views, respectively. Aside from the main source of variance associated with the subject (S) in most cases, the evaluator (E) is the systematic error with the highest variance percentage, up to 13.7 % of the total variance. The other systematic errors (visit and image) are negligible and vary from 0 to 1.9 % of the total variance. A significant proportion of random error is attributable to sources of variance that involve an interaction between the subject and the evaluator (SE, SEV, SEI) with proportions of up to 32.9 %, 40.0 %, and 16.3 % of the total variance, respectively. The contribution of the other errors (SV, SI, EV, EI VI, SVI, EVI) is lower, with percentages ranging from 0 to 10.8 %, where 10.8 % represents SVI interaction. The unexplained residual error (SEVI) is variable (1.5 to 27.3 %) for all of the measurements, with the exception of kurtosis (21.1 to 43.9 %) and skewness (22.7 to 39.2 %), which remains slightly higher.

Reliability and minimal detectable change
The reliability and MDC of different hypothetical QUS measurement acquisition protocol designs are described in Tables 5 and 6 for the images in longitudinal and transverse views, respectively. Different trends in reliability and MDC of the QUS measurements are observed for the three main measurement categories.

The reliability and MDC of the results for the protocol design in which the QUS measurement results of three images taken by a single evaluator in a single visit (E = 1, V = 1, I = 3) are averaged were compared for the three main measurement categories. This measurement scenario is compatible with clinical practice.

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 8 of 17

Table 2 Averages, standard deviations and confidence intervals of the different QUS measurements

Outcome measures (QUS Measurements)										
	All tendons *n* = 86			Symptomatic *n* = 23			Asymptomatic *n* = 63			Difference between groups (%)
	Average	Stand. Dev.	95 % CI[a]	Average	Stand. Dev.	95 % CI[a]	Average	Stand. Dev.	95 % CI[a]	
Longitudinal view										
Geometric										
Thickness (cm)	0.56	0.16	0.53–0.59	0.68	0.21	0.60–0.76	0.52	0.10	0.49–0.54	31.88
Grayscale histogram										
Echogenicity (0–255)	82.37	14.44	79.31–85.42	79.32	14.22	73.51–85.13	83.48	13.32	80.19–86.77	−4.98
Variance	618.72	216.98	572.86–664.58	654.93	213.52	567.66–742.19	605.50	160.62	565.83–645.16	8.16
Skewness	0.52	0.34	0.45–0.59	0.65	0.25	0.55–0.75	0.47	0.21	0.42–0.53	38.13
Kurtosis	3.82	1.28	3.54–4.09	4.18	1.02	3.76–4.59	3.68	0.64	3.52–3.84	13.41
Entropy	6.50	0.24	6.45–6.55	6.51	0.24	6.41–6.61	6.50	0.19	6.46–6.55	0.10
Co –occurrence matrix										
Contrast	0.64	0.13	0.61–0.67	0.61	0.11	0.57–0.66	0.65	0.11	0.62–0.68	−5.67
Energy	0.21	0.05	0.20–0.22	0.21	0.04	0.19–0.22	0.21	0.04	0.20–0.22	−2.76
Homogeneity	0.76	0.03	0.76–0.77	0.77	0.02	0.76–0.77	0.76	0.02	0.75–0.77	0.82
Transverse view										
Geometric										
Area (CSA) (cm²)	0.61	0.23	0.56–0.66	0.76	0.33	0.63–0.90	0.56	0.14	0.52–0.59	36.61
Thickness (cm)	0.57	0.15	0.54–0.60	0.67	0.20	0.59–0.75	0.53	0.09	0.51–0.55	26.73
Width (cm)	1.32	0.16	1.28–1.35	1.37	0.19	1.29–1.44	1.30	0.13	1.27–1.33	5.30
Grayscale histogram										
Echogenicity (0–255)	90.98	13.42	88.14–93.81	86.62	14.85	80.55–92.68	92.57	10.97	89.86–95.28	−6.43
Variance	767.57	207.20	723.77–811.36	789.63	180.44	715.88–863.37	759.51	150.15	722.44–796.59	3.96
Skewness	0.54	0.25	0.49–0.60	0.63	0.20	0.55–0.71	0.51	0.20	0.46–0.56	24.15
Kurtosis	3.61	0.83	3.43–3.78	3.78	0.75	3.48–4.09	3.55	0.58	3.40–3.69	6.68
Entropy	6.74	0.15	6.71–6.77	6.73	0.11	6.68–6.78	6.74	0.12	6.71–6.77	−0.15
Co-occurrence matrix										
Contrast	0.90	0.16	0.86–0.93	0.84	0.16	0.78–0.91	0.92	0.13	0.89–0.95	−8.33
Energy	0.14	0.04	0.13–0.15	0.15	0.03	0.13–0.16	0.14	0.03	0.13–0.14	5.76
Homogeneity	0.73	0.03	0.72–0.73	0.74	0.03	0.72–0.75	0.72	0.02	0.72–0.73	1.95

[a]Confidence intervals (CI) are calculated at a 95 % confidence level

Geometric measurements

In general, these QUS measurements have shown good to excellent reliability, with dependability coefficients ranging from 0.88 to 0.98 and good accuracy with a MDC 90%$_{NORM}$ <15 % obtained in most cases. Only the thickness of symptomatic ATs in longitudinal view had a MDC 90%$_{NORM}$ value greater than 15 % (MDC 90 %$_{NORM}$ = 23.66 %) which still remains acceptable.

Measurements computed from a grayscale histogram

Echogenicity stands out in this category by its excellent results with dependability coefficients ranging from 0.88 to 0.92 and a MDC90%$_{NORM}$ ranging from 8.56 to 15.51 %. The entropy also seems promising with a

tendency for slightly better reliability than other QUS measurements in this category (all Φ =0.77 except for one Φ = 0.34), and an excellent MDC 90%$_{NORM}$ ranging from 2.18 to 4.95 %. However, the other QUS measurements in this category have shown only weak to moderate reliability, with most dependability coefficients below the threshold of 0.75 (Φ range = 0.49–0.79). These measurements also showed a large MDC90%$_{NORM}$ (with the exception of entropy) ranging from 26.09 to 76.02 %.

Measurements computed from a co-occurrence matrix

In general, these QUS measurements have shown moderate to excellent reliability with Φ ranging from Φ = 0.69 to Φ = 0.92 and a variable MDC90%$_{NORM}$ ranging from

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 9 of 17

Table 3 Magnitude of variance components for QUS measurements computed for the images recorded in longitudinal view

QUS measurements	Relative variance proportions for images recorded in longitudinal view (%)														
	S	E	V	I	SE	SV	SI	EV	EI	VI	SEV	SEI	SVI	EVI	SEVI
Geometric															
Thickness															
Sympt	83.0	2.2	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	6.2	4.0	1.7	0.0	1.9
Asympt	86.2	0.8	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	3.3	1.4	0.8	0.0	4.8
Grayscale histogram															
Echogenicity															
Sympt	75.7	0.0	0.0	0.0	7.7	0.0	0.0	0.0	0.0	0.1	8.8	1.2	0.5	0.0	5.8
Asympt	85.2	0.0	0.0	0.0	3.4	1.6	0.0	0.1	0.0	0.0	3.4	0.0	0.0	0.0	6.2
Variance															
Sympt	35.1	1.1	0.0	0.0	32.9	0.0	0.0	0.0	0.0	0.0	12.2	0.0	0.0	0.0	18.7
Asympt	61.7	0.0	0.1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	8.7	0.0	1.3	0.0	27.3
Skewness															
Sympt	21.8	13.2	0.0	1.9	15.0	0.0	0.0	0.2	0.3	1.8	0.0	0.0	7.4	0.0	38.5
Asympt	26.8	7.6	0.0	0.0	12.7	0.0	0.0	0.0	0.4	0.1	7.3	11.0	3.9	0.0	30.2
Kurtosis															
Sympt	10.9	5.8	0.0	0.0	17.4	0.0	0.0	0.0	0.2	2.0	11.0	0.0	8.8	0.0	43.9
Asympt	16.3	7.3	0.0	0.0	19.1	0.5	0.3	0.0	1.1	0.0	9.3	13.4	1.9	0.0	30.8
Entropy															
Sympt	44.0	0.0	0.0	0.0	26.3	0.0	0.0	0.0	0.0	0.0	16.4	1.6	2.1	0.0	9.6
Asympt	58.8	0.2	0.2	0.0	5.7	0.0	0.0	0.0	0.0	0.0	10.2	0.0	5.5	0.0	19.5
Co-occurrence matrix															
Contrast															
Sympt	55.3	0.0	0.0	0.0	8.1	0.0	0.0	0.4	0.5	0.0	12.8	7.0	6.1	0.0	10.0
Asympt	61.3	0.0	0.0	0.0	4.5	0.8	0.0	0.9	0.0	0.2	5.1	0.0	4.2	0.0	22.9
Energy															
Sympt	40.9	0.0	0.0	0.4	19.6	0.0	0.4	0.0	0.0	0.0	17.8	0.9	0.0	0.0	20.0
Asympt	55.3	0.0	0.0	0.0	9.2	0.4	0.0	0.7	0.0	0.0	8.1	0.0	4.4	0.0	22.0
Homogeneity															
Sympt	56.3	0.0	0.0	0.0	6.3	0.0	0.0	0.0	0.0	0.0	12.5	4.7	7.8	0.0	12.5
Asympt	63.9	0.0	0.0	0.0	6.9	1.4	0.0	1.4	0.0	0.0	4.2	0.0	4.2	0.0	18.1

2.36 to 33.67 %. Of these measurements, homogeneity stands out with good to excellent reliability ($\Phi = 0.75$ to $\Phi = 0.92$) and excellent MDC 90%$_{NORM}$ ranging from 2.36 to 3.52 %.

Impact of averaging results from multiple images or visits
D-study reliability and error measurement estimates were computed for 6 experimental designs (Tables 5 and 6). Improved reliability and decreased MDC were obtained in all cases by increasing the number of recorded images. Slightly larger improvements in reliability and MDC were observed by recording images from two visits, in all cases.

Discussion
Evaluator as a significant source of variability
The evaluator represents a significant source of variability in the recording of UIs. The high level of technical skills and manual dexterity during UI recording requires extensive clinical experience and may contribute to variability associated with the evaluator [44]. In the present study, the limited experience of both ultrasonographers (i.e., evaluators) might have increased the variability of the evaluator (E) facet. However, Gellhorn et al. revealed excellent inter-rater reliability for the cross sectional area (CSA) measurement of the patellar tendon (comparable to the AT in terms of shape, content and

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 10 of 17

Table 4 Magnitude of variance components for QUS measurements computed for the images recorded in transverse view

QUS measurements	Relative variance proportions for images recorded in transverse view (%)														
	S	E	V	I	SE	SV	SI	EV	EI	VI	SEV	SEI	SVI	EVI	SEVI
Geometric															
Area (CSA)															
Sympt	92.8	0.0	0.0	0.0	3.0	0.1	0.2	0.0	0.0	0.1	0.9	0.0	0.4	0.0	2.4
Asympt	86.8	0.0	0.0	0.0	3.9	0.0	0.1	0.0	0.0	0.0	1.6	1.3	0.0	0.1	6.3
Thickness															
Sympt	92.6	0.0	0.0	0.0	3.3	0.2	0.1	0.0	0.0	0.0	1.9	0.4	0.0	0.1	1.5
Asympt	74.5	0.0	0.0	0.0	8.7	0.2	0.0	0.0	0.0	0.0	2.1	3.9	0.0	0.0	10.5
Width															
Sympt	80.4	0.0	0.0	0.0	2.0	0.0	0.2	0.0	0.0	0.4	2.1	0.0	0.0	0.0	15.0
Asympt	71.0	2.3	0.1	0.0	7.3	0.0	0.0	0.0	0.1	0.0	6.8	1.6	1.6	0.1	9.2
Grayscale histogram															
Echogenicity															
Sympt	81.3	4.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	6.1	1.5	1.5	0.4	4.6
Asympt	73.2	5.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	2.3	0.0	1.6	0.0	13.9
Variance															
Sympt	31.2	13.7	0.0	0.0	9.4	0.0	0.0	0.0	0.1	0.8	26.3	5.1	4.5	0.2	8.6
Asympt	49.9	10.1	0.0	0.0	6.7	1.4	0.0	0.0	0.0	0.0	10.7	2.8	2.0	0.0	16.5
Skewness															
Sympt	42.9	7.5	0.0	0.0	1.0	7.2	0.0	0.0	0.0	0.0	2.2	0.0	0.0	0.0	39.2
Asympt	58.9	0.0	0.0	0.0	0.8	1.0	0.1	0.0	0.2	0.0	9.0	7.0	0.0	0.3	22.7
Kurtosis															
Sympt	37.8	0.3	0.5	0.0	0.0	6.7	0.0	0.0	0.3	0.0	13.5	6.9	0.0	0.0	34.1
Asympt	44.6	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.8	16.3	4.7	0.8	21.1
Entropy															
Sympt	25.8	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.7	2.1	40.0	5.9	10.8	0.0	10.5
Asympt	48.8	3.6	0.0	0.0	12.3	4.2	3.6	0.0	0.2	0.2	5.7	1.8	0.0	0.0	19.5
Co-occurrence matrix															
Contrast															
Sympt	73.8	6.8	0.0	0.0	1.3	0.0	1.4	0.0	0.0	0.0	8.3	0.1	0.0	0.4	7.8
Asympt	59.6	2.2	0.0	0.6	4.0	3.0	0.0	0.0	0.0	0.6	5.2	3.3	2.5	0.0	19.0
Energy															
Sympt	58.2	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	21.8	3.0	2.4	0.6	7.3
Asympt	68.9	4.9	0.0	0.0	4.9	2.5	0.8	0.0	0.0	0.0	3.3	0.0	0.0	0.0	14.8
Homogeneity															
Sympt	80.8	6.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	1.0	8.1
Asympt	67.2	4.9	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	1.6	1.6	1.6	0.0	18.0

superficial location) between a novice and an experienced sonographer [45]. The important interplay between the evaluator and the participants during an UI recording is highlighted by the high proportion of variance in all QUS measurements attributed to sources of error involving interaction between the subject and the evaluator (SE, SEV, SEI). Every AT is different and

those anatomical and physiological differences (oblique orientation of the tendon, the subject's weak natural echogenicity, blurred outline of the tendon, etc.) are expressed by in the subject facet (S). These dissimilarities make capturing an image of certain tendons more challenging than others and may explain why an evaluator may have more difficulty in assessing some subjects than

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 11 of 17

Table 5 Impact of different hypothetical protocols (D-Study) for QUS measurements recorded in longitudinal view

QUS measurements in longitudinal view			E = 1					
			V = 1			V = 2		
			I = 1	I = 2	I = 3	I = 1	I = 2	I = 3
Geometric								
Thickness	SYMPT	Φ	0.85	0.88	0.90	0.90	0.93	0.94
		SEM	0.09	0.07	0.07	0.07	0.06	0.05
		MDC90%$_{NORM}$	29.21	25.17	23.66	23.28	19.35	17.84
	ASYMPT	Φ	0.90	0.93	0.94	0.94	0.96	0.97
		SEM	0.03	0.03	0.02	0.02	0.02	0.02
		MDC90%$_{NORM}$	14.52	11.80	10.74	10.94	8.76	7.90
Grayscale histogram								
Echogenicity	SYMPT	Φ	0.83	0.87	0.88	0.90	0.93	0.93
		SEM	6.35	5.56	5.27	4.65	4.03	3.79
		MDC90%$_{NORM}$	18.68	16.36	15.51	13.69	11.84	11.16
	ASYMPT	Φ	0.89	0.91	0.92	0.94	0.96	0.96
		SEM	4.76	4.05	3.78	3.37	2.87	2.68
		MDC90%$_{NORM}$	13.31	11.32	10.57	9.43	8.01	7.48
Variance	SYMPT	Φ	0.69	0.76	0.79	0.81	0.86	0.88
		SEM	157.27	131.36	121.51	111.20	92.89	85.92
		MDC90%$_{NORM}$	56.03	46.80	43.29	39.62	33.09	30.61
	ASYMPT	Φ	0.63	0.73	0.77	0.77	0.84	0.87
		SEM	118.97	93.45	83.22	84.12	66.08	58.85
		MDC90%$_{NORM}$	45.85	36.01	32.07	32.42	25.47	22.68
Skewness	SYMPT	Φ	0.42	0.59	0.69	0.58	0.74	0.81
		SEM	0.29	0.21	0.17	0.21	0.15	0.12
		MDC90%$_{NORM}$	104.67	74.18	60.70	75.60	53.58	43.84
	ASYMPT	Φ	0.43	0.57	0.64	0.55	0.69	0.75
		SEM	0.24	0.18	0.15	0.18	0.14	0.12
		MDC90%$_{NORM}$	116.65	87.95	76.02	90.96	67.84	58.13
Kurtosis	SYMPT	Φ	0.30	0.42	0.49	0.46	0.59	0.66
		SEM	1.50	1.15	1.00	1.06	0.81	0.71
		MDC90%$_{NORM}$	83.94	64.13	55.99	59.44	45.40	39.63
	ASYMPT	Φ	0.38	0.51	0.58	0.50	0.63	0.70
		SEM	0.81	0.62	0.54	0.64	0.48	0.42
		MDC90%$_{NORM}$	51.31	39.27	34.34	40.69	30.67	26.51
Entropy	SYMPT	Φ	0.70	0.75	0.77	0.82	0.85	0.87
		SEM	0.17	0.15	0.14	0.12	0.10	0.10
		MDC90%$_{NORM}$	5.92	5.21	4.95	4.30	3.75	3.55
	ASYMPT	Φ	0.65	0.74	0.77	0.78	0.85	0.87
		SEM	0.13	0.11	0.10	0.09	0.08	0.07
		MDC90%$_{NORM}$	4.79	3.85	3.48	3.39	2.72	2.46
Co-occurrence matrix								
Contrast	SYMPT	Φ	0.63	0.72	0.75	0.74	0.82	0.84
		SEM	0.08	0.07	0.06	0.06	0.05	0.05
		MDC90%$_{NORM}$	30.57	25.20	23.14	23.72	19.11	17.30

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 12 of 17

Table 5 Impact of different hypothetical protocols (D-Study) for QUS measurements recorded in longitudinal view *(Continued)*

	ASYMPT	Φ	0.66	0.76	0.80	0.79	0.87	0.89
		SEM	0.08	0.06	0.05	0.06	0.04	0.04
		MDC90%$_{NORM}$	28.23	21.88	19.30	19.96	15.47	13.65
Energy	SYMPT	Φ	0.60	0.68	0.71	0.75	0.80	0.83
		SEM	0.03	0.03	0.02	0.02	0.02	0.02
		MDC90%$_{NORM}$	33.74	28.73	26.85	24.39	20.63	19.21
	ASYMPT	Φ	0.64	0.74	0.78	0.78	0.85	0.88
		SEM	0.03	0.02	0.02	0.02	0.02	0.02
		MDC90%$_{NORM}$	34.26	27.17	24.35	24.22	19.21	17.22
Homogeneity	SYMPT	Φ	0.63	0.71	0.75	0.75	0.82	0.85
		SEM	0.02	0.01	0.01	0.01	0.01	0.01
		MDC90%$_{NORM}$	4.72	3.85	3.52	3.54	2.85	2.58
	ASYMPT	Φ	0.71	0.80	0.83	0.83	0.89	0.91
		SEM	0.01	0.01	0.01	0.01	0.01	0.01
		MDC90%$_{NORM}$	4.45	3.50	3.12	3.15	2.48	2.21

others (SE). Consequently, it is recommended that a single evaluator record US images and extract QUS measurements, particularly when the goal is to monitor treatment effects over time.

Superiority of geometric measurements and echogenicity

The excellent results obtained in this study, in terms of reliability and accuracy, of geometric QUS measurements of area, thickness and width (Φ obtained mostly at the top of the clinical acceptable threshold of 0.90 and a MDC90%$_{NORM}$ <15 %) are similar or better than those obtained in comparable studies targeting the AT [31–33, 46]. Although echogenicity is a measurement computed from a grayscale histogram, it behaves as a geometric measurement and has also shown excellent reliability and accuracy (mostly all Φ > 0.90 and MDC$_{NORM}$ < 15 %). Echogenicity has been previously studied, mainly on muscles, and has shown good reliability for repeated measurements (variation coefficients ranging from 5 to 11 %) [47–49]. Strict compliance of a standardized measurement taking protocol and the use of software to extract the geometric QUS measurements may explain, among other reasons, the favourable results obtained in this study. Continued use of the geometric QUS measurements, as well as echogenicity, is therefore encouraged in quantifying AT integrity.

The poor results, in terms of reliability and MDC, of the QUS measurements computed from a grayscale histogram (variance, skewness, kurtosis) obtained in this study confirm the need for refinement and further study before advocating the use of these QUS measurements in the assessment of AT integrity. For a hypothetical protocol in which the evaluator averages the results of three images recorded during a single visit (E = 1, V = 1, I = 3), the

dependability coefficients obtained for variance, skewness and kurtosis range from 0.49 to 0.79, with the majority of measurements falling under the threshold established to ensure good reliability of 0.75. The accuracy results are also disappointing (MDC90%$_{NORM}$ ranging from 26.09 to 76.02 %). Only entropy stands out for its general good reliability (all Φ =0.77 except for one Φ = 0.34) and high accuracy (MDC 90%$_{NORM}$ ranging from 2.18 to 4.95 %) in this study when using the above-described protocol.

The reliability of QUS measurements computed from a grayscale histogram was also studied previously by two research teams that obtained similar results to those of the present study. Collinger et al. assessed the reliability and accuracy of various QUS measurements extracted from longitudinal images of the long head of the biceps and the supraspinatus tendons based on the generalizability theory [39]. The reliability and accuracy scores, determined with a D-study for an E = 1, V = 1, I = 2 protocol, are similar to the present study. Good reliability and accuracy of thickness (Φ ranging from 0.92 to 0.94; MDC90%$_{NORM}$ ranging from 9.42 to 14.49 %) and echogenicity (Φ ranging from 0.79 to 0.85; MDC90%$_{NORM}$ ranging from 16.03 to 18.72 %), as well as low reliability and weak accuracy of the variance, skewness and kurtosis have been found (Φ ranging from 0.57 to 0.69; MDC90% greater than 15 % and up to 297.35 %). Entropy demonstrated moderate reliability and good accuracy (Φ ranging from 0.64 to 0.68; MDC90% ranging from 5.56 to 5.58 %). Slightly better results in terms of reliability of the QUS measurements were obtained in the present study compared to the study by Collinger. The superior reliability of the AT QUS measurements is possibly explained by the fact that this tendon is easier to assess than shoulder tendons due to its superficial position, alignment

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 13 of 17

Table 6 Impact of different hypothetical protocols (D-Study) for QUS measurements recorded in transverse view

QUS measurements in transverse view			E = 1					
			V = 1			V = 2		
			I = 1	I = 2	I = 3	I = 1	I = 2	I = 3
Geometric								
Area (CSA)	SYMPT	Φ	0.96	0.97	0.98	0.98	0.99	0.99
		SEM	0.07	0.05	0.05	0.05	0.04	0.03
		MDC90%$_{NORM}$	20.91	16.58	14.85	15.18	11.97	10.69
	ASYMPT	Φ	0.91	0.94	0.96	0.94	0.97	0.98
		SEM	0.05	0.03	0.03	0.03	0.03	0.02
		MDC90%$_{NORM}$	18.96	14.50	12.67	14.38	10.89	9.45
Thickness	SYMPT	Φ	0.96	0.97	0.97	0.98	0.98	0.99
		SEM	0.04	0.04	0.03	0.03	0.03	0.03
		MDC90%$_{NORM}$	14.79	12.85	12.13	11.02	9.41	8.81
	ASYMPT	Φ	0.83	0.90	0.92	0.89	0.94	0.95
		SEM	0.04	0.03	0.03	0.03	0.02	0.02
		MDC90%$_{NORM}$	17.46	13.18	11.41	13.72	10.24	8.78
Width	SYMPT	Φ	0.82	0.89	0.92	0.90	0.94	0.96
		SEM	0.09	0.06	0.06	0.06	0.05	0.04
		MDC90%$_{NORM}$	14.69	10.98	9.42	10.44	7.80	6.69
	ASYMPT	Φ	0.80	0.86	0.88	0.88	0.92	0.93
		SEM	0.06	0.05	0.05	0.05	0.04	0.04
		MDC90%$_{NORM}$	11.61	9.55	8.75	8.56	6.97	6.35
Grayscale histogram								
Echogenicity	SYMPT	Φ	0.85	0.89	0.90	0.91	0.94	0.95
		SEM	6.11	5.16	4.81	4.55	3.79	3.50
		MDC90%$_{NORM}$	16.47	13.91	12.95	12.25	10.20	9.42
	ASYMPT	Φ	0.81	0.88	0.91	0.90	0.94	0.95
		SEM	5.25	3.94	3.39	3.71	2.79	2.40
		MDC90%$_{NORM}$	13.23	9.93	8.56	9.35	7.02	6.05
Variance	SYMPT	Φ	0.47	0.53	0.55	0.61	0.68	0.70
		SEM	179.83	159.58	152.23	134.27	116.89	110.49
		MDC90%$_{NORM}$	53.14	47.16	44.99	39.68	34.54	32.65
	ASYMPT	Φ	0.63	0.71	0.75	0.76	0.82	0.85
		SEM	112.00	92.43	84.91	82.42	67.32	61.47
		MDC90%$_{NORM}$	34.41	28.40	26.09	25.32	20.68	18.89
Skewness	SYMPT	Φ	0.47	0.60	0.66	0.64	0.75	0.80
		SEM	0.19	0.15	0.13	0.13	0.10	0.09
		MDC90%$_{NORM}$	69.88	53.98	47.51	49.41	38.17	33.59
	ASYMPT	Φ	0.60	0.70	0.75	0.72	0.81	0.84
		SEM	0.15	0.12	0.11	0.12	0.09	0.08
		MDC90%$_{NORM}$	69.85	55.19	49.35	53.67	41.75	36.94
Kurtosis	SYMPT	Φ	0.38	0.48	0.52	0.52	0.63	0.67
		SEM	0.83	0.68	0.62	0.62	0.50	0.45
		MDC90%$_{NORM}$	51.01	41.66	38.04	38.09	30.71	27.81

Table 6 Impact of different hypothetical protocols (D-Study) for QUS measurements recorded in transverse view *(Continued)*

	ASYMPT	Φ	0.45	0.57	0.63	0.56	0.68	0.74
		SEM	0.58	0.45	0.40	0.47	0.36	0.31
		MDC90%$_{NORM}$	38.08	29.68	26.29	30.68	23.42	20.44
Entropy	SYMPT	Φ	0.27	0.32	0.34	0.40	0.47	0.50
		SEM	0.15	0.13	0.13	0.11	0.10	0.09
		MDC90%$_{NORM}$	5.17	4.58	4.37	3.82	3.34	3.16
	ASYMPT	Φ	0.63	0.73	0.77	0.75	0.83	0.86
		SEM	0.09	0.07	0.06	0.07	0.05	0.05
		MDC90%$_{NORM}$	3.02	2.42	2.18	2.30	1.81	1.62
Co-occurrence matrix								
Contrast	SYMPT	Φ	0.81	0.85	0.87	0.88	0.92	0.93
		SEM	0.08	0.07	0.06	0.06	0.05	0.04
		MDC90%$_{NORM}$	21.25	18.16	17.01	15.64	13.20	12.29
	ASYMPT	Φ	0.65	0.75	0.79	0.77	0.85	0.88
		SEM	0.09	0.07	0.06	0.07	0.05	0.05
		MDC90%$_{NORM}$	22.63	17.82	15.90	16.89	13.17	11.67
Energy	SYMPT	Φ	0.62	0.67	0.69	0.75	0.79	0.81
		SEM	0.02	0.02	0.02	0.02	0.02	0.02
		MDC90%$_{NORM}$	39.31	35.16	33.67	28.93	25.50	24.25
	ASYMPT	Φ	0.78	0.85	0.87	0.87	0.91	0.93
		SEM	0.02	0.01	0.01	0.01	0.01	0.01
		MDC90%$_{NORM}$	27.37	21.80	19.60	19.72	15.65	14.03
Homogeneity	SYMPT	Φ	0.86	0.90	0.92	0.92	0.95	0.96
		SEM	0.01	0.01	0.01	0.01	0.01	0.01
		MDC90%$_{NORM}$	3.61	2.92	2.65	2.55	2.07	1.87
	ASYMPT	Φ	0.76	0.85	0.89	0.85	0.92	0.94
		SEM	0.01	0.01	0.01	0.01	0.01	0.01
		MDC90%$_{NORM}$	3.82	2.80	2.36	2.80	2.04	1.72

and surrounding structures. Nielsen et al. assessed the test-retest reliability for repeated measurements of the QUS measurements of variance, skewness and kurtosis of the supraspinatus muscle. A single evaluator repeated 30 images during a single visit (E = 1, V = 1, I = 30) [49]. This team also obtained results consistent with those in this study, that is, low reliability in these QUS measurements, with variation coefficients (analogous to SEM expressed as a percentage of the grand mean) ranging from 14 to 35 %.

Knowledge of the theoretical foundations and the underlying calculation of the different QUS measurements computed from a grayscale histogram is essential in understanding the disappointing results, in terms of reliability and accuracy, of these measurements. The skewness and kurtosis QUS measurements are calculated directly from the shape of the grey level frequency distribution curve (grayscale histogram), while the shape of this curve also influences variance and entropy. For

its part, the QUS measurement of echogenicity is an average of the grey scale values for all of the pixels in the ROI and does not take into account the shape of the distribution curve. The appearance of an anatomical structure on an ultrasound image can vary significantly depending upon the angle and the pressure applied to the tissues with the probe, both in terms of shape and echogenicity [50]. The edges of the ROI are sensitive to these variations in echogenicity, which can change the shape of the grey level frequency distribution curve without having a significant influence on the echogenicity's average value for the ROI.

Differences between transverse and longitudinal AT images

In this study, reliability and MDC of QUS measurements are generally similar between transverse and longitudinal images. The QUS measurement of thickness in a longitudinal view of symptomatic ATs is the exception to the rule

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 15 of 17

as it tends to be less reliable (−8.58 %) and less accurate (−48.74 %) than in a transverse view. This difference might be explained by the fact that when the probe is positioned longitudinally to the AT fibres, it can be repositioned at different locations or angles on the tendon's sagittal plane for each image (Fig. 1e). In a longitudinal view, the thickness measurement is captured only for a slice located directly under the probe and it is difficult to ensure that it is located on the thickest portion of the tendon. Therefore, it appears preferable that AT QUS thickness measurements are also taken in the transverse view, at a location considered relevant and determined following a full excursion of the transducer along the AT in both planes. Other thickness measures previously reported in the literature (e.g., true thickness measure) may also deserve to be explored in the future, especially with regard to the thickness measured in the transverse view [32]. In addition, since our QUS thickness measurement in the longitudinal view reflects the average thickness of a targeted area, it is likely that its value is reduced in comparison to the maximum thickness found in this region captured in the transverse view of the AT.

Co-occurrence matrix shows promise in quantitative ultrasound imaging

The co-occurrence matrix is an image analysis method that considers the spatial organization of the pixels, as opposed to the grayscale histogram that only considers the grey scale values of the pixels, without taking into account their position on the image or their interaction with the surrounding pixels [18, 19, 51]. In our study, better reliability was achieved for QUS measurements drawn from a co-occurrence matrix in comparison to the reliability of the QUS measurements computed from a grayscale histogram. Collinger et al. obtained similar results [39]. Superior reliability may be explained by the fact that the co-occurrence matrix studies pairs of pixels, and not the isolated value of each pixel's grey level.

Proposing a measurement collection protocol for clinical practice

In clinical practice, it is difficult to consider having more than one assessment visit in which additional images would be recorded (V = 2). Even though one or more additional visits positively influences the reliability and accuracy of QUS measurements, productivity constraints should be considered. A protocol in which the averaged result obtained from three images collected by a unique evaluator during a single visit seems to represent a good compromise. The clinical applicability of AT QUS measurements becomes highly relevant since the time required for recording them is, in the latter protocol, at most 10 min, which can help in the clinical decision making process and in practice.

Study limitations

This study has several limitations. Within the context of this reliability study, all measurements were taken in an identical location across all participants according to a standardized protocol. The location selected was set at 6 cm proximal to the enthesis of the AT considering that AT tendinopathy typically occurs between 2–6 cm proximal to its enthesis on the calcaneus [52, 53]. Hence, measurements of the symptomatic tendons were not necessarily done exactly at the pathology's precise location for all participants which, in turn, may have minimized the variance between the pathological tendons and underestimated the reliability and accuracy of the measurements. Two separate tasks, which are both dependent upon the evaluator, must be performed when obtaining QUS measurements. The first task involves obtaining the ultrasound image (image acquiring) and the second consists of processing the acquired image in order to extract the desired quantitative values (image analyzing). Both "image acquiring" and "image analyzing" can affect reliability separately. For example, when the evaluator plots the delineation of the ROI while analyzing the UIs, the sometimes-blurred outline of the AT increases this task's difficulty. Syha et al. found that reliability of thickness measurements of the AT was more reliable when the ROI was traced automatically compared to manual tracing [54, 55]. In the present study, it is impossible to differentiate the error related to the recording of the image from that of the analysis of the image. Further studies are required to isolate these potential sources of variability that are currently encompassed within the evaluator facet. Another source of variability and potential limitation of the study is that we are quantifying the integrity of a three dimensional tendon using two dimensional UIs.

Conclusions

This study focused on the reliability of three types of QUS measurements: geometric QUS measurements, QUS measurements computed from a grayscale histogram, and QUS measurements computed from a co-occurrence matrix. Even though additional validity and responsiveness studies are necessary, the favourable results of geometric QUS measurements and of the echogenicity further support their use in clinical practice and research protocols. These measurements could be used in longitudinal follow-up to capture the progress of the severity of AT tendinopathy and the impact of follow-up treatment in clinical practice or in rehabilitation research protocols. These measurements may also be useful in a transversal context in order to compare individuals between themselves or against standards established in clinical practice or in future studies. Furthermore, it is imperative that particular emphasis be given to adhering to a rigorous, standardized measurement-taking protocol

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 16 of 17

when acquiring UIs to reach an acceptable level of reliability and accuracy. With respect to QUS measurements computed from grayscale histograms, in light of the results obtained in this study, the use of these QUS measurements in evaluating AT integrity should be reconsidered. Lastly, QUS measurements computed from a co-occurrence matrix are promising and additional studies on this emerging method of image analysis are necessary.

Availability of data and material
All data and materials can be made available from the corresponding author upon request.

Authors' contributions
MJN, AD, ML, CL, DHG participated in the conception and design of the study. MJN and AD undertook data acquisition. MJN and DHG processed data, carried out the statistical analysis and the interpretation of the results, and drafted the initial manuscript. AD, ML and CL assisted in the drafting of the manuscript. All authors were involved in revising the initial manuscript, and read and approved the final version to be published.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
This study was approved by the Research Ethics Committee of the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR) (Certificate: CRIR-557–1110). All participants were fully informed of the nature of the study and signed a consent form before entering the study.

Author details
[1]Pathokinesiology Laboratory, Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR) – Institut de réadaptation Lindsay-Gingras de Montréal, Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Centre-Sud-de-l'Île-de-Montréal, Montreal, Canada. [2]School of Rehabilitation, Université de Montréal, Pavillon 7077 Avenue du Parc, Station Centre-Ville, P.O. Box 6128, Montreal, Quebec H3C 3J7, Canada. [3]Faculty of Medicine, Université de Montréal, Montreal, Canada. [4]Occupational Health and Safety Research Institute Robert-Sauvé (IRSST), Montreal, Canada.

References

1. De Jonge S, Van den Berg C, De Vos R, Van der Heide H, Weir A, Verhaar J, Bierma-Zeinstra S, Tol J. Incidence of midportion Achilles tendinopathy in the general population. Br J Sports Med. 2011;45(13):1026–8.
2. Alfredson H, Lorentzon R. Chronic achilles tendinosis. Sports Med. 2000; 29(2):135–46.
3. Rolf C, Movin T. Etiology, histopathology, and outcome of surgery in achillodynia. Foot Ankle Int. 1997;18(9):565–9.
4. Kvist M. Achilles tendon injuries in athletes. Sports Med. 1994;18(3):173–201.
5. Krolo I, Višković K, Ikić Đ, Klarić-Ćustović R, Marotti M, Cicvara T. The risk of sports activities–the injuries of the Achilles tendon in sportsmen. Coll Antropol. 2007;31(1):275–8.
6. Lopes AD, Junior MLCH, Yeung SS, Costa LOP. What are the Main Running-Related Musculoskeletal Injuries? Sports Med. 2012;42(10):891–905.
7. Maffulli N, Sharma P, Luscombe KL. Achilles tendinopathy: aetiology and management. J R Soc Med. 2004;97(10):472–6.
8. Alfredson H, Cook J. A treatment algorithm for managing Achilles tendinopathy: new treatment options. Br J Sports Med. 2007;41(4):211–6.
9. McShane JM, Ostick B, McCabe F. Noninsertional Achilles tendinopathy: pathology and management. Curr Sports Med Rep. 2007;6(5):288–92.
10. Magnussen RA, Dunn WR, Thomson AB. Nonoperative treatment of midportion Achilles tendinopathy: a systematic review. Clin J Sport Med. 2009;19(1):54–64.
11. Brasseur J, Zeitoun-Eiss D, Renoux J, Bach G. Apport de l'échographie dans l'exploration du pied et de la cheville. EM-Consulte Podologie. 2010. doi:10.1016/S0292-062X(10)53770-X.
12. Bianchi S, Martinoli C. Ultrasound of the Musculoskeletal System. 1st ed. Berlin Heidelberg: Springer; 2007.
13. Jacobson JA. Fundamentals of musculoskeletal ultrasound. 1st ed. Philadelphia: Elsevier Health Sciences; 2007.
14. Sharma P, Maffulli N. Biology of tendon injury: healing, modeling and remodeling. J Musculoskelet Neuronal Interact. 2006;6(2):181.
15. Khan KM, Cook JL, Bonar F, Harcourt P, Åstrom M. Histopathology of common tendinopathies. Sports Med. 1999;27(6):393–408.
16. Hertzberg BS, Kliewer MA, Bowie JD, Carroll BA, DeLong DH, Gray L, Nelson RC. Physician Training Requirements in Sonography: How Many Cases Are Needed for Competence? AJR Am J Roentgenol. 2000;174(5):1221–7.
17. Krupinski EA. Computer-aided Detection in Clinical Environment: Benefits and Challenges for Radiologists1. Radiology. 2004;231(1):7–9.
18. Castellano G, Bonilha L, Li L, Cendes F. Texture analysis of medical images. Clin Radiol. 2004;59(12):1061–9.
19. Bharati MH, Liu JJ, MacGregor JF. Image texture analysis: methods and comparisons. Chemometr Intell Lab Syst. 2004;72(1):57–71.
20. Schie HT, Bakker EM, Jonker AM, Weeren P. Ultrasonographic tissue characterization of equine superficial digital flexor tendons by means of gray level statistics. Am J Vet Res. 2000;61(2):210–9.
21. Sipila S, Suominen H. Ultrasound imaging of the quadriceps muscle in elderly athletes and untrained men. Muscle Nerve. 1991;14(6):527–33.
22. Sipilä S, Suominen H. Muscle ultrasonography and computed tomography in elderly trained and untrained women. Muscle Nerve. 1993;16(3):294–300.
23. Sipilä S, Suominen H. Quantitative ultrasonography of muscle: detection of adaptations to training in elderly women. Arch Phys Med Rehabil. 1996;77(11):1173–8.
24. Nielsen PK, Jensen BR, Darvann T, Jørgensen K, Bakke M. Quantitative ultrasound tissue characterization in shoulder and thigh muscles–a new approach. BMC Musculoskelet Disord. 2006;7(1):2.
25. Pillen S, Verrips A, Van Alfen N, Arts I, Sie L, Zwarts M. Quantitative skeletal muscle ultrasound: diagnostic value in childhood neuromuscular disease. Neuromuscul Disord. 2007;17(7):509–16.
26. Bashford GR, Tomsen N, Arya S, Burnfield JM, Kulig K. Tendinopathy discrimination by use of spatial frequency parameters in ultrasound B-mode images. IEEE Trans Med Imaging. 2008;27(5):608–15.
27. van Schie H, De Vos R, De Jonge S, Bakker E, Heijboer M, Verhaar J, Tol J, Weinans H. Ultrasonographic tissue characterisation of human Achilles tendons: quantification of tendon structure through a novel non-invasive approach. Br J Sports Med. 2010;44(16):1153–9.
28. Tuthill TA, Rubin JM, Fowlkes JB, Jamadar DA, Bude RO. Frequency analysis of echo texture in tendon. Ultrasound Med Biol. 1999;25(6):959–68.

Nadeau *et al. Journal of Foot and Ankle Research* (2016) 9:30

Page 17 of 17

29. O'Connor PJ, Grainger AJ, Morgan S, Smith K, Waterton J, Nash A. Ultrasound assessment of tendons in asymptomatic volunteers: a study of reproducibility. Eur Radiol. 2004;14(11):1968–73.

30. 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;110(5):1384–9.

31. Schmidt W, Schmidt H, Schicke B, Gromnica-Ihle E. Standard reference values for musculoskeletal ultrasonography. Ann Rheum Dis. 2004;63(8):988–94.

32. Fredberg U, Bolvig L, Andersen NT, Stengaard-Pedersen K. Ultrasonography in evaluation of Achilles and patella tendon thickness. Ultraschall Med. 2008;29(01):60–5.

33. Brushøj C, Henriksen B, Albrecht-Beste E, Hölmich P, Larsen K, Nielsen MB. Reproducibility of ultrasound and magnetic resonance imaging measurements of tendon size. Acta Radiol. 2006;47(9):954–9.

34. Ying M, Yeung E, Li B, Li W, Lui M, Tsoi C-W. Sonographic evaluation of the size of Achilles tendon: the effect of exercise and dominance of the ankle. Ultrasound Med Biol. 2003;29(5):637–42.

35. Larivière C, Gagnon D, De Oliveira JE, Henry SM, Mecheri H, Dumas J-P. Reliability of Ultrasound Measures of the Transversus Abdominis: Effect of Task and Transducer Position. PM R. 2013;5(2):104–13.

36. Lariviere C, Gagnon D, De Oliveira JE, Henry SM, Mecheri H, Dumas J-P. Ultrasound Measures of the Lumbar Multifidus: Effect of Task and Transducer Position on Reliability. PM R. 2013;5(8):678–87.

37. Robinson J, Cook JL, Purdam C, Visentini P, Ross J, Maffulli N, Taunton J, Khan K. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. Br J Sports Med. 2001;35(5):335–41.

38. Vestergrd Iversen J, Bartels EM, Langberg H. The Victorian Institute of Sports Assessment-Achilles Questionnaire (VISA-A)-reliable tool for measuring Achilles tendinopathy. Int J Sports Phys Ther. 2012;7(1):76–84.

39. Collinger JL, Gagnon D, Jacobson J, Impink BG, Boninger ML. Reliability of quantitative ultrasound measures of the biceps and supraspinatus tendons. Acad Radiol. 2009;16(11):1424–32.

40. Collinger JL, Fullerton B, Impink BG, Koontz AM, Boninger ML. Validation of greyscale-based quantitative ultrasound in manual wheelchair users: relationship to established clinical measures of shoulder pathology. Am J Phys Med Rehabil. 2010;89(5):390.

41. Collinger JL, Impink BG, Ozawa H, Boninger ML. Effect of an intense wheelchair propulsion task on quantitative ultrasound of shoulder tendons. PM R. 2010;2(10):920–5.

42. Roebroeck ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther. 1993;73(6):386–95.

43. Portney LG, Watkins MP. Foundations of clinical research: Applications to practice. 3rd ed. Prentice Hall; 2007

44. McCreesh KM, Crotty JM, Lewis JS. Acromiohumeral distance measurement in rotator cuff tendinopathy: is there a reliable, clinically applicable method? A systematic review. Br J Sports Med. 2015;49(5):298–305.

45. Gellhorn AC, Carlson MJ. Inter-rater, intra-rater, and inter-machine reliability of quantitative ultrasound measurements of the patellar tendon. Ultrasound Med Biol. 2013;39(5):791–6.

46. Tweedell AJ, Ryan ED, Scharville MJ, Rosenberg JG, Sobolewski EJ, Kleinberg CR. The influence of ultrasound measurement techniques on the age-related differences in Achilles tendon size. Exp Gerontol. 2016;76:68–71.

47. Scholten R, Pillen S, Verrips A, Zwarts M. Quantitative ultrasonography of skeletal muscles in children: normal values. Muscle Nerve. 2003;27(6):693–8.

48. Reimers CD, Schlotter B, Eicke BM, Witt TN. Calf enlargement in neuromuscular diseases: a quantitative ultrasound study in 350 patients and review of the literature. J Neurol Sci. 1996;143(1):46–56.

49. Nielsen PK, Jensen BR, Darvann T, Jørgensen K, Bakke M. Quantitative ultrasound image analysis of the supraspinatus muscle. Clin Biomech (Bristol, Avon). 2000;15:S13–6.

50. Arend CF. Top ten pitfalls to avoid when performing musculoskeletal sonography: what you should know before entering the examination room. Eur J Radiol. 2013;82(11):1933–9.

51. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern Syst. 1973;3(6):610–21.

52. Lagergren C, Lindholm Å. Vascular distribution in the Achilles tendon; an angiographic and microangiographic study. Acta Chir Scand. 1959;116(5–6):491.

53. Clain MR, Baxter DE. Achilles tendinitis. Foot Ankle Int. 1992;13(8):482–7.

54. Syha R, Peters M, Birnesser H, Niess A, Hirschmueller A, Dickhuth H, Sandrock M. Computer-based quantification of the mean Achilles tendon thickness in ultrasound images–effect of tendinosis. Br J Sports Med. 2007;41(12):897–902.

55. Syha R, Grau S, Nieß AM, Hein T, Werner M, Ketelsen D, Schick F, Best R. Computer-based quantification of the Achilles tendon thickness in sequential B-mode ultrasound images: a study of feasibility and reliability. Arch Orthop Trauma Surg. 2014;134(10):1443–9.